# PARTE 1 - PROBABILIDAD

## *1- Probabilidad*

### *1.1 - Espacios muestrales y eventos.*

La Teoría de Probabilidades estudia los llamados ***experimentos aleatorios***.

Ejemplos clásicos de experimentos aleatorios son los juegos de azar:

a) tirar un dado y observar el número en la cara de arriba.
b) tirar una moneda
c) lanzar una moneda cuatro veces y contar el número total de caras obtenidas.
d) lanzar una moneda cuatro veces y observar la sucesión de caras y cecas obtenidas.

Simbolizamos con $\varepsilon$ a un experimento aleatorio.

Un experimento aleatorio tiene las siguientes características:

1-Se lo puede repetir bajo las mismas condiciones tantas veces como se desee.

2- No se puede predecir con exactitud el resultado de dicho experimento, pero se puede decir cuáles son los posibles resultados del mismo.

3- A medida que el experimento se repite, los resultados individuales parecen ocurrir en forma caprichosa. Pero si el experimento se repite un gran número de veces, y registramos la ***proporción*** de veces que ocurre un determinado resultado, veremos que esa proporción tiende a estabilizarse en un valor determinado a medida que aumenta el número de veces que se repite el experimento.

Por ejemplo, consideremos el experimento de lanzar un dado y observar el número de la cara superior. Supongamos que tiramos el dado N veces, y sea $n$ el número de veces que sale el número 5 en los N tiros del dado. Entonces $\frac{n}{N}$ es la proporción de veces que sale el número 5 en los N tiros.

Si el dado es normal a medida que N aumenta, $\frac{n}{N}$ tiende a estabilizarse en un número que es 1/6.

<u>Observación</u>: en los experimentos no aleatorios o ***deterministas*** se puede predecir con exactitud el resultado del experimento, es decir, las condiciones en las que se verifica un experimento determinan el resultado del mismo. Por ejemplo, si colocamos una batería en un circuito simple, el modelo matemático que posiblemente describiría el flujo observable de corriente sería I = E/R, que es la ley de Ohm. El modelo predice el valor de I al dar E y R. O sea, si se repite el experimento anterior cierto número de veces, empleando cada vez el mismo circuito, es decir manteniendo fijas E y R, esperaríamos observar el mismo valor de I.

A veces sucede que un experimento no es aleatorio estrictamente, pero resulta mucho más sencillo estudiarlo como si fuera aleatorio. Por ejemplo, si tiramos una moneda y observamos qué lado queda hacia arriba, el resultado sería predecible conociendo en forma precisa las velocidades iniciales de traslación y rotación, y las elasticidades de los materiales del piso y de la moneda. Pero la precisión con la que se necesitan conocer estos datos es casi imposible de obtener en la realidad, por lo que es más conveniente tratar al experimento como aleatorio.

El conjunto de todos los resultados posibles de un experimento aleatorio es el ***espacio muestral***. Al espacio muestral lo anotamos con la letra *S*.

Por ejemplo,

a) Si $\varepsilon$: tirar un dado y observar el número en la cara de arriba, entonces podemos tomar como espacio muestral a $S=\{1,2,3,4,5,6\}$
b) Si $\varepsilon$: tirar una moneda, entonces $S=\{c,s\}$
c) Si $\varepsilon$: lanzar una moneda tres veces y contar el número total de caras obtenidas entonces podemos considerar $S=\{0,1,2,3\}$
d) Si $\varepsilon$: lanzar una moneda tres veces y observar la sucesión de caras y cecas obtenidas, entonces $S=\{(c,c,c);(c,c,s);(c,s,c);(s,c,c);(c,s,s);(s,s,c);(c,s,c);(s,s,s)\}$
e) Si $\varepsilon$: tirar un dado las veces necesarias hasta que sale un 6 por primera vez, y contar el número de tiros realizados, entonces $S=\{1,2,3,4,.....\}=N$, donde $N$ es el conjunto de los números naturales.
f) Si $\varepsilon$: medir el tiempo de vida de una lamparita eléctrica, entonces $S=\{t\in R, t\geq 0\}$ donde $R$ es el conjunto de los números reales.

Observaciones:

1- la elección de $S$ no es única, depende de lo que se quiera observar del experimento aleatorio.
2- El espacio muestral puede ser un conjunto finito, o infinito. A su vez si es infinito puede ser infinito numerable o no numerable. En e) el conjunto $S$ es infinito numerable, en f) el conjunto $S$ es infinito no numerable.

Se llama ***evento*** o ***suceso*** a todo subconjunto del espacio muestral. Por ejemplo,

a) En el experimento dado en el ejemplo a), un evento de $S$ sería $A=\{2,4,6\}$ pues $A\subset S$.
Podemos expresar al evento $A$ con palabras de la siguiente manera $A$: "sale un número par"

También $B=\{1,2,3\}$ es un evento al que podemos expresar verbalmente como $B$: "sale un número menor o igual que 3"
b) En el experimento dado en el ejemplo d), un evento de $S$ sería $C=\{(c,c,c);(c,c,s);(c,s,c);(s,c,c)\}$, el que en palabras se puede expresar como $C$: "salen por lo menos dos caras"
c) En el experimento dado en el ejemplo f), un evento de $S$ sería $D$: "la lamparita dura más de 1200 horas", en notación de conjuntos $D=\{t\in R;\ \ t>1200\}$

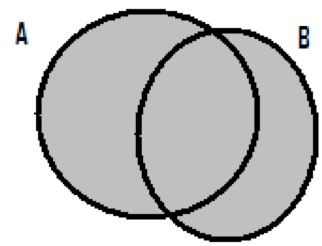
Observaciones:

1- en el ejemplo a) anterior, si al tirar el dado sale el número 2, entonces podemos decir que ***A ocurrió*** pues $2\in A$. Pero también ***B ocurrió*** pues $2\in B$. En cambio si al tirar el dado sale el número 4, entonces el evento $A$ ocurrió pero ***B no ocurrió***, pues $4\notin B$.
2- el conjunto $\varnothing$ es un evento (pues el conjunto $\varnothing$ está incluido en todo conjunto, en particular $\varnothing\subset S$). Es el evento que ***nunca ocurre.***
El espacio muestral $S$ es un evento (pues todo conjunto está incluido en sí mismo), y $S$ ***siempre ocurre.***

Las operaciones habituales entre conjuntos se pueden aplicar a los eventos, dando como resultado nuevos eventos. Específicamente
1- Si $A$ y $B$ son eventos, entonces $A\cup B$ es otro evento. $A\cup B$ ocurre si y solo si ocurre $A$ **o** si ocurre $B$
2- Si $A$ y $B$ son eventos, entonces $A\cap B$ es otro evento. $A\cap B$ ocurre si y solo si ocurre $A$ **y** ocurre $B$

3- Si $A$ es un evento $A^C$ es un evento. $A^C$ ocurre si y solo si no ocurre $A$

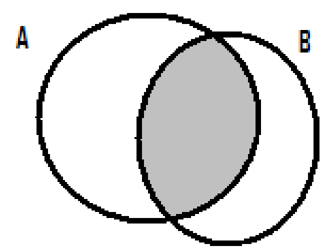
Nota: recordar que las operaciones de unión, intersección diferencia y complemento se definen de la siguiente manera:
1- $A \cup B$ es el conjunto formado por los elementos que están en $A$ ***o*** en $B$ (donde el ***o*** está en sentido inclusivo), en notación de conjuntos
$A \cup B = \{x;\ x \in A \vee x \in B\}$




En la figura la zona en gris simboliza $A \cup B$

2- $A \cap B$ es el conjunto formado por los elementos que están en $A$ ***y*** en $B$, en notación de conjuntos

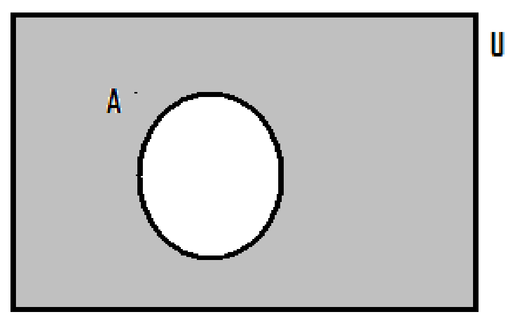
$A \cap B = \{x;\ x \in A \wedge x \in B\}$




En la figura la zona en gris simboliza $A \cap B$

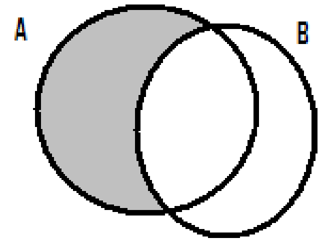
3- $A^C$ es el conjunto formado por los elementos que están en el conjunto universal $U$ y no están en $A$, en notación de conjuntos
$A^C = \{x;\ x \in U \wedge x \notin A\}$

La zona en gris simboliza $A^C$




4- $A - B$ es el conjunto formado por los elementos que están en $A$ ***y no están*** en $B$, en notación de conjuntos $A - B = \{x;\ x \in A \wedge x \notin B\}$,

La zona en gris simboliza $A - B$
Notar que $A - B = A \cap B^C$



Es útil recordar las siguientes propiedades sobre el álgebra de conjuntos:

1- Leyes de idempotencia

a) $A \cup A = A$ b) $A \cap A = A$

2- Leyes asociativas

a) $A \cup (B \cup C) = (A \cup B) \cup C$ b) $A \cap (B \cap C) = (A \cap B) \cap C$

3- Leyes conmutativas

a) $A \cup B = B \cup A$ b) $A \cap B = B \cap A$

4- Leyes distributivas

a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

5- Leyes de identidad

a) $A \cup \varnothing = A$ b) $A \cap \varnothing = \varnothing$

c) $A \cup U = U$ d) $A \cap U = A$

6- Leyes de complemento

a) $A \cup A^C = U$ b) $A \cap A^C = \varnothing$

c) $\left(A^C\right)^C = A$ d) $U^C = \varnothing$, $\varnothing^C = U$

7- Leyes de De Morgan

a) $(A \cup B)^C = A^C \cap B^C$ b) $(A \cap B)^C = A^C \cup B^C$

La relación de inclusión entre un conjunto y otro y las operaciones anteriores con conjuntos lleva al siguiente

Teorema:

a) $A \subset B$ es equivalente a $A \cap B = A$

b) $A \subset B$ es equivalente a $A \cup B = B$

c) $A \subset B$ es equivalente a $B^C \subset A^C$

d) $A \subset B$ es equivalente a $A \cap B^C = \varnothing$,

e) $A \subset B$ es equivalente a $B \cup A^C = U$

Sean $A$ y $B$ dos conjuntos. El ***conjunto producto*** de $A$ y $B$, expresado $A \times B$, está formado por todos los pares ordenados $(a,b)$ donde $a \in A$ y $b \in B$, es decir

$$A \times B = \{(a,b);\ \ a \in A \wedge b \in B\}$$

Se anota al producto $A \times A = A^2$

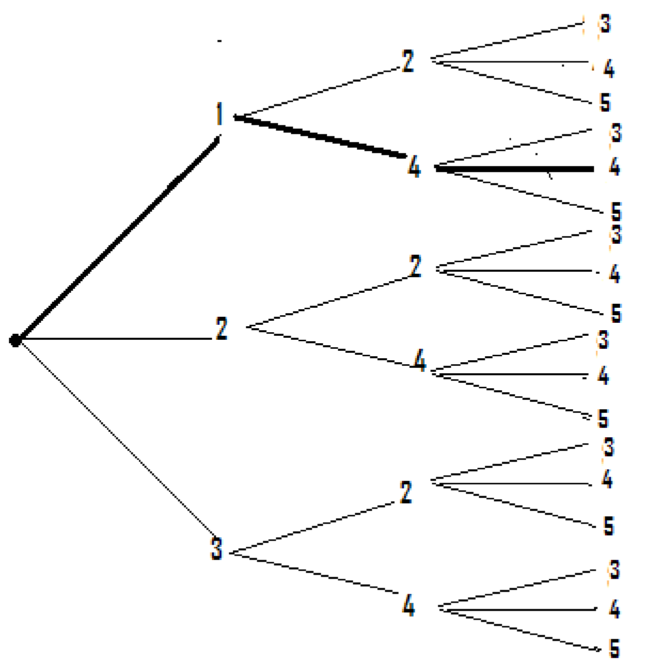
Por ejemplo, si $A = \{1,2,3\}$ y $B = \{7,8\}$, entonces $A \times B = \{(1,7);(1,8);(2,7);(2,8);(3,7);(3,8)\}$

El concepto de conjunto producto se extiende a un número finito de conjuntos en forma natural. El conjunto producto de los conjuntos $A_1, A_2, ..., A_n$ se anota $A_1 \times A_2 \times .. \times A_n$ y es igual a

$A_1 \times A_2 \times .. \times A_n = \{(a_1, a_2, ..., a_n),\ a_i \in A_i,\ i = 1,2,...,n\}$, es decir es el conjunto de todas las *n-uplas* $(a_1, a_2, ..., a_n)$ donde $a_i \in A_i$ para cada *i.*

Por ejemplo, si $A = \{1,2,3\},\ B = \{2,4\},\ C = \{3,4,5\}$, entonces un método conveniente para hallar el producto $A \times B \times C$ es por medio del denominado "diagrama de árbol" que se muestra a continuación, el cual se construye de izquierda a derecha.

$A \times B \times C$ consta de todas las ternas (o 3-uplas) formadas al seguir cada "camino" del árbol




Por ejemplo el camino con trazo más grueso indicaría la terna
(1, 4, 4)

Las operaciones de unión e intersección también se pueden generalizar a más de dos conjuntos
Si $A_1, A_2, ..., A_n$ son *n* conjuntos, entonces

a) la unión de todos ellos se anota $A_1 \cup A_2 \cup .. \cup A_n$ o también $\bigcup_{i=1}^{n} A_i$ y es igual a $A_1 \cup A_2 \cup .. \cup A_n = \{x;\ \text{existe } i \text{ tal que } x \in A_i\}$, es decir un elemento $x$ pertenece a $A_1 \cup A_2 \cup .. \cup A_n$ si $x$ pertenece ***a alguno de los conjuntos*** $A_i$**.**

De forma análoga se define la unión de una secuencia infinita de conjuntos $A_1, A_2, ...$, y se la anota con el símbolo $\bigcup_{i=1}^{\infty} A_i$

b) la intersección de todos ellos se anota $A_1 \cap A_2 \cap .. \cap A_n$ o también $\bigcap_{i=1}^{n} A_i$ y es igual a $\bigcap_{i=1}^{n} A_i = \{x;\ x \in A_i \ \text{ para todo } i\}$, es decir un elemento *x* pertenece a $\bigcap_{i=1}^{n} A_i$ si $x$ pertenece ***a todos los conjuntos*** $A_i$

De forma análoga se define la intersección de una secuencia infinita de conjuntos $A_1, A_2, \ldots$, y se la anota con el símbolo $\bigcap_{i=1}^{\infty} A_i$

Si $A$ y $B$ son dos eventos tales que $A \cap B = \emptyset$, se dice que son disjuntos o ***mutuamente excluyentes.***
Es decir si $A$ y $B$ son mutuamente excluyentes no pueden ocurrir a la vez. Por ejemplo, si se tira un dado y se observa el número de la cara superior, los eventos $A = \{1,3,5\}$ y $B = \{2,4,6\}$ son mutuamente excluyentes. Al tirar un dado sale un número par o un número impar, no pueden darse ambas cosas a la vez.
Los eventos con un solo elemento son eventos ***elementales o simples***. Por ejemplo, volviendo al experimento $\varepsilon$: tirar un dado y observar el número en la cara de arriba, y $S = \{1,2,3,4,5,6\}$, entonces los conjuntos unitarios $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ son eventos simples. Notar que dos eventos simples cualesquiera son mutuamente excluyentes.

Dado un evento $A$ asociado a un experimento aleatorio $\varepsilon$. Supongamos que se repite $n$ veces el experimento $\varepsilon$, y anotamos $n_A$ al número de veces que ocurre $A$ en la $n$ repeticiones de $\varepsilon$. Se define la ***frecuencia relativa de A***, y se simboliza $f_A$, al cociente $\frac{n_A}{n}$. Es decir que $f_A$ es la ***proporción de veces que ocurre A en las n repeticiones de*** $\varepsilon$.
La frecuencia relativa $f_A$ tiene las siguientes propiedades:
1- $0 \le f_A \le 1$
2- $f_A = 1$ si y solo si $A$ ocurre ***cada vez*** en las $n$ repeticiones
3- $f_A = 0$ si y solo si $A$ no ocurre ***nunca*** en las $n$ repeticiones
4- si $A$ y $B$ son dos eventos mutuamente excluyentes entonces $f_{A \cup B} = f_A + f_B$

Dado un evento $A$, se quiere asignar al mismo un número que indique qué tan ***probable*** es que $A$ ocurra. A ese número lo definiríamos como la ***probabilidad de A***. En ese sentido parecería que la frecuencia relativa $f_A$ sería una buena elección. Pero nos encontramos con el siguiente problema, ¿cuántas veces deberíamos repetir el experimento aleatorio para definir $f_A$, es decir qué valor de $n$ tomaríamos?

Por ejemplo en el experimento de tirar un dado consideremos el evento A: "sale el número 4", si lo lanzamos 100 veces podríamos encontrar que $n_A = 14$, y si lo lanzamos nuevamente 100 veces podría ocurrir que $n_A$ sea diferente del anterior por ejemplo podría $A$ ocurrir 20 veces. Entonces, tendríamos dos valores diferentes para $f_A$, 0.14 y 0.2

Se dijo antes que en un experimento aleatorio a medida que $n$ aumenta la frecuencia relativa de $A$ tiende a estabilizarse en un número, pero no podemos en la práctica repetir el experimento infinitas veces.

Se quiere asignar a cada evento $A$ un número que no dependa de la experimentación. Por este motivo procedemos como sigue:

**Definición axiomática de probabilidad**.

Sea $\varepsilon$ un experimento aleatorio y *S* un espacio muestral asociado con $\varepsilon$. Con cada evento *A* asociamos un número real llamado probabilidad de *A*, que anotamos *P(A)*, el cual satisface las siguientes propiedades básicas o *axiomas*

1- $0 \le P(A) \le 1$
2- $P(S) = 1$
3- Si *A* y *B* son eventos mutuamente excluyentes entonces $P(A \cup B) = P(A) + P(B)$
4- Si $A_1, A_2, ..., A_n, A_{n+1}, ...$ es una secuencia de eventos tales que

$A_i \cap A_j = \varnothing$ si $i \neq j$, entonces $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

La elección de estas propiedades está motivada por las características correspondientes de la frecuencia relativa.

Observación: supongamos el experimento de lanzar una moneda y el espacio muestral $S = \{c, s\}$. Notar que podemos escribir $S = \{c\} \cup \{s\}$, es decir como unión de eventos simples. Por los axiomas 2 y 3 de la definición de probabilidad tenemos que

$$1 = P(S) = P(\{c\}) + P(\{s\})$$

Es decir $1 = P(\{c\}) + P(\{s\})$, lo que implica que $P(\{c\}) = 1 - P(\{s\})$. No podemos deducir el valor de $P(\{c\})$ ni el valor de $P(\{s\})$ de las propiedades anteriores. Necesitamos información adicional, por ejemplo si el dado es normal. De ser así podemos plantear que $P(\{c\}) = P(\{s\})$, entonces

$$\left.\begin{array}{l} 1 = P(\{c\}) + P(\{s\}) \\ P(\{c\}) = P(\{s\}) \end{array}\right\} \Rightarrow P(\{c\}) = P(\{s\}) = 0.5$$

Podemos deducir de los axiomas otras propiedades útiles para el cálculo de probabilidades.

**Propiedades de la probabilidad**

1- $P(\varnothing) = 0$

*Dem.*) Siendo *A* un evento cualquiera podemos escribir $A \cup \varnothing = A$
Además *A* y $\varnothing$ son mutuamente excluyentes, por lo tanto por axioma 3

$$P(A) = P(A \cup \varnothing) = \mathrm{P(A)} + \mathrm{P}(\varnothing)$$

O sea que

$$P(A) = \mathrm{P(A)} + \mathrm{P}(\varnothing) \Rightarrow \mathrm{P}(\varnothing) = 0$$

2- Si $A^C$ es el evento complementario de $A$, entonces $P(A^C) = 1 - P(A)$

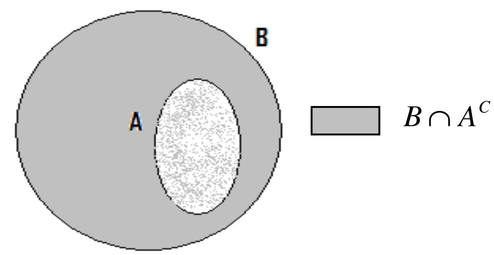
*Dem.*) Si $A$ es un evento cualquiera, entonces podemos escribir $A \cup A^C = S$
Además, por definición de complemento de un conjunto, $A$ y $A^C$ son mutuamente excluyentes.
Por lo tanto, por axioma 2 y por axioma 3

$$1 = P(S) = P(A) + P(A^C)$$

Despejando

$$P(A^C) = 1 - P(A)$$

3- $\boxed{\text{Si } A \subset B \text{ entonces } P(A) \le P(B)}$




*Dem.*) Sean $A$ y $B$ dos eventos tales que $A \subset B$.
De la figura vemos que $B = A \cup (B \cap A^C)$
Y además $A$ y $B \cap A^C$ son mutuamente excluyentes. Entonces

$P(B) = P(A) + P(B \cap A^C)$

Y como por axioma 1 tenemos que $P(B \cap A^C) \ge 0$, entonces $P(B) \ge P(A)$

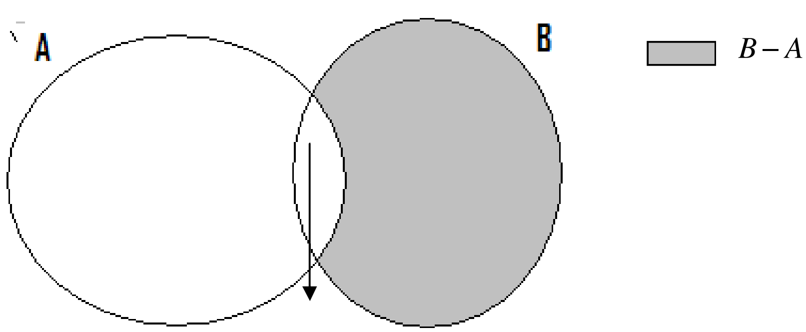
4- $\boxed{\text{Si } A \subset B, \text{ entonces } P(B - A) = P(B) - P(A)}$

*Dem.*) Siguiendo los pasos de la demostración anterior llegamos a

$P(B) = P(A) + P(B \cap A^C)$, lo que implica que $P(B \cap A^C) = P(B) - P(A)$

Y como $B \cap A^C = B - A$, entonces $P(B - A) = P(B) - P(A)$.

<u>Observación</u>: ***en general*** vale la siguiente propiedad $\boxed{P(B - A) = P(B) - P(A \cap B)}$

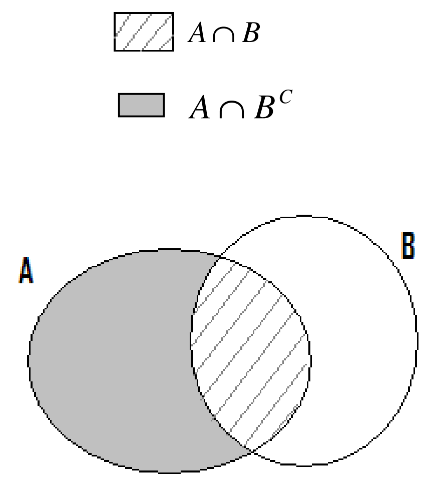



$A \cap B$

5- Si $A$ y $B$ son dos eventos ***cualesquiera***, entonces

$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

*Dem.*) Escribimos a $A \cup B$ como unión de partes disjuntas de la siguiente manera, (observar la figura)




$$A \cup B = B \cup (A \cap B^C)$$

$$\therefore P(A \cup B) = P(B) + P(A \cap B^C) \quad (1)$$

Y por la observación anterior

$$P(B \cap A^C) = P(B - A) = P(B) - P(A \cap B) \quad (2)$$

Por lo tanto, reemplazando (2) en (1): $P(A \cup B) = P(B) + P(A) - P(A \cap B)$

Con lo que queda demostrada la propiedad.

Observaciones:

1- La propiedad anterior se puede generalizar para expresar la probabilidad de tres eventos

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Se llega a la igualdad anterior escribiendo $A \cup B = D$ y aplicando la propiedad 5, es decir:

$$P(A \cup B \cup C) = P(D \cup C) = P(D) + P(C) - P(D \cap C) =$$
$$= P(A \cup B) + P(C) - P((A \cup B) \cap C) \quad (3)$$

Nuevamente aplicamos 5: $P(A \cup B) = P(B) + P(A) - P(A \cap B)$

Y además, aplicando las leyes distributivas $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

Entonces:

$$P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) = P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C))$$

Como $(A \cap C) \cap (B \cap C) = A \cap B \cap C$ , reemplazando en (3) se llega a :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

En general, se puede probar que si $A_1, A_2, ..., A_n$ son *n* eventos cualesquiera entonces

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - ..... + (-1)^{n-1} P(A_1 \cap A_2 \cap .... \cap A_n)$$

2- Si $A_1, A_2, ..., A_n$ son *n* eventos tales que $A_i \cap A_j = \emptyset$ con $i \neq j$ entonces $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Notar que la igualdad anterior es una generalización del axioma 3.

## 1.2 - El espacio muestral finito

Sea $\varepsilon$ un experimento aleatorio y $S$ un espacio muestral asociado a $\varepsilon$. Supongamos que $S$ es un conjunto finito al que anotamos $S=\{a_1,a_2,...,a_n\}$, es decir consideramos que tiene $n$ elementos.

Podemos escribir a $S$ como unión de eventos elementales de la siguiente forma.

Si $A_i=\{a_i\}$ entonces $S=A_1\cup A_2\cup...\cup A_n$. Además sabemos que $P(S)=1$, por lo tanto tenemos el resultado:

$$P(A_1)+P(A_2)+...+P(A_n)=1 \qquad (4)$$

Es decir: ***la suma de las probabilidades de los eventos elementales es igual a 1***

Si ahora tomamos un evento cualquiera $B$ de $S$, lo anotamos $B=\{a_{i_1},a_{i_2},...,a_{i_k}\}$, entonces podemos escribir a $B$ como unión de eventos simples: $B=A_{i_1}\cup A_{i_2}\cup...\cup A_{i_k}$. Por lo tanto

$$P(B)=P(A_{i_1}\cup A_{i_2}\cup...\cup A_{i_k})=P(A_{i_1})+P(A_{i_2})+...+P(A_{i_k})$$

***Es decir para calcular la probabilidad de B se suman las probabilidades de los eventos elementales que lo componen.***

Si asumimos que $P(A_1)=P(A_2)=...=P(A_n)=p$, es decir todos los eventos elementales tienen la misma probabilidad de ocurrir, entonces reemplazando en (4):

$$\underbrace{p+p+...+p}_{n \text{ veces } p}=1 \quad\Rightarrow\quad np=1 \quad\Rightarrow\quad p=\frac{1}{n}=\frac{1}{\#S}$$

Donde el símbolo $\#S$ se lee: número de elementos de $S$, (en general $\#S$ indica el ***cardinal*** de $S$)

Cuando $P(A_1)=P(A_2)=...=P(A_n)=p$, se dice que $S$ es ***equiprobable.***

Si $B$ es un evento cualquiera de $S$, lo anotamos $B=\{a_{i_1},a_{i_2},...,a_{i_k}\}$ ( o sea que $\#B=k$), entonces escribimos a $B$ como unión de eventos simples: $B=A_{i_1}\cup A_{i_2}\cup...\cup A_{i_k}$, y nuevamente

$$P(B)=P(A_{i_1}\cup A_{i_2}\cup...\cup A_{i_k})=P(A_{i_1})+P(A_{i_2})+...+P(A_{i_k})$$

Como ya vimos que $P(A_i)=\dfrac{1}{n}$ para todo $i=1,2,\ldots,n$, se tiene que:

$$P(B)=P(A_{i_1})+P(A_{i_2})+...+P(A_{i_k})=\underbrace{\frac{1}{n}+\frac{1}{n}+...+\frac{1}{n}}_{k \text{ veces}}=\frac{k}{n}=\frac{\#B}{\#S}$$

Es decir ***si B es un evento de un espacio muestral equiprobable entonces*** $P(B)=\dfrac{\#B}{\#S}$

Ejemplos:

1- Supongamos que se tira una moneda normal tres veces, y se cuenta el número de caras obtenido luego de los tres tiros. Tomemos como espacio muestral a $S=\{0,1,2,3\}$

¿Cuál es la probabilidad de que salgan al menos dos caras?

En este caso *S* ***no es equiprobable***, pues
el evento $\{0\}$ ocurre de una sola forma: cuando en los tres tiros sale ceca
El evento $\{3\}$ ocurre de una sola forma: cuando en los tres tiros sale cara
Pero el evento $\{1\}$ ocurre cuando en los tres tiros sale una sola cara y eso puede darse de tres formas distintas $(c,s,s)$; $(s,s,c)$ o $(c,s,c)$
Análogamente para el evento $\{2\}$ puede darse de tres formas distintas:
$(c,c,s)$; $(c,s,c)$ o $(s,c,c)$
Por lo tanto $P(\{0\})=P(\{3\})$ y $P(\{1\})=P(\{2\})$. Si anotamos $P(\{0\})=P(\{3\})=p$ y $P(\{1\})=P(\{2\})=q$ entonces $q=3p$ (5)
Además se cumple $P(\{0\})+P(\{1\})+P(\{2\})+P(\{3\})=1$ (6)
Entonces de (5) y (6)

$$\left.\begin{matrix} q=3p \\ p+q+q+p=1 \end{matrix}\right\} \Rightarrow \left.\begin{matrix} q=3p \\ 2p+2q=1 \end{matrix}\right\} \Rightarrow p=\frac{1}{8}\ ,\ \ q=\frac{3}{8}$$

***De esta forma conocemos las probabilidades de todos los eventos elementales.***

Si *B*: " salen al menos dos caras" , entonces $B=\{2,3\}$
Para calcular *P(B)* hacemos

$$P(B)=P(\{2\})+P(\{3\})=\frac{3}{8}+\frac{1}{8}=\frac{4}{8}=0.5$$

Si tomamos como espacio muestral a

$$S=\{(c,c,c);(c,c,s);(c,s,c);(s,c,c);(c,s,s);(s,s,c);(c,s,c);(s,s,s)\}$$

Entonces *S* ***es equiprobable*** pues cada resultado (cada terna) se de una sola forma.

En este caso ***podemos plantear directamente*** $P(B)=\dfrac{\#B}{\#S}=\dfrac{4}{8}=0.5$

2- En el caso de espacios muestrales equiprobables es necesario recordar las ***técnicas de conteo*** para calcular el número de elementos de un conjunto sin enumerar sus elementos.
Por ejemplo:

Se tiene una urna con 15 bolillas distinguibles (pueden estar numeradas), de las cuales 10 son blancas y 5 son rojas. Se extraen ***al azar*** dos bolillas de la urna,

a) ¿cuál es la probabilidad de extraer todas blancas?
b) ¿cuál es la probabilidad de extraer exactamente una bolilla blanca?
c) ¿cuál es la probabilidad de extraer al menos una bolilla blanca?

Supongamos que se extraen las dos bolillas ***sin reemplazo*** y ***no importa el orden*** en que son extraídas.

Al decir que se extraen al azar, se entiende que el espacio muestral del experimento es equiprobable.
Al decir que se extraen las bolillas sin reemplazo, significa que se extrae una bolilla y sin regresarla a la urna se extrae la siguiente.
El espacio muestral de este experimento se puede tomar como $S = \{\{1,4\},\{2,5\},\{2,10\},...\}$, es decir como el conjunto cuyos elementos son conjuntos, todos los conjuntos de dos elementos que se puedan formar a partir de un conjunto de 15 elementos. Por ejemplo, el resultado $\{1,4\}$ significa que se extrajeron las bolillas 1 y 4. ¿Cuántos elementos tiene $S$?, recordando las técnicas de conteo el número de elementos de $S$ es igual al número combinatorio $\binom{15}{2} = \frac{15!}{2!(15-2)!}$

Para responder la pregunta **a)** anotemos $A$: "las dos bolillas son blancas"
Entonces $P(A) = \frac{\# A}{\# S}$. ¿Cuántos elementos tiene el conjunto $A$?, tantos como formas haya de extraer dos bolillas blancas. Se tiene que $\# A = \binom{10}{2} = \frac{10!}{2!(10-2)!}$.

Por lo tanto $P(A) = \frac{\# A}{\# S} = \frac{\binom{10}{2}}{\binom{15}{2}} = \frac{\frac{10!}{2!(10-2)!}}{\frac{15!}{2!(15-2)!}} = \frac{3}{7}$

Veamos ahora el inciso **b)**. Anotamos $B$: "se extrae exactamente una bolilla blanca"
Entonces

$$P(B) = \frac{\# B}{\# S} = \frac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}}$$

Y por último el inciso c). Anotamos C: "se extrae al menos una bolilla blanca"
Podemos resolverlo de dos maneras:

La forma más fácil es calcular $P(C^C)$ donde $C^C$: "ninguna bolilla es blanca"

$$P(C^C) = \frac{\# C^C}{\# S} = \frac{\binom{5}{2}}{\binom{15}{2}} \Rightarrow P(C) = 1 - P(C^C) = 1 - \frac{\binom{5}{2}}{\binom{15}{2}} = 1 - \frac{2}{21} = \frac{19}{21}$$

Otra forma de resolverlo podría ser la siguiente, escribimos $C = C_1 \cup C_2$ donde

$C_i$ :"se extraen $i$ bolillas blancas exactamente" $\quad i = 1,2$

Notar que $C_1 \cap C_2 = \varnothing$. Por lo tanto $P(C) = P(C_1) + P(C_2)$

Y $\quad P(C_1) = \frac{\# C_1}{\# S} = \frac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}} \quad ; \quad P(C_2) = \frac{\# C_2}{\# S} = \frac{\binom{10}{2}\binom{5}{0}}{\binom{15}{2}}$

Por lo tanto $P(C)=\dfrac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}}+\dfrac{\binom{10}{2}\binom{5}{0}}{\binom{15}{2}}=\dfrac{10}{21}+\dfrac{9}{21}=\dfrac{19}{21}$

3- Si en el ejemplo anterior importa el orden en que son extraídas las bolillas, entonces $S=\{(1,4);(2,5),(2,10),.(4,1);(5,2)....\}$, es decir que $S$ es el conjunto de todos los ***pares*** $(a,b)$ donde $a\neq b$, por ejemplo (1,4) expresa el resultado: primero se extrajo la bolilla 1 y en segundo término se extrajo la bolilla 2. El número de elementos de $S$ es $15x14 = 210$.
En este caso
$P(A)=\dfrac{10\times 9}{15\times 14}=\dfrac{3}{7}$;
Para calcular la probabilidad de $B$ hay que distinguir si la bolilla blanca se extrae primero o se extrae en segundo término:
$P(B)=\dfrac{10\times 5+5\times 10}{15\times 14}=\dfrac{2(10\times 5)}{15\times 14}$
Para calcular la probabilidad de C nuevamente podemos pensarlo de dos formas, pero si consideramos $C=C_1\cup C_2$ en cada caso hay que tener en cuenta el orden en que se extraen las bolillas.
Es mas práctico hacer $P(C)=1-P(C^C)=1-\dfrac{5\times 4}{15\times 14}$

4- Si ahora ***importa el orden*** en que son extraídas las bolillas y además la extracción se hace ***con re emplazo*** (se extrae una bolilla y se la devuelve a la urna antes de extraer la siguiente), entonces $S$ es el conjunto de pares *(a,b)* donde ahora puede ocurrir que $a = b$.
En este caso $\#S$ = 15 x 15, y calculamos

$$P(A)=\frac{10\times 10}{15\times 15}\quad;\quad P(B)=\frac{10\times 5+5\times 10}{15\times 15}=\frac{2(10\times 5)}{15\times 15}\quad;\quad P(C)=1-P(C^C)=1-\frac{5\times 5}{15\times 15}$$

## *1.3 - Espacios muestrales infinitos*

Sea $S$ un espacio muestral ***infinito numerable*** es decir $S=\{a_1;a_2;a_3;......\}$. Como en el caso finito, a cada $a_i$ asignamos un número $p_i=P(\{a_i\})$ tal que

$a)\ p_i\geq 0 \qquad b)\ \sum_{i=1}^{\infty}p_i=1$

La probabilidad $P(A)$ de un evento $A$ es entonces la suma de las probabilidades de los eventos elementales que lo componen.

Los únicos espacios muestrales ***infinitos no numerables*** que consideraremos son aquellos de medida geométrica finita $m(S)$ tales como longitud, área o volumen, y en los cuales un punto se selecciona al azar. La probabilidad de un evento $A$, esto es aquella en la que el punto seleccionado pertenece a $A$, es entonces la relación entre $m(A)$ a $m(S)$, es decir

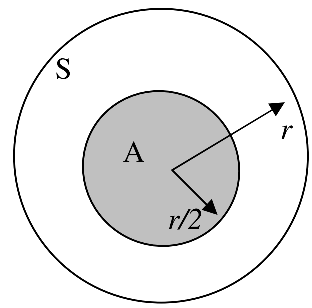
$$P(A)=\frac{\text{longitud de } A}{\text{longitud de } S} \quad \text{ó} \quad P(A)=\frac{\text{área de } A}{\text{área de } S} \quad \text{ó} \quad P(A)=\frac{\text{volumen de } A}{\text{volumen de } S}$$

Ejemplo:

En el interior de un círculo se selecciona un punto al azar. Hallar la probabilidad de que el punto quede más cercano al centro que a la circunferencia.

Tomamos como espacio muestral a $S = \{(x, y),\ \ x^2 + y^2 \le r^2\}$, entonces

$A = \left\{(x, y),\ \ x^2 + y^2 \le \left(\frac{r}{2}\right)^2\right\}$




Por lo tanto $P(A) = \dfrac{\text{área de } A}{\text{área de } S} = \dfrac{\pi\left(\frac{r}{2}\right)^2}{\pi\, r^2} = \dfrac{1}{4}$

Observación: si $A = \left\{(x, y),\ \ x^2 + y^2 = \left(\frac{r}{2}\right)^2\right\}$,

entonces $P(A) = \dfrac{\text{área de } A}{\text{área de } S} = \dfrac{0}{\pi\, r^2} = 0$ , pero $A \ne \varnothing$

Por lo tanto si un evento tiene probabilidad 0 eso no implica que sea el evento vacío.

# 2 - Probabilidad condicional

## 2.1- Definición

Supongamos el experimento aleatorio de extraer al azar sin reemplazo dos bolillas de una urna que contiene 7 bolillas rojas y 3 blancas. Asumimos que las bolillas de un mismo color son distinguibles.

Consideramos los eventos A: "la primer bolilla extraída es blanca", y B: "la segunda bolilla extraída es blanca".

El espacio muestral *S* se puede pensar como el conjunto
$S = \{(a,b);\ \ a = 1,2,...10;\ \ b = 1,2,...10;\ \ a \neq b\}$ y $\# S = 10 \times 9$ .

Es claro que $P(A) = \dfrac{3 \times 9}{10 \times 9} = \dfrac{3}{10}$. Pero si queremos calcular $P(B)$ no es tan directo. Podemos calcular la ***probabilidad de B sabiendo que A ocurrió***: es igual a $\dfrac{2}{9}$, ya que si *A* ocurrió, entonces en la urna quedaron 9 bolillas de las cuales 2 son blancas. La probabilidad anterior la anotamos $P(B/A)$ y se lee: probabilidad condicional de *B* dado *A*. Es decir $P(B/A) = \dfrac{2}{9}$.

Notar que podemos interpretar lo anterior de la siguiente forma: el espacio muestral original *S* se ha ***reducido*** al evento *A*, es decir ***se toma a A como nuevo espacio muestral*** para calcular la probabilidad de *B*.

También podemos interpretar que la probabilidad condicional de *B* dado *A* debería ser la proporción de veces que ocurre $A \cap B$ con respecto al número de veces que ocurre *A*. Pensando en términos de frecuencia relativa: $P(B/A) \approx \dfrac{n_{A\cap B}}{n_A}$.

Esta idea motiva la siguiente definición:

Sean *A* y *B* dos eventos de un espacio muestral *S*. La probabilidad condicional de *B* dado *A* se define como

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \qquad \text{si } P(A) \neq 0$$

Análogamente

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad \text{si } P(B) \neq 0$$

En algunos casos se puede calcular $P(B/A)$ directamente reduciendo el espacio muestral. En otros será necesario aplicar la definición anterior.

Observación: si *A* y *B* son eventos de un espacio muestral *S* ***equiprobable***, entonces

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\#(A \cap B)}{\# A}$$

Ejemplos:

1- En el experimento de extraer dos bolillas

$$P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{\frac{3\times 2}{10\times 9}}{\frac{3}{10}}=\frac{2}{9}$$

2- Se tira un dado normal dos veces. Sean los eventos $A$: "la suma de los números obtenidos es 6" y $B$: "el primer número es igual a 4" Tenemos que $A=\{(1,5);(2,4);(3,3);(4,2);(5,1)\}$ y $B=\{(4,1);(4,2);(4,3);(4,4);(4,5);(4,6)\}$ Entonces para calcular $P(A/B)$ mediante la definición de probabilidad condicional

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{36}}{\frac{6}{36}}=\frac{1}{6}$$

También podemos calcularlo en forma directa, reduciendo el espacio muestral, de todos los pares del evento $B$, observamos cuáles cumplen con lo requerido por $A$, es decir de todos los pares de *B,* solo uno tiene la propiedad de que sus componentes suman 6, por lo tanto

$$P(A/B)=\frac{1}{6}$$

3- Se lanza una moneda normal tres veces.
Hallar la probabilidad de que salgan todas caras si sale alguna cara.
El espacio muestral reducido es el evento A: "sale alguna cara"
Tenemos que $A=\{(c,c,c);(c,c,s);(c,s,c);(s,c,c);(c,s,s);(s,s,c);(s,c,s)\}$
Y $B=\{(c,c,c)\}$
Por lo tanto $P(B/A)=\frac{1}{7}$

4- En cierta ciudad, 40% de la población tiene cabellos castaños, 25% tiene ojos castaños y 15% tiene cabellos y ojos castaños. Se escoge una persona al azar
a) si tiene cabellos castaños, ¿cuál es la probabilidad de que también tenga ojos castaños?
b) Si tiene ojos castaños, ¿cuál es la probabilidad de que tenga cabellos castaños?
c) ¿Cuál es la probabilidad de que no tenga ni cabellos ni ojos castaños?

Sean los eventos $A$: "la persona elegida al azar tiene ojos castaños", $B$: "la persona elegida al azar tiene cabellos castaños"
Entonces $P(A)=0.25$, $P(B)=0.40$ y $P(A\cap B)=0.15$
a) Se pide calcular $P(A/B)$:
$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{0.15}{0.40}$$
b) $P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{0.15}{0.25}$
c) $P(A^C\cap B^C)=P((A\cup B)^C)=1-P(A\cup B)=1-(0.25+0.40-0.15)$

5- Sean los eventos $A$ y $B$ con $P(A)=0.5$, $P(B)=\frac{1}{3}$ y $P(A\cap B)=\frac{1}{4}$.

Hallar: a) $P(A/B)$, b) $P(B/A)$ , c) $P(A \cup B)$, d) $P(A^C / B^C)$ , e) $P(B^C / A^C)$

a) $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{1}{4}}{\frac{1}{3}} = \dfrac{3}{4}$ b) $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{\frac{1}{4}}{\frac{1}{2}} = \dfrac{1}{2}$

c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \dfrac{1}{2} + \dfrac{1}{3} - \dfrac{1}{4} = \dfrac{7}{12}$

d) $P(A^C / B^C) = \dfrac{P(A^C \cap B^C)}{P(B^C)}$

Calculamos: $P(B^C) = 1 - P(B) = 1 - \dfrac{1}{3} = \dfrac{2}{3}$ ;

$$P(A^C \cap B^C) = P((A \cup B)^C) = 1 - P(A \cup B) = 1 - \frac{7}{12} = \frac{5}{12}$$

Por la ley de De Morgan

Entonces $P(A^C / B^C) = \dfrac{P(A^C \cap B^C)}{P(B^C)} = \dfrac{\frac{5}{12}}{\frac{2}{3}} = \dfrac{15}{24} = \dfrac{5}{8}$

e) $P(B^C / A^C) = \dfrac{P(A^C \cap B^C)}{P(A^C)} = \dfrac{P(A^C \cap B^C)}{1 - P(A)} = \dfrac{\frac{5}{12}}{1 - \frac{1}{2}} = \dfrac{5}{6}$

<u>Observaciones</u>:

a) Si $A \subset B$ entonces $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(A)}{P(A)} = 1$

b) Si $A \cap B = \emptyset$ entonces $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(\emptyset)}{P(A)} = 0$

c) Es fácil comprobar que $P(B(/A)$ para $A$ fijo, satisface los axiomas de la probabilidad, esto es:

1- $0 \le P(B/A) \le 1$

2- $P(S/A) = 1$

3-Si $B_1$ y $B_2$ son eventos mutuamente excluyentes entonces

$$P\big((B_1 \cup B_2)/A\big) = P(B_1/A) + P(B_2/A)$$

4-Si $B_1, B_2, ..., B_n, B_{n+1}, ...$ es una secuencia de eventos tales que $B_i \cap B_j = \emptyset$ si $i \neq j$ entonces

$$P\left(\left(\bigcup_{i=1}^{\infty} B_i\right)/A\right) = \sum_{i=1}^{\infty} P(B_i/A)$$

**Teorema de la multiplicación**

Si $A$ y $B$ son dos eventos entonces $P(B/A)=\dfrac{P(A\cap B)}{P(A)}$ si $P(A)\neq 0$

Por lo tanto

$$P(A\cap B)=P(B/A)\,P(A) \qquad (6)$$

Análogamente de $P(A/B)=\dfrac{P(A\cap B)}{P(B)}$ si $P(B)\neq 0$ , se deduce

$$P(A\cap B)=P(A/B)\,P(B) \qquad (7)$$

(6)y (7) se conocen como ***teorema de la multiplicación***.
Consideremos otra vez el ejemplo de extraer dos bolillas al azar sin reemplazo de una urna que contiene 3 bolillas blancas y 7 rojas. Si $A$: "la primer bolilla extraída es blanca", y $B$: "la segunda bolilla extraída es blanca", entonces

$$P(A\cap B)=P(A)P(B/A)=\frac{3}{10}\times\frac{2}{9}$$

Si $A_1$, $A_2$, $A_3$ son tres eventos entonces

$$P(A_1\cap A_2\cap A_3)=P(A_1)P(A_2/A_1)P(A_3/A_1\cap A_2)$$

pues:

$$P(A_1\cap A_2\cap A_3)=P(B\cap A_3)=P(B)P(A_3/B)=P(A_1\cap A_2)P(A_3/A_1\cap A_{23})=$$

$B=A_1\cap A_2$ Teorema de la multiplicación

$$=P(A_1)P(A_2/A_1)P(A_3/A_1\cap A_2)$$

Teorema de la multiplicación

El teorema de la multiplicación se puede generalizar a $n$ eventos $A_1,A_2,\ldots,A_n$:

$$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)P(A_2/A_1)P(A_3/A_1\cap A_2)\ldots.P(A_n/A_1\cap A_2\cap\ldots,A_{n-1}) \qquad (8)$$

Ejemplos:

1- Una clase tiene 12 niños y 4 niñas. Si se escogen tres estudiantes de la clase al azar, ¿cuál es la probabilidad de que sean todos niños?

Solución:
Anotamos $A_i$ : "el i-ésimo estudiante elegido es un niño" $i=1,2,3$
Entonces la probabilidad pedida es

$$P(A_1\cap A_2\cap A_3)=P(A_1)P(A_2/A_1)P(A_3/A_1\cap A_2)=\frac{12}{16}\times\frac{11}{15}\times\frac{10}{14}$$

2- Los estudiantes de una clase se escogen al azar, uno tras otro, para presentar un examen.
   a) Si la clase consta de 4 niños y 3 niñas, ¿cuál es la probabilidad de que niños y niñas queden alternados?
   b) Si la clase consta de 3 niños y 3 niñas, ¿cuál es la probabilidad de que niños y niñas queden alternados?

Solución:

a) Nuevamente anotamos $A_i$ : "el i-ésimo estudiante elegido es un niño"

Entonces la probabilidad pedida es, aplicando (8):

$$P(A_1 \cap A_2^C \cap A_3 \cap A_4^C \cap A_5 \cap A_6^C \cap A_7) = \frac{4}{7}\times\frac{3}{6}\times\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{1} = \frac{1}{35}$$

b) Hay dos casos mutuamente excluyentes: el primer estudiante es un niño, y el primero es una niña.
Si el primero es un niño, entonces por (8) la probabilidad de que los estudiantes se alternen es

$$P(A_1 \cap A_2^C \cap A_3 \cap A_4^C \cap A_5 \cap A_6^C) = \frac{3}{6}\times\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{1} = \frac{1}{20}$$

Si el primero es una niña, entonces por (8) la probabilidad de que los estudiantes se alternen es

$$P(A_1^C \cap A_2 \cap A_3^C \cap A_4 \cap A_5^C \cap A_6) = \frac{3}{6}\times\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{1} = \frac{1}{20}$$

Entonces la probabilidad pedida es $\frac{1}{20}+\frac{1}{20}=\frac{1}{10}$

En el ejemplo inicial de extraer dos bolillas de una urna, todavía queda por resolver cómo calculamos la $P(B)$, siendo *A*: "la primer bolilla extraída es blanca", y *B*: "la segunda bolilla extraída es blanca"

Podemos expresar al espacio muestral *S* como $S = A \cup A^C$
Además *A* y $A^C$ son mutuamente excluyentes.
Por otro lado podemos escribir
$B = B \cap S = B \cap \left(A \cup A^C\right) = \left(B \cap A\right) \cup \left(B \cap A^C\right)$ por la ley distributiva
Entonces
$P(B) = P(B \cap A) + P(B \cap A^C) = P(B/A)P(A) + P(B/A^C)P(A^C)$

$\left(B \cap A\right) \cap \left(B \cap A^C\right) = \emptyset$ teorema de la multiplicación

Lo hecho en este ejemplo se generaliza en el siguiente teorema

## 2.2 - Teorema de la probabilidad total

Sean $A_1, A_2, \dots, A_n$ eventos de un espacio muestral *S* que cumplen:

a) $A_1 \cup A_2 \cup \dots \cup A_n = S$

b) $A_i \cap A_j = \emptyset$ si $i \neq j$

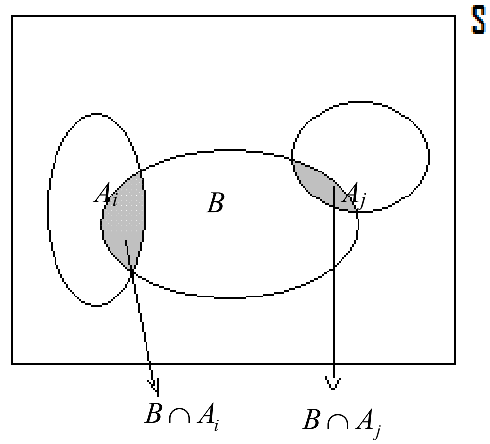
c) $P(A_i) > 0 \quad \forall i = 1,2,\dots,n$ Se dice que $A_1, A_2, \dots, A_n$ forman una ***partición de S*** Entonces para cualquier evento *B* de *S*

$$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + \dots + P(B/A_n)P(A_n)$$

Dem.) Podemos escribir

$$B = B \cap S = B \cap (A_1 \cup A_2 \cup \ldots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$$

Ley distributiva de la $\cap$ con respecto a la $\cup$




Además si $A_i \cap A_j = \emptyset$ con $i \neq j$, entonces $(B \cap A_i) \cap (B \cap A_j) = \emptyset$ si $i \neq j$ (ver figura)

Por lo tanto $P(B) = P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_n) =$

Teorema de la multiplicación

$$= P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + \ldots + P(B/A_n)P(A_n)$$

**<u>Teorema de Bayes</u>**

Sean $A_1, A_2, \ldots, A_n$ eventos de un espacio muestral $S$ que cumplen:

a) $A_1 \cup A_2 \cup \ldots \cup A_n = S$

b) $A_i \cap A_j = \emptyset$ si $i \neq j$

c) $P(A_i) > 0 \quad \forall i = 1, 2, \ldots, n$

Entonces para cualquier evento $B$ de $S$ tal que $P(B) > 0$

$$P(A_k / B) = \frac{P(B/A_k)P(A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)} \qquad k = 1, \ldots, n$$

Dem.)

$$P(A_k / B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B/A_k)P(A_k)}{P(B)} = \frac{P(B/A_k)P(A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$

def. de prob. Condicional teor, de la probabilidad total

teor. de multiplicación

<u>Ejemplos:</u>

1- Tres máquinas A, B, y C producen respectivamente 60%, 30% y 10% del número total de artículos de una fábrica. Los porcentajes de desperfectos de producción de estas máquinas son respectivamente 2%, 3% y 4%. Se selecciona un artículo al azar

a) ¿Cuál es la probabilidad de que sea defectuoso?

b) Si al seleccionar un artículo al azar resulta defectuoso, ¿cuál es la probabilidad de que el artículo hubiera sido producido por la máquina C?

<u>Solución</u>:

a) Sean los eventos

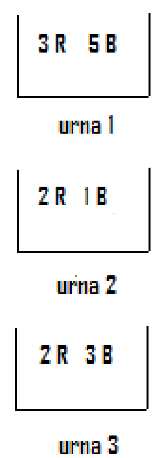
*A:* “el artículo seleccionado fue producido por la máquina A”
*B*: “el artículo seleccionado fue producido por la máquina B”
*C*: “el artículo seleccionado fue producido por la máquina C”
D: “el artículo seleccionado es defectuoso”

Los datos que tenemos son los siguientes
$P(A)=0.6 \quad P(B)=0.3 \quad P(C)=0.1 \quad P(D/A)=0.02 \quad P(D/B)=0.03 \quad P(D/C)=0.04$
Se pide hallar la $P(D)$.
Se aplica el teorema de la probabilidad total tomando como partición de *S* a los eventos *A*, *B* y *C*.
Entonces
$$P(D)=P(D/A)P(A)+P(D/B)P(B)+P(D/C)P(C)=0.02\times 0.6+0.03\times 0.3+0.04\times 0.1$$

b) Se pide hallar $P(C/D)$. Aplicamos el teorema de Bayes

$$P(C/D)=\frac{P(C\cap D)}{P(D)}=\frac{P(D/C)P(C)}{P(D)}=\frac{0.04\times 0.1}{0.02\times 0.6+0.03\times 0.3+0.04\times 0.1}=\frac{4}{25}$$

2- Se nos dan tres urnas como sigue:
Una urna 1 contiene 3 bolas rojas y 5 blancas.
Una urna 2 contiene 2 bolas rojas y 1 blanca.
Una urna 3 contiene 2 bolas rojas y 3 blancas
Se selecciona una urna al azar y se saca una bola de la urna. Si la bola es roja, ¿cuál es la probabilidad de que proceda de la urna 1?




Solución:
Sean los eventos $A_i$: “se elige la urna *i*” $i = 1, 2, 3$
Entonces $P(A_i)=1/3 \quad i=1,2,3$
Además podemos tomar a $A_1, A_2, A_3$ como una partición del espacio muestral *S*.
Sea el evento $B$ :"extraer bolilla roja"
Se pide calcular $P(A_1/B)$
Entonces

$$P(A_1/B)=\frac{P(A_1\cap B)}{P(B)}=\frac{P(B/A_1)P(A_1)}{P(B/A_1)P(A_1)+P(B/A_2)P(A_2)+P(B/A_3)P(A_3)}=$$

def. de prob. condicional

Teorema de Bayes

$$=\frac{\frac{3}{8}\times\frac{1}{3}}{\frac{3}{8}\times\frac{1}{3}+\frac{2}{3}\times\frac{1}{3}+\frac{2}{5}\times\frac{1}{3}}=\frac{\frac{1}{8}}{\frac{173}{360}}=\frac{45}{173}$$

## 2.3 - Independencia

Dados dos eventos $A$ y $B$ , puede ocurrir que $P(B/A)$ y $P(B)$ sean diferentes, eso significa que saber que $A$ ocurrió modifica la probabilidad de ocurrencia de $B$

En el ejemplo anterior $P(B/A_1)=\frac{3}{8}\neq P(B)=\frac{173}{360}$

Pero puede suceder que $P(B/A)$ y $P(B)$ sean iguales, en ese caso $A$ y $B$ son eventos independientes, saber que $A$ ocurrió no afecta la probabilidad de ocurrencia de $B$.

Entonces, dos eventos $A$ y $B$ son ***independientes*** si $P(B/A)=P(B)$, y son ***dependientes*** de otro modo.

Notar que por el teorema de la multiplicación $P(A\cap B)=P(B/A)P(A)$ si $P(A)>0$
Entonces
$A$ y $B$ son independientes $\Rightarrow P(A\cap B)=P(B/A)P(A)=P(B)P(A)$
Recíprocamente
Si $P(A\cap B)=P(B)P(A)$ entonces

$$P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{P(B)P(A)}{P(A)}=P(B) \qquad \therefore \qquad A \text{ y } B \text{ son independientes}$$

si $P(A)>0$

Por lo tanto:

$$A \text{ y } B \text{ son independientes si y solo si } P(A\cap B)=P(A)P(B) \qquad (9)$$

***Es decir podemos usar (9) como definición de independencia de eventos.***

Ejemplos:

1-Se tira un dado normal dos veces, sean los eventos
$A$: “la suma de los números obtenidos es igual a 7”
$B$: “el primer número obtenido es 4”
¿Son $A$ y $B$ independientes?

Sabemos que el espacio muestral es el conjunto de 36 pares ordenados $(a,b)$ donde tanto $a$ como $b$ pueden tomar los valores 1, 2, 3, 4, 5, 6.
Además $A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ y $B=\{(4,1);(4,2);(4,3);(4,4);(4,5);(4,6)\}$
Entonces

$$P(A\cap B)=\frac{1}{36} \qquad P(A)=\frac{6}{36}=\frac{1}{6} \qquad P(B)=\frac{6}{36}=\frac{1}{6}$$

Como $P(A\cap B)=\frac{1}{36}=P(A)P(B)=\frac{1}{6}\times\frac{1}{6}$ entonces $A$ y $B$ son independientes

Observación: si $A$ fuera el evento $A$: “la suma de los números obtenidos es igual a 6”, entonces $A$ y $B$ son dependientes

2-Se tiene una urna con 10 bolillas blancas y 5 rojas. Se extraen al azar dos bolillas ***con reemplazo*** de la urna. Entonces los eventos
$A$: “la primer bolilla es blanca”
$B$: “la segunda bolilla es roja”
Son independientes

$$P(A \cap B) = \frac{10 \times 5}{15 \times 15} \qquad P(A) = \frac{10 \times 15}{15 \times 15} = \frac{10}{15} \qquad P(B) = \frac{5 \times 15}{15 \times 15} = \frac{5}{15}$$

Pero si la extracción se hace ***sin reemplazo***, entonces $A$ y $B$ son ***dependientes*** pues

$$P(A \cap B) = P(B / A)P(A) = \frac{5}{14} \times \frac{10}{15} \qquad P(A) = \frac{10}{15}$$

y $P(B) = P(B / A)P(A) + P(B / A^C)P(A^C) = \frac{5}{14} \times \frac{10}{15} + \frac{4}{14} \times \frac{5}{15} = \frac{5}{15}$

por lo tanto $P(A \cap B) \neq P(B)P(A)$

Notar que la diferencia está en que $P(B / A) = \frac{5}{14} \neq P(B) = \frac{5}{15}$

<u>Observación</u>: si en el ejemplo anterior la urna tiene 1000 bolillas blancas y 500 rojas, y se extraen ***sin reemplazo*** dos bolillas al azar entonces

$$P(B / A) = \frac{500}{1499} = 0.3335557... \qquad \text{y} \qquad P(B) = \frac{5}{15} = 0.3333333...$$

O sea que $P(B / A)$ y $P(B)$ son ***casi*** iguales.

Por lo tanto podemos asumir que $A$ y $B$ son independientes, aunque la extracción se haga sin reemplazo.

En la práctica, si $N$ es el tamaño de la población y $n$ el tamaño de la muestra extraída sin reemplazo, si $\frac{n}{N} < 0.05$ entonces podemos operar como si la extracción se hubiera hecho con reemplazo.

Si dos eventos $A$ y $B$ son independientes entonces $A$ y $B^C$ son independientes (10)

*Dem*.) se debe probar que $P(A \cap B^C) = P(A)P(B^C)$

Para esto escribimos

$A = (A \cap B) \cup (A \cap B^C)$

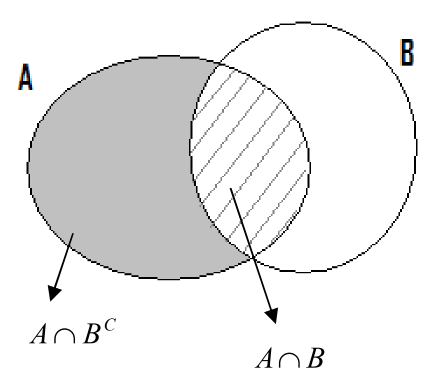
Como $A \cap B$ y $A \cap B^C$ son mutuamente excluyentes entonces

$$P(A) = P(A \cap B) + P(A \cap B^C) \Rightarrow$$

$$\Rightarrow P(A \cap B^C) = P(A) - P(A \cap B) \underset{A \text{ y } B \text{ independientes}}{=} P(A) - P(A)P(B) =$$

$$= P(A)(1 - P(B)) = P(A)P(B^C)$$

Con lo que queda demostrada la propiedad.



Observación: si A y B son eventos independientes, entonces $A^C$ y $B^C$ son independientes. Se llega a este resultado aplicando (10) sobre $A$ y $B$, y luego se aplica (10) nuevamente a $A$ y $B^C$.

**Independencia de más de dos eventos.**

La noción de independencia de eventos se puede ampliar a $n$ eventos de la siguiente manera:

Sean $A_1, A_2, \ldots, A_n$ eventos, se dice que son ***independientes*** si

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k}) \qquad k = 2,\ldots,n \qquad (11)$$

Observaciones:

1- si $n = 2$ entonces (11) se reduce a la definición de dos eventos independientes.
2- si $n = 3$ entonces (11) significa que se deben cumplir las siguientes 4 condiciones:

a) $P(A_1 \cap A_2) = P(A_1)P(A_2)$
b) $P(A_1 \cap A_3) = P(A_1)P(A_3)$
c) $P(A_2 \cap A_3) = P(A_2)P(A_3)$
d) $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

La condición d) significa que un evento es independiente de la intersección de los otros dos, por ejemplo $P(A_1 / A_2 \cap A_3) = P(A_1)$

Esto es porque en general por el teorema de la multiplicación vale que

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2)$$

y por d)

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

entonces

$$P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2) = P(A_1)P(A_2)P(A_3) \Rightarrow P(A_3 / A_1 \cap A_2) = P(A_3)$$

Ejemplos:

1- Las probabilidades de que tres hombres peguen en el blanco son, respectivamente, $\frac{1}{6}$, $\frac{1}{4}$, y $\frac{1}{3}$.

Cada uno dispara una vez al blanco.
a) ¿Cuál es la probabilidad de que exactamente uno de ellos pegue en el blanco?
b) Si solamente uno pega en el blanco, ¿cuál es la probabilidad de que sea el primer hombre?

Solución:
a) consideremos los eventos $A_i$: "el hombre i-ésimo pega en el blanco" $i = 1, 2, 3$

$$P(A_1) = \frac{1}{6} \qquad P(A_2) = \frac{1}{4} \qquad P(A_3) = \frac{1}{3}$$

Sea el evento $B$: "exactamente un hombre pega en el blanco"

Entonces $B = \left(A_1^C \cap A_2^C \cap A_3\right) \cup \left(A_1^C \cap A_2 \cap A_3^C\right) \cup \left(A \cap A_2^C \cap A_3^C\right)$

Por lo tanto $P(B) = P\left(A_1^C \cap A_2^C \cap A_3\right) + P\left(A_1^C \cap A_2 \cap A_3^C\right) + P\left(A \cap A_2^C \cap A_3^C\right)$

Y por independencia $P(B) = P\left(A_1^C)P(A_2^C)P(A_3\right) + P\left(A_1^C)P(A_2)P)(A_3^C\right) + P\left(A_1)P(A_2^C)P)(A_3^C\right) =$

$$=\left(1-\frac{1}{6}\right)\left(1-\frac{1}{4}\right)\frac{1}{3}+\left(1-\frac{1}{6}\right)\frac{1}{4}\left(1-\frac{1}{3}\right)+\frac{1}{6}\left(1-\frac{1}{4}\right)\left(1-\frac{1}{3}\right)=\frac{1}{12}+\frac{5}{36}+\frac{5}{24}=\frac{31}{72}$$

b) Se pide calcular $P(A_1 / B)$

$$P(A_1 / B)=\frac{P(A_1\cap B)}{P(B)}=\frac{P(A_1\cap A_2^C\cap A_3^C)}{P(B)}=\frac{\frac{1}{6}\left(1-\frac{1}{4}\right)\left(1-\frac{1}{3}\right)}{\frac{31}{72}}=\frac{6}{31}$$

2- Cierto tipo de proyectil da en el blanco con probabilidad 0.3, ¿ cuántos proyectiles deberán ser disparados para que haya al menos un 80% de probabilidad de pegar en el blanco?

<u>Solución</u>:
Escribimos $A_i$: "el proyectil i-ésimo da en el blanco" i = 1, 2, …, $n$
Se quiere que

$$P(A_1\cup A_2\cup\ldots\cup A_n)>0.8$$

Asumiendo independencia esto es equivalente a

$$P(A_1\cup A_2\cup\ldots\cup A_n)=1-P\left(\left(A_1\cup A_2\cup\ldots\cup A_n\right)^C\right)=1-P(A_1{}^C\cap A_2{}^C\cap\ldots\cap A_n{}^C)=$$
$$=1-P(A_1{}^C)P(A_2{}^C)\ldots P(A_n{}^C))=1-0.7^n>0.8$$

Por lo tanto $0.7^n<0.2 \Rightarrow n\ln(0.7)<\ln(0.2) \Rightarrow n>\frac{\ln(0.2)}{\ln(0.7)}=4.5123$

Es decir se deben hacer por lo menos 5 disparos

# 3 - VARIABLES ALEATORIAS

## 3.1- Generalidades

En muchas situaciones experimentales se quiere asignar un número real a cada uno de los elementos del espacio muestral. Al describir el espacio muestral de un experimento un resultado individual no tiene que ser necesariamente un número, por ejemplo, al tirar una moneda y tomar como espacio muestral $S = \{c, s\}$ , o al tirar un dado dos veces tomamos como espacio muestral a $S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$ , aquí $S$ es un conjunto de pares ordenados.

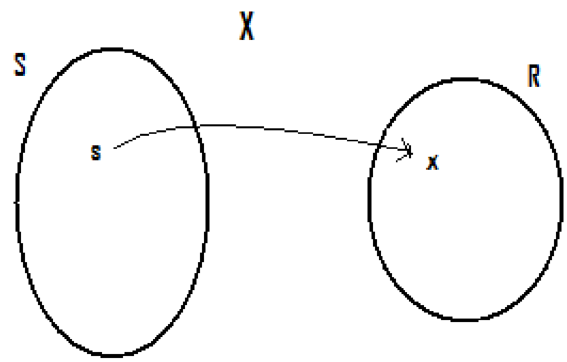
Definición: Sea $\varepsilon$ un experimento aleatorio y $S$ un espacio muestral asociado a él. Una ***variable aleatoria*** es una función que asigna a cada elemento de $S$ un número real.

Notación: se anota a una variable aleatoria con letras mayúsculas *X, Y, Z, W,…*

Entonces, si $X$ es una variable aleatoria de $S$ en $R$

$$X : S \to R \quad \text{tal que} \quad X(s) = x$$

Con diagramas de Venn




*Desde ahora en lugar de escribir variable aleatoria, escribiremos v.a.*

Ejemplos:

1- Se tira una moneda tres veces

Sea $X$ la v.a. X: "número de caras obtenidas luego de los tres tiros"

Si tomamos como espacio muestral

$$S = \{(c,c,c);(c,c,s);(c,s,c);(s,c,c);(c,s,s);(s,s,c);(s,c,s);(s,s,s)\}$$

entonces

$$X((c,c,c)) = 3$$

$$X((c,c,s)) = X((s,c,c)) = X((c,s,c)) = 2$$

$$X((c,s,s)) = X((s,c,s)) = X((s,s,c)) = 1$$

$$X((s,s,s)) = 0$$

La imagen de esta función es el conjunto $\{0,1,2,3\}$

Dada una v.a. X a su imagen se la anota $R_X$ y se la denomina ***rango o recorrido de X***

En el ejemplo anterior $R_X = \{0,1,2,3\}$

2- Se tira un dado tantas veces como sean necesarias hasta que sale el número 1 por primera vez.

Podemos simbolizar el espacio muestral de la siguiente manera $S = \{1,\ 01,\ 001,\ 0001,\ \ldots\}$, por ejemplo 001 simboliza el resultado que en los dos primeros tiros no salió el número 1 y en el tercer tiro salió el 1.
Sea $Y$ la v.a.:
$Y$: "número de tiros necesarios hasta que sale el 1 por primera vez"
Entonces $R_Y = \{1,2,3,4,\ldots\}$, es decir el rango de Y es el conjunto de los números naturales.

3- En el interior de un círculo de radio $r$ y centro el origen de coordenadas, se elige un punto al azar.
Tomamos como espacio muestral a $S = \{(x,y),\ \ x^2 + y^2 \le r^2\}$. Aquí $S$ es infinito no numerable
Definimos la v.a. $Z$: "distancia del punto elegido al origen"
Entonces $R_Z = \{z;\ \ 0 \le z \le r\}$

Las variables aleatorias se clasifican según su rango.
Sea X es una v.a. con rango $R_X$. Si $R_X$ es un conjunto ***finito o infinito numerable*** entonces se dice que $X$ es una ***v.a. discreta***. Si $R_X$ es un conjunto ***infinito no numerable*** entonces $X$ es una ***v.a. continua***.

El rango $R_X$ es considerado un nuevo ***espacio muestral***, y sus subconjuntos son ***eventos***.
Por ejemplo:
En el ejemplo 1, los eventos unitarios o elementales son $\{0\};\ \{1\};\ \{2\};\ \{3\}$, ***pero los anotamos*** $\{X = 0\};\ \{X = 1\};\ \{X = 2\};\ \{X = 3\}$
Otros eventos son, por ejemplo:
$\{X \le 1\}$, es decir, salió a lo sumo una cara.
Notar que podemos escribir $\{X \le 1\} = \{X = 0\} \cup \{X = 1\}$, o sea escribimos al evento como ***unión de eventos elementales***
$\{X > 0\}$, es decir, salieron una o más caras. Tenemos que $\{X > 0\} = R_X - \{X = 0\}$
En el ejemplo 2, $\{Y \ge 4\}$ sería el evento "al menos 4 tiros son necesarios para que salga por primera vez el numero 1"
$\{4 \le Y \le 6\}$ sería el evento "se necesitan entre 4 y 6 tiros para que salga el 1 por primera vez"
Notar que $\{4 \le Y \le 6\} = \{Y = 4\} \cup \{Y = 5\} \cup \{Y = 6\}$
En el ejemplo 3, $\left\{\frac{1}{3}r < Z < \frac{2}{3}r\right\}$ sería el evento "el punto elegido se encuentra a una distancia del centro mayor que $\frac{1}{3}r$, pero menor que $\frac{2}{3}r$

Volviendo al ejemplo 1, notar que $B = \{X = 0\}$ ocurre en $R_X$ si y solo si el evento $A = \{(s,s,s)\}$ ocurre en $S$. Se dice que A y B son ***eventos equivalentes***.
De la misma forma los eventos
$A = \{(c,c,c)\}$ y $B = \{X = 3\}$ son equivalentes
$A = \{(c,c,s);(c,s,c);(s,c,c)\}$ y $B = \{X = 2\}$ son equivalentes
En general

siendo $A \subset S$ y $B \subset R_X$, $A$ y $B$ ***son equivalentes*** si $A = \{s \in S;\ \ X(s) \in B\}$

Si $X$ es una v.a. de $S$ en $R$, y $R_X$ es el rango de $X$, para calcular la probabilidad de un evento $B$ de $R_X$ se busca el evento $A$ en $S$ equivalente a $B$ y entonces $P(B)=P(A)$

Por ejemplo,

En el ejemplo 1, $P(B)=P(X=0)=P(A)=P(\{(s,s,s)\})=\dfrac{1}{8}$ si la moneda es normal.

$P(B)=P(X\le 1)=P(\{(s,s,s);(c,s,s);(s,c,s);(s,s,c)\})=\dfrac{4}{8}=0.5$ si la moneda es normal

También podríamos haber planteado

$$P(B)=P(X\le 1)=P(X=0)+P(X=1)=\frac{1}{8}+\frac{3}{8}=0.5$$

En el ejemplo 3, si $B=\left\{\frac{1}{3}r\le Z\le \frac{2}{3}r\right\}$, entonces $B$ es equivalente a

$A=\left\{(x,y);\ \frac{1}{3}r\le\sqrt{x^2+y^2}\le\frac{2}{3}r\right\}$, por lo tanto

$$P(B)=P(A)=\frac{\text{area de }A}{\text{area de }S}=\frac{\pi\left(\frac{2}{3}r\right)^2-\pi\left(\frac{1}{3}r\right)^2}{\pi r^2}=\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2=\frac{3}{9}=\frac{1}{3}$$

<u>Observación</u>: en este ejemplo si $B=\left\{Z=\frac{2}{3}r\right\}$, entonces $P(B)=P(A)=\dfrac{\text{area de }A}{\text{area de }S}=\dfrac{0}{\pi r^2}=0$

### 3.2 - <u>Variables aleatorias discretas</u>

Sea $X$ una v.a. discreta. Anotamos su rango como $R_X=\{x_1,x_2,\dots,x_n\}$ si el rango es un conjunto finito de $n$ elementos, y anotamos $R_X=\{x_1,x_2,\dots\dots\}$ si el rango es un conjunto infinito numerable.

A cada $x_i$ se le asigna un número $p(x_i)=P(X=x_i)$. Estos números deben satisfacer las condiciones siguientes

a) $p(x_i)\ge 0$ para todo $i$

b) $\sum_i p(x_i)=1$

La función $p(x)$ que antes se definió, se llama ***función de probabilidad o de frecuencia de la v.a. X***.
El conjunto de pares $(x_i,p(x_i))$ $i=1,2,\dots$ es la ***distribución de probabilidad de X.***

Por ejemplo

1-Se tira una moneda normal tres veces, sea la v.a. $X$: "número de caras obtenidas"
Entonces $R_X=\{0,1,2,3\}$

Para hallar la distribución de probabilidad de $X$ supongamos que la probabilidad de salir cara es ½ entonces

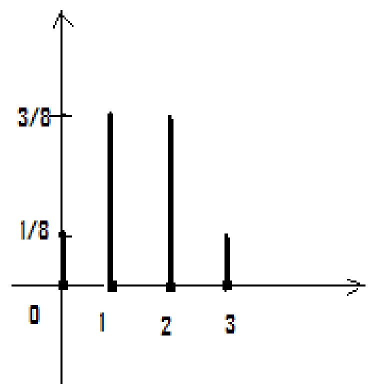
$$P(X=0)=P(\{(s,s,s)\})=\frac{1}{8}$$

$$P(X=1)=P(\{(c,s,s);(s,c,s);(s,s,c)\})=\frac{3}{8}$$
$$P(X=2)=P(\{(c,c,s);(s,c,c);(c,s,c)\})=\frac{3}{8}$$
$$P(X=3)=P(\{(c,c,c)\})=\frac{1}{8}$$

Se puede presentar la distribución de probabilidad de *X* en una tabla de la siguiente forma

| *x* | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| *p(x)* | 1/8 | 3/8 | 3/8 | 1/8 |

Un gráfico de la distribución de probabilidad de *X* sería




2-Se tira un dado normal. Sea X: "número que queda en la cara superior" Entonces $R_X=\{1,2,3,4,5,6\}$
La función de distribución de *X* es

| *x* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *p(x)* | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

<u>Observación</u>:

Sea X una v.a. discreta con rango finito $R_X=\{x_1,x_2,\dots,x_n\}$, donde cada $x_i$ es un número entero y $x_{i+1}$ es el consecutivo de $x_i$ .

Si $P(X=x_i)=\frac{1}{n}$ para cada $i$ entonces se dice que X tiene ***distribución uniforme discreta***.

Por ejemplo podría ser $R_X=\{1,2,3,\dots,n-1,n\}$, en este caso X es uniforme discreta en el intervalo natural $[1,n]$.

La v.a. del ejemplo 2 anterior es uniforme discreta en el intervalo natural $[1,6]$

## Función de distribución acumulada

Sea X una v.a. con rango $R_X$ . Se define la ***función de distribución acumulada de X*** (abreviamos F.d.a de *X*) como

$$F(x)=P(X\le x) \qquad -\infty<x<\infty \tag{12}$$

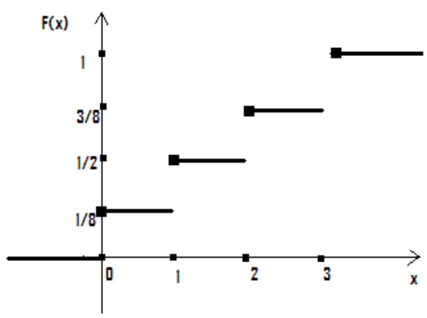
En el caso de ser $X$ una v.a. discreta

$$F(x) = P(X \le x) = \sum_{x_i \le x} p(x_i) \qquad -\infty < x < \infty$$

Volviendo al ejemplo 1 anterior, la F.d.a. de $X$ es

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{8} & \text{si } 0 \le x < 1 \\ \frac{1}{8} + \frac{3}{8} & \text{si } 1 \le x < 2 \\ \frac{1}{8} + \frac{3}{8} + \frac{3}{8} & \text{si } 2 \le x < 3 \\ 1 & \text{si } x > 3 \end{cases} \quad \Rightarrow \quad F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{8} & \text{si } 0 \le x < 1 \\ \frac{1}{2} & \text{si } 1 \le x < 2 \\ \frac{7}{8} & \text{si } 2 \le x < 3 \\ 1 & \text{si } x > 3 \end{cases}$$

La gráfica de la F.d.a. de $X$ es




Observación: la F.d.a. de $X$ es una función ***escalonada***, los puntos de "salto" coinciden con los puntos del rango de $X$, y la magnitud del salto en $x_i$ es igual a $P(X = x_i)$

***En general si X es una v.a. discreta cualquiera, su F.d.a. será una función escalonada.***

Además si $x_1, x_2, \dots, x_n$ son los valores del rango de $X$ ordenados de menor a mayor entonces

$$P(X = x_1) = F(x_1)$$
$$P(X = x_i) = F(x_i) - F(x_{i-1}) \qquad i = 2, \dots, n$$

***Es decir, se puede obtener la función de distribución de X a partir de su F.d.a.***

Para números cualesquiera $a$ y $b$

1- Si $a \le b$ entonces $P(a \le X \le b) = P(X \le b) - P(X < a)$

2- Si $a < b$ entonces $P(a < X \le b) = P(X \le b) - P(X \le a)$

3- Si $a < b$ entonces $F(a) \le F(b)$ (es decir $F(x)$ es una función creciente)

Además se cumple que

$$1- \lim_{x\to\infty} F(x) = \lim_{x\to\infty} \sum_{x_i \le x} P(X = x_i) = \sum_{x_i} P(X = x_i) = 1$$

$$2- \lim_{x\to -\infty} F(x) = \lim_{x\to -\infty} \sum_{x_i \le x} P(X = x_i) = \sum_{x_i \le -\infty} P(X = x_i) = 0$$

## 3.3 – Esperanza de una variable aleatoria discreta

Sea $X$ una v.a. discreta con rango $R_X$ . La ***esperanza , valor medio o valor esperado de X*** , lo anotamos $E(X)$ , y se define como

$$E(X) = \sum_{x_i \in R_X} x_i P(X = x_i)$$

La sumatoria se hace sobre todos los posibles valores de $X$
Otra notación usual es $\mu_X$ o $\mu$

Ejemplos:
1- Sea la v.a. X: “número que queda en la cara de arriba al tirar un dado normal” $R_X = \{1,2,3,4,5,6\}$

Entonces $E(X) = \sum_{x=1}^{6} xP(X = x) =$

$$= 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + 4P(X = 4) + 5P(X = 5) + 6P(X = 6) =$$

$$= 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = \frac{7}{2} = 3.5$$

2- Se tira una moneda normal tres veces, sea la v.a. $X$: “número de caras obtenidas”

Entonces $R_X = \{0,1,2,3\}$

Calculamos la esperanza de $X$

$$E(X) = \sum_{x=0}^{3} xP(X = x) = 0\times\frac{1}{8} + 1\times\frac{3}{8} + 2\times\frac{3}{8} + 3\times\frac{1}{8} = \frac{3}{2} = 1.5$$

Observaciones:
1- La esperanza de una v.a. no tiene que coincidir necesariamente con algún valor del rango de la variable
2- En el ejemplo 1 donde el rango es finito y equiprobable, la esperanza de X coincide con el ***promedio de los valores del rango de X***
3- Se puede interpretar a la esperanza de una v.a. como un ***promedio “pesado” o “ponderado” de los valores del rango de la variable***, donde el “peso” de cada $x_i$ es la probabilidad $P(X = x_i)$
4- Otra interpretación que se puede hacer de la esperanza es la siguiente: consideremos el ejemplo1, supongamos que tiramos el dado muchas veces, $N$ veces, y entonces obtenemos una secuencia de $N$ valores $x_1, x_2, ....., x_N$ donde cada $x_i$ es un número natural del 1 al 6. Supongamos además que hacemos un promedio de esos $N$ valores, y si llamamos $n_i$ al número de veces que sale el número $i$ tenemos que

$$\frac{x_1 + x_2 + \dots + x_N}{N} = \frac{n_1 1 + n_2 2 + \dots + n_6 6}{N} =$$

$$= \frac{n_1}{N}\times 1 + \frac{n_2}{N}\times 2 + \dots + \frac{n_6}{N}\times 6 \approx 1\times P(X=1) + 2\times P(X=2) + \dots + 6\times P(X=6) = E(X)$$

Es decir si promediamos los *N* valores medidos de *X*, ese promedio tiende a *E(X)* cuando $N \to \infty$, pues $\frac{n_i}{N} \approx P(X=i)$ cuando *N* es grande.

**Esperanza de una función**

A veces importa hallar la esperanza de una ***función de X*** y no de *X* misma. Veamos un ejemplo.
Un instructor de escritura técnica ha solicitado que cierto reporte sea entregado a la semana siguiente, agregando la restricción de que cualquier reporte que sobrepase las cuatro páginas será rechazado.
Sea *X*: "número de páginas del reporte de cierto estudiante seleccionado al azar"
Supongamos que *X* tenga la siguiente distribución de probabilidad

| *x* | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *p(x)* | 0.01 | 0.19 | 0.35 | 0.45 |

Suponga que el instructor tarda $\sqrt{X}$ minutos calificando un trabajo que consiste en *X* páginas. Claramente $\sqrt{X}$ es ***otra variable aleatoria***. ¿Cuál será su esperanza?, es decir ¿a qué es igual $E\left(\sqrt{X}\right)$?
Para calcular la esperanza de una v.a. se necesita conocer su función de distribución de probabilidad, por lo tanto habría que hallar previamente la distribución de probabilidad de la v.a. $Y = \sqrt{X}$ .
Está claro que si el rango de *X* es $R_X = \{1,2,3,4\}$ entonces el rango de *Y* será $R_Y = \{\sqrt{1}, \sqrt{2}, \sqrt{3}, \sqrt{4}\}$.
Además

$$P(Y=\sqrt{1}) = P(X=1) = 0.01$$

$$P(Y=\sqrt{2}) = P(X=2) = 0.19$$

$$P(Y=\sqrt{3}) = P(X=3) = 0.35$$

$$P(Y=\sqrt{4}) = P(X=4) = 0.45$$

Por lo tanto

$$E(Y) = \sqrt{1}\times P(Y=\sqrt{1}) + \sqrt{2}\times P(Y=\sqrt{2}) + \sqrt{3}\times P(Y=\sqrt{3}) + \sqrt{4}\times P(Y=\sqrt{4}) =$$

$$= \sqrt{1}\times P(X=1) + \sqrt{2}\times P(X=2) + \sqrt{3}\times P(X=3) + \sqrt{4}\times P(X=4) = 1.78491$$

O sea

$$E(Y) = \sum_x \sqrt{x} P(X=x)$$

Lo visto en este ejemplo se puede generalizar en el siguiente

Teorema: Si *X* es una v.a. discreta con rango $R_X$ y distribución de probabilidad *p(x)*, entonces la esperanza de cualquier función *h(X)* es igual a

$$E(h(X)) = \sum_x h(x)p(x)$$

Ejemplo:
Un negocio de computadoras ha comprado tres computadoras de cierto tipo a \$500 cada una y las ven-

derá a \$1000 cada una. El fabricante ha aceptado volver a comprar en \$200 cualquier computadora que no se haya vendido en un tiempo especificado.
Sea $X$: "número de computadoras vendidas" , y supongamos que la distribución de probabilidad de $X$ es

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p(x)$ | 0.1 | 0.2 | 0.3 | 0.4 |

Si consideramos la v.a. $Y$: "utilidad obtenida", entonces $Y$ es una función de $X$, es decir $Y = h(X)$
Específicamente

$$Y = 1000X + 200(3 - X) - 1500 = 800X - 900$$

La utilidad esperada, es decir la $E(Y)$ será

$$E(Y) = \sum_{x=0}^{3} (800x - 900)P(X = x) =$$

$$= (800\times 0 - 900)P(X = 0)(800\times 1 - 900)P(X = 1) + (800\times 2 - 900)P(X = 2) + (800\times 3 - 900)P(X = 3) =$$

$$= (-900)\times 0.1 + (-100)\times 0.2 + 700\times 0.3 + 1500\times 0.4 = \$700$$

Notar que aplicando propiedades de la notación $\Sigma$ se puede plantear

$$E(Y) = \sum_{x=0}^{3} (800x - 900)P(X = x) = 800\sum_{x=0}^{3} xP(X = x) - 900\sum_{x=0}^{3} P(X = x) = 800E(X) - 900$$

y calculando la esperanza de $X$ , se llega al mismo resultado

**<u>Propiedades de la esperanza</u>**

En el ejemplo anterior tenemos que ***Y es una función lineal de X*** , es decir $Y = aX + b$ con $a$ y $b$ números reales.
En este caso vale entonces la siguiente propiedad

$$E(aX + b) = aE(X) + b$$

La demostración sigue los mismos pasos que en el ejemplo anterior

$$E(aX + b) = \sum_{x} (ax + b)P(X = x) = a\underbrace{\boxed{\sum_{x} xP(X = x)}}_{= E(X)} + b\underbrace{\boxed{\sum_{x} xP(X = x)}}_{= 1} = aE(X) + b$$

<u>Ejemplo</u>:
En el ejemplo anterior donde $Y = 800X - 900$
Directamente calculamos

$$E(Y) = 800E(X) - 900$$

Y

$$E(X) = 0\times 0.1 + 1\times 0.2 + 2\times 0.3 + 3\times 0.4 = 2$$

En consecuencia

$$E(Y) = 800E(X) - 900 = 800\times 2 - 900 = 700$$

Observaciones:
1- Para cualquier constante $a$, $E(aX) = aE(X)$
2- Para cualquier constante $b$, $E(X+b) = E(X)+b$

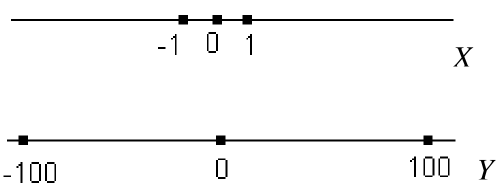
**Varianza de una variable aleatoria**

La esperanza de una v.a. mide dónde está centrada la distribución de probabilidad. Pero supongamos el siguiente ejemplo
Sean X e Y dos variables aleatorias con distribuciones dadas por

| *x* | -1 | 1 |
|---|---|---|
| *p(x)* | 0.5 | 0.5 |

| *y* | -100 | 100 |
|---|---|---|
| *p(y)* | 0.5 | 0.5 |

Es fácil verificar que $E(X) = E(Y) = 0$, pero los valores que toma la v.a. *Y* están más "alejados" de su esperanza que los valores de *X*.




Se busca una medida que refleje este hecho, se define entonces la varianza de una v.a.

Sea X una v.a. discreta con rango $R_X$, función de distribución de probabilidad *p(x)* y esperanza $E(X) = \mu$,
Entonces la ***varianza de X***, que anotamos $V(X)$, $\sigma^2$ o $\sigma_X^2$ es

$$V(X) = E\left[(X-\mu)^2\right] = \sum_{x \in R_X} (x-\mu)^2\, p(x)$$

La ***desviación estándar de X*** es $\sigma_X = \sqrt{V(X)}$

Observaciones:
1- La varianza de una v.a. nunca es negativa
2- La cantidad $h(X) = (X-\mu)^2$ es el cuadrado de la desviación de *X* desde su media, y la varianza de *X* es la esperanza de la desviación al cuadrado. Si la mayor parte de la distribución de probabilidad está cerca de $\mu$, entonces $\sigma^2$ será relativamente pequeña. Si hay valores de la variable alejados de $\mu$ que tengan alta probabilidad, entonces $\sigma^2$ será grande.
3- $\sigma^2$ está expresado en las unidades de medida de *X* al cuadrado, mientras que $\sigma$ está expresada en las mismas unidades de medida que *X*.

Ejemplo:
En el caso de las variables aleatorias *X* e *Y* nombradas anteriormente,
$V(X) = (-1-0)^2 \times 0.5 + (1-0)^2 \times 0.5 = 1$ y $\sigma_X = 1$

$V(Y) = (-100-0)^2 \times 0.5 + (100-0)^2 \times 0.5 = 100^2$ y $\sigma_Y = 100$

Otra forma de escribir la varianza de una v.a., que facilita los cálculos es

$$V(X) = \sum_{x \in R_X} (x-\mu)^2 p(x) = \sum_{x \in R_X} (x^2 - 2\mu x + \mu^2) p(x) = \sum_{x \in R_X} x^2 p(x) - 2\mu \sum_{x \in R_X} x p(x) + \mu^2 \sum_{x \in R_X} p(x) =$$

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

Por lo tanto

$$V(X) = E(X^2) - \mu^2$$

**Propiedades de la varianza**

Las propiedades de la varianza de una v.a. son consecuencia de las propiedades de la esperanza de una v.a.

Si *X* es una v.a. discreta con rango $R_X$ y distribución de probabilidad *p(x)*, entonces la varianza de cualquier función *h(X)* es igual a

$$V(h(X)) = \sum_{x \in R_X} \left(h(x) - E(h(X))\right)^2 p(x)$$

Si *h(X)* es una función lineal, entonces

$$V(aX+b) = a^2 V(X) \quad \text{y} \quad \sigma_{aX+b} = \sqrt{V(aX+b)} = |a|\sigma_X$$

Observaciones:

1- $V(aX) = a^2 V(X)$

2- $V(X+b) = V(X)$

Ejemplo:

En un ejemplo anterior donde *X*: "número de computadoras vendidas" y *Y*: "utilidad obtenida", la *V(Y)* sería $V(Y) = 800^2 V(X)$

Necesitamos calcular *V(X)*

$V(X) = E(X^2) - \mu^2$

Sabemos ya que $\mu = E(X) = 2$

Calculamos $E(X^2) = 0^2 \times 0.1 + 1^2 \times 0.2 + 2^2 \times 0.3 + 3^2 \times 0.4 = 5$

En consecuencia

$V(Y) = 800^2 V(X) = 800^2 \left(E(X^2) - \mu^2\right) = 800^2 \left(5 - 2^2\right) = 800$

## 3.4- Variables aleatorias discretas importantes

### *Distribución binomial*

Sea $\varepsilon$ un experimento aleatorio. Sea A un evento asociado a $\varepsilon$ y anotamos $P(A) = p$.

Supongamos un experimento aleatorio $\varepsilon_0$ que cumple los siguientes requisitos:

1- se realizan ***n*** repeticiones ***independientes*** de $\varepsilon$, donde $n$ se fija de antemano.
2- las repeticiones son idénticas, y en cada repetición de $\varepsilon$ observamos ***si ocurre A*** o ***no ocurre A*** (cuando *A* ocurre se dice que se obtuvo un "éxito", caso contrario se obtuvo un "fracaso")
3- la probabilidad de éxito es constante de una repetición a otra de $\varepsilon$, y es igual a *p*

Se dice entonces que $\varepsilon_0$ es un ***experimento binomial***

Ejemplos:

1- Se tira una moneda 4 veces en forma sucesiva e independiente, y observamos en cada tiro si sale cara o no sale cara.
Entonces este es un experimento binomial pues:
$\varepsilon$ sería el experimento "tirar una moneda"
*A* sería el evento "sale cara"
$\varepsilon$ se repite en forma sucesiva e independiente $n = 4$ veces
$P(A) = p$ es la misma en cada tiro.

2- Se tiene una urna con 15 bolillas blancas y 5 verdes. Se extraen al azar ***con reemplazo*** tres bolillas y se observa si la bolilla extraída es blanca.
Entonces este es un experimento binomial pues:
$\varepsilon$ sería el experimento "extraer al azar una bolilla de la urna"
*A* sería el evento "se extrae bolilla blanca"
$\varepsilon$ se repite en forma sucesiva e independiente $n = 3$ veces
$P(A) = \frac{15}{20} = \frac{3}{4}$ es la misma en cada extracción.

3- Si en el ejemplo anterior se extraen las bolillas ***sin reemplazo*** entonces el experimento no es binomial, pues ***falla la independencia***:

Si anotamos $A_i$ :"se extrae bolilla blanca en la $i-$ésima extracción" , entonces

$$P(A_1) = \frac{15}{20} \quad ; \quad P(A_2) = P(A_2 / A_1)P(A_1) + P(A_2 / A_1^{C})P(A_1^{C}) = \frac{14}{19}\times\frac{15}{20} + \frac{15}{19}\times\frac{5}{20} = \frac{15}{20}$$

Pero $P(A_2) = \frac{15}{20} \neq P(A_2 / A_1) = \frac{14}{19}$ por lo tanto las extracciones no son independientes

Observación: si en la urna hubiese 1500 bolillas blancas y 500 verdes y se extraen dos bolillas al azar sin reemplazo, entonces

$$P(A_2) = \frac{15}{20} = 0.75 \approx P(A_2 / A_1) = \frac{1499}{1999} = 0.74987$$

Por lo tanto en estas condiciones podemos asumir que el experimento es binomial

### La variable aleatoria binomial y su distribución

En la mayoría de los experimentos binomiales, interesa el número total de éxitos, más que saber exactamente cuáles repeticiones produjeron los éxitos

Sea la v.a. *X*: "número de éxitos en las *n* repeticiones de $\varepsilon$"

Entonces se dice que X es una v.a. binomial

Veamos cuál es la distribución de probabilidad de *X*, para esto primero tomamos un caso concreto: el ejemplo 1 anterior en el que se tira una moneda 4 veces. Supongamos que la probabilidad de cara es ¾

Aquí el rango de *X* sería $R_X = \{0,1,2,3,4\}$
Para facilitar la notación escribimos $A_i$ :"sale cara en el $i-$ésimo tiro" $i = 1,2,3,4$
Por lo tanto

$$P(X=0) = P\left(A_1^C \cap A_2^C \cap A_3^C \cap A_4^C\right) = P(A_1^C)P(A_2^C)P(A_3^C)P(A_4^C) = \frac{1}{4}\times\frac{1}{4}\times\frac{1}{4}\times\frac{1}{4} = \left(\frac{1}{4}\right)^4$$

por independencia

Para calcular la $P(X=1)$ pensamos que hay cuatro casos posibles en los que se puede obtener exactamente una cara, que la cara salga en el 1º tiro, o en el 2º o en el 3º o en el 4º tiro. Notar que tenemos cuatro casos y eso ***es igual a la cantidad de formas en que podemos elegir entre los 4 tiros uno de ellos en el cual sale cara, es decir tenemos*** $\binom{4}{1} = \frac{4!}{1!3!} = 4$ ***casos diferentes.***

$$P(X=1) =$$
$$= P\left(A_1 \cap A_2^C \cap A_3^C \cap A_4^C\right) + P\left(A_1^C \cap A_2 \cap A_3^C \cap A_4^C\right) + P\left(A_1^C \cap A_2^C \cap A_3 \cap A_4^C\right) +$$
$$+ P\left(A_1^C \cap A_2^C \cap A_3^C \cap A_4\right)$$

Cada término es igual a $p(1-p)^3$ por lo tanto $P(X=1) = 4p(1-p)^3$

Análogamente, para calcular $P(X=2)$ tenemos $\binom{4}{2} = \frac{4!}{2!2!} = 6$ casos en los que salen exactamente dos caras, por lo tanto

$$P(X=2) = P\left(A_1 \cap A_2 \cap A_3^C \cap A_4^C\right) + P\left(A_1 \cap A_2^C \cap A_3 \cap A_4^C\right) + \ldots\ldots = \binom{4}{2}p^2(1-p)^2$$

Pensando de la misma forma los otros casos se llega a

$$P(X=3) = \binom{4}{3}p^3(1-p) \quad ; \quad P(X=4) = p^4$$

En general con un argumento análogo tenemos que $R_X = \{0,1,2,\ldots,n\}$ y

$$P(X=k) = \binom{n}{k}p^k 1-p)^{n-k} \qquad k = 0,1,2,...,n$$

<u>Notación</u>: indicamos que *X* es una v.a. binomial ***con parámetros n y p*** con el símbolo $X \sim B(n,p)$

Dado que los números $P(X=k)$ corresponden a la distribución de una v.a., automáticamente cumplen que $\sum_{k=0}^{n} P(X=k) = 1$

De todas formas se podría hacer una verificación algebraica utilizando la fórmula del binomio de Newton

$$\sum_{k=0}^{n} P(X=k)=\sum_{k=0}^{n}\binom{n}{k}p^{k}(1-p)^{n-k}=\left(p+(1-p)\right)^{n}=1$$

<u>Ejemplos</u>:

1- En el ejemplo anterior en el que se tira una moneda 4 veces, calcular la probabilidad de obtener:
   a) exactamente una cara
   b) al menos una cara
   c) a lo sumo una cara

<u>Solución</u>:

a) tenemos que la v.a. X: “número de caras obtenido” es $B(4,0.25)$

se pide $P(X=1)=\binom{4}{1}\left(\frac{1}{4}\right)^{1}\left(1-\frac{1}{4}\right)^{3}=4\times\frac{1}{4}\times\left(\frac{3}{4}\right)^{3}=0.421875$

b) la probabilidad de obtener ***al menos una cara*** es

$$P(X\geq 1)=P(X=1)+P(X=2)+P(X=3)+P(X=4)=\sum_{k=1}^{4}\binom{4}{k}\left(\frac{1}{4}\right)^{k}\left(\frac{3}{4}\right)^{4-k}$$

Pero más fácil es hacer

$$P(X\geq 1)=1-P(X=0)=1-\binom{4}{0}\left(\frac{1}{4}\right)^{0}\left(1-\frac{1}{4}\right)^{4-0}=1-\left(\frac{3}{4}\right)^{4}=1-0.421875=0.578125$$

c) la probabilidad de obtener ***a lo sumo una cara*** es

$$P(X\leq 1)=P(X=0)+P(X=1)=\binom{4}{0}\left(\frac{1}{4}\right)^{0}\left(1-\frac{1}{4}\right)^{4-0}+\binom{4}{1}\left(\frac{1}{4}\right)^{1}\left(1-\frac{1}{4}\right)^{4-1}=0.84375$$

<u>Observación</u>: si $X \sim B(n,p)$ para calcular $P(X\leq k)$ en general se debe hacer

$$P(X\leq k)=\sum_{i=0}^{k}P(X=i)=P(X=0)+P(X=1)+\ldots\ldots+P(X=k)$$

***Notar que*** $P(X\leq k)$ ***es la F.d.a. de X evaluada en k, es decir*** $F(k)=P(X\leq k)$

Existen tablas de la función de distribución acumulada de la binomial para diferentes valores de $n$ y $p$

Consultando estas tablas se puede obtener directamente el resultado del inciso c) buscando para $n=4$ y $p=0.25$

Además consultando las tablas podemos evaluar $P(X=k)$ haciendo

$$P(X=k)=F(k)-F(k-1) \qquad k=1,2,\ldots,n$$

2- Supongamos que el 20% de todos los ejemplares de un texto en particular fallan en una prueba de resistencia a la encuadernación. Se seleccionan 15 ejemplares al azar.
   Sea la v.a. X: “número de ejemplares que fallan en la prueba entre los 15 seleccionados”
   a) ¿cuál es la probabilidad de que a lo sumo 8 fallen en la prueba?
   b) ¿cuál es la probabilidad de que exactamente 8 fallen en la prueba?

c) ¿cuál es la probabilidad de que al menos 8 fallen en la prueba?

Solución:

a) Tenemos que $X \sim B(15{,}0.2)$

$$P(X \leq 8) = \sum_{k=0}^{8} P(X = k) = F(8) = 0.999$$

por tabla de la F.d.a.

b) $P(X = 8) = F(8) - F(7) = 0.999 - 0.996 = 0.003$

por tabla de la F.d.a.

c) $P(X \geq 8) = 1 - P(X \leq 7) = 1 - F(7) = 1 - 0.996 = 0.004$

por tabla de la F.d.a.

Observaciones:

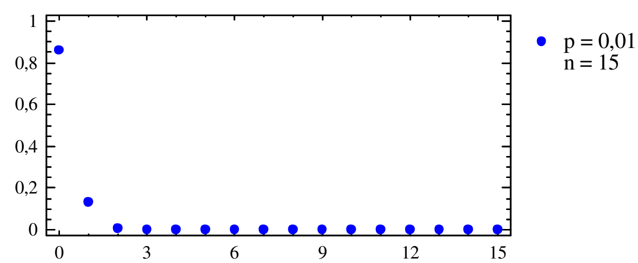
1- Si $X \sim B(1, p)$ entonces la v.a. $X$ toma sólo dos valores 0 y 1 con probabilidades $p$ y *1-p* es decir podemos escribir

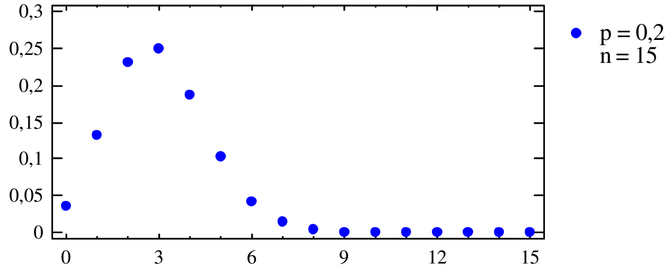
$$X = \begin{cases} 1 & \text{si al ejecutar } \varepsilon \text{ ocurre éxito} \\ 0 & \text{caso contrario} \end{cases} \qquad \begin{array}{l} P(X = 1) = P(A) = p \\ P(X = 0) = P(A^C) = 1 - p \end{array}$$

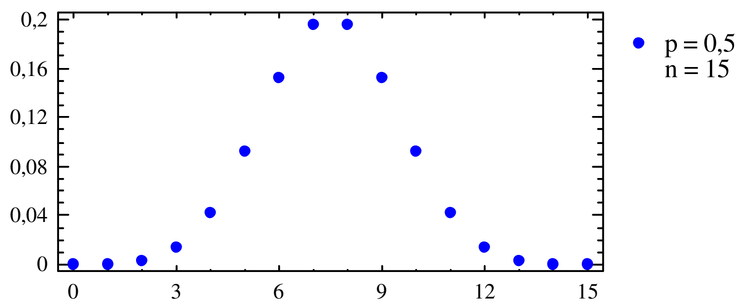
En este caso se dice que $X$ tiene distribución de Bernoulli

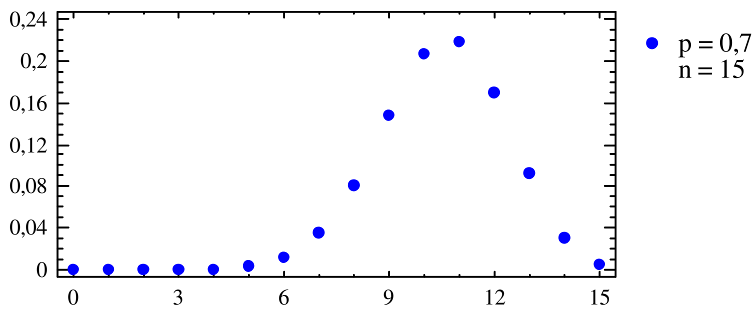
En el caso de ser $X \sim B(n, p)$ se dice que se tienen ***"n ensayos de Bernoulli"***

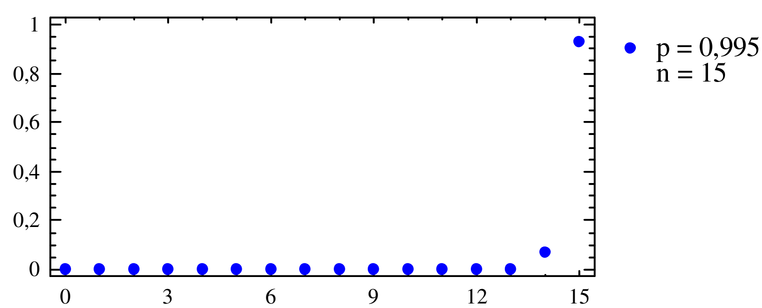
2- A continuación se muestra cómo varía la forma de la distribución a medida que $p$ aumenta manteniendo $n$ fijo en 15. Se grafica la distribución de frecuencia para $p = 0.01$; $0.2$, $0.5$, $0.7$ y $0.995$. Observar que para $p = 0.5$ la distribución de frecuencia es simétrica.






**Esperanza y varianza**

Sea $X \sim B(n,p)$, entonces $E(X)=np$ y $V(X)=np(1-p)$

Dem.)
Aplicamos la definición:

$$E(X)=\sum_{k=0}^{n} kp(k)=\sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

El primer término es cero, por lo tanto podemos comenzar la suma en $k$=1:

$$E(X)=\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}=\sum_{k=1}^{n}\frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k}$$

Ponemos todo en términos de $k$-1:

$$E(X)=\sum_{(k-1)=0}^{n}\frac{n(n-1)!}{(k-1)![(n-1)-(k-1)]!}pp^{k-1}(1-p)^{[(n-1)-(k-1)]}$$

Sacamos fuera de la suma $n$ y $p$ que no dependen del índice $k$ y hacemos el cambio de índice: $(k-1)\rightarrow s$:

$$E(X)=np\sum_{s=0}^{n-1}\frac{(n-1)!}{s![(n-1)-s]!}p^s(1-p)^{[(n-1)-s]}=np\sum_{s=0}^{n-1}\binom{n-1}{s}p^s(1-p)^{[(n-1)-s]}$$

Recordando el desarrollo del binomio de Newton

$(a+b)^r=\sum_{s=0}^{n-1}\binom{r}{s}a^rb^{[r-s]}$ , tenemos (con $r$=n-1):

$$E(X)=np[p+(1-p)]^{n-1}=np[1]^{n-1}=np$$

Veamos el cálculo de la varianza

$V(X)=E(X^2)-[E(X)]^2$.

Luego:

$V(X)=E(X^2)-[np]^2$. Nos queda calcular $E(X^2)$.

$$E(X^2)=\sum_{k=0}^{n}k^2.p(k)=\sum_{k=0}^{n}k^2.\binom{n}{k}.p^k(1-p)^{n-k}=\sum_{k=0}^{n}k^2.\frac{n!}{k!(n-k)!}.p^k(1-p)^{n-k}.$$

Para calcular esta suma tratamos de llevarla a la forma del desarrollo de un binomio de Newton. Como el primer término es cero comenzamos a sumar desde $k$=1. Además simplificamos $k$ en el numerador con el factor $k$ en $k!$ del denominador:

$$E(X^2)=\sum_{k=1}^{n}k.\frac{n!}{(k-1)!(n-k)!}.p^k(1-p)^{n-k}=\sum_{k=1}^{n}k.\frac{n(n-1)!}{(k-1)!(n-k)!}.p^k(1-p)^{n-k}.$$

Separamos en dos sumas:

$$E(X^2)=\sum_{k=1}^{n}[1+(k-1)].\frac{n(n-1)!}{(k-1)!(n-k)!}.p^k(1-p)^{n-k}$$

$$=np\sum_{(k-1)=0}^{n-1}\frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!}.p^{(k-1)}(1-p)^{[(n-k)-(k-1)]}+$$

$$+n(n-1)p^2\sum_{(k-2)=0}^{n-2}\frac{(n-2)!}{(k-1)![(n-2)-(k-2)]!}.p^{(k-2)}(1-p)^{[(n-2)-(k-2)]}$$

Esto es:

$$E(X^2)=n.p\sum_{s=0}^{n-1}k.\binom{n-1}{s}.p^s(1-p)^{[(n-1)-s]}+n.(n-1)p^2\sum_{r=0}^{n-2}k.\binom{n-2}{r}.p^r(1-p)^{[(n-2)-r]}$$

Las sumas corresponden al desarrollo de un binomio de Newton:

$E(X^2) = np[p + (1-p)]^{n-1} + n(n-1)p^2[p + (1-p)]^{n-2} = np[1]^{n-1} + n(n-1)p^2[1]^{n-2}$, es decir

$E(X^2) = np + n(n-1)p^2$. Entonces:

$V(X) = E(X^2) - [np]^2 = np + n(n-1)p^2 - [np]^2 = np - np^2$ o sea:

$V(X) = np(1-p)$

## ***Distribución geométrica***

Sea $\varepsilon$ un experimento aleatorio. Sea *A* un evento asociado a $\varepsilon$ y anotamos $P(A) = p$.

Supongamos un experimento aleatorio $\varepsilon_0$ que cumple los siguientes requisitos:

1- se realizan repeticiones independientes de $\varepsilon$, hasta que ocurre A por primera vez inclusive.
2- las repeticiones son idénticas, y en cada repetición de $\varepsilon$ observamos ***si ocurre A*** o ***no ocurre A*** (cuando *A* ocurre se dice que se obtuvo un "éxito", caso contrario se obtuvo un "fracaso")
3- la probabilidad de éxito es constante de una repetición a otra de $\varepsilon$, y es igual a *p*

Sea la v.a. *X*: "número de repeticiones de $\varepsilon$ hasta que ocurre *A* por primera vez inclusive"
Veamos cuál es la distribución de probabilidad de $X$
El rango es el conjunto $R_X = \{1,2,3,....\} = N$
Además si anotamos $A_i$ :"ocurre $A$ en la $i-$ésima repetición de $\varepsilon$" , entonces

$$P(X=k) = P(A_1^C \cap A_2^C \cap \ldots \cap A_{k-1}^C \cap A_k) = P(A_1^C)P)(A_2^C)\ldots P(A_{k-1}^C)P(A_k) = (1-p)^{k-1}p$$

por independencia

En consecuencia

$$\boxed{P(X=k) = (1-p)^{k-1}p \qquad k = 1,2,.....}$$

Notación: $\boxed{X \sim G(p)}$

Para verificar que $\sum_{k=1}^{n} P(X=k) = 1$ recordar que $\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$ si $|a| < 1$

$$\sum_{k=1}^{n} P(X=k) = \sum_{k=1}^{n} (1-p)^{k-1}p = p\sum_{k=1}^{\infty}(1-p)^{k-1} = p\frac{1}{1-(1-p)} = 1$$

Observaciones:

1- $P(X=k) = (1-p)^{k-1}p$ y $P(X=k+1) = (1-p)^k p$

Es decir $P(X=k+1) = (1-p)^{k-1}p(1-p) = P(X=k)(1-p) \qquad \forall k = 1,2,.....$

Podemos interpretar que para cada *k*, los $a_k = P(X=k)$ son los términos de una sucesión geométrica con razón *1-p* y primer término $a_1 = p$

2- La F.d.a. sería

$$F(x) = P(X \le x) = \sum_{k=1}^{[x]} P(X = k)$$

donde $[x]$ indica parte entera de $x$

Si $x$ es un entero positivo entonces recordando que la suma de los $n$ primeros términos de una sucesión geométrica con razón $r$ y término general $a_k = pr^{k-1}$ es $\sum_{k=1}^{n} a_k = \frac{a_1(1-r^n)}{1-r}$ tenemos que

$$F(x) = P(X \le x) = \sum_{k=1}^{[x]} P(X = k) = \frac{p(1-r^{[x]})}{1-r} \underset{r=1-p}{=} \frac{p(1-(1-p)^{[x]})}{1-(1-p)} = 1-(1-p)^{[x]}$$

Por lo tanto

$$F(x) = \begin{cases} 1-(1-p)^{[x]} & \text{si} \quad x \ge 1 \\ 0 & \text{caso contrario} \end{cases}$$

3- Una variante en la definición de distribución geométrica es definir la v.a. *Y*: "***número de fracasos*** hasta el primer éxito", en este caso $R_Y = \{0,1,2,....\}$, es decir se incluye al cero.
La distribución de probabilidad o frecuencia sería en este caso

$$P(Y = k) = (1-p)^k p \qquad k = 0,1,2,.....$$

Notar que la relación entre entre $X$ e $Y$ sería: $X = Y + 1$.

Es decir si adoptamos esta última definición ***no incluimos*** la repetición del experimento en el cual ocurre el primer éxito.

Ejemplos:

1- La probabilidad de que una computadora que corre cierto sistema operativo se descomponga en determinado día es de 0.1. Determinar la probabilidad de que la máquina se descomponga por primera vez en el duodécimo día, después de la instalación del sistema operativo

Solución:
Definimos la v.a. X: "número de días hasta que la computadora se descompone por primera vez"
Entonces $X \sim G(0.1)$
Se pide calcular la $P(X = 12)$
$P(X = 12) = (1-p)^{12-1} p = 0.9^{11} \times 0.1 = 0.031381$

2- Una prueba de resistencia a la soldadura consiste en poner carga en uniones soldadas hasta que se dé una ruptura. Para cierto tipo de soldadura, 80% de las rupturas ocurre en la propia soldadura, mientras que otro 20% se da en las vigas. Se prueba cierto número de soldaduras.
Sea la v.a. X: "número de pruebas hasta que se produce la ruptura de la viga"
¿Qué distribución tiene *X*?. ¿Cuál es la probabilidad que en la tercera prueba se produzca la primera ruptura de la viga?

Solución:
Cada prueba es un ***"ensayo de Bernoulli"***, con un éxito definido como la ruptura de una viga. Por lo tanto, la probabilidad de éxito es $p = 0.2$.
La v.a. $X$ tiene una distribución geométrica con parámetro $p = 0.2$ es decir $X \sim G(0.2)$
Para calcular la probabilidad pedida hacemos $P(X = 3) = (1-p)^{3-1} p = 0.8^2 \times 0.2 = 0.128$

**Esperanza y varianza**

| |
|---|
| Sea $X \sim G(p)$ entonces $E(X) = \dfrac{1}{p}$ y $V(X) = \dfrac{1-p}{p^2}$ |

Dem.) Llamamos $1-p = q$
Planteamos

$$\sum_{k=1}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} =$$

Notar que $\dfrac{d}{dq}\left(q^k\right) = kq^{k-1}$ , por lo tanto como $\displaystyle\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$ si $|a| < 1$

$$= p\sum_{k=1}^{\infty} \frac{d}{dq}\left(q^k\right) = p\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p} \qquad \therefore\ E(X) = \frac{1}{p}$$

Calculamos ahora la varianza

$V(X) = E(X^2) - \mu^2$ donde

$$\mu = E(X) = \frac{1}{p}$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2 P(X = k) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=1}^{\infty} k(k-1+1)p(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(k-1)q^{k-1} + \sum_{k=1}^{\infty} kpq^{k-1}$$

Pero

$$p\sum_{k=1}^{\infty} k(k-1)q^{k-1} = pq\sum_{k=1}^{\infty} k(k-1)q^{k-2} = pq\sum_{k=1}^{\infty} \frac{d^2}{dq^2}\left(q^k\right) = pq\frac{d^2}{dq^2}\left(\sum_{k=1}^{\infty} q^k\right) = pq\frac{d^2}{dq^2}\left(\frac{q}{1-q}\right) =$$

$$= pq\frac{2}{(1-q)^3} = q\frac{2}{p^2}$$

Y $\displaystyle\sum_{k=1}^{\infty} kpq^{k-1} = \frac{1}{p}$

Por lo tanto

$$V(X) = E(X^2) - \frac{1}{p^2} = q\frac{2}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q + p - 1}{p^2} = \frac{2(1-p) + p - 1}{p^2} = \frac{1-p}{p^2}$$

## ***Distribución binomial negativa***

La distribución binomial negativa constituye una extensión de la distribución geométrica.
Sea $r$ un entero positivo.
Sea $\varepsilon$ un experimento aleatorio. Sea $A$ un evento asociado a $\varepsilon$ y anotamos $P(A) = p$.
Supongamos un experimento aleatorio $\varepsilon_0$ que cumple los siguientes requisitos:

1- se realizan repeticiones independientes de $\varepsilon$, hasta que ocurre A por r-ésima vez inclusive.
2- las repeticiones son idénticas, y en cada repetición de $\varepsilon$ observamos ***si ocurre A*** o ***no ocurre A*** (cuando $A$ ocurre se dice que se obtuvo un "éxito", caso contrario se obtuvo un "fracaso")
3- la probabilidad de éxito es constante de una repetición a otra de $\varepsilon$, y es igual a $p$

Sea la v.a. $X$: "número de repeticiones de $\varepsilon$ hasta que ocurre $A$ por $r$-ésima vez, incluyendo la r-ésima vez que ocurre $A$"

Veamos cuál es la distribución de probabilidad de $X$
El rango es el conjunto $R_X = \{r, r+1, r+2, r+3, ....\}$
Para obtener una expresión genérica de la $P(X = k)$, notar que si en el $k$-ésimo ensayo ocurre éxito por $r$-ésima vez, entonces en los $k$-$1$ primeros ensayos ocurrieron $r$-$1$ éxitos. Si anotamos $B$: " en los primeros $k$-$1$ ensayos ocurran $r$-$1$ éxitos", entonces

$$P(B) = \binom{k-1}{r-1} p^{r-1}(1-p)^{(k-1)-(r-1)} = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}$$

Además si anotamos $A$:"ocurre $A$ en la $r-$ésima repetición de $\varepsilon$" , entonces $P(A) = p$ y $A$ y $B$ son independientes, por lo tanto

$$P(X = k) = P(B \cap A) = P(B)P(A) = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \times p = \binom{k-1}{r-1} p^{r}(1-p)^{k-r}$$

En resumen

$$P(X = k) = \binom{k-1}{r-1} p^{r}(1-p)^{k-r} \qquad k = r, r+1, r+2, .....$$

<u>Notación</u>: $X \sim BN(r, p)$

Para verificar que $\sum_{k=0}^{n} P(X = k) = 1$ se deriva $r$ veces la igualdad $\sum_{k=0}^{\infty} (1-p)^{k-1} = \frac{1}{p}$ la que se deduce de $p\sum_{k=0}^{\infty} (1-p)^{k-1} = 1$

<u>Ejemplo:</u>

En una prueba de fuerza de soldadura, 80% de las pruebas da como resultado ruptura de soldadura, mientras que otro 20% da ruptura de la viga. Sea la v.a. $X$:"número de pruebas hasta la tercera ruptura de la viga inclusive". ¿Cuál es la distribución de $X$?. Determinar la $P(X = 8)$

<u>Solución</u>:
Tenemos que $X \sim BN(3,\ 0.2)$
Por lo tanto $P(X = 8) = \binom{8-1}{3-1} p^{3}(1-p)^{8-3} = \binom{7}{2} 0.2^{3} \times 0.8^{5} = 0.05505$

Observación: la distribución geométrica puede verse como un caso particular de la distribución binomial negativa con $r = 1$

**Esperanza y varianza**

Si $X \sim BN(r, p)$ entonces $E(X) = \dfrac{r}{p}$ y $V(X) = \dfrac{r(1-p)}{p^2}$

Dem.) Se hará mas adelante

## ***Distribución hipergeométrica***

Supongamos que tenemos una población o conjunto de *N* objetos o individuos (es decir tenemos una población finita).

Clasificamos a los objetos de la población en dos categorías. Hay *M* objetos de una categoría y *N-M* de la otra categoría. Se suele decir que tenemos *M* "éxitos" y *N-M* "fracasos".

Se extraen al azar y sin reemplazo *n* objetos de dicha población. Es decir se extrae una muestra de *n* objetos de la población, de manera tal que es igualmente probable que se seleccione cada subconjunto de tamaño *n*.

Consideramos la v.a. *X*: "número de éxitos en la muestra extraída"

Se dice que X tiene una distribución hipergeométrica con parámetros *n*, *M* y *N*

Notación: $X \sim H(n,\ M,\ N)$

Veamos cuál es la distribución de *X*

Primero notar que una expresión para la $P(X = k)$, usando combinatoria es

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

donde para que los números combinatorios estén bien definidos debe cumplirse $0 \le k \le M$ y $0 \le n-k \le N-M$ . Pero estas condiciones son equivalentes a

$\max(0, n-N+M) \le k \le \min(n, M)$

Por lo tanto la distribución de probabilidad de *X* es

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad \max(0, n-N+M) \le k \le \min(n, M)$$

Se verifica que $\sum_{k=0}^{n} P(X = k) = 1$ pues los números $P(X = k)$ corresponden a la distribución de una v.a.

Ejemplo:

1- De 50 edificios en un parque industrial, 12 no cumplen el código eléctrico. Si se seleccionan aleatoriamente 10 edificios para inspeccionarlos, ¿cuál es la probabilidad de que exactamente tres de los diez no cumplan el código?

Solución:
Sea la v.a. X: "número de edificios seleccionados que violan el código", entonces $X \sim H(10,\ 12,\ 50)$ . se pide calcular la $P(X=3)$

$$P(X=3)=\frac{\binom{12}{3}\binom{50-12}{10-3}}{\binom{50}{10}}=0.2703$$

2- Un cargamento contiene 40 elementos. Se seleccionará de forma aleatoria y se probará 5 elementos. Si dos o más están defectuosos, se regresará el cargamento.
a) si de hecho el cargamento contiene cinco elementos defectuosos, ¿cuál es la probabilidad de que sean aceptados?
b) si de hecho el cargamento contiene diez elementos defectuosos, ¿cuál es la probabilidad de que no sean aceptados?

Solución:
a) Sea la v.a. X: "número de elementos defectuosos en la muestra"
En este caso $X \sim H(5,\ 5,\ 40)$. Hay que calcular la $P(X \le 1)$

$$P(X \le 1)=P(X=0)+P(X=1)$$

$$P(X=0)=\frac{\binom{5}{0}\binom{40-5}{5-0}}{\binom{40}{5}}=0.4933557 \qquad P(X=1)=\frac{\binom{5}{1}\binom{40-5}{5-1}}{\binom{40}{5}}=0.3978675$$

$$\therefore \quad P(X \le 2)=0.4933557+0.3978675=0.8912232$$

b) Sea la v.a. X: "número de elementos defectuosos en la muestra"
En este caso $X \sim H(5,\ 10,\ 40)$. Hay que calcular la $P(X \ge 2)$

$$P(X \ge 2)=1-P(X \le 1)=1-P(X=0)-P(X=1)$$

$$P(X=0)=\frac{\binom{10}{0}\binom{40-10}{5-0}}{\binom{40}{5}}=0.2165718 \qquad P(X=1)=\frac{\binom{10}{1}\binom{40-10}{5-1}}{\binom{40}{5}}=0.416484$$

$$\therefore \quad P(X \ge 2)=0.3669442$$

Observación:

En el ejemplo anterior si el cargamento hubiese tenido 400 elementos se podría haber considerado en la parte a) a $X$ con distribución binomial con parámetros $n=5$ y $p=\frac{5}{400}$

En general, si el tamaño de la población $N$ y el número de éxitos $M$ crecen pero de manera tal que $\frac{M}{N} \to p$ y $n$ es chico comparado con $N$, se puede verificar que

$$\lim_{N\to\infty}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\binom{n}{k}p^k(1-p)^{n-k} \qquad \text{donde } \frac{M}{N}=p$$

Por lo tanto, para una fracción fija de defectuosos $\frac{M}{N}=p$ la función de probabilidad hipergeométrica converge a la función de probabilidad binomial cuando $N$ se hace grande.

**<u>Esperanza y varianza</u>**

Si $X \sim H(n,\ M,\ N)$ entonces $E(X)=\frac{nM}{N}$ y $V(X)=n\frac{M}{N}\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$

Dem.) La demostración se hará mas adelante.

## ***<u>Distribución de Poisson</u>***

Una v.a. $X$ con rango $R_X=\{0,1,2,....\}$ se dice tener ***distribución de Poisson con parámetro*** $\lambda$, si para algún $\lambda>0$

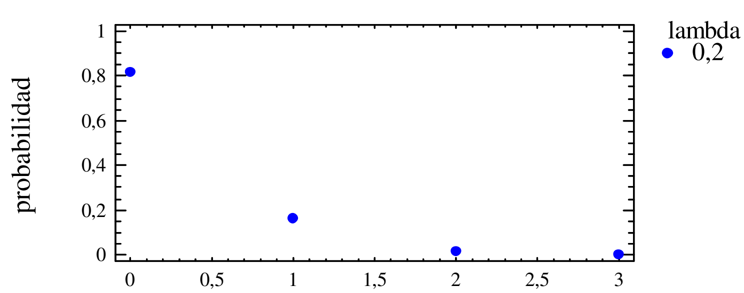
$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!} \qquad k=0,1,2,.....$$

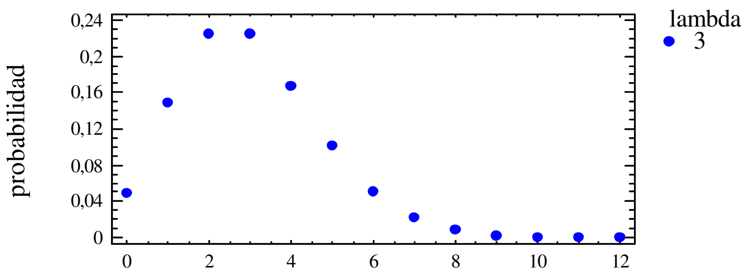
Es fácil verificar que $\sum_{k=0}^{n}P(X=k)=1$ usando el hecho que $\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{\lambda}$

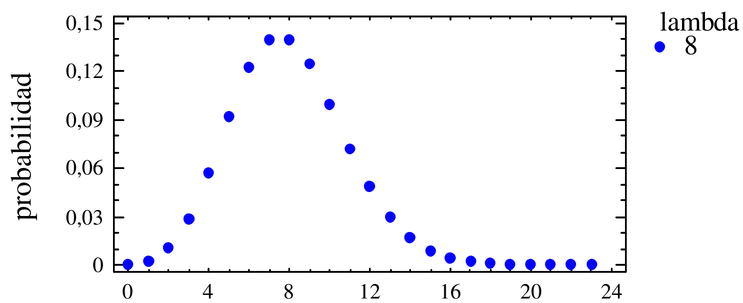
$$\sum_{k=0}^{n}P(X=k)=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$$

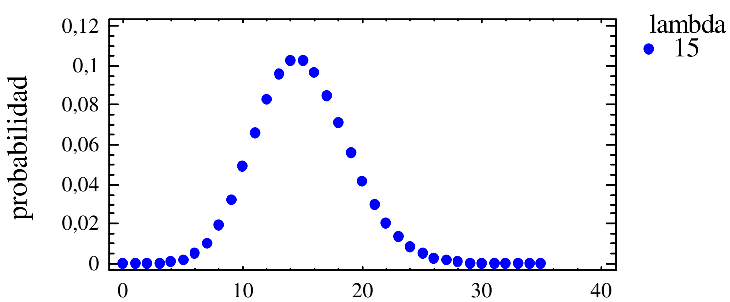
En los siguientes gráficos se ve como varía la forma de la distribución con los valores de $\lambda$

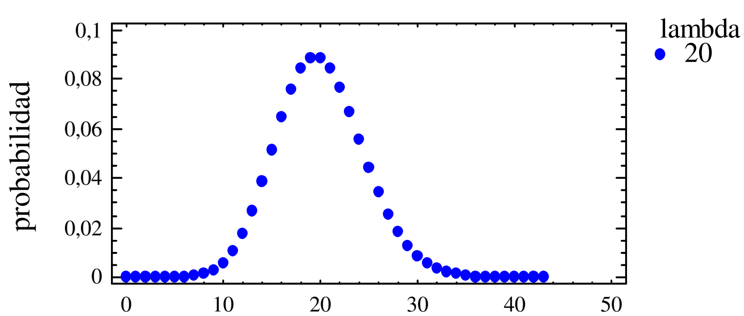
Notar que para valores de $\lambda$ "pequeños" la distribución es asimétrica, a medida que $\lambda$ aumenta, la distribución tiende a ser cada vez más simétrica

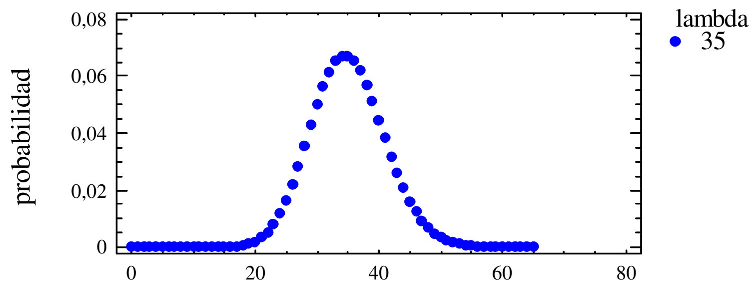



probabilidad
lambda
3
0,24
0,2
0,16
0,12
0,08
0,04
0
0 2 4 6 8 10 12


probabilidad
lambda
8
0,15
0,12
0,09
0,06
0,03
0
0 4 8 12 16 20 24


probabilidad
lambda
15
0,12
0,1
0,08
0,06
0,04
0,02
0
0 10 20 30 40


probabilidad
lambda
20
0,1
0,08
0,06
0,04
0,02
0
0 10 20 30 40 50




Ejemplo:

Considere escribir en un disco de computadora y luego enviar el escrito por un certificador que cuenta el número de pulsos faltantes. Suponga que este número $X$ tiene una distribución de Poisson con parámetro igual a 0.2

a) ¿Cuál es la probabilidad de que un disco tenga exactamente un pulso faltante?
b) ¿Cuál es la probabilidad de que un disco tenga al menos dos pulsos faltantes?
c) Si dos discos se seleccionan independientemente, ¿cuál es la probabilidad de que ninguno contenga algún pulso faltante?

Solución:
a) Sea la v.a. $X$: "número de pulsos faltantes en un disco"
Entonces $X \sim P(0.2)$

Se pide $P(X=1) = e^{-0.2}\dfrac{0.2^1}{1!} = 0.163746$

b) Siendo $X$ como en a) se pide calcular $P(X \geq 2)$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X=0) - P(X=1) = 1 - e^{-0.2}\frac{0.2^0}{0!} - e^{-0.2}\frac{0.2^1}{1!} = 0.01752$$

c) Sea la v.a. Y: "número de discos sin pulsos faltantes"
Entonces $Y \sim B(2,p)$ donde $p = P(X=0) = e^{-0.2}$
Por lo tanto se pide calcular $P(Y=2)$

$$P(Y=2) = \binom{2}{2} p^2 (1-p)^0 = p^2 = \left(e^{-0.2}\right)^2 = 0.67032$$

**Esperanza y varianza**

$$\boxed{\text{Si } X \sim P(\lambda) \quad \text{entonces} \quad E(X) = \lambda \quad \text{y} \quad V(X) = \lambda}$$

Dem.)

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k-1)!} = \lambda\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

$V(X) = E(X^2) - \mu^2$ donde $\mu = E(X) = \lambda$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{\infty} k^2\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{(k-1)!} = \sum_{k=1}^{\infty}(k-1+1)\frac{e^{-\lambda}\lambda^k}{(k-1)!} =$$

$$=\sum_{k=1}^{\infty}(k-1)\frac{e^{-\lambda}\lambda^k}{(k-1)!}+\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k-1)!}=\sum_{k=2}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k-2)!}+\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k-1)!}=\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^{k+2}}{k!}+\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^{k+1}}{k!}=$$

$$=\lambda^2\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}+\lambda\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}=\lambda^2+\lambda$$

Por lo tanto

$$V(X)=\lambda^2+\lambda-\lambda^2=\lambda$$

**<u>Aplicaciones de la distribución de Poisson</u>**

La v.a. Poisson tiene un gran rango de aplicaciones, una de ellas es la aproximación para una v.a. binomial con parámetros $n$ y $p$ cuando $n$ es grande y $p$ es pequeño de manera tal que $np \to \lambda$, específicamente, sea $X \sim B(n,p)$ y sea $\lambda = np$, entonces

$$P(X=k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}=\frac{n(n-1)....(n-k+1)}{n^k}\frac{\lambda^k}{k!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

Para $n$ grande y $p$ chico

$$\left(1-\frac{\lambda}{n}\right)^n\approx e^{-\lambda}\quad;\quad\frac{n(n-1)....(n-k+1)}{n^k}\approx 1\quad;\quad\left(1-\frac{\lambda}{n}\right)^k\approx 1$$

Entonces, para $n$ grande y $p$ chico

$$P(X=k)\approx e^{-\lambda}\frac{\lambda^k}{k!}$$

Es decir cuando $n$ es grande, $p$ chico y $np$ es "moderado" entonces la v.a. binomial con parámetros n y p tiene una distribución que se ***aproxima*** a la de una Poisson con parámetro $\lambda = np$

<u>Ejemplo</u>:

Supongamos que la probabilidad de que un artículo producido por cierta máquina sea defectuoso es 0.1. Hallar la probabilidad que una muestra de 10 artículos contenga a lo sumo un defectuoso.

Sea $X$: "número de artículos defectuosos en la muestra"
Podemos asumir que $X \sim B(10,\ 0.1)$
La probabilidad ***exacta*** pedida es

$$P(X\le 1)=P(X=0)+P(X=1)=\binom{10}{0}(0.1)^0(0.9)^{10-0}+\binom{10}{1}(0.1)^1(0.9)^{10-1}=0.7361$$

La ***aproximación*** de Poisson da

$$P(X\le 1)=P(X=0)+P(X=1)\approx e^{-1}\frac{\lambda^0}{0!}+e^{-1}\frac{\lambda^1}{1!}=e^{-1}+e^{-1}\approx 0.7358$$

$$\lambda=np=10\times 0.1=1$$

Algunos autores sostienen que la aproximación de Poisson funciona bien cuando $n$ es grande, $p$ es chico y $np < 7$ (*Mendenhall*, Estadística matemática con aplicaciones), otros recomiendan usar la aproximación de Poisson a la binomial cuando $n \geq 100$ y $p \leq 0.01$ y $np \leq 20$ (*Devore*, Probabilidad para ingeniería y ciencias)

En la siguiente tabla se da un ejemplo de una aproximación de Poisson a la función de distribución de la binomial. Se tabula la $P(X = k)$ para algunos valores de $k$ para las distribuciones binomial y Poisson con los parámetros que se indican

| k | $X = B(10^5, 2\times10^{-5})$ | $X = B(5\times10^3, 4\times10^{-5})$ | P(2) |
|---|---|---|---|
| 0 | 0.135281 | 0.135281 | 0.135335 |
| 1 | 0.270671 | 0.270671 | 0.270671 |
| 2 | 0.270725 | 0.270725 | 0.270671 |
| 3 | 0.180483 | 0.180483 | 0.180447 |
| 4 | 0.0902235 | 0.0902235 | 0.0902235 |
| 5 | 0.036075 | 0.036075 | 0.0360894 |
| 6 | 0.0120178 | 0.0120178 | 0.0120298 |
| 7 | 0.0034309 | 0.0034309 | 0.00343709 |
| 8 | 0.000856867 | 0.000856867 | 0.000859272 |
| 9 | 0.000190186 | 0.000190186 | 0.000190949 |
| 10 | 0.000037984 | 0.000037984 | 0.0000381899 |
| 11 | $6.89513\times10^{-6}$ | $6.89513\times10^{-6}$ | $6.94361\times10^{-6}$ |
| 12 | $1.14712\times10^{-6}$ | $1.14712\times10^{-6}$ | $1.15727\times10^{-6}$ |
| 13 | $1.76127\times10^{-7}$ | $1.76127\times10^{-7}$ | $1.78041\times10^{-7}$ |
| 14 | $2.51056\times10^{-8}$ | $2.51056\times10^{-8}$ | $2.54345\times10^{-8}$ |
| 15 | $3.33937\times10^{-9}$ | $3.33937\times10^{-9}$ | $3.39126\times10^{-9}$ |

<u>Ejemplo</u>:

En una prueba de tarjetas de circuitos, la probabilidad de que un diodo en particular falle es 0.01. Suponga que una tarjeta contiene 200 diodos.

a) ¿Cuál es la probabilidad aproximada de que por lo menos 4 diodos fallen en una tarjeta seleccionada al azar?

b) Si se embarcan cinco tarjetas a un cliente en particular, ¿cuál es la probabilidad de que por lo menos cuatro de ellas funcionen bien? (Una tarjeta funciona bien solo si todos sus diodos funcionan bien)

<u>Solución</u>:

a) Sea la v.a. X: "número de diodos en una tarjeta que fallan"

Entonces $X \sim B(200,\ 0.01)$. Como $n$ es grande y $p$ chico aplicamos la aproximación Poisson con $np = 200 \times 0.01 = 2$

Se pide calcular la $P(X \geq 4)$

$P(X \geq 4) = 1 - P(X \leq 3) \approx 1 - 0.857 = 0.143$

por tabla de la acumulada de la Poisson

b) Sea la v.a. Y: "número de tarjetas entre 5 que funcionan bien"

Tenemos que $Y \sim B(5,\ p)$ donde $p = P(X = 0) \approx e^{-2}\dfrac{2^0}{0!} = e^{-2}$

Se pide calcular $P(Y \geq 4)$

$$P(Y \geq 4) = P(Y=4) + P(Y=5) = \binom{5}{4}\left(e^{-2}\right)^4\left(1-e^{-2}\right)^{5-4} + \binom{5}{5}\left(e^{-2}\right)^5\left(1-e^{-2}\right)^{5-5} = 5e^{-8}\left(1-e^{-2}\right)+5e^{-10}$$

## Proceso de Poisson

Una aplicación importante de la distribución de Poisson se presenta en relación con el acontecimiento de eventos de un tipo particular en el tiempo. Por ejemplo, un evento podría ser un individuo entrando en un establecimiento en particular, o pulsos radiactivos registrados por un contador Geiger, o automóviles pasando por un cruce determinado.

Supongamos que tenemos eventos que ocurren en ciertos puntos aleatorios de tiempo, y asumimos que para alguna constante positiva $\lambda$ las siguientes suposiciones se sostienen:

1- La probabilidad que exactamente 1 evento ocurra en un intervalo de longitud $t$ es la misma para todos los intervalos de longitud $t$ y es igual a $\lambda t + o(t)$ (donde $o(t)$ simboliza una función $f(t)$ tal que $\lim\limits_{t \to 0} \dfrac{f(t)}{t} = 0$, por ejemplo $f(t) = t^2$ es $o(t)$, pero $f(t) = t$ no lo es)

2- La probabilidad que 2 o mas eventos ocurran en un intervalo de longitud $t$ es la misma para todos los intervalos de longitud $t$ y es igual a $o(t)$.

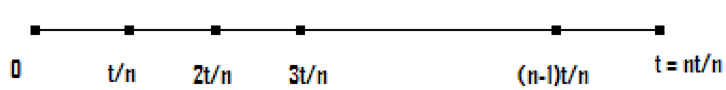
3- Para cualesquiera enteros $n, k_1, k_2, \ldots, k_n$ y cualquier conjunto de $n$ intervalos $I_1, I_2, \ldots, I_n$ que no se superpongan, si definimos los eventos $E_i$ :"en el intervalo $I_i$ ocurren exactamente $k_i$ eventos" $i = 1,2,\ldots,n$, entonces los eventos $E_1, E_2, \ldots, E_n$ son independientes.

Las suposiciones 1 y 2 establecen que para pequeños valores de $t$, la probabilidad de que exactamente un evento ocurra en un intervalo de longitud $t$ es igual a $\lambda t$ mas algo que es chico comparado con $t$, mientras que la probabilidad de que 2 o mas eventos ocurran es pequeño comparado con $t$. la suposición 3 establece que lo que ocurra en un intervalo no tiene efecto (en la probabilidad) sobre lo que ocurrirá en otro intervalo que no se superponga.

Bajo las suposiciones 1, 2 y 3 se puede probar que la v.a.
$X$ :"número de eventos que ocurren en cualquier intervalo de longitud $t$" , tiene distribución Poisson con parámetro $\lambda t$
Específicamente

$$P(X = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad k = 0,1,2,\ldots$$

La idea de la demostración es la siguiente, partimos al intervalo $[0,t]$ en n subintervalos que no se superpongan cada uno de longitud $\dfrac{t}{n}$



Elegimos $n$ suficientemente grande para que en cada subintervalo se tenga una ocurrencia exactamente o ninguna ocurrencia.

Sea la v.a. $Y$: "número de subintervalos en los que hay exactamente una ocurrencia", entonces podemos asumir que $Y \sim B(n,\ p)$ donde $p$ es la probabilidad que en un subintervalo hay exactamente una ocurrencia y si $n$ es grande entonces la longitud del subintervalo $\frac{t}{n}$ es chica con lo cual por suposición 1 tenemos que $p \approx \lambda \frac{t}{n}$.

Entonces, utilizando la aproximación de Poisson a la binomial con parámetro $np = n\lambda \frac{t}{n} \lambda t$ tenemos

$$P(X = k) = P(Y = k) \approx e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0,1,2,...$$

Observaciones:

1- Un ***proceso temporal de Poisson*** consiste en eventos que ocurren en el tiempo en forma aleatoria que cumplen con las suposiciones 1, 2 y 3.
2- El parámetro $\lambda$ es la ***tasa o rapidez del proceso.***
3- Si en lugar de observar eventos en el tiempo, consideramos observar eventos de algún tipo que ocurren en una región de dos o tres dimensiones, por ejemplo, podríamos seleccionar de un mapa una región R de un bosque, ir a esa región y contar el número de árboles. Cada árbol representaría un evento que ocurre en un punto particular del espacio. Bajo las suposiciones 1, 2 y 3, se puede demostrar que el número de eventos que ocurren en la región R tiene una distribución de Poisson con parámetro $\lambda a$ donde $\boldsymbol{a}$ es el área de R ( $\lambda$ se interpreta como la densidad del proceso). Se trata ahora de un ***proceso espacial de Poisson.***

Ejemplos:

1- Suponga que aviones pequeños llegan a cierto aeropuerto según un proceso de Poisson, con tasa $\lambda = 8$ aviones por hora, de modo que el número de llegadas durante un período de $t$ horas es una v.a. Poisson con parámetro $\lambda = 8t$ .
   a) ¿Cuál es la probabilidad de que exactamente 5 aviones pequeños lleguen durante un período de una hora? ¿Por lo menos 5?
   b) Cuál es la probabilidad de que por lo menos 20 aviones pequeños lleguen durante un período de 2 ½ hs? ¿De que a lo sumo 10 lleguen en ese período?

   Solución:
   a) Sea la v.a. $X$: "número de aviones pequeños que llegan a cierto aeropuerto en una hora" Entonces $X \sim P(8)$ . Por lo tanto

   $$P(X = 5) = e^{-8} \frac{8^5}{5!} = 0.0916$$

   O también usando la tabla de distribución acumulada para la Poisson

   $$P(X = 5) = P(X \le 5) - P(X \le 4) = 0.191 - 0.099 = 0.092$$

   Y la probabilidad de que lleguen al menos 5 aviones será

   $$P(X \ge 5) = 1 - P(X \le 4) = 1 - 0.099 = 0.901$$

   b) Sea la v.a. $X$: "número de aviones pequeños que llegan a cierto aeropuerto en 2 ½ horas"

Entonces $X \sim P(8 \times 2.5)$ es decir ahora $\lambda t = 8 \times 2.5 = 20$

$$P(X \geq 20) = 1 - P(X \leq 19) = 1 - 0.470 = 0.53$$

por tabla de F.d.a.

Y por último calculamos por tabla $P(X \leq 10) = 0.010$

2- Se supone que el número de defectos en los rollos de tela de cierta industria textil es una v.a. Poisson con tasa 0.1 defectos por metro cuadrado..

a) ¿Cuál es la probabilidad de tener dos defectos en un metro cuadrado de tela?
b) ¿Cuál es la probabilidad de tener un defecto en 10 metros cuadrados de tela?
c) ¿Cuál es la probabilidad de que no halla defectos en 20 metros cuadrados de tela?
d) supongamos que el número de defectos está relacionado con la máquina que produce la tela, debido a desperfectos de la máquina el número de defectos varía en ciertos tramos del rollo. ¿Se puede asumir que el número de defectos sigue una distribución de Poisson?

Solución:

a) Sea $X$: "número de defectos en un metro cuadrado". Entonces $X \sim P(0.1)$ pues $\lambda a = \lambda \times 1 = 0.1 \times 1 = 0.1$

$$P(X = 2) = e^{-0.1} \frac{0.1^2}{2!} = 0.004524$$

b) Si X: "número de defectos en 10 metros cuadrados". Entonces $X \sim P(1)$ pues $\lambda a = \lambda \times 10 = 0.1 \times 10 = 1$

$$P(X = 1) = e^{-1} \frac{1^1}{1!} = 0.3678794$$

c) $X$: "número de defectos en 20 metros cuadrados". Entonces $X \sim P(2)$ pues

$\lambda a = \lambda \times 20 = 0.1 \times 20 = 2$

$$P(X = 0) = e^{-2} \frac{2^0}{0!} = 0.135$$

d) NO se puede asumir que el número de defectos sigue una distribución de Poisson, ya que las suposiciones que debe satisfacer un proceso de Poisson no se cumplirían.

### 3.5 – Variables aleatorias continuas

En la sección anterior se consideraron variables aleatorias discretas, o sea variables aleatorias cuyo rango es un conjunto finito o infinito numerable. Pero hay variables aleatorias cuyo rango son todos los números reales de un intervalo dado, (es decir es un conjunto infinito no numerable). Ejemplos de variables continuas podrían ser

*X*: "tiempo que tarda en llegar un colectivo a una parada"
*Y*: "tiempo de vida de un fusible"
Como ahora los valores de una v.a. continua no son contables no se puede hablar del *i*-ésimo valor de la v.a. *X* y por lo tanto $p(x_i) = P(X = x_i)$ pierde su significado. Lo que se hace es sustituir la función $p(x)$ definida sólo para $x_1, x_2, ....$, por una función $f(x)$ definida para todos los valores $x$ del rango de *X*. Por lo tanto se da la siguiente definición de v.a. continua

Sea *X* una v.a.. Decimos que es ***continua*** si existe una función no negativa $f$, definida sobre todos los reales $x \in (-\infty, \infty)$, tal que para cualquier conjunto *B* de números reales

$$P(X \in B) = \int_B f(x)dx$$

O sea que la probabilidad de que *X* tome valores en *B* se obtiene al integrar la función *f* sobre el conjunto *B*.

***A la función f la llamamos función densidad de probabilidad (f.d.p.).***

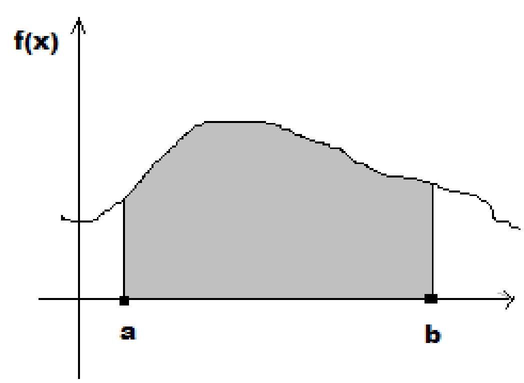
Observaciones:

1- Como *X* debe tomar algún valor real, entonces debe cumplirse que

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx$$

2- Si *B* es el intervalo real $[a, b] = \{x \in R;\ a \le x \le b\}$ entonces

$$P(X \in B) = P(a \le X \le b) = \int_a^b f(x)dx$$




Notar que en este caso la probabilidad de que *X* tome valores en el intervalo $[a, b]$ es ***el área bajo f entre a y b***

3- Si en la observación anterior $a = b$ entonces

$$P(X = a) = \int_a^a f(x)dx = 0$$

Es decir la probabilidad que una v.a. continua tome algún valor fijado es cero. Por lo tanto, para una v.a. continua

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b) = \int_a^b f(x)dx$$

**Función de distribución acumulada**

Sea X una v.a. continua. Se define la ***función de distribución acumulada de X*** (abreviamos F.d.a de *X*) como

$$F(x) = P(X \le x) \qquad -\infty < x < \infty$$

*Si X tiene f.d.p. f(x) entonces*

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \qquad -\infty < x < \infty$$

Además

$$P(a \le X \le b) = \int_a^b f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a)$$

Observaciones:

1- Si $X$ es una v.a. con *f.d.p. f(x)* y función de distribución acumulada $F(x)$ entonces

$$\frac{dF(x)}{dx} = \frac{d}{dx}\left(\int_{-\infty}^{x} f(t)dt\right) = f(x) \quad \text{donde} \quad F(x) \text{ sea derivable}$$

***Es decir, se puede obtener la función de densidad de X a partir de su F.d.a.***

2- Como en el caso discreto vale

Si $a < b$ entonces $F(a) \le F(b)$ (es decir *F(x)* es una función creciente)

Y además se cumple que

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{\infty} f(t)dt = 1$$

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{-\infty} f(t)dt = 0$$

Ejemplos:

1- Supongamos que *X* es una v.a. continua con *f.d.p.* dada por

$$f(x) = \begin{cases} C\left(4x - 2x^2\right) & \text{si } 0 \le x \le 2 \\ 0 & \text{caso contrario} \end{cases}$$

a) ¿Cuál es el valor de C?

b) Hallar $P(X > 1)$

c) Hallar la F.d.a. de *X*

Solución:

a) Por lo dicho en la observación 1, se debe cumplir que $1=\int_{-\infty}^{\infty} f(x)dx$ , por lo tanto

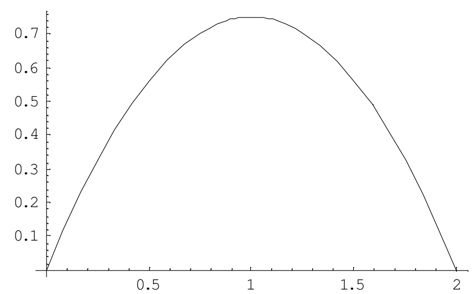
$$1=\int_{-\infty}^{\infty} f(x)dx=\int_{-\infty}^{0} 0dx+\int_{0}^{2} C\left(4x-2x^2\right)dx+\int_{2}^{\infty} 0dx=\int_{0}^{2} C\left(4x-2x^2\right)dx$$

Entonces

$$C\int_{0}^{2}\left(4x-2x^2\right)dx=C\left(4\frac{x^2}{2}-2\frac{x^3}{3}\right)\Bigg|_0^2=C\frac{8}{3}=1$$

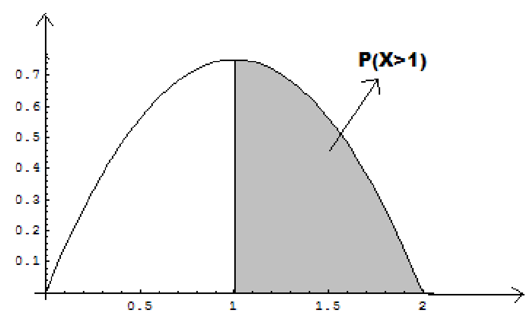
$$\therefore \quad C=\frac{3}{8}$$

Es útil hacer un gráfico de la densidad




b) Para calcular la probabilidad que $X$ sea mayor que 1, planteamos

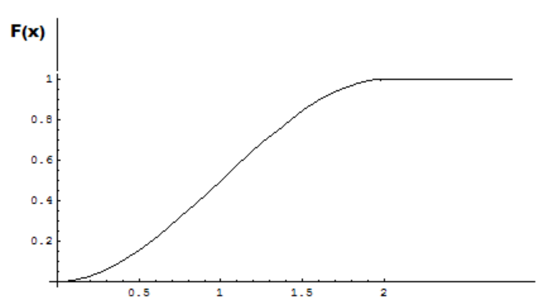
$$P(X>1)=$$
$$=\int_{1}^{2}\frac{3}{8}\left(4x-2x^2\right)dx=\frac{1}{2}$$




c) Para calcular la F.d.a. notar que tenemos tres casos: $x<0$, $0\le x\le 2$ y $x>2$, por lo tanto

$$F(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ \int_0^x \frac{3}{8}\left(4t-2t^2\right)dt & \text{si} \quad 0\le x\le 2 \\ 1 & \text{si} \quad x>2 \end{cases} \quad \text{es decir} \quad F(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ \frac{3}{4}x^2\left(1-\frac{x}{3}\right) & \text{si} \quad 0\le x\le 2 \\ 1 & \text{si} \quad x>2 \end{cases}$$

El gráfico de la F.d.a. es




Se podría haber calculado la $P(X>1)$ a partir de la F.d.a. de la siguiente forma

$$P(X>1)=1-P(X\le 1)=1-F(1)=1-\frac{3}{4}1^2\left(1-\frac{1}{3}\right)=1-\frac{3}{4}\times\frac{2}{3}=\frac{1}{2}$$

2- El tiempo de vida en horas que una computadora funciona antes de descomponerse es una v.a. continua con *f.d.p.* dada por

$$f(x)=\begin{cases}0.01e^{-0.01x} & \text{si } x\ge 0\\ 0 & \text{si } x<0\end{cases}$$

a) hallar la F.d.a. de $X$

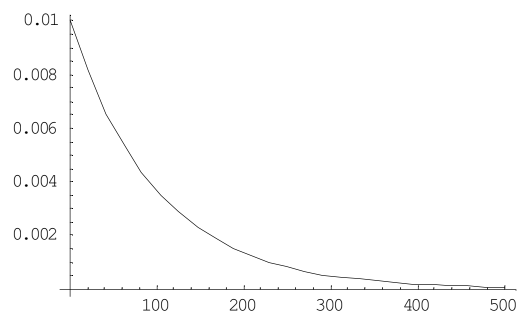
b) ¿Cuál es la probabilidad que la computadora funcione entre 50 y 150 horas antes de descomponerse?

c) ¿Cuál es la probabilidad que una computadora se descomponga antes de registrar 100 horas de uso?

d) ¿Cuál es la probabilidad que exactamente 2 de 5 computadoras se descompongan antes de registrar 100 horas de uso?. Asumir que las computadoras trabajan en forma independiente.

Solución:

a) Hacemos un gráfico de la densidad y entonces observamos claramente que hay dos casos a considerar para calcular la F.d.a.: $x\ge 0$ y $x<0$




Si $x<0$ entonces

$$F(x)=P(X\le x)=\int_{-\infty}^{0}0dx=0$$

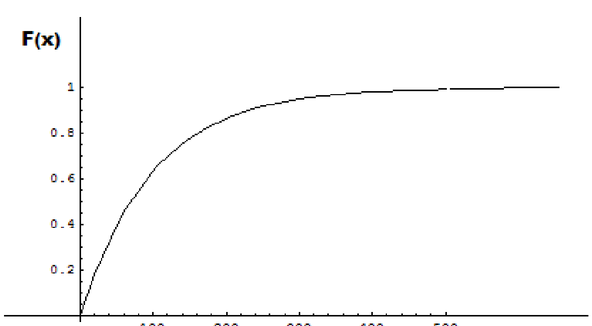
Y si $x\ge 0$ tenemos

$$F(x)=P(X\le x)=\int_{-\infty}^{0}0dt+\int_{0}^{x}0.01e^{-0.01t}dt=$$

$$=0+0.01\left(\frac{e^{-0.01t}}{-0.01}\right)\Bigg|_0^x=1-e^{-0.01x}$$

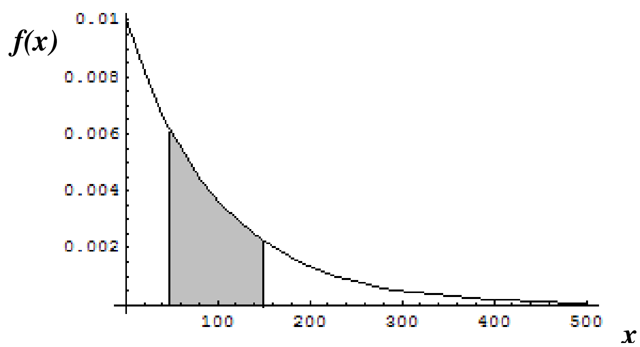
$$\therefore\quad F(x)=\begin{cases}1-e^{-0.01x} & \text{si } x\ge 0\\ 0 & \text{caso contrario}\end{cases}$$

La gráfica de la F.d.a. es




Se pide calcular $P(50\le X\le 150)$, lo hacemos con la F.d.a.:

$$P(50 \le X \le 150) = F(150) - F(50) = e^{-0.01\times 50} - e^{-0.01\times 150} = 0.3834$$




c) Hay que calcular $P(X < 100)$

$$P(X < 100) = F(100) = 1 - e^{-0.01\times 100} = 1 - e^{-1} \approx 0.63212$$

d) Podemos definir la v.a.

*Y*: "número de computadores entre 5 que se descomponen antes de las 100 horas de uso"
Entonces $Y \sim B(5,p)$ donde $p = P(X < 100)$
Por lo tanto hay que calcular

$$P(Y = 2) = \binom{5}{2} p^2 (1-p)^3 = \frac{5!}{2!3!}\left(1 - e^{-1}\right)^2 \left(e^{-1}\right)^3 \approx 0.198937$$

## 3.6 – Esperanza de una variable aleatoria continua

Para una v.a. discreta la $E(X)$ se definió como la suma de los $x_i\, p(x_i)$. Si *X* es una v.a. continua con *f.d.p.* *f(x)*, se define $E(X)$ sustituyendo la sumatoria por integración y $p(x_i)$ por *f(x)*.

La esperanza de una v.a. continua *X* con *f.d.p. f(x)* se define como

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

Ejemplo:
Para ciertas muestras de minerales la proporción de impurezas por muestra, es una v.a. *X* con f.d.p. dada por

$$f(x) = \begin{cases} \frac{3}{2}x^2 + x & \text{si } \; 0 < x < 1 \\ 0 & \text{caso contrario} \end{cases}$$

Hallar la esperanza de *X*

Solución:
Se plantea

$$E(X)=\int_0^1 x\left(\frac{3}{2}x^2+x\right)dx=\left(\frac{3}{2}\frac{x^4}{4}+\frac{x^3}{3}\right)\Bigg|_0^1=\left(\frac{3}{2}\times\frac{1}{4}+\frac{1}{3}\right)=\frac{17}{24}$$

A menudo se desea calcular la esperanza de una función de *X*, *Y = h(X)*, esto se puede hacer hallando previamente la densidad de *Y* y luego calcular *E(Y)* aplicando la definición anterior.
Otra forma de calcular *E(Y)* sin hallar la densidad de *Y* está dada por el siguiente

Teorema: Si *X* es una v.a. continua con *f.d.p. f(x)* y *h(X)* es cualquier función de *X*, entonces

$$E(h(X))=\int_{-\infty}^{\infty} h(x)f(x)dx$$

Dem.) sin demostración

Ejemplo: En el ejemplo anterior supongamos que el valor en dólares de cada muestra es $Y=h(X)=5-0.5X$ . Encontrar la esperanza de *Y*

Podemos hallar la esperanza de *Y* encontrando previamente su *f.d.p.*
Para esto se encuentra la *F.d.a.* de *Y* , para luego hallar la densidad de *Y* derivando la *F.d.a*
Anotamos $G(y)$ y $g(y)$ a la *F.d.a* de *Y* y a la densidad de *Y* respectivamente

$$G(y)=P(Y\le y)=P(5-0.5X\le y)=P\left(X\ge\frac{5-y}{0.5}\right)=1-P\left(X\le\frac{5-y}{0.5}\right)=1-F\left(\frac{5-y}{0.5}\right)$$

Donde $F(x)=P(X\le x)$

Entonces

$$g(y)=\frac{d}{dy}\left(1-F\left(\frac{5-y}{0.5}\right)\right)=-f\left(\frac{5-y}{0.5}\right)\left(-\frac{1}{0.5}\right)=f\left(\frac{5-y}{0.5}\right)\left(\frac{1}{0.5}\right)=$$

$$=\begin{cases}\left(\dfrac{3}{2}\left(\dfrac{5-y}{0.5}\right)^2+\dfrac{5-y}{0.5}\right)\left(\dfrac{1}{0.5}\right) & \text{si} \quad 0<\dfrac{5-y}{0.5}<1\\ 0 & \text{caso contrario}\end{cases}$$

O sea

$$g(y)=\begin{cases}\left(\dfrac{3}{2}\left(\dfrac{5-y}{0.5}\right)^2+\dfrac{5-y}{0.5}\right)\left(\dfrac{1}{0.5}\right) & \text{si} \quad \dfrac{9}{2}<y<5\\ 0 & \text{caso contrario}\end{cases}$$

Ahora calculamos la $E(Y)$

$$E(Y)=\int_{9/2}^{5} y\left(\frac{3}{2}\left(\frac{5-y}{0.5}\right)^2+\frac{5-y}{0.5}\right)\frac{1}{0.5}dy=\frac{223}{46}$$

***Aplicando el teorema anterior los cálculos se reducen:***

$$E(Y) = E(h(X)) = \int_0^1 \underbrace{(5-0.5x)}_{h(x)} \underbrace{\left(\frac{3}{2}x^2 + x\right)}_{f(x)} dx = \frac{223}{46}$$

Notar que de la misma forma que en el caso discreto, si $h(x) = ax + b$, es decir si $h$ es una función lineal, aplicando las propiedades de linealidad de la integral tenemos

$$E(aX + b) = aE(X) + b$$

En el ejemplo anterior se podía encontrar la esperanza de $Y$ haciendo

$$E(Y) = E(h(X)) = E(5 - 0.5X) = 5 - 0.5E(X) = 5 - 0.5 \times \frac{17}{24} = \frac{223}{46}$$

**<u>Varianza de una variable aleatoria continua</u>**

Sea $X$ una v.a. continua con *f.d.p. f(x)* y sea $E(X) = \mu$, entonces la varianza de $X$ es

$$V(X) = \sigma_X^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

La interpretación de la varianza de una v.a. continua es la misma que para el caso discreto.
Además sigue valiendo la igualdad

$$V(X) = E(X^2) - \mu^2$$

Pues en la demostración hecha para el caso discreto si sustituyen las sumatorias por integrales.
Por la misma razón, también vale que

$$V(aX + b) = a^2 V(X) \qquad \text{y} \qquad \sigma_{aX+b} = |a|\sigma_X$$

<u>Ejemplo</u>:
Calculamos la varianza de $Y = 5 - 0.5X$

$$V(Y) = 0.5^2 V(X)$$

$$V(X) = E(X^2) - \mu^2 = E(X^2) - \left(\frac{17}{24}\right)^2$$

$$E(X^2) = \int_0^1 x^2 \left(\frac{3}{2}x^2 + x\right) dx = \left(\frac{3}{2}\frac{x^5}{5} + \frac{x^4}{4}\right)\Bigg|_0^1 = \left(\frac{3}{2} \times \frac{1}{5} + \frac{1}{4}\right) = \frac{11}{20}$$

$$\therefore \quad V(X) = \frac{11}{20} - \left(\frac{17}{24}\right)^2 \quad \Rightarrow \quad V(Y) = 0.5^2 \left(\frac{11}{20} - \left(\frac{17}{24}\right)^2\right) = \frac{139}{11520}$$

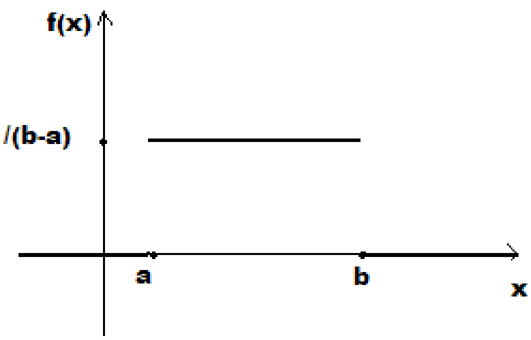
## 3.7 - Variables aleatorias continuas importantes

### *Distribución uniforme*

La distribución continua más sencilla es análoga a su contraparte discreta.

Una v.a. continua $X$ se dice que tiene ***distribución uniforme en el intervalo*** $[a,b]$, con $a < b$, si tiene función de densidad de probabilidad dada por

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{si} \quad a \le x \le b\\ 0 & \text{caso contrario}\end{cases}$$




La figura muestra la gráfica de la *f.d.p.*

Notación: $X \sim U[a,b]$

Es fácil verificar que *f(x)* es una *f.d.p.* pues $f(x) \ge 0$ para todo $x$, y además

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = \frac{1}{b-a}\, x\Big|_a^b = \frac{1}{b-a}(b-a) = 1$$

La F.d.a. para $a \le x \le b$ sería

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt = \int_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}$$

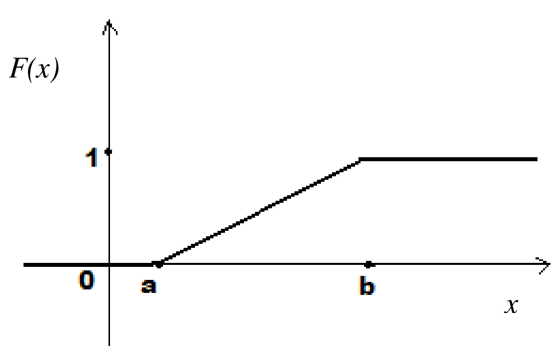
Para $x < a$, tenemos que $P(X \le x) = \int_{-\infty}^{x} 0dt = 0$

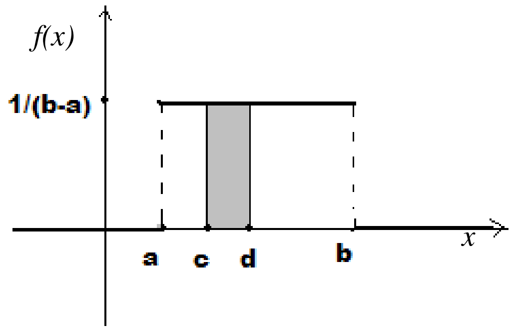
Y para $x > b$ tenemos que $P(X \le x) = \int_{-\infty}^{a} 0dt + \int_a^b \frac{1}{b-a}dt + \int_b^x 0dt = 1$

Por lo tanto la F.d.a. es

$$F(x)=\begin{cases}0 & \text{si} \quad x<a\\ \dfrac{x-a}{b-a} & \text{si} \quad a \le x \le b\\ 1 & \text{si} \quad x>b\end{cases}$$

Y su gráfica es






Observación:

Si $X \sim U[a,b]$ y $a \le c < d \le b$, entonces

$$P(c \le X \le d) = F(d) - F(c) = \frac{d-a}{b-a} - \frac{c-a}{b-a} = \frac{d-c}{b-a}$$

Ejemplo:

Los colectivos de una determinada línea llegan a una parada en particular en intervalos de 15 minutos comenzando desde las 7 A.M. Esto es, ellos llegan a la parada a las 7, 7:15, 7:30, 7:45 y así siguiendo. Si un pasajero llega a la parada en un tiempo que se puede considerar una v.a. distribuida uniformemente entre 7 y 7:30, encontrar la probabilidad de que

a) el pasajero espere menos de 5 minutos al colectivo

b) el pasajero espere más de 10 minutos al colectivo

Solución:

Sea X: "tiempo en minutos desde las 7 hs en que el pasajero llega a la parada"

Entonces podemos considerar que $X \sim U[0,30]$

a) si el pasajero espera menos de 5 minutos al colectivo, entonces llega a la parada entre las 7:10 y 7:15 o entre las 7:25 y 7:30, entonces la probabilidad pedida es

$$P(10 \le X \le 15) + P(25 \le X \le 30) = \int_{10}^{15} \tfrac{1}{30} dx + \int_{25}^{30} \tfrac{1}{30} dx = \frac{1}{3}$$

O también se puede plantear directamente

$$P(10 \le X \le 15) + P(25 \le X \le 30) = \frac{15-10}{30} + \frac{30-25}{30} = \frac{2 \times 5}{30} = \frac{1}{3}$$

b) Análogamente si debe esperar más de 10 minutos, deberá llegar entre las 7 y las 7:05 o entre las 7:15 y 7:20, por lo tanto la probabilidad pedida es

$$P(0 \le X \le 5) + P(15 \le X \le 20) = \frac{5}{30} + \frac{5}{30} = \frac{2 \times 5}{30} = \frac{1}{3}$$

**Esperanza y varianza**

Sea una $X$ variable aleatoria continua distribuida uniformemente en el intervalo $[a,b]$, es decir $X \sim U[a,b]$. Entonces, $E(X) = \dfrac{a+b}{2}$ y $V(X) = \dfrac{(b-a)^2}{12}$.

Dem.)

Recordamos que la *f.d.p.* de una v.a. $X \sim U[a,b]$ es

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{si } \ a\le x\le b\\ 0 & \text{caso contrario}\end{cases}$$

Entonces:

$$E(X)=\int_{-\infty}^{\infty} f(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{b+a}{2}$$

Notar que (*a*+*b*)/2 representa el punto medio del intervalo $[a,b]$ (como es de esperar por el significado de la distribución y de la esperanza):
Calculamos ahora la varianza de *X*

Deseamos calcular $V(X)=E(X^2)-[E(X)]^2$, es decir

$V(X)=E(X^2)-\left(\dfrac{a+b}{2}\right)^2$. Debemos obtener $E(X^2)$:

$E(X^2)=\displaystyle\int_a^b x^2\frac{1}{b-a}dx=\frac{1}{b-a}.\left.\frac{x^3}{3}\right|_a^b=\frac{1}{3}\frac{b^3-a^3}{b-a}=\frac{b^2+b.a+a^2}{3}$. Entonces:

$$V(X)=\frac{b^2+b.a+a^2}{3}-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}$$

## ***Distribución normal o gaussiana***

Sea *X* una v.a. Decimos que tiene distribución normal con parámetros $\mu$ y $\sigma$ si su *f.d.p.* es de la forma

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty<x<\infty \qquad (13)$$

Donde $\mu\in R$ y $\sigma>0$

Notación: $X \sim N(\mu,\sigma^2)$

Para darse una idea de la forma de la gráfica notar que:

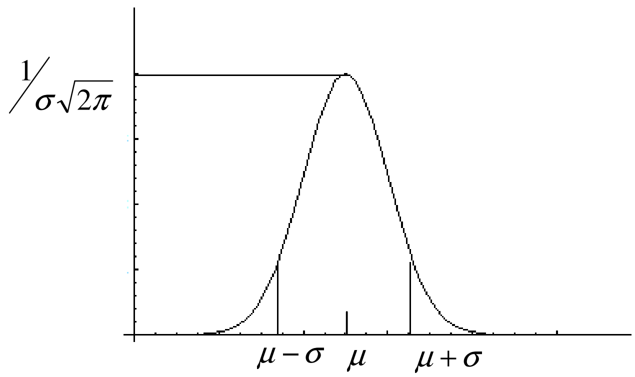
1- $f(x)$ es simétrica alrededor de $\mu$, es decir $f(\mu+x)=f(\mu-x)$ para todo *x*

2- $\lim_{x\to\pm\infty} f(x)=0$ (eje *x* asíntota horizontal)

3- Si planteamos $\dfrac{d}{dx}f(x)=0 \ \Rightarrow \ x=\mu$. Se pude verificar que en $x=\mu$ la función tiene un máximo absoluto, $f(\mu)=\dfrac{1}{\sigma\sqrt{2\pi}}$

4- Si planteamos $\frac{d^2}{dx^2} f(x) = 0 \Rightarrow x = \mu \pm \sigma$. Se puede verificar que en $x = \mu - \sigma$ y en $x = \mu + \sigma$ la función tiene dos puntos de inflexión, y además en el intervalo $(\mu - \sigma, \mu + \sigma)$ la función es cóncava hacia abajo y fuera de ese intervalo es cóncava hacia arriba

La gráfica de $f(x)$ tiene forma de campana

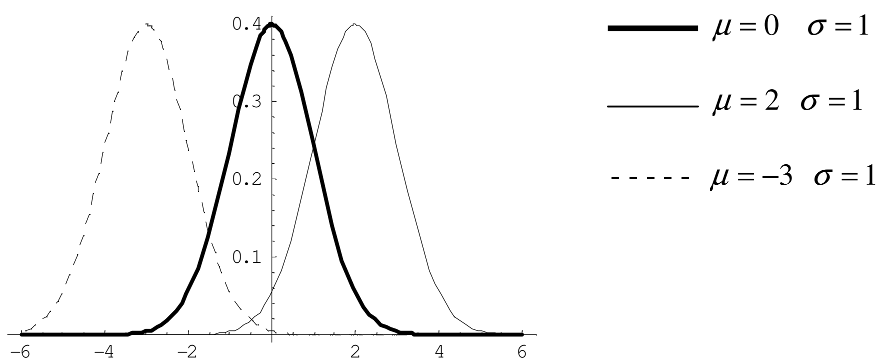


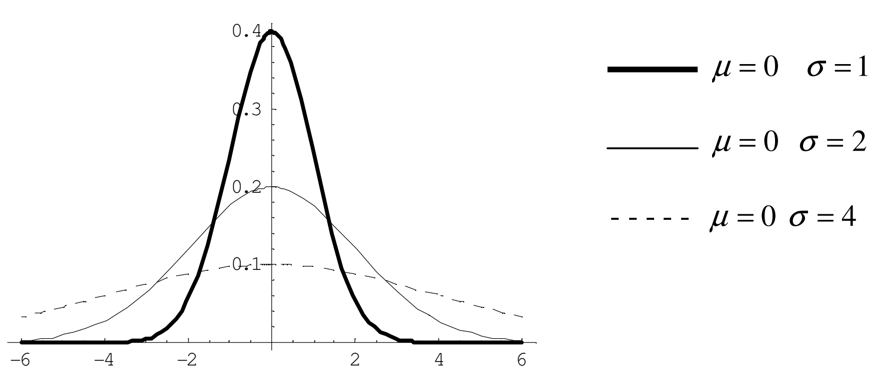

Observación:

Cuando $\mu$ varía la gráfica de la función se ***traslada***, $\mu$ ***es un parámetro de posición***.

Cuando $\sigma$ aumenta, la gráfica se "achata", cuando $\sigma$ disminuye la gráfica se hace mas "puntiaguda", se dice que $\sigma$ ***es un parámetro de escala***.

En las siguientes figuras vemos cómo varía la gráfica de *f(x)* con la variación de los parámetros



Se puede probar que f(x) es una *f.d.p.* es decir que

a) $f(x) \geq 0$ para todo $x$

b) $\int_{-\infty}^{\infty} f(x)dx = 1$

Que a) es cierta es obvio; para probar b) es necesario recurrir al cálculo en dos variables (no lo demostramos).

Si $\mu = 0$ y $\sigma = 1$ entonces se dice que *X* tiene ***distribución normal estándar***. Se anota $X \sim N(0,1)$

En este caso la *f.d.p.* se simboliza con $\varphi(x)$, es decir

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$$

En este caso la gráfica de la densidad es simétrica con respecto al origen.
La *F.d.a.* de una v.a. normal estándar se anota $\Phi(x)$

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

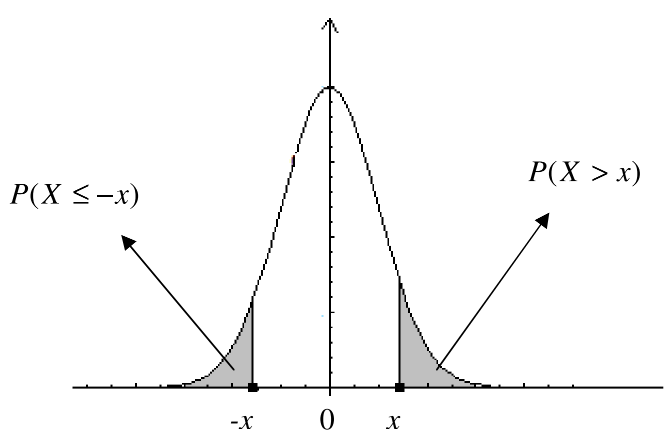
Esta integral no puede expresarse en términos de funciones elementales, por lo tanto se calcula $\Phi(x)$

Para valores específicos de *x* mediante una aproximación numérica.

Esto ya está hecho, existen tablas de la función de distribución acumulada de la normal estándar para valores de *x* que oscilan en general entre -4 y 4, pues para valores de *x* ***menores que -4***, $\Phi(x) \approx 0$, y para valores de *x* ***mayores que 4***, $\Phi(x) \approx 1$

Notar que como la $\varphi(x)$ es simétrica con respecto al origen entonces

$$\Phi(-x) = P(X \leq -x) = P(X > x) = 1 - P(X \leq x) = 1 - \Phi(x)$$



Por ejemplo, si $X \sim N(0,1)$ entonces utilizando la tabla de la *F.d.a.* de $X$

a) $P(X \leq 1.26) = \Phi(1.26) = 0.89616$

b) $P(X > 1.26) = 1 - P(X \leq 1.26) = 1 - \Phi(1.26) = 1 - 0.89616 = 0.10384$

c) $P(X > -1.37) = P(X \leq 1.37) = \Phi(1.37) = 0.91465$

d) $P(-1.25 < X < 0.37) = P(X < 0.37) - P(X < -1.25) = \Phi(0.37) - \Phi(-1.25) =$

$= \Phi(0.37) - (1 - \Phi(1.25)) = 0.64431 - (1 - 0.89435) = 0.53866$

e) ¿Para qué valor $x$ se cumple que $P(-x < X < x) = 0.95$?

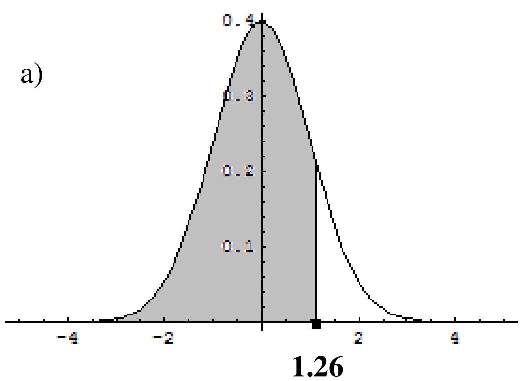
Tenemos que $P(-x < X < x) = \Phi(x) - \Phi(-x) = \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1$

Por lo tanto $2\Phi(x) - 1 = 0.95 \quad \Rightarrow \quad \Phi(x) = \dfrac{0.90 + 1}{2} = 0.975$

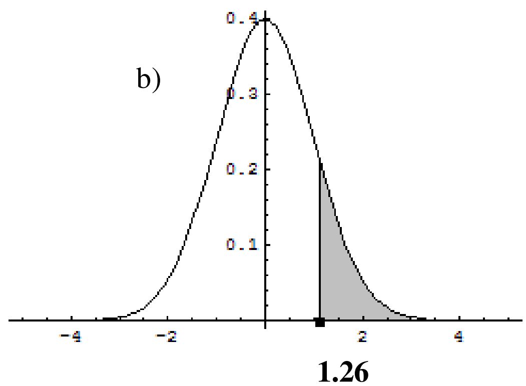
Observamos en la tabla de la F.d.a. que $x = 1.96$, pues $\Phi(1.96) = 0.975$

Para los incisos a) , b) y c) se grafican las regiones correspondientes

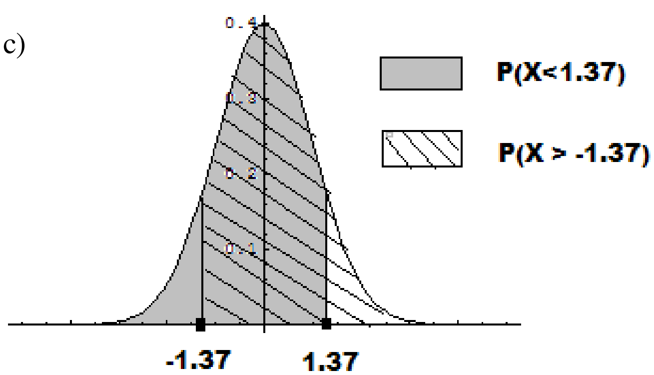
a)




b)




c)




Una propiedad importante de la distribución normal es que si $X \sim N(\mu, \sigma^2)$ entonces la v.a. $Y = aX + b$ con $a$ y $b$ números reales, $a \neq 0$, tiene también distribución normal pero con parámetros $a\mu + b$ y $a^2\sigma^2$, es decir

$$X \sim N(\mu, \sigma^2) \ \Rightarrow \ aX + b \sim N(a\mu + b, a^2\sigma^2) \qquad (14)$$

Podemos demostrar lo anterior primero hallando la *F.d.a.* de $Y$

$$G(y)=P(Y\le y)=P(aX+b\le y)\underset{\substack{\downarrow\\ a>0}}{=}P\left(X\le\frac{y-b}{a}\right)=F\left(\frac{y-b}{a}\right)$$ donde $F$ es la F.d.a. de $X$

Por lo tanto, la f.d.p. de $Y$ la obtenemos derivando $G(y)$

$$g(y)=\frac{d}{dy}G(y)=\frac{d}{dy}F\left(\frac{y-b}{a}\right)=f\left(\frac{y-b}{a}\right)\frac{1}{a}=\frac{1}{a}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2}$$ donde $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Operando en el exponente se llega a

$$g(y)=\frac{1}{a}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2}$$

Si $a<0$ entonces

$$G(y)=P(Y\le y)=P(aX+b\le y)=P\left(X\ge\frac{y-b}{a}\right)=1-F\left(\frac{y-b}{a}\right)$$

Y derivando con respecto a $y$ obtenemos

$$g(y)=\frac{d}{dy}G(y)=\frac{d}{dy}\left(1-F\left(\frac{y-b}{a}\right)\right)=-f\left(\frac{y-b}{a}\right)\frac{1}{a}=-\frac{1}{a}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2}$$

O sea, para $a\ne 0$ la *f.d.p.* de $Y$ es $g(y)=\frac{1}{|a|}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2}$

Y comparando con (13) se deduce que $Y=aX+b\sim N(a\mu+b,a^2\sigma^2)$

Una consecuencia importante del resultado (14) es que

$$\boxed{\text{si}\quad X\sim N(\mu,\sigma^2)\quad\text{entonces}\quad Y=\frac{X-\mu}{\sigma}\sim N(0,1)}\qquad(15)$$

Notar que $Y$ se pude escribir como $Y=\frac{1}{\sigma}X+\left(-\frac{\mu}{\sigma}\right)$ es decir claramente $Y$ es una función lineal de $X$

Por lo tanto aplicamos el resultado (14) con $a=\frac{1}{\sigma}$ y $b=-\frac{\mu}{\sigma}$ y llegamos a (15).

Si $X\sim N(\mu,\sigma^2)$ entonces la *F.d.a.* de X es

$$F(x)=P(X\le x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}dt$$

*F(x)* no puede expresarse en términos de funciones elementales y sólo hay tablas de la F.d.a. de la normal estándar.
Para calcular *F(x)* procedemos de la siguiente forma

$$F(x)=P(X\le x)=P\left(\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right)=P\left(Y\le\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$Y\sim N(0,1)$$

Ejemplos:

1- Si $X\sim N(3,9)$ entonces

a) $P(2<X<5)=P\left(\frac{2-3}{3}<\frac{X-3}{3}<\frac{5-3}{3}\right)=\Phi\left(\frac{5-3}{3}\right)-\Phi\left(\frac{2-3}{3}\right)=\Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{1}{3}\right)=0.3779$

b) $P(X>0)=P\left(\frac{X-3}{3}>\frac{0-3}{3}\right)=1-\Phi(-1)=\Phi(1)=0.8413$

c) $P(|X-3|>6)=1-P(|X-3|\le 6)=1-[P(-6\le X-3\le 6)]=1-\left[P\left(\frac{-6}{3}\le\frac{X-3}{3}\le\frac{6}{3}\right)\right]=$

$=1-[\Phi(2)-\Phi(-2)]=2[1-\Phi(2)]=0.0456$

2- Hay dos máquinas para cortar corchos destinados para usarse en botellas de vino. La primera produce corchos con diámetros que están normalmente distribuidos con media de 3 cm y desviación estándar de 0.1 cm. La segunda máquina produce corchos con diámetros que tienen una distribución normal con media de 3.04 cm y desviación estándar de 0.02 cm. Los corchos aceptables tienen diámetros entre 2.9 cm y 3.1 cm. ¿Cuál máquina tiene mas probabilidad de producir un corcho aceptable?

Solución:

Sean las variables aleatorias

*X*: “diámetro de un corcho producido por la máquina 1”

*Y:* “diámetro de un corcho producido por la máquina 2”

Entonces $X\sim N(3,0.1^2)$ y $Y\sim N(3.04,0.02^2)$

Calculamos cuál es la probabilidad que la máquina 1 produzca un corcho aceptable

$P(2.9\le X\le 3.1)=P\left(\frac{2.9-3}{0.1}\le\frac{X-3}{0.1}\le\frac{3.1-3}{0.1}\right)=\Phi\left(\frac{3.1-3}{0.1}\right)-\Phi\left(\frac{2.9-3}{0.1}\right)=$

$=\Phi(1)-\Phi(-1)=2\Phi(1)-1=0.6826$

Análogamente para la máquina 2

$P(2.9\le Y\le 3.1)=P\left(\frac{2.9-3.04}{0.02}\le\frac{Y-3.04}{0.02}\le\frac{3.1-3.04}{0.02}\right)=\Phi\left(\frac{3.1-3.04}{0.02}\right)-\Phi\left(\frac{2.9-3.04}{0.02}\right)=$

$=\Phi(3)-\Phi(-7)=0.9987-0=0.9987$

Entonces es más probable que la máquina 2 produzca corchos aceptables.

3- El dispositivo automático de apertura de un paracaídas militar de carga se ha diseñado para abrir el paracaídas cuando éste se encuentre a 200 m de altura sobre el suelo. Supongamos que la altitud de apertura en realidad tiene una distribución normal con valor medio de 200 m y desviación estándar de 30 m. Habrá un daño al equipo si el paracaídas se abre a una altitud de menos de 100 m. ¿Cuál es la probabilidad de que haya daño a la carga en al menos uno de cinco paracaídas lanzados independientemente?

Solución:

Sea la v.a. *X*: “altitud de apertura en metros de un paracaídas”

Entonces $X \sim N(200,30^2)$

Calculamos $P(X < 100) = \Phi\left(\dfrac{100-200}{30}\right) = \Phi(-3.33) = 0.0004$

Consideramos ahora la v.a. $Y$: "número de paracaídas entre 5 que se abren a menos de 100 metros"
Podemos considerar que $Y \sim B(5,\ 0.0004)$
Por lo tanto hay que calcular $P(Y \geq 1)$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{5}{0} 0.0004^0 (1-0.0004)^{5-0} = 1-(1-0.0004)^5 = 0.0019984$$

4- Supóngase que la resistencia a romperse (en Kgr) de fibras de yute está descrita por una v.a. continua $X$ normalmente distribuida con $\mu = E(X) = 165$ Kgr y $\sigma^2 = V(X) = 9$ (Kgr)$^2$. suponiendo además que una muestra de esta fibra se considera defectuosa si $X < 162$. Cuál es la probabilidad de que una fibra elegida al azar sea defectuosa?

Solución:
Deseamos conocer $P(X < 162)$

$$P(X < 162) = P\left(\frac{X-165}{3} < \frac{162-165}{3}\right) = P\left(\frac{X-165}{3} < -1\right) = \Phi(-1),$$

puesto que $Z = \dfrac{X-165}{3} \sim N(0,1)$. Entonces

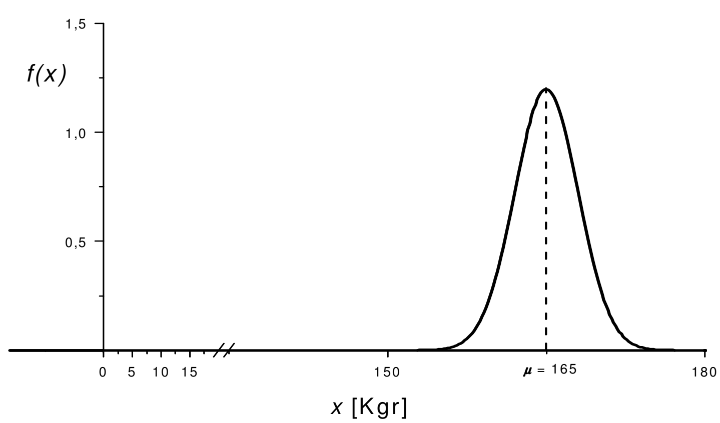
$P(X < 162) = \Phi(-1) = 1 - \Phi(1).$

De la tabla tenemos $\Phi(1) = 0.8413 \rightarrow \Phi(-1) = 1 - \Phi(1) = 0.1587$.

Es decir $P(X < 162) = 0.1587$.

Observación:
Uno puede objetar el usar una distribución normal para describir a la v.a. $X$ que representa la resistencia a romperse de la fibra ya que ésta es, obviamente, una cantidad no negativa, mientras que una v.a. normalmente distribuida puede tomar valores que varían entre $-\infty$ y $+\infty$. Sin embargo al modelar el problema con una normal (que aparentemente debería ser invalidada como modelo por lo señalado) vemos que les estamos asignando al suceso $\{X < 0\}$ una probabilidad prácticamente nula (ver también la figura siguiente):

$$P(X < 0) = P\left(\frac{X-165}{3} < \frac{0-165}{3}\right) = \Phi(-55) = 1 - \Phi(55) \approx 1 - 1 = 0.$$




En casos como estos se justifica usar la distribución normal para modelar situaciones en que la variable aleatoria considerada puede tomar, por su significado, sólo valores positivos, aún cuando la normal permita tomar valores tanto positivos como negativos por cuanto la probabilidad de que la v.a. tome valores negativos es prácticamente nula.

**Esperanza y varianza de una variable aleatoria con distribución normal**

Sea $X \sim N(\mu, \sigma^2)$ entonces $E(X) = \mu$ y $V(X) = \sigma^2$

Dem.) Usaremos el resultado:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = 1$$

Calculamos la esperanza de $X$

$$E(X) = \frac{1}{\sqrt{2\pi}.\sigma} \int_{-\infty}^{+\infty} x.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$ hacemos la sustitución

$$t = \frac{x-\mu}{\sigma} \quad \rightarrow \quad x = (\sigma.t + \mu), \quad dt = \frac{dx}{\sigma} \quad y \quad -\infty < t < +\infty .$$

Luego:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma.t + \mu).e^{-\frac{t^2}{2}} dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}} dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Considerado como función de $t$, el integrando de la primera integral $f(t) = te^{-\frac{t^2}{2}}$ es una función impar de $t$, es decir, verifica $f(-t) = -f(t)$. En consecuencia la integral a lo largo de un intervalo simétrico con respecto al origen, como lo es $(-\infty, \infty)$ se anula.

El segundo sumando, por su parte, es justamente $\mu$ veces la integral $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = 1$. Entonces $E(X) = \mu$.

Veamos el cálculo de la varianza de $X$

$V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$. Nos queda evaluar $E(X^2)$,

$$E(X^2) = \frac{1}{\sqrt{2\pi}.\sigma} \int_{-\infty}^{+\infty} x^2.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$. Hacemos nuevamente la sustitución

$$t = \frac{x-\mu}{\sigma} \quad \rightarrow \quad x = (\sigma.t + \mu), \quad dt = \frac{dx}{\sigma} \quad y \quad -\infty < t < +\infty .$$ Reemplazando:

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma.t + \mu)^2.e^{-\frac{t^2}{2}} dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2.e^{-\frac{t^2}{2}} dt + \frac{2\sigma.\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t.e^{-\frac{t^2}{2}} dt + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Ahora el integrando de la segunda integral es impar y por lo tanto se anula. Entonces

$$E\left(X^2\right)=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t^2.e^{-\frac{t^2}{2}}dt+\mu^2 .$$

Debemos calcular la integral $\int_{-\infty}^{+\infty} t^2.e^{-\frac{t^2}{2}}dt$. Integrando por partes

con $\begin{cases} u=t & du=dt \\ dv=te^{-\frac{t^2}{2}}dt & v=-e^{-\frac{t^2}{2}} \end{cases}$ $\rightarrow$ $\int_{-\infty}^{+\infty} t^2.e^{-\frac{t^2}{2}}dt=-te^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}dt$ .

El corchete de Barrow se anula en ambos extremos y la integral es justamente $\sqrt{2\pi}I=\sqrt{2\pi}$ . Entonces : $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t^2.e^{-\frac{t^2}{2}}dt=1$. Por lo tanto $E\left(X^2\right)=\sigma^2+\mu^2$ y, en consecuencia,

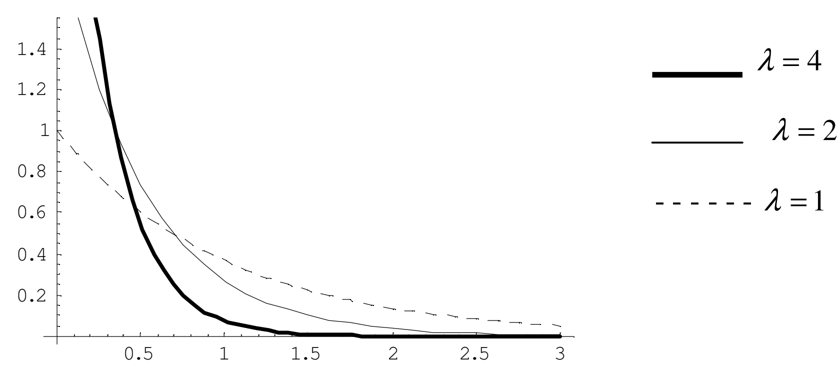
$V(X)=E\left(X^2\right)-\mu^2=\sigma^2+\mu^2-\mu^2=\sigma^2$

## ***Distribución exponencial***

Sea X una v.a. continua. Se dice que tiene distribución exponencial con parámetro $\lambda$ si su f.d.p. es de la forma

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{si } \ x\geq 0\\ 0 & \text{caso contrario}\end{cases} \qquad \text{donde} \quad \lambda>0$$

La distribución exponencial se utiliza algunas veces para modelar el tiempo que transcurre antes de que ocurra un evento. A menudo se lo llama ***tiempo de espera***.
La siguiente figura muestra la gráfica de la densidad para diferentes valores del parámetro




Es fácil verificar que *f(x)* es una densidad bien definida
a) Claramente $f(x)\geq 0$ para todo $x$

b) $\int_0^{\infty}\lambda e^{-\lambda t}dt=\lambda\left(\frac{e^{-\lambda t}}{-\lambda}\right)\Bigg|_0^{\infty}=1$

La F.d.a. de una v.a. exponencial es sencilla:

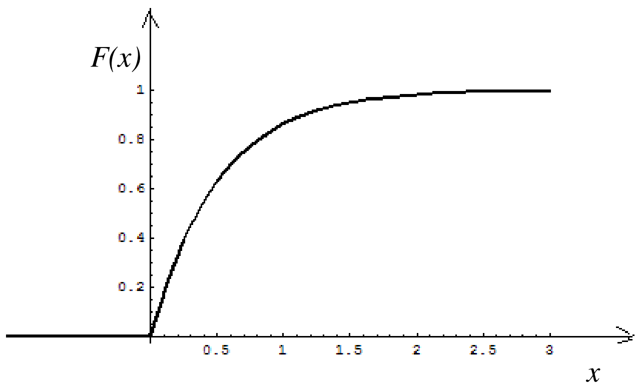
Si $x<0$ entonces $F(x)=P(X\leq x)=0$

Si $x \geq 0$ entonces $F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = \lambda \left( \dfrac{e^{-\lambda t}}{-\lambda} \right) \Bigg|_0^x = 1 - e^{-\lambda x}$

Por lo tanto

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{caso contrario} \end{cases}$$

Su gráfica es




<u>Notación</u>: $X \sim Exp(\lambda)$

<u>Ejemplo</u>:
Supongamos que el tiempo, en segundos, de respuesta en cierta terminal de computadora en línea (es decir el tiempo transcurrido entre el fin de la consulta del usuario y el principio de la respuesta del sistema a esa consulta) se puede considerar como una variable aleatoria con distribución exponencial con parámetro $\lambda = 0.2$. Calcular
a) la probabilidad de que el tiempo de respuesta sea a lo sumo 10 segundos.
b) la probabilidad de que el tiempo de respuesta esté entre 5 y 10 segundos.

<u>Solución</u>:
Sea $X$ la v.a., entonces $X \sim Exp(0.2)$
a) Se pide calcular $P(X \leq 10)$
Por lo tanto
$P(X \leq 10) = F(10) = 1 - e^{-0.2 \times 10} = 1 - 0.135 = 0.865$
b) $P(5 \leq X \leq 10) = F(10) - F(5) = \left(1 - e^{-0.2 \times 10}\right) - \left(1 - e^{-0.2 \times 5}\right) = 0.233$

**<u>Esperanza y varianza de la distribución exponencial</u>**

$$\boxed{\text{Sea } X \sim Exp(\lambda) \text{ entonces } E(X) = \frac{1}{\lambda} \quad \text{y} \quad V(X) = \frac{1}{\lambda^2}}$$

Dem.)
Es útil calcular en general $E(X^k)$ el ***momento de orden k con respecto al origen*** siendo $k$ un número natural

$$E(X^k) = \int_0^{+\infty} x^k f(x)dx = \lambda \int_0^{+\infty} x^k e^{-\lambda x} dx \qquad (k = 0,1,2,...).$$

Integrando por partes:

$$\begin{cases} u = x^k \\ dv = \lambda e^{-\lambda x} dx \end{cases} \qquad \begin{cases} du = kx^{k-1}dx \\ v = -e^{-\lambda x} \end{cases} \quad \rightarrow \quad \mu_k = -x^k e^{-\lambda x}\Big|_0^{+\infty} + k\int_0^{+\infty} x^{k-1}e^{-\lambda x}dx$$. El corchete de Barrow se anula y queda $E(X^k) = k\lambda \frac{1}{\lambda}\int_0^{+\infty} x^{k-1}e^{-\lambda x}dx$. Aplicando reiteradamente se llega finalmente a

$$E(X^k) = k(k-1)...2.1\left(\frac{1}{\lambda}\right)^k \lambda \int_0^{+\infty} e^{-\lambda x}dx = k!\left(\frac{1}{\lambda}\right)^k$$

Por lo tanto tenemos para la esperanza y la varianza:

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = E(X^2) - \mu^2 = 2\left(\frac{1}{\lambda}\right)^2 - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

**Propiedades de la distribución exponencial**

1- **Relación entre la distribución exponencial y el proceso de Poisson**.

Sea *T* el tiempo de espera hasta el siguiente evento en un proceso de Poisson con parámetro $\lambda$.
Veamos cuál es la *F.d.a*. de *T*.
Si $t < 0$ entonces claramente $F(t) = P(T \le t) = 0$
Si $t \ge 0$ entonces para hallar $F(t) = P(T \le t)$ consideramos el evento complementario de $\{T \le t\}$.
Notar que $\{T > t\}$ si y solo si no ocurre ningún evento durante las siguientes $t$ unidades de tiempo.
Si *X*: "número de eventos que ocurren en las siguientes *t* unidades de tiempo", entonces $\{T > t\}$ ocurre si y solo si $\{X = 0\}$ ocurre, por lo tanto $P(T > t) = P(X = 0)$
Como $X \sim P(\lambda t)$ entonces

$$P(T > t) = P(X = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

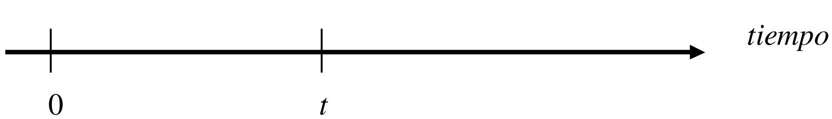
Por lo tanto

$$F(t) = P(T \le t) = 1 - P(T > t) = 1 - e^{-\lambda t}$$

Como *F(t)* es la F.d.a. de una v.a. exponencial, entonces $T \sim Exp(\lambda)$
Por lo tanto

> Si los eventos siguen un proceso de Poisson con parámetro $\lambda$, y si *T* representa el tiempo de espera desde cualquier punto inicial hasta el próximo evento, entonces $T \sim Exp(\lambda)$

Consideremos un aplicación hidrológica de estas ideas



i) $X$ = Nº de inundaciones en un período $[0,t] \rightarrow X \sim P(\lambda t)$

ii) $\lambda$ = Nº medio de inundaciones por unidad de tiempo $\rightarrow \lambda = \dfrac{E(X)}{t}$

iii) $T$ = Tiempo transcurrido entre inundaciones $\rightarrow T \sim Exp(\lambda)$

iv) $E(T)$ = Tiempo medio de retorno de las inundaciones $\rightarrow E(T) = \dfrac{1}{\lambda}$.

2- **<u>Propiedad falta de memoria</u>**

La distribución exponencial tiene una propiedad conocida como falta de memoria, que se muestra en el siguiente ejemplo:
El tiempo de vida, en años, de un circuito integrado particular tiene una distribución exponencial con parámetro 0.5. Encuentre la probabilidad de que el circuito dure más de tres años
Sea la v.a. X : “tiempo de vida, en años, de un circuito integrado particular”, entonces $X \sim Exp(0.5)$
Y $P(X > 3) = 1 - F(3) = e^{-1.5} = 0.223$
Supongamos ahora que actualmente un circuito tiene cuatro años y aún funciona. Se quiere hallar la probabilidad de que funcione tres años más.
Por lo tanto planteamos una probabilidad condicional

$$P / X > 7 / X > 4) = \frac{P(X > 7 \text{ y } X > 4)}{P(X > 4)} = \frac{P(X > 7)}{P(X > 4)} = \frac{e^{-0.5\times 7}}{e^{-0.5\times 4}} = e^{-0.5\times(7-4)} = e^{-1.5} = 0.223$$

Observamos que $P / X > 7 / X > 4) = P(X > 7 - 4) = P(X > 3)$
En general, la probabilidad que se tenga que esperar *t* unidades adicionales, dado que ya se han esperado *s* unidades, es la misma que la probabilidad de que se tenga que esperar *t* unidades desde el inicio. La distribución exponencial no “recuerda” cuánto tiempo se ha esperado.
En particular, si el tiempo de vida de un componente sigue una distribución exponencial, entonces la probabilidad de que un componente que tiene *s* unidades de tiempo dure *t* unidades de tiempo adicionales es la misma que la probabilidad de que un componente nuevo dure t unidades de tiempo. En otras palabras, un componente cuyo tiempo de vida siga una distribución exponencial no muestra ningún síntoma de los años o del uso.
Los cálculos hechos en el ejemplo anterior se pueden repetir para valores cualesquiera *s* y *t* y entonces se pude probar que

si $X \sim Exp(\lambda)$ y $t$ y $s$ son números positivos, entonces $P(X > t + s / X > s) = P(X > t)$

# *4 - DESIGUALDAD DE CHEBYSHEV- LEY DE LOS GRANDES NUMEROS*

La desigualdad de Chebyshev es una importante herramienta teórica. Entre otras aplicaciones constituirá un medio para comprender cómo la varianza mide la variabilidad de una dada variable aleatoria, con respecto a su esperanza matemática. También nos permitirá establecer con más precisión el hecho, reiteradamente señalando, de que la frecuencia relativa $f_A$ de un suceso $A$ asociado a un experimento aleatorio $\varepsilon$ tiende, cuando el número de repeticiones de $\varepsilon$ se hace infinitamente grande, a la probabilidad $P(A)$ (resultado conocido como la Ley de los grandes números). Pero además es de utilidad práctica pues, al constituir una cota de ciertas probabilidades, nos podrá servir como una estimación de esas mismas probabilidades.

## 4.1-Desigualdad de Chebyshev

Sea $X$ una variable aleatoria cuya esperanza es $E(X)$, sea $c$ un número real cualquiera y supongamos que $E\left[(X-c)^2\right]$ existe y es finito. Entonces

$$\boxed{\forall \varepsilon > 0 \qquad P(|X-c| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} E\left[(X-c)^2\right]}$$

Dem.) Consideraremos el caso en que $X$ es una v.a. continua. El caso de una v.a. discreta se demuestra en forma similar cambiando integrales por sumas.

Sea, entonces, $f(x)$ la *fdp* de *X*. Tenemos:

$$P(|X-c| \geq \varepsilon) = \int_{x:|x-c|\geq\varepsilon} f(x)dx$$

Ahora bien, los valores de $x$ que verifican $|x-c| \geq \varepsilon$ son los mismos que verifican $\frac{(x-c)^2}{\varepsilon^2} \geq 1$. Entonces, puesto que $f(x) \geq 0$ y que ambos miembros en la desigualdad anterior son también no negativos es:

$$P(|X-c| \geq \varepsilon) = \int_{x:|x-c|\geq\varepsilon} 1.f(x)dx \leq \int_{x:|x-c|\geq\varepsilon} \frac{(x-c)^2}{\varepsilon^2}.f(x)dx \leq \int_{-\infty}^{\infty} \frac{(x-c)^2}{\varepsilon^2}.f(x)dx$$

donde la última desigualdad proviene del hecho de que simplemente extiendo los límites de integración a todos los reales pero siendo siempre el integrando positivo. De manera que estoy agregando una contribución positiva sobre el valor de la integral $\int_{x:|x-c|\geq\varepsilon} \frac{(x-c)^2}{\varepsilon^2}.f(x)dx$.

Si tenemos presente la expresión para la esperanza de una función $H(X)$ de una variable aleatoria *X*:

$\int_{-\infty}^{\infty} H(x) f(x) dx = E[H(X)]$ y lo aplicamos a $H(x) = \dfrac{(x-c)^2}{\varepsilon^2}$, tenemos finalmente:

$$P(|X-c| \geq \varepsilon) \leq E\left[\frac{(X-c)^2}{\varepsilon^2}\right] = \frac{1}{\varepsilon^2} E\left[(X-c)^2\right]$$

**Observación:** La desigualdad lleva el nombre del matemático ruso que la descubrió. Su nombre aparece en una variedad de formas en la literatura: Chebyshev, Chebychev, Tchebyshev, etc.

## Formas alternativas de la desigualdad de Chebyshev.

Podemos escribir la desigualdad de Chebyshev en una serie de formas alternativas:

$a_1$) Tenemos en primer lugar la forma que acabamos de demostrar:

$$P(|X-c| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} E\left[(X-c)^2\right]$$

$a_2$) Si consideramos el suceso complementario a $|X-c| \geq \varepsilon$, es decir $|X-c| < \varepsilon$, podemos escribir, recordando que $P(\overline{A}) = 1 - P(A)$:

$P(|X-c| < \varepsilon) = 1 - P(|X-c| \geq \varepsilon) \geq 1 - \dfrac{1}{\varepsilon^2} E\left[(X-c)^2\right]$, esto es:

$$P(|X-c| < \varepsilon) \geq 1 - \frac{1}{\varepsilon^2} E\left[(X-c)^2\right]$$

$b_1$) Si en $a_1$) elegimos $c = E(X)$ tenemos $P(|X-E(X)| \geq \varepsilon) \leq \dfrac{1}{\varepsilon^2} E\left([X-E(X)]^2\right)$ y recordando la definición de la varianza: $V(X) = E\left([X-E(X)]^2\right)$ tenemos:

$$P(|X-E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

$b_2$) La correspondiente expresión para el suceso complementario es:

$$P(|X-E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2}$$

$c_1$) Si en $a_1$) elegimos $c = E(X)$ y además elegimos $\varepsilon = k\sigma_X = k\sqrt{V(X)}$, tenemos $P(|X-E(X)| \geq k\sigma_X) \leq \dfrac{V(X)}{(k\sigma_X)^2} = \dfrac{\sigma_X^2}{(k\sigma_X)^2}$, es decir:

$$P(|X-E(X)|\geq k\sigma_X)\leq \frac{1}{k^2}$$

$c_2$) Finalmente a correspondiente expresión para el suceso complementario en $c_1$) es:

$$P(|X-E(X)|< k\sigma_X)\geq 1-\frac{1}{k^2}$$

A continuación daremos un ejemplo en el que podremos apreciar cómo la desigualdad de Chebyshev nos permite tener estimaciones de ciertas probabilidades que en algunos casos mejoran las "estimaciones" triviales dadas por el axioma i) de las probabilidades, esto es, que $0\leq P(A)\leq 1$.

<u>Ejemplo</u> Deseamos estimar la probabilidad $P\left(|X-E(X)|\geq \frac{3}{2}\sigma_X\right)$

a) Sin conocer la distribución
b) Sabiendo que $X \sim U\left[1-\frac{1}{\sqrt{3}},\ 1+\frac{1}{\sqrt{3}}\right]$.

a) Una estimación de la probabilidad en consideración cuando no se conoce la distribución, es decir que vale cualquiera sea la distribución puede tenerse usando la desigualdad de Chebyshev en la forma $c_1$) : $P(|X-E(X)|\geq k\sigma_X)\leq \frac{1}{k^2}$ con $k=\frac{3}{2}$.
Tenemos en consecuencia:
$P\left(|X-E(X)|\geq \frac{3}{2}\sigma_X\right)\leq \frac{1}{(3/2)^2}=\frac{4}{9}=0.44$ . Vemos que, en este caso, la estimación mejora sustancialmente la cota superior trivial 1:

$$P\left(|X-E(X)|\geq \frac{3}{2}\sigma_X\right)\leq 0.44<1$$

Notemos que esta estimación es aplicable cualquiera se la distribución de $X$ y vale sea ésta discreta o continua.
b)Si sabemos que $X \sim U\left[1-\frac{1}{\sqrt{3}},\ 1+\frac{1}{\sqrt{3}}\right]$ tenemos toda la información y podemos calcular la probabilidad exactamente.

Tenemos:

$$E(X)=\frac{\left(1+\frac{1}{\sqrt{3}}\right)+\left(1`-\frac{1}{\sqrt{3}}\right)}{2}=1$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{\frac{\left[\left(1+\frac{1}{\sqrt{3}}\right)-\left(1`-\frac{1}{\sqrt{3}}\right)\right]^2}{12}} = \sqrt{\frac{\left(2/\sqrt{3}\right)^2}{12}} = \sqrt{\frac{4}{36}} = \sqrt{\frac{1}{9}} = \frac{1}{3}$$. Entonces:

$$P\left(|X-E(X)| \geq \frac{3}{2}\sigma_X\right) = P\left(|X-1| \geq \frac{1}{2}\right) = 1 - P\left(|X-1| < \frac{1}{2}\right) = 1 - P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right)$$

$$P\left(|X-E(X)| \geq \frac{3}{2}\sigma_X\right) = 1 - \frac{\frac{3}{2}-\frac{1}{2}}{\left(1+\frac{1}{\sqrt{3}}\right)-\left(1-\frac{1}{\sqrt{3}}\right)} = 1 - \frac{1}{\frac{2}{\sqrt{3}}} = 1 - \frac{\sqrt{3}}{2} = 0.134$$. Este valor exacto es menor que la cota superior dada por la desigualdad de Chebyshev:

$$P\left(|X-E(X)| \geq \frac{3}{2}\sigma_X\right) = 0.134 \leq 0.44 < 1.$$

## La varianza como una medida de la concentración de la *fdp* de una v.a. alrededor de la esperanza.

Podemos usar las formas *b*) de la desigualdad de Chebyshev para interpretar a la varianza $V(X)$ como una medida de la variabilidad de la variable aleatoria $X$ con respecto a su esperanza o en otras palabras de cómo la distribución de la v.a. $X$ se concentra o dispersa con respecto a la esperanza $E(X)$.

De la expresión $P(|X-E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$ vemos que, para un $\varepsilon$ dado, si $V(X)$ es muy pequeño entonces la probabilidad de que $X$ tome valores lejos de $E(X)$ es muy chica, es decir hay una gran probabilidad de que $X$ tome valores próximos a $E(X)$. Inversamente si $V(X)$ es grande, la probabilidad de que $X$ tome valores alejados de $E(X)$ puede ser también grande. Podemos precisar un poco más esto considerando el siguiente

**Teorema.** Si $V(X)=0$ entonces $P[X=E(X)]=1$. Decimos que $X=E(X)$ con probabilidad 1 ($X$ es igual a su esperanza con probabilidad 1).

Dem.) Para cualquier $\varepsilon > 0$ , si $V(X)=0$ tenemos de $b_2$) :

$P(|X-E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2} \rightarrow P(|X-E(X)| < \varepsilon) = 1$, donde me quedo sólo con la igualdad porque la probabilidad no pude superar a 1.

Puesto que $\varepsilon$ puede hacerse arbitrariamente pequeño, el teorema queda demostrado.

## 4.2 - La ley de los grandes números.

La ley de los grandes números establece en forma precisa el hecho que cuando el número de repeticiones de un experimento se hace muy grande, la frecuencia relativa $f_A$ de un suceso *A* relacionado con el experimento converge en sentido probabilístico a la probabilidad $P(A)$. Daremos una versión de la Ley de los grandes números conocida como la ***forma de Bernoulli***.

**Teorema** (Forma de Bernoulli de la ley de los grandes números)
Sea un experimento probabilístico y sea *A* un suceso asociado con él. Consideremos *n* repeticiones independientes del experimento. Sea $n_A$ el número de veces que ocurre *A* en las *n* repeticiones de forma tal que $f_A = \dfrac{n_A}{n}$ es la frecuencia relativa. Sea $P(A) = p$ (que se supone igual para todas las repeticiones). Entonces, para cualquier número $\varepsilon > 0$ se cumple
$P(|f_A - p| < \varepsilon) \geq 1 - \dfrac{p(1-p)}{n\varepsilon^2}$.

Dem.)
De acuerdo con su significado $n_A$, el número de veces que ocurre el suceso *A* en las *n* repeticiones, es una variable aleatoria distribuida binomialmente con parámetros *n* y *p*: $n_A \sim B(n, p)$ . Luego:

$$\begin{cases} E(n_A) = np \\ V(n_A) = np(1-p) \end{cases}$$

Por lo tanto

$$\begin{cases} E(f_A) = E\left(\dfrac{n_A}{n}\right) = \dfrac{1}{n}E(n_A) = p \\ V(f_A) = V\left(\dfrac{n_A}{n}\right) = \dfrac{1}{n^2}V(n_A) = \dfrac{p(1-p)}{n} \end{cases}$$

En consecuencia, aplicando a la v.a. $f_A$ la desigualdad de Chebyshev en la forma $b_2$)
$P(|X - E(X)| < \varepsilon) \geq 1 - \dfrac{V(X)}{\varepsilon^2}$, es decir,
$P(|f_A - E(f_A)| < \varepsilon) \geq 1 - \dfrac{V(f_A)}{\varepsilon^2}$ llegamos a lo propuesto por el teorema:
$P(|f_A - p| < \varepsilon) \geq 1 - \dfrac{p(1-p)}{n\varepsilon^2}$.

Es evidente que el *teorema* anterior implica que
$\lim_{n\to\infty} P(|f_A - p| < \varepsilon) = 1$ para todo $\varepsilon > 0$.

Entonces decimos que, en este sentido, la frecuencia relativa $f_A$ converge a la probabilidad $P(A)$.

Observación: Esta convergencia, llamada ***convergencia en probabilidad*** difiere de la convergencia normalmente usada en Cálculo (límite aritmético).
Recordemos que una sucesión $\alpha_1, \alpha_2, ..., \alpha_n, ...$ tiene límite $\alpha$ o también que $\lim_{n \to \infty} \alpha_n = \alpha$ si $\forall \varepsilon > 0 \quad \exists \nu$ tal que $|\alpha_n - \alpha| < \varepsilon$ para todo $n > \nu$. Esto significa que, desde un $n$ en adelante, $\alpha_n$ se aproxima permanentemente al valor límite $\alpha$. En cambio cuando decimos que $f_A = n_A / n$ converge a $P(A)$, estamos significando que la probabilidad del suceso $\left\{ |f_A - P(A)| < \varepsilon \right\}$ puede hacerse arbitrariamente próximo a uno tomando un $n$ suficientemente grande. Pero estamos hablando de probabilidad y no de certeza como en el caso del límite aritmético. Es decir, no significa que al tomar un $n$ grande ocurra ciertamente que nos aproximemos más al valor de la probabilidad sino que existe una gran probabilidad de que eso ocurra.

# *5- VARIABLES ALEATORIAS BIDIMENSIONALES*

## 5.1 – Generalidades

Hasta ahora hemos considerado el caso de variables aleatorias unidimensionales. Esto es, el resultado del experimento de interés se registra como un único número real.
En muchos casos, sin embargo, nos puede interesar asociar a cada resultado de un experimento aleatorio, dos o más características numéricas. Por ejemplo, de los remaches que salen de una línea de producción nos puede interesar el diámetro $X$ y la longitud $Y$. Teniendo en cuenta la inevitable variabilidad en las dimensiones de los remaches debido a las numerosas causas presentes en el proceso de fabricación, los podemos representar asociándoles dos variables aleatorias $X$ e $Y$ que pueden pensarse como una ***variable aleatoria bidimensional***: $(X,Y)$.

Sea $\varepsilon$ un experimento aleatorio y $S$ un espacio muestral asociado a él. Sean $X : S \to R$ , $Y : S \to R$ , que a cada resultado $s \in S$ le asignan el par de números reales $(x, y)$
Llamaremos a $(X,Y)$ ***variable aleatoria bidimensional***.
Si en lugar de dos variables aleatorias, tenemos $n$ variables aleatorias $X_1, X_2,..., X_n$, llamaremos a $(X_1, X_2,..., X_n)$ ***variable aleatoria n-dimensional***

En lo que sigue nos referiremos en particular a variables aleatorias $n$-dimensionales con $n$=2, es decir nos concentraremos en ***variables aleatorias bidimensionales*** por cuanto son las más simples de describir, fundamentalmente en relación a la notación. Pero debemos tener presente que las propiedades que estudiemos para ellas se pueden extender sin demasiada dificultad al caso general.

Al conjunto de valores que toma la variable aleatoria bidimensional $(X,Y)$ lo llamaremos ***recorrido*** de la v.a. $(X,Y)$ y lo indicaremos $R_{XY}$. En otras palabras $R_{XY} = \left\{(x, y) : x = X(s) \ \ e \ \ y = Y(s) \ \ con \ \ s \in S\right\}$, es decir, es la imagen por $(X,Y)$ del espacio muestral $S$.
Notar que el recorrido de $(X,Y)$ es un subconjunto del espacio Euclidiano: $R_{XY} \subseteq R^2$. Como antes, puede considerarse al recorrido $R_{XY}$ como un espacio muestral cuyos elementos son ahora pares de números reales.

Como con cualquier espacio muestral, según el número de elementos que lo constituyen, podemos clasificar a los recorridos $R_{XY}$ en numerables (finitos o infinitos) y no-numerables.
Los recorridos numerables son, en general, de la forma

$$R_{XY} = \left\{(x_i, y_j) \ \ con \ \ i = 1,2,...,n \ \ y \ \ j = 1,2,..m\right\} = \{(x_1, y_1),(x_1, y_2),...,(x_n, y_m)\} \qquad \text{(finito)}$$

$$R_{XY} = \left\{(x_i, y_j) \ \ con \ \ i = 1,2,... \ \ y \ \ j = 1,2,..\right\} = \{(x_1, y_1),(x_1, y_2),...\} \qquad \text{(infinito numerable)}$$

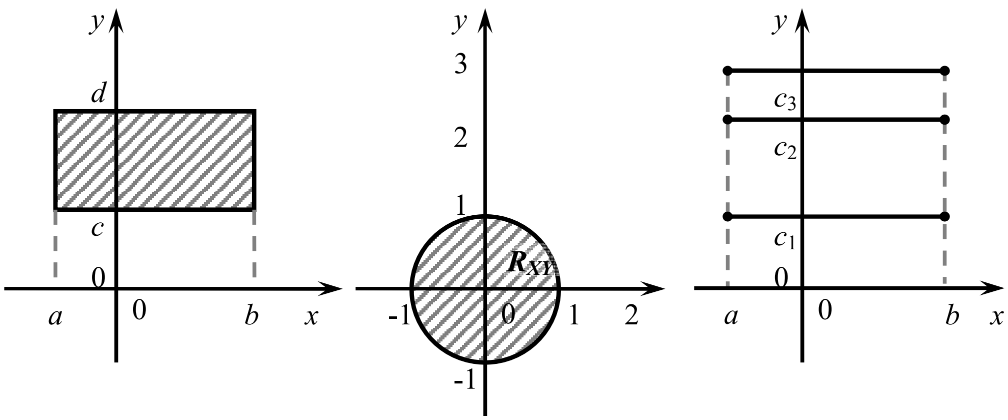
Los recorridos no numerables son regiones o subconjuntos no numerables del plano Euclidiano. Por ejemplo:

$$R_{XY} = \left\{(x, y) : a \le x \le b; \ \ c \le y \le d\right\} \qquad \text{(no numerable)}$$

$R_{XY} = \{(x, y) : x^2 + y^2 \leq 1\}$ (no numerable)

$R_{XY} = \left\{(x, y_j) : a \leq x \leq b, \;\; y_j = c_1, c_2, c_3\right\}$ (no numerable "mixto")

cuyas gráficas se pueden apreciar en la figura siguiente. Notar en el último recorrido, $X$ es v.a. continua e $Y$ discreta.




Clasificaremos a las variables aleatorias bidimensionales de la siguiente manera:
$(X,Y)$ es v.a. ***bidimensional discreta*** si $X$ e $Y$ son discretas
$(X,Y)$ es v.a. ***bidimensional continua*** si $X$ e $Y$ son continuas
El caso X continua, Y discreta (o viceversa) no lo consideramos.

Sea $(X,Y)$ una variable aleatoria bidimensional discreta y sea $R_{XY}$ su recorrido (numerable). Sea $p : R_{XY} \to R$ una función que a cada elemento $(x_i, y_j)$ le asigna un número real $p(x_i, y_j)$ tal que $P\left(X = x_i, \; Y = y_j\right) = p(x_i, y_j) \;\; \forall (x_i, y_j) \in R_{XY}$ y que verifica.

a) $p(x_i, y_j) \geq 0 \quad \forall (x_i, y_j) \in R_{XY}$

b) $\sum_{(x_i, y_j) \in R_{XY}} p(x_i, y_j) = \sum_i \sum_j p(x_i, y_j) = 1$

A esta función la llamaremos ***función de probabilidad puntual conjunta*** de la variable aleatoria bidimensional $(X,Y)$. En forma abreviada la designaremos *fdp conjunta*.

Ejemplos:

1-Dos líneas de producción, señaladas I y II, manufacturan cierto tipo de artículo a pequeña escala. Supóngase que la capacidad máxima de producción de la línea I es cinco artículos por día, mientras que para la línea II es 3 artículos/día. Debido a los innumerables factores presentes en todo proceso de producción, el número de artículos realmente producido por cada línea puede pensarse como una variable aleatoria. En conjunto podemos pensar en una variable aleatoria bidimensional $(X,Y)$ discreta, donde la primera componente $X$ corresponde a la producción de la línea I y la segunda componente $Y$ a los artículos que salen de la línea II. La *fdp* conjunta correspondiente a variables aleatorias bidimensionales suele presentarse, por comodidad, como una tabla. Supongamos que la para la v.a. $(X,Y)$ que nos interesa aquí la tabla correspondiente a $p(x_i, y_j)$ es

| $Y/X$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 0 | 0.01 | 0.03 | 0.05 | 0.07 | 0.09 |
| 1 | 0.01 | 0.02 | 0.04 | 0.05 | 0.06 | 0.08 |
| 2 | 0.01 | 0.03 | 0.05 | 0.05 | 0.05 | 0.06 |
| 3 | 0.01 | 0.02 | 0.04 | 0.06 | 0.06 | 0.05 |

¿Cuál es la probabilidad de qué salgan más artículos de la línea I que de la línea II?

Antes de calcular la probabilidad que nos pide el problema, hagamos algunas consideraciones sobre la tabla que representa a $p(x_i, y_j)$.

Se trata de una tabla a doble entrada donde en la primera fila se indican los valores que puede tomar la v.a. $X$ (en este caso $X$=0,1,2,3,4,5) y la primera columna indica los valores que puede tomar la variable $Y$ ( 0,1,2,3). Para determinar el valor de la $p(x_i, y_j)$ cuando la v.a. $(X,Y)$ toma el valor $(x_i, y_j)$ consideramos el número que se encuentra en la columna correspondiente a $X = x_i$ y la fila correspondiente a $Y = y_j$. Por ejemplo: $p(4,2) = P(X = 4, Y = 2) = 0.05$.

Podemos verificar fácilmente que la *fdp* conjunta definida por esta bien definida. En efecto verifica las condiciones a) $p(x_i, y_j) \geq 0 \quad \forall (x_i, y_j) \in R_{XY}$ y b) $\sum_{(x_i, y_j) \in R_{XY}} p(x_i, y_j) = 1$.

Para contestar la pregunta del enunciado, consideremos el suceso $B \subset R_{XY}$ definido

$B$: "es el suceso que ocurre cuando la línea I produce más artículos que la línea II" o, $B = \{X > Y\}$. Luego:

$$P(B) = P(X > Y) = \sum_{y_j=0}^{3} \sum_{x_i > y_j} p(x_i, y_j) = 0.01+0.03+0.05+0.07+0.09+0.04+0.05+0.06+0.08+$$

$$+0.05+0.05+0.06+0.06+0.05=0.75.$$

2- Hay tres cajas registradoras a la salida de un supermercado. Dos clientes llegan a las cajas en diferentes momentos cuando no hay otros clientes ante aquellas. Cada cliente escoge una caja al azar e independientemente del otro.
Sean las variables aleatorias X: " nº de clientes que escogen la caja 1" e Y: "nº de clientes que escogen la caja 2". Hallar la *fdp* conjunta de *(X,Y)*

Podemos suponer que el espacio muestral original $S$ es el conjunto de pares ordenados $S = \{(1,1);(1,2);(1,3);(2,1);(2,2);(2,3);(3,1);(3,2);(3,3)\}$ donde la primera componente del par indica la caja elegida por el cliente 1 y la segunda componente del par indica la caja elegida por el cliente 2.
Además notar que $X$ como $Y$ pueden tomar los valores 0, 1, 2
El punto muestral (3,3) es el único punto muestral que corresponde al evento $\{X = 0, Y = 0\}$
Entonces

$P(X = 0, Y = 0) = \frac{1}{9}$; pensando de forma análoga los otros casos:

$P(X = 1, Y = 0) = \frac{2}{9}$; $P(X = 2, Y = 0) = \frac{1}{9}$; $P(X = 0, Y = 1) = \frac{2}{9}$, $P(X = 1, Y = 1) = \frac{2}{9}$,

$P(X = 0, Y = 2) = \frac{1}{9}$; $P(X = 1, Y = 2) = P(X = 2, Y = 2) = 0$

Disponemos estas probabilidades en una tabla de la siguiente forma

| **Y \ X** | **0** | **1** | **2** |
|---|---|---|---|
| **0** | 1/9 | 2/9 | 1/9 |
| **1** | 2/9 | 2/9 | 0 |
| **2** | 1/9 | 0 | 0 |

## 5.2 - Funciones de distribución marginales de una v.a. (*X*,*Y*) discreta

En el ejemplo 1, supongamos que queremos saber cuál es la probabilidad de que el número de artículos producidos por la línea I sea 2, o sea $P(X=2)$
Como el evento $\{X=2\}$ es igual a $\{X=2\}\cap(\{Y=0\}\cup\{Y=1\}\cup\{Y=2\}\cup\{Y=3\})$, y a su vez
$\{X=2\}\cap(\{Y=0\}\cup\{Y=1\}\cup\{Y=2\}\cup\{Y=3\})=$
$=(\{X=2\}\cap\{Y=0\})\cup(\{X=2\}\cap\{Y=1\})\cup(\{X=2\}\cap\{Y=2\})\cup(\{X=2\}\cap\{Y=3\})$
Entonces

$$P(X=2)=$$
$$=P(\{X=2\}\cap\{Y=0\})+P(\{X=2\}\cap\{Y=1\})+P(\{X=2\}\cap\{Y=2\})+P(\{X=2\}\cap\{Y=3\})=$$
$$=P(X=2,Y=0)+P(X=2,Y=1)+P(X=2,Y=2)+P(X=2,Y=3)=\sum_{j=0}^{3}P(X=2,Y=j)$$

Razonando de la misma forma podemos escribir

$$P(X=i)=\sum_{j=0}^{3}P(X=i,Y=j) \qquad i=0,1,...,5$$

Es decir obtenemos la ***función de distribución de probabilidad de X***

Análogamente obtenemos

$$P(Y=j)=\sum_{i=0}^{5}P(X=i,Y=j) \qquad j=0,1,2,3$$

Que es la ***función de distribución de probabilidad de Y***

En general se las denomina ***distribuciones marginales de X e Y***, y su definición sería la siguiente

Sea (*X*,*Y*) discreta y sea $p(x_i,y_j)$ (*i*=1,2,…*n*, *j*=1,2,…,*m*) su función de probabilidad conjunta (Eventualmente *n* y/o *m* pueden ser $\infty$).

La función de probabilidad marginal de *X* es

$$p(x_i)=P(X=x_i)=\sum_{j=1}^{m}p(x_i,y_j) \qquad (i=1,2,\ldots,n)$$

La función de probabilidad marginal de *Y* es

$$q(y_j)=P(Y=y_j)=\sum_{i=1}^{n}p(x_i,y_j) \qquad (j=1,2,\ldots,m)$$

Observación: Remarcamos que la función de probabilidad marginal de *X*, es decir $p(x_i)$ calculada a partir de $p(x_i,y_j)$ en la forma indicada, coincide con la función de probabilidad de la variable aleatoria unidimensional *X* considerada en forma aislada. Análogamente la función de probabilidad marginal de

*Y*, es decir $q(y_j)$ calculada a partir de $p(x_i, y_j)$ en la forma indicada, coincide con la función de probabilidad de variable aleatoria unidimensional *Y* considerada en forma aislada.

Ejemplo:
Siguiendo con el ejemplo 1,

$$p(5) = P(X = 5) = p(5,0) + p(5,1) + p(5,2) + p(5,3) = 0.09 + 0.08 + 0.06 + 0.05$$
$$= 0.28$$

$$q(1) = P(Y = 1) = p(0,1) + p(1,1) + p(2,1) + p(3,1) + p(4,1) + p(5,1) = 0.01 + 0.02 + 0.04 + 0.05 + 0.06$$
$$= 0.26$$

Observemos que se verifica la condición de normalización para cada una de las marginales:

$$\sum_{x_i=0}^{5} p(x_i) = 0.03 + 0.08 + 0.16 + 0.21 + 0.24 + 0.28 = 1$$

$$\sum_{y_j=0}^{3} q(y_j) = 0.25 + 0.26 + 0.25 + 0.24 = 1$$

## 5.3 - Funciones de probabilidades condicionales

Consideremos nuevamente el ejemplo de las dos líneas I y II que producen cierto artículo a pequeña escala. Definimos la v.a. $(X,Y)$ cuya función de probabilidad conjunta $p(x_i, y_j)$ está dada por la tabla anterior que repetimos

| $Y/X$ | 0 | 1 | 2 | 3 | 4 | 5 | $q(y_j)$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.01 | 0.03 | 0.05 | 0.07 | 0.09 | 0.25 |
| 1 | 0.01 | 0.02 | 0.04 | 0.05 | 0.06 | 0.08 | 0.26 |
| 2 | 0.01 | 0.03 | 0.05 | 0.05 | 0.05 | 0.06 | 0.25 |
| 3 | 0.01 | 0.02 | 0.04 | 0.06 | 0.06 | 0.05 | 0.24 |
| $p(x_i)$ | 0.03 | 0.08 | 0.16 | 0.21 | 0.24 | 0.28 | 1 |

Supongamos que deseamos conocer la probabilidad de que la línea I produzca tres artículos sabiendo que la línea II ha fabricado dos. Tenemos que calcular una probabilidad condicional. Entonces

$$P(X = 3 | Y = 2) = \frac{P\left( X = 3 \ , \ Y = 2 \right)}{P(Y = 2)} = \frac{p(3,2)}{q(2)} = \frac{0.05}{0.25} = 0.2$$

En general definimos la ***función de probabilidad puntual de X condicional a Y*** como sigue:

$$p(x_i | y_j) = P(X = x_i | Y = y_j) = \frac{p(x_i, y_j)}{q(y_j)}$$ , es decir como el cociente de la función de probabilidad conjunta de $(X,Y)$ y la función de probabilidad puntual marginal de *Y*.

Análogamente, definimos la ***función de probabilidad puntual de Y condicional a X*** :

$q(y_j|x_i) = P(Y = y_j|X = x_i) = \dfrac{p(x_i, y_j)}{p(x_i)}$ , es decir como el cociente de la función de probabilidad puntual conjunta de $(X,Y)$ y la función de probabilidad puntual marginal de *X*.

## 5.4– Variables aleatorias independientes

Ya se discutió el concepto de independencia entre dos eventos *A* y *B*. Esas mismas ideas podemos trasladarlas en relación a dos variables aleatorias *X* e *Y* que, eventualmente, podemos considerarlas como las componentes de una variable aleatoria bidimensional $(X,Y)$.
De acuerdo con esto, intuitivamente decimos que dos variables, *X* e *Y*, son independientes si el valor que toma una de ellas no influye de ninguna manera sobre el valor que toma la otra. Esto lo establecemos más formalmente:

Sea $(X,Y)$ una variable aleatoria bidimensional discreta. Sea $p(x_i, y_j)$ su *fdp* conjunta y $p(x_i)$ y $q(y_j)$ las correspondientes *fdp* marginales de *X* e *Y*. Decimos que *X* e *Y* son ***variables aleatorias independientes*** si y sólo si

$$p(x_i, y_j) = p(x_i)q(y_j) \quad \forall (x_i, y_j) \in R_{XY}$$

Observación: Notar que para poder afirmar la independencia de *X* e *Y* debe cumplirse la factorización de la *fdp* conjunta como producto de las *fdp* marginales para ***todos*** los pares de valores de la v.a. $(X,Y)$. Por lo tanto, para verificar la independencia es necesario demostrar la validez de la factorización para todos los pares. En cambio, es suficiente encontrar un solo par que no la verifica, para afirmar, de acuerdo con la definición, que las variables *X* e *Y* son no independientes, es decir, que son dependientes. Esto es, para demostrar la dependencia es suficiente con encontrar un solo par que no verifique la factorización señalada.

Vimos que dos sucesos *A* y *B* son independientes si y sólo si $P(A|B) = P(A)$ y $P(B|A) = P(B)$ (donde por supuesto debía ser $P(A) \neq 0$ y $P(B) \neq 0$). En términos de variables aleatorias, esta forma de ver la independencia se manifiesta en la igualdad entre las *fdp* condicionales y las correspondientes *fdp* marginales, como demostramos en este

Teorema
Sea $(X,Y)$ una variable aleatoria bidimensional discreta cuyas *fdp* conjunta, condicionales y marginales son, respectivamente, $p(x_i, y_j)$; $p(x_i|y_j)$, $q(y_j|x_i)$ y $p(x_i)$, $q(y_j)$.
Entonces, *X* e *Y* son variables aleatorias independientes si y sólo si

1) $p(x_i|y_j) = p(x_i) \quad \forall (x_i, y_j) \in R_{XY}$ , o

2) $q(y_j|x_i) = q(y_j) \quad \forall (x_i, y_j) \in R_{XY}$ , que es equivalente a lo anterior

Dem.)
Demostraremos solamente 1). La equivalencia entre1) y 2) la dejamos como ejercicio.
Para demostrar 1) verificaremos la doble equivalencia entre ésta y la definición de v.a. independientes.
$\Rightarrow$)
Sean $X$ e $Y$ variables aleatorias independientes. Entonces $\forall(x_i, y_j) \in R_{XY}$

$$p(x_i|y_j) = \frac{p(x_i, y_j)}{q(y_j)} = \frac{p(x_i)q(y_j)}{q(y_j)} = p(x_i)$$

Aquí la primera igualdad es la definición de *fdp* condicional y la segunda sale de la definición de independencia al suponer que $X$ e $Y$ son independientes.
$\Leftarrow$)
Supongamos que se verifica 1). Entonces $\forall(x_i, y_j) \in R_{XY}$

$$p(x_i|y_j) = p(x_i) \to \frac{p(x_i, y_j)}{q(y_j)} = p(x_i) \to p(x_i, y_j) = p(x_i)q(y_j) \to X \text{ e } Y \text{ independientes}$$

Aquí, la primera implicación se debe a la definición de *fdp* condicional y la tercera a la definición de v.a. independientes.

Ejemplo:
1- Supongamos que una máquina se usa para un trabajo específico a la mañana y para uno diferente en la tarde. Representemos por $X$ e $Y$ el número de veces que la máquina falla en la mañana y en la tarde respectivamente. Supongamos que la tabla siguiente da la función de probabilidad conjunta $p(x_i, y_j)$ de la variable aleatoria bidimensional discreta $(X, Y)$.

| $Y/X$ | 0 | 1 | 2 | $q(y_j)$ |
|---|---|---|---|---|
| 0 | 0.1 | 0.2 | 0.2 | 0.5 |
| 1 | 0.04 | 0.08 | 0.08 | 0.2 |
| 2 | 0.06 | 0.12 | 0.12 | 0.3 |
| $P(x_i)$ | 0.2 | 0.4 | 0.4 | 1 |

Deseamos saber si las variables aleatorias $X$ e $Y$ son independientes o dependientes.
Para demostrar que son independientes debemos probar que se verifica $\forall(x_i, y_j) \in R_{XY}$
$p(x_i, y_j) = p(x_i)q(y_j)$ Verificamos directamente que

$$p(0,0) = 0.1 = p(0)q(0) = 0.2 \times 0.5$$
$$p(0,1) = 0.04 = p(0)q(1) = 0.2 \times 0.2$$
$$p(0,2) = 0.06 = p(0)q(2) = 0.2 \times 0.3$$
$$p(1,0) = 0.2 = p(1)q(0) = 0.4 \times 0.5$$
$$p(1,1) = 0.08 = p(1)q(1) = 0.4 \times 0.2$$
$$p(1,2) = 0.12 = p(1)q(2) = 0.4 \times 0.3$$
$$p(2,0) = 0.2 = p(2)q(0) = 0.4 \times 0.5$$
$$p(2,1) = 0.08 = p(2)q(1) = 0.4 \times 0.2$$
$$p(2,2) = 0.12 = p(2)q(2) = 0.4 \times 0.3$$

Luego $X$ e $Y$ son independientes.

Podríamos haber usado las condiciones 1) $p(x_i|y_j) = p(x_i) \ \ \forall(x_i, y_j) \in R_{XY}$ , o su equivalente

2) $q(y_j|x_i) = q(y_j) \ \ \forall(x_i, y_j) \in R_{XY}$ . Veamos, como muestra para un solo valor, que se verifica

$p(2|1) = \dfrac{p(2,1)}{q(1)} = \dfrac{0.08}{0.2} = 0.4 = p(2)$ . Para demostrar la independencia por este camino habría que demostrar que se cumple la condición para el resto de los pares de valores. Se deja este cálculo como ejercicio optativo.

Observaciones

1- De la definición de las *fdp* marginales, vemos que tanto en el caso discreto como en el continuo, la *fdp* conjunta determina unívocamente las *fdp* marginales. Es decir, si $(X,Y)$ es discreta del conocimiento de la función de probabilidad conjunta $p(x_i, y_j)$ podemos determinar unívocamente las funciones de probabilidad $p(x_i)$ y $q(y_j)$. Sin embargo la inversa no se cumple en general. Es decir del conocimiento de $p(x_i)$ y $q(y_j)$ no se puede, en general, reconstruir $p(x_i, y_j)$ a menos que *X* e *Y* sean variables independientes en cuyo caso es $p(x_i, y_j) = p(x_i)q(y_j)$ .

2- El concepto de independencia entre dos variables aleatorias se puede generalizar a *n* variables aleatorias $X_1, X_2, ..., X_n$

## 5.5 - Función de una variable aleatoria bidimensional

Existen muchas situaciones en las que dado una variable aleatoria bidimensional nos interesa considerar otra variable aleatoria que es función de aquélla. Por ejemplo, supongamos que las variables aleatorias *X* e *Y* denotan la longitud y el ancho, respectivamente, de una pieza, entonces $Z = 2X + 2Y$ es una v.a. que representa el perímetro de la pieza, o la v.a. $W = X.Y$ representa el área de la pieza. Tanto *Z* como *W* ***son variables aleatorias***.

En general, sea $S$ un espacio muestral asociado a un experimento probabilístico $\varepsilon$, sean $X : S \to R$ e $Y : S \to R$ dos variables aleatorias que definen una variable aleatoria bidimensional $(X,Y)$ cuyo recorrido es $R_{XY}$, y sea una función de dos variables reales $H : R_{XY} \to R$ que a cada elemento $(x, y)$ del recorrido $R_{XY}$ le hace corresponder un número real $z = H(x, y)$, entonces la función compuesta $Z = H(X,Y): S \to R$ es una variable aleatoria, puesto que a cada elemento $s \in S$ le hace corresponder un número real $z = H[X(s), Y(s)]$. Diremos que la variable aleatoria *Z* es ***función de la variable aleatoria bidimensional (X,Y).***

Algunas variables aleatorias que son función de variables aleatorias bidimensionales son $Z = X + Y$, $Z = X.Y$, $Z = X / Y$, $Z = mín(X,Y)$, $Z = máx(X,Y)$, etc.

Lo anterior se puede generalizar si en lugar de dos variables aleatorias tenemos *n* variables aleatorias $X_1, X_2, ..., X_n$, y $z = H(x_1, x_2, ...x_n)$ es una función de $n$ variables a valores reales.

Ejemplos:
1- Sea $Z \sim B(n, p)$
Podemos escribir a $Z$ como suma de variables aleatorias de la siguiente forma.
Recordar que $Z$ cuenta el número de éxitos en $n$ repeticiones o ensayos del experimento $\varepsilon$
Si definimos

$$X_i = \begin{cases} 1 \quad si \quad en \quad la \quad í-ésima \quad repetición \quad de \quad \varepsilon \quad ocurre \quad éxito \\ \\ \qquad\qquad 0 \qquad caso \quad contrario \end{cases} \qquad\qquad i = 1,2,...,n$$

Notar que a cada $X_i$ se la puede considerar $B(1, p)$, y además $X_1, X_2,..., X_n$ son independientes
Podemos escribir $Z = X_1 + X_2 + ... + X_n$

2- Sea $Z$ v.a. binomial negativa con parámetros $r$ y $p$, es decir $Z \sim BN(r, p)$
Si definimos
$X_1$: "número de repeticiones del experimento requeridos hasta el 1º éxito"
$X_2$: "número de repeticiones del experimento adicionales requeridos hasta el 2º éxito"
$X_3$: "número de repeticiones del experimento adicionales requeridos hasta el 3º éxito"
Y en general
$X_i$: "número de repeticiones del experimento adicionales después del *(i-1)*–ésimo éxito requeridos hasta el *i*-ésimo éxito"
Entonces cada variable tiene ***distribución geométrica con parámetro p*** y $Z = X_1 + X_2 + ... + X_r$
Notar además que $X_1, X_2,..., X_r$ son independientes

## Esperanza de una v.a. que es función de una v.a. bidimensional

Sea una variable aleatoria bidimensional $(X,Y)$ cuya *fdp* conjunta es la función de probabilidad conjunta $p(x_i, y_j)$ si es discreta o la función de densidad de probabilidad conjunta $f(x, y)$ si es continua y sea una función real de dos variables $z = H(x, y)$ de manera que podemos definir una variable aleatoria $Z$ que es función de la variable aleatoria bidimensional $(X,Y)$ de la forma $Z = H(X,Y)$. Si la *fdp* de $Z$ es $q(z_i)$, siendo $Z$ discreta, entonces la esperanza matemática de $Z$ es, de acuerdo con la definición general,

$$E(Z) = \sum_{x_i \in R_X} z_i . q(z_i) \qquad\qquad (Z \text{ discreta})$$

Nuevamente lo interesante es considerar la posibilidad de evaluar $E(Z)$ sin tener que calcular previamente la *fdp* de $Z$. El siguiente teorema nos muestra cómo hacerlo.

Teorema Sea $(X,Y)$ una variable aleatoria bidimensional y sea $Z=H(X,Y)$ una variable aleatoria que es función de $(X,Y)$.

Si $Z$ es variable aleatoria discreta que proviene de la variable aleatoria bidimensional discreta $(X,Y)$ cuyo recorrido es $R_{XY}$ y su *fdp* conjunta es $p(x_i, y_j)$, entonces:

$$E(Z) = E[H(X,Y)] = \sum_{(x_i, y_j) \in R_{XY}} H(x_i, y_j) p(x_i, y_j)$$

Dem.) sin demostración

## Esperanza de una suma de variables aleatorias

> Sean $X$ e $Y$ dos variables aleatorias arbitrarias. Entonces $E(X+Y)=E(X)+E(Y)$.

Dem.) en el teorema anterior consideramos $H(x,y)=x+y$
Si *(X,Y)* es discreta

$$E(Z)=E[H(X,Y)]=\sum_{(x_i,y_j)\in R_{XY}} H(x_i,y_j)p(x_i,y_j)=\sum_{(x_i,y_j)\in R_{XY}}(x_i+y_j)p(x_i,y_j)=$$

Aplicando la propiedad distributiva y separando en dos sumas

$$E(Z)=\sum_{(x_i,y_j)\in R_{XY}}(x_i+y_j)p(x_i,y_j)=\sum_{(x_i,y_j)\in R_{XY}}x_i\,p(x_i,y_j)+\sum_{(x_i,y_j)\in R_{XY}}y_j\,p(x_i,y_j)=$$

$$=\sum_i\sum_j x_i\,p(x_i,y_j)+\sum_i\sum_j y_j\,p(x_i,y_j)=\sum_i x_i\sum_j p(x_i,y_j)+\sum_j y_j\sum_i p(x_i,y_j)=$$

Pero $\sum_j p(x_i,y_j)=p(x_i)$ y $\sum_i p(x_i,y_j)=q(y_j)$, por lo tanto

$$=\sum_i x_i\,p(x_i)+\sum_j y_j q(y_j)=E(X)+E(Y)$$

**Para el caso *(X,Y)* continua <u>sigue siendo válida esta propiedad</u>.**

Podemos generalizar la propiedad anterior a un número finito cualquiera de variables aleatorias:

> Sean $X_1,X_2,...,X_n$ $n$ variables aleatorias arbitrarias. Entonces:
>
> $E(X_1+X_2+...+X_n)=E(X_1)+E(X_2)+...+E(X_n)$ o, en notación más concentrada,:
>
> $$E\left(\sum_{i=1}^{n}X_1\right)=\sum_{i=1}^{n}E(X_i)$$

(leeremos: "la esperanza de la suma es la suma de las esperanzas")

Dem.) Se deduce por inducción completa sobre el número $n$ de variables aleatorias.

Observación: se deduce que la esperanza verifica la ***propiedad lineal***:

$$E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i).$$

<u>Ejemplos</u>:
1- Vamos a aplicar algunas de las propiedades anteriores para calcular de una manera alternativa la esperanza matemática de una variable aleatoria $X$ distribuida binomialmente.
Sea entonces una v.a. $X$~B($n$,$p$).

Ya vimos que podemos escribir $X = X_1 + X_2 + ... + X_n$ donde cada $X_i$ se la puede considerar $B(1, p)$, y además $X_1, X_2,..., X_n$ son independientes

Entonces

$E(X_i) = 1 \times P(X_i = 1) + 0 \times P(X_i = 0) = P(X_i = 1) = p$ para cualquier $i$

Por lo tanto

$$E(X) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n) = \underbrace{p + p + ... + p}_{n \ \ veces} = np$$

<u>Observación</u>: muchas veces es conveniente descomponer una variable aleatoria como suma de otras más simples para facilitar los cálculos

2- ***Esperanza de una v.a. binomial negativa***

Cuando se trató la v.a. binomial negativa se dijo cuál era su esperanza. Ahora damos una demostración

Sea *X* v.a. binomial negativa con parámetros *r* y *p*, es decir $X \sim BN(r, p)$

Si definimos

$X_1$: "número de repeticiones del experimento requeridos hasta el 1º éxito"

$X_2$: "número de repeticiones del experimento adicionales requeridos hasta el 2º éxito"

$X_3$: "número de repeticiones del experimento adicionales requeridos hasta el 3º éxito"

Y en general

$X_i$: "número de repeticiones del experimento adicionales después del *(i-1)*–ésimo éxito requeridos hasta el *i*-ésimo éxito"

Entonces cada variable tiene ***distribución geométrica con parámetro p*** y $X = X_1 + X_2 + ... + X_r$

Por lo tanto

$$E(X) = E(X_1 + X_2 + ... + X_r) = E(X_1) + E(X_2) + ... + E(X_r) = \underbrace{\frac{1}{p} + \frac{1}{p} + ... + \frac{1}{p}}_{r \ \ veces} = r\frac{1}{p} = \frac{r}{p}$$

3- ***Esperanza de una v.a. hipergeométrica***

Si $X \sim H(n,\ M,\ N)$ entonces $E(X) = \dfrac{nM}{N}$

Para facilitar la demostración supongamos que tenemos *N* bolillas en una urna de las cuales *M* son rojas y *N-M* son blancas. Queremos hallar el número esperado de bolillas rojas extraídas

Definimos las variables

$$X_i = \begin{cases} 1 \ \ si \ \ la \ \ i-ésima \ \ bolilla \ \ roja \ \ es \ \ extraída \\ \\ 0 \ \ caso \ \ contrario \end{cases}$$

Las variables $X_1, X_2,...X_M$ ***no son independientes***

Se puede escribir $X = X_1 + X_2 + ... + X_M$, además

$$E(X_i) = P(X_i = 1) = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Por lo tanto

$$E(X) = E(X_1 + X_2 + ... + X_M) = E(X_1) + E(X_2) + ... + E(X_{Mr}) = \underbrace{\frac{n}{N} + \frac{n}{N} + ... + \frac{n}{N}}_{M \quad veces} = M\frac{n}{N} = \frac{nM}{N}$$

Ejemplo
El espesor $X$ de una cuña de madera (en milímetros) tiene una función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{3}{4} - \frac{3(x-5)^2}{4} & 4 \le x \le 6 \\ 0 & \text{en otro lado} \end{cases}$$

**a)** Determine E($X$)
**b)** Si Y denota el espesor de una cuña en pulgadas (1mm = 0.0394 pulgadas), determine E($Y$)
**c)** Si se seleccionan tres cuñas de manera independiente y las apilamos una encima de otra, encuentre la media y la varianza del espesor total.

**a)** Verifique el lector que

$$E(X) = \int_4^6 x\left(\frac{3}{4} - \frac{3}{4}(x-5)^2\right)dx = 5$$

**b)** $Y = 0.0394X$ entonces $E(Y) = E(0.0394X) = 0.0394E(X) = 0.197$

**c)** Notar que si $X_i$: "espesor de cuña $i$", $i = 1, 2, 3$ entonces $X = X_1 + X_2 + X_3$ es el espesor total

Por lo tanto $E(X) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 5 + 5 + 5 = 15$

En general ***la esperanza de un producto de variables aleatorias no es igual al producto de las esperanzas***

> Si $(X,Y)$ es una variable aleatoria bidimensional tal que $X$ e $Y$ son variables aleatorias ***independientes***, entonces: $E(X.Y) = E(X).E(Y)$

(leeremos:" la esperanza del producto es el producto de las esperanzas").

Dem.) análoga a la demostración de la propiedad anterior.

**Para el caso *(X,Y)* continua sigue siendo válida esta propiedad.**

Ejemplo:
Supongamos que debido a innumerables causas incontrolables la corriente $i$ y la resistencia $r$ de un circuito varían aleatoriamente de forma tal que pueden considerarse como variables aleatorias $I$ y $R$ independientes. Supongamos que las correspondientes *fdp* son:

$$g(i) = \begin{cases} 2i & 0 \le i \le 1 \\ 0 & demás \quad valores \end{cases} \qquad h(r) = \begin{cases} \frac{r^2}{9} & 0 \le r \le 3 \\ 0 & demás \quad valores \end{cases}$$

Nos interesa considerar el voltaje $v = i.r$ de manera que podemos definir la variable aleatoria $V = I.R$.
Hallar el valor esperado o esperanza matemática del voltaje: $E(V)$.
Como I y R son independientes, usando la propiedad anterior

$$E(V) = E(I)E(R)$$

$$E(I)=\int_0^1 i(2i)di = 2\frac{i^3}{3}\Big|_0^1 = \frac{2}{3} \qquad E(R)=\int_0^3 r\left(\frac{r^2}{9}\right)dr = \frac{1}{9}\frac{r^4}{4}\Big|_0^3 = \frac{1}{9}\times\frac{3^4}{9}=1$$

$$\therefore \quad E(V)=\frac{2}{3}\times 1=\frac{2}{3}$$

## Varianza de una suma de variables aleatorias

.

$$\boxed{V(X+Y)=V(X)+V(Y)+2\sigma_{XY} \quad \text{con} \quad \sigma_{XY}=E(X.Y)-E(X).E(Y)}$$

Dem.) Escribimos la varianza en su forma alternativa

$V(X+Y)=E\left([X+Y]^2\right)-[E(X+Y)]^2$. Desarrollamos los cuadrados y aplicamos la propiedad lineal de la esperanza:

$$\begin{aligned} V(X+Y) &= E\left(X^2+2.X.Y+Y^2\right)-[E(X)+E(Y)]^2 \\ &= E\left(X^2\right)+2E(X.Y)+E\left(Y^2\right)-\left\{[E(X)]^2+2E(X)E(Y)+[E(Y)]^2\right\} \end{aligned}$$

Agrupando convenientemente:

$$\begin{aligned} V(X+Y) &= \left\{E\left(X^2\right)-[E(X)]^2\right\}+\left\{E\left(Y^2\right)-[E(Y)]^2\right\}+2\{E(X.Y)-E(X)E(Y)\} \\ &= V(X)+V(Y)+2\{E(X.Y)-E(X)E(Y)\} \end{aligned}$$

, es decir

$$V(X+Y)=V(X)+V(Y)+2\sigma_{XY}$$

$\boxed{\sigma_{XY}=E(X.Y)-E(X).E(Y) \text{ se la llama la } \textbf{\textit{covarianza de X e Y.}}}$

Observaciones:

1- Teniendo presente la definición de la desviación estándar de una v.a. *X:* $\sigma_X=\sqrt{V(X)}$, vemos que a la propiedad anterior la podemos escribir:

$V(X+Y)=\sigma_X^2+\sigma_Y^2+2\sigma_{XY}$

2- Análogamente se prueba que $V(X-Y)=\sigma_X^2+\sigma_Y^2-2\sigma_{XY}$

3- $X$ e $Y$ son independientes, entonces $V(X+Y)=V(X-Y)=V(X)+V(Y)$

Esto es porque si las variables aleatorias $X$ e $Y$ son independientes, entonces $E(X.Y)=E(X).E(Y)$.

Por lo tanto la covarianza vale cero : $\sigma_{XY}=E(X.Y)-E(X).E(Y)=0$ .

4- Podemos generalizar, usando el principio de inducción completa, al caso de $n$ variables aleatorias independientes:
Si $X_1, X_2, ..., X_n$ son n variables aleatorias independientes entonces:

$V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ... + V(X_n)$ o, en forma más compacta, $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$.

5- Vemos que la esperanza de la suma de dos variables aleatorias $X$ e $Y$ es igual a la suma de las esperanzas $E(X+Y) = E(X) + E(Y)$ cualesquiera sean $X$ e $Y$. En cambio la varianza de la suma de las variables aleatorias $X$ e $Y$ es, en general, igual a la suma de las varianzas, $V(X+Y) = V(X) + V(Y)$, sólo si $X$ e $Y$ son variables independientes.

<u>Ejemplos</u>:
1- Podemos ejemplificar la aplicación de las propiedades de la varianza, calculando nuevamente la varianza de una v.a. $X$ distribuida binomialmente con parámetros $n$ y $p$.
Sea entonces una v.a. $X$~B($n$,$p$). Vimos que se puede escribir:
$X = X_1 + X_2 + ... + X_n$, donde las $n$ variables aleatorias son independientes entre sí y tienen todas la misma distribución:
$X_i \sim B(1, p) \ \forall i = 1,2,...,n$
Entonces, tratándose de $n$ variables aleatorias independientes
$V(X) = V(X_1) + V(X_2) + ... + V(X_n)$ todas la varianzas son iguales y podemos escribir la suma como $n$ veces una cualquiera de ellas:
$V(X) = nV(X_i)$. Pero
$V(X_i) = E(X_i^2) - [E(X_i)]^2$.
Ya vimos que $E(X_i) = 1.p + 0(1-p) = 0$
Además es: $E(X_i^2) = 1^2.p + 0^2(1-p) = p$
Entonces: $V(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p)$.
Luego:
$V(X) = nV(X_i) = np(1-p)$
que es el resultado que habíamos obtenido a partir de la definición y llevando las sumas involucradas a la forma del desarrollo de un binomio de Newton.

2- ***Varianza de una v.a. binomial negativa***
Ya vimos que podemos escribir $X = X_1 + X_2 + ... + X_r$, donde cada variable $X_i$ tiene ***distribución geométrica con parámetro p***
Por lo tanto
$$V(X) = V(X_1) + V(X_2) + ... + V(X_r) = r\frac{1-p}{p^2}$$

## 5.6 - <u>Covarianza</u>

Sean $X$ e $Y$ dos variables aleatorias. La ***covarianza de X e Y*** se define:

$$\boxed{Cov(X,Y) = E\{[X - E(X)].[Y - E(Y)]\} = E(X.Y) - E(X).E(Y)}$$

Notación: la notación usual para la covarianza de $X$ e $Y$ es $\sigma_{XY}$ o $Cov(X,Y)$

La última igualdad surge de desarrollar el producto y aplicar las propiedades de la esperanza:
$E\{[X-E(X)][Y-E(Y)]\} = E\{X.Y - X.E(Y) - E(X).Y + E(X)E(Y)\}$

Teniendo presente que $E(X)$ y $E(Y)$ son constantes:
$E\{[X-E(X)][Y-E(Y)]\} = E(X.Y) - E(X).E(Y) - E(X)E(Y) + E(X)E(Y) = E(X.Y) - E(X).E(Y)$.

Si $X$ e $Y$ son variables aleatorias independientes, entonces $Cov(X,Y)=0$.

Dem. )
Según vimos, si $X$ e $Y$ son variables aleatorias independientes, entonces $E(X.Y)=E(X).E(Y)$, de donde se sigue la propiedad.

**Propiedades de la covarianza**

Las siguientes propiedades son útiles y su verificación se deja como ejercicio
1- $Cov(a+bX, c+dY) = bdCov(X,Y)$
2- $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$
3- $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$
4- $Cov(X,X) = V(X)$

Ejemplos:
1)***Varianza de una v.a. hipergeométrica***
Si $X \sim H(n,\ M,\ N)$ entonces $V(X) = n\dfrac{M}{N}\left(\dfrac{N-M}{N}\right)\left(\dfrac{N-n}{N-1}\right)$
Para facilitar la demostración supongamos que tenemos $N$ bolillas en una urna de las cuales $M$ son rojas y $N$-$M$ son blancas. Queremos hallar la varianza del número de bolillas blancas extraídas
Como antes definimos las variables

$$X_i = \begin{cases} 1 \quad si \quad la \quad i-ésima \quad bolilla \quad roja \quad es \quad extraída \\ \qquad\qquad 0 \quad caso \quad contrario \end{cases}$$

Las variables $X_1, X_2, ...X_M$ ***no son independientes***
Se puede escribir $X = X_1 + X_2 + ... + X_M$, además

$$E(X_i) = P(X_i = 1) = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N} \qquad \text{y}$$

$$V(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n}{N}\left(1-\frac{n}{N}\right)$$

Por lo tanto $V(X) = V(X_1 + X_2 + ... + X_M) = \sum_{i=1}^{M} V(X_i) + 2 \sum_{i1 \le < j \le M} Cov(X_i, X_j)$

Por otro lado $Cov(X_i; X_j) = E(X_i X_j) - E(X_i)E(Y_j)$

Y $E(X_i X_j) = \dfrac{n(n-1)}{N(N-1)}$, entonces $Cov(X_i; X_j) = \dfrac{n(n-1)}{N(N-1)} - \left(\dfrac{n}{N}\right)^2$

Aplicando algunos pasos algebraicos se llega a $Cov(X_i; X_j) = \dfrac{n(n-1)}{N(N-1)} - \left(\dfrac{n}{N}\right)^2 = -\dfrac{n}{N}\left(\dfrac{1}{N-1}\right)\left(1 - \dfrac{n}{N}\right)$

Reemplazando

$$V(X) = \sum_{i=1}^{M} V(X_i) + 2 \sum_{i1 \le < j \le M} Cov(X_i, X_j) = M\,\frac{n}{N}\left(1 - \frac{n}{N}\right) + 2\left[\binom{M}{2}\left(-\frac{1}{N-1}\right)\frac{n}{N}\left(1 - \frac{n}{N}\right)\right]$$

Nuevamente, luego de algunos cálculos algebraicos se llega a

$$V(X) = n\frac{M}{N}\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

2) De una caja con frutas que contiene 3 naranjas, 2 manzanas y 3 plátanos se selecciona una muestra de 4 frutas.

Sean las variables aleatorias

*X*: " nº de naranjas extraídas"

*Y*: "nº de manzanas extraídas"

Notar que la *f.d.p.* conjunta de *(X,Y)* es

$$P(X = x, Y = y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{4-x-y}}{\binom{8}{4}} \qquad x = 0,\ 1,\ 2,\ 3\ ;\ y = 0,\ 1,\ 2\ \ ;\ 1 \le x + y \le 4$$

Es un ejemplo de ***v.a. hipergeométrica bidimensional***.

También se podría haber presentado la *f.d.p.* conjunta en una tabla, donde también figuran las distribuciones marginales de *X* e *Y*.

| ***X/Y*** | **0** | **1** | **2** | |
|---|---|---|---|---|
| **0** | 0 | 2/70 | 3/70 | 5/70 |
| **1** | 3/70 | 18/70 | 9/70 | 30/70 |
| **2** | 9/70 | 18/70 | 3/70 | 30/70 |
| **3** | 3/70 | 2/70 | 0 | 5/70 |
| | 15/70 | 40/70 | 15/70 | |

a) ¿Cuales son *E(X)* , *V(X)*, *E(Y)* y *V(Y)?*

$$E(X) = 0 \times \frac{0}{} + 1 \times \frac{3}{70} + 2 \times \frac{9}{70} + 3 \times \frac{3}{70} =$$

$$= \frac{105}{70} = \frac{3}{2}$$

$$E(Y) = 0 \times \frac{15}{70} + 1 \times \frac{40}{70} + 2 \times \frac{15}{70} = 1$$

Verifique el lector que $V(X) = \dfrac{15}{28}$ y $V(Y) = \dfrac{3}{7}$

b) ¿Son *X* e *Y* independientes?

$$P(X=0,Y=0)=0 \quad \text{pero} \quad P(X=0)P(Y=0)=\frac{5}{70}\frac{15}{70}\neq 0$$

Por lo tanto $X$ e $Y$ son dependientes, lo que implica que $Cov(X,Y)\neq 0$

c) ¿Cuál es la *Cov(X,Y)?*

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

$$\begin{aligned}E(XY)=&\,0\times 0\times 0+0\times 1\times\frac{2}{70}+0\times 2\times\frac{3}{70}+\\&+1\times 0\times\frac{3}{70}+1\times 1\times\frac{18}{70}+1\times 2\times\frac{9}{70}+\\&+2\times 0\times\frac{9}{70}+2\times 1\times\frac{18}{70}+2\times 2\times\frac{3}{70}+\\&+3\times 0\times\frac{3}{70}+3\times 1\times\frac{2}{70}+3\times 2\times 0=\frac{90}{70}\end{aligned}$$

Entonces $Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{9}{7}-\frac{3}{2}\times 1=-\frac{3}{14}$

d) Z = $X$+$Y$ simboliza el total de naranjas y manzanas extraídas

¿Cuál es la *E(Z)* y *V(Z)* ?

$$E(Z)=E(X+Y)=E(X)+E(Y)=\frac{3}{2}+1=2.5$$

$$V(Z)=V(X)+V(Y)+2Cov(X,Y)=\frac{15}{18}+\frac{3}{7}+2\times\left(-\frac{3}{14}\right)=\frac{15}{28}$$

e) Supongamos que cada naranja cuesta 2$ y cada manzana cuesta 1.5$ entonces

$W=2X+1.5Y$ es el costo del total de frutas extraídas. Hallar *E(W)* y *V(W)*

$$E(W)=E(2X+1.5Y)=2E(X)+1.5E(Y)=4.5\$$$

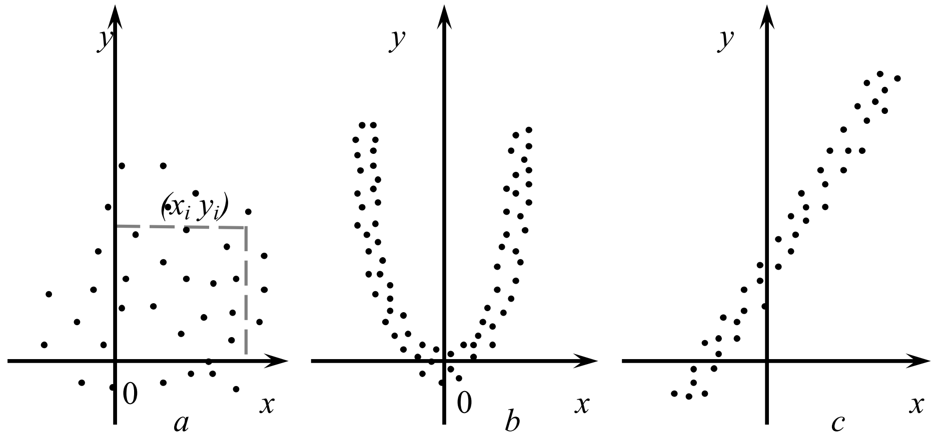
$$V(W)=V(2X+1.5Y)=2^2V(X)+1.5^2V(Y)+2\times 2\times 1.5Cov(X,Y)=\frac{51}{28}$$

### 5.7 - Coeficiente de correlación lineal.

En realidad más que la covarianza aquí nos interesa considerar una cantidad relacionada con $\sigma_{XY}$ y que según veremos nos dará información sobre el grado de asociación que existe entre $X$ e $Y$. Más concretamente nos contará si existe algún grado de relación lineal entre $X$ e $Y$. Esa cantidad es el coeficiente de correlación lineal.

En el mismo sentido en que podemos tener una idea aproximada sobre la probabilidad de un suceso $A$ si repetimos el experimento y consideramos las ocurrencias de $A$ en las $n$ repeticiones,

así podemos tener también una primera idea sobre la existencia de una relación funcional, específicamente una relación lineal, entre $X$ e $Y$ si consideramos un ***diagrama de dispersión***. Consiste en dibujar pares de valores $(x_i, y_j)$ medidos de la variable aleatoria $(X,Y)$ en un sistema de coordenadas. En la figura mostramos diversas situaciones posibles.




De la figura *a* se deduciría que entre $X$ e $Y$ no hay ningún tipo de relación funcional. La figura *b* sugiere la posibilidad de que exista una relación funcional que corresponde a una parábola. La figura *c*, por su parte, sugiere una relación lineal entre $X$ e $Y$. Este último es el comportamiento que nos interesa caracterizar. Con ese fin definimos el coeficiente de correlación lineal como sigue:

Sea $(X,Y)$ una variable aleatoria bidimensional. Definimos el ***coeficiente de correlación lineal entre X e Y*** como $\rho_{XY} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}$

En consecuencia:

$$\rho_{XY} = \frac{E\{[X - E(X)].[Y - E(Y)]\}}{\sqrt{V(X).V(Y)}} = \frac{E(X.Y) - E(X).E(Y)}{\sqrt{V(X).V(Y)}}.$$

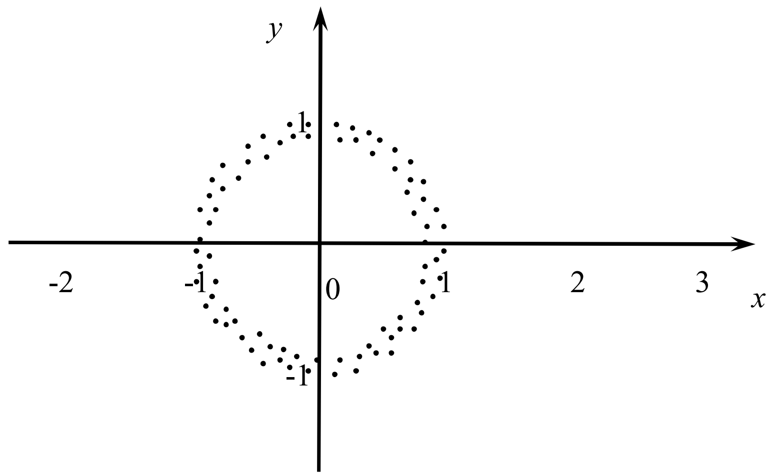
Daremos una serie de propiedades de $\rho_{XY}$ que nos permitirán establecer más concretamente su significado.

**<u>Propiedad 1</u>**

Si $X$ e $Y$ son variables aleatorias independientes entonces $\rho_{XY} = 0$.

Dem.) inmediata a partir del hecho que si X e Y son independientes entonces $E(XY) = E(X)E(Y)$

<u>Observación:</u> La inversa no es necesariamente cierta. Puede ser que $\rho_{XY} = 0$ y sin embargo $X$ e $Y$ no sean variables aleatorias independientes. En efecto si tenemos una v.a. bidimensional $(X,Y)$ que da lugar a un diagrama de dispersión como el que se muestra en la figura, veremos que correspondería a un coeficiente de correlación lineal $\rho_{XY} = 0$ y sin embargo la figura sugiere que entre $X$ e $Y$ existe la relación funcional $X^2 + Y^2 = 1$, es decir $X$ e $Y$ son v.a. dependientes. En realidad, como veremos, $\rho_{XY}$ es una medida de la existencia de una relación lineal entre $X$ e $Y$ y una circunferencia se aleja mucho de una línea recta.




**Propiedad 2** :

$$\boxed{-1 \le \rho_{XY} \le 1}$$

Dem.)

Si consideramos la v.a. $\dfrac{X}{\sigma_X} + \dfrac{Y}{\sigma_Y}$ entonces

$$0 \le V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{V(X)}{\sigma_X{}^2} + \frac{V(Y)}{\sigma_Y{}^2} + \frac{2Cov(X,Y)}{\sigma_X \sigma_Y} = 2(1 + \rho_{XY})$$

Implicando que $-1 \le \rho_{XY}$

Por otro lado: $0 \le V\left(\dfrac{X}{\sigma_X} - \dfrac{Y}{\sigma_Y}\right) = \dfrac{V(X)}{\sigma_X{}^2} + \dfrac{V(Y)}{\sigma_Y{}^2} - \dfrac{2Cov(X,Y)}{\sigma_X \sigma_Y} = 2(1 - \rho_{XY})$

Implicando que $\rho_{XY} \le 1$

$$\therefore \quad -1 \le \rho_{XY} \le 1$$

**Propiedad 3** :

Si $\rho_{XY}^2 = 1$, entonces con probabilidad 1 es $Y = A.X + B$ donde $A$ y $B$ son constantes.

Dem.) Si $\rho_{XY}^2 = 1$ entonces $\rho_{XY} = 1$ o $\rho_{XY} = -1$

Si $\rho_{XY} = -1$ entonces de la demostración anterior se deduce que

$V\left(\dfrac{X}{\sigma_X}+\dfrac{Y}{\sigma_Y}\right)=2(1+\rho_{XY})=0$, lo que implica que la v.a. $Z=\dfrac{X}{\sigma_X}+\dfrac{Y}{\sigma_Y}$ tiene varianza cero. Según la interpretación de varianza podemos deducir (en forma intuitiva) que la v.a. ***no tiene dispersión con respecto a su esperanza***, es decir la v.a. ***Z es una constante con probabilidad 1***

Por lo tanto esto implica que $Y = A.X + B$ con $A=-\dfrac{\sigma_Y}{\sigma_X}<0$

Análogamente $\rho_{XY}=1$ implica que $Y = A.X + B$ con $A=\dfrac{\sigma_Y}{\sigma_X}>0$

**Propiedad 4** :

> Si *X* e *Y* son dos variables aleatorias tales que $Y = A.X + B$, donde *A* y *B* son constantes, entonces $\rho_{XY}^2=1$. Si $A>0$ es $\rho_{XY}=1$ y si $A<0$ es $\rho_{XY}=-1$.

Dem.) se deja como ejercicio

Observación: Claramente las propiedades anteriores establecen que el coeficiente de correlación lineal es una medida del grado de linealidad entre *X* e *Y*.

Ejemplo

En el ejemplo anterior

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{-\dfrac{3}{14}}{\sqrt{\dfrac{15}{28}\dfrac{3}{7}}}=-\frac{\sqrt{5}}{5}=-0.44721$$

# 6- SUMA DE VARIABLES ALEATORIAS Y TEOREMA CENTRAL DEL LÍMITE

## 6.1 – Suma de variables aleatorias independientes

Cuando se estudiaron las variables aleatorias bidimensionales se habló de una ***función de variable aleatoria bidimensional.*** En particular se nombró la suma de $n$ variables aleatorias**,** pero no se dijo nada sobre la ***distribución*** de esa v.a. suma.
Es a menudo importante saber cuál es la distribución de una suma de variables aleatorias independientes.
Consideramos algunos ejemplos en el caso discreto

1- ***Suma de variables aleatorias independientes con distribución Poisson***

$$X \sim P(\lambda_1)\ ;\ Y \sim P(\lambda_2)\ ;\ X \text{ y } Y \text{ independientes} \quad \Rightarrow \quad X+Y \sim P(\lambda_1+\lambda_2)$$

Dem.)
Consideramos el evento $\{X+Y=n\}$ como unión de eventos excluyentes
$\{X=k, Y=n-k\} \quad 0 \le k \le n$ , entonces

$$P(X+Y=n)=\sum_{k=0}^{n} P(X=k, Y=n-k) = \sum_{k=0}^{n} P(X=k)P(Y=n-k) = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^{\,k}}{k!}e^{-\lambda_2}\frac{\lambda_2^{\,n-k}}{(n-k)!} =$$

$X$ e $Y$ independientes

$$= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^{\,k}}{k!}\frac{\lambda_2^{\,n-k}}{(n-k)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^{\,k}\lambda_2^{\,n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n$$

Binomio de Newton

O sea $X+Y$ tiene distribución Poisson con parámetro $\lambda_1+\lambda_2$

2- ***Suma de variables aleatorias binomiales independientes***

$$X \sim B(n_1, p)\ ;\ Y \sim B(n_2, p)\ ;\ X \text{ y } Y \text{ independientes} \quad \Rightarrow \quad X+Y \sim B(n_1+n_2, p)$$

Dem.)
Nuevamente consideramos el evento $\{X+Y=k\}$ como unión de eventos excluyentes $\{X=i, Y=k-i\} \quad 0 \le i \le n_1$ , entonces

$$P(X+Y=k)=\sum_{i=0}^{n_1} P(X=i, Y=k-i) = \sum_{i=0}^{n_1} P(X=i)P(Y=k-i) = \sum_{k=0}^{n_1}\binom{n_1}{i}p^i(1-p)^{n_1-i}\binom{n_2}{k-i}p^{k-i}(1-p)^{n_2-k+i} =$$

$X$ e $Y$ independientes

$$= p^k(1-p)^{n_1+n_2-k}\sum_{i=0}^{n_1}\binom{n_1}{i}\binom{n_2}{k-i}$$

En la expresión anterior si $j > r$ entonces $\binom{r}{j} = 0$

Por último usamos la siguiente identidad combinatoria $\sum_{i=0}^{n_1}\binom{n_1}{i}\binom{n_2}{k-i}=\binom{n_1+n_2}{k}$

Y entonces

$$P(X+Y=k)=\binom{n_1+n_2}{k}p^k(1-p)^{n_1+n_2-k}$$

O sea *X+Y* tiene distribución binomial con parámetros $n_1+n_2$ y $p$

<u>Observación</u>: en los dos casos anteriores se puede generalizar el resultado a *n* variables aleatorias independientes, usando el principio de inducción completa, es decir

1- Si $X_1, X_2, \ldots, X_n$ son $n$ variables aleatorias independientes donde $X_i \sim P(\lambda_i)$ para todo $i=1,2,\ldots,n$ entonces $\sum_{i=0}^{n} X_i \sim P(\sum_{i=0}^{n}\lambda_i)$

2- Si $X_1, X_2, \ldots, X_n$ son $n$ variables aleatorias independientes donde $X_i \sim B(n_i, p)$ para todo $i=1,2,\ldots,n$ entonces $\sum_{i=0}^{n} X_i \sim B(\sum_{i=0}^{n}n_i, p)$

## <u>Suma de variables aleatorias normales independientes</u>

Si *X* e *Y* son dos variables aleatorias continuas independientes con densidades *g(x)* y *h(y)* respectivamente se puede probar (no lo demostraremos aquí) que la v.a. $Z=X+Y$ tiene densidad dada por $f_{X+Y}(z)=\int_{-\infty}^{\infty} g(z-y)h(y)dy$

Usando esto se puede demostrar el siguiente importante resultado:

| |
|---|
| Si *X* e *Y* son variables aleatorias independientes donde $X \sim N(\mu_1, \sigma_1{}^2)$ y $Y \sim N(\mu_2, \sigma_2{}^2)$ entonces $X+Y \sim N(\mu_1+\mu_2, \sigma_1{}^2+\sigma_2{}^2)$ |

Por inducción completa se puede generalizar este resultado a *n* variables:

| |
|---|
| Si $X_1, X_2, \ldots, X_n$ son *n* variables aleatorias independientes donde $X_i \sim N(\mu_i, \sigma_i^2)$ para todo $i=1,2,\ldots,n$ entonces $\sum_{i=0}^{n} X_i \sim N(\sum_{i=0}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2)$ |

De lo anterior y del hecho que $X \sim N(\mu, \sigma^2) \Rightarrow aX+b \sim N(a\mu+b, a^2\sigma^2)$ tenemos:

| |
|---|
| Si $X_1, X_2, \ldots, X_n$ son *n* variables aleatorias independientes donde $X_i \sim N(\mu_i, \sigma_i^2)$ para todo $i=1,2,\ldots,n$ entonces $\sum_{i=0}^{n} a_i X_i \sim N(\sum_{i=0}^{n}a_i\mu_i, \sum_{i=1}^{n}a_i^2\sigma_i^2)$ donde $a_1, a_2, \ldots, a_n$ son números reales |

Se dice que $\sum_{i=0}^{n} a_i X_i$ es una ***combinación lineal de variables aleatorias***.

Ejemplos:

1- La envoltura de plástico para un disco magnético está formada por dos hojas. El espesor de cada una tiene una distribución normal con media 1.5 milímetros y desviación estándar de 0.1 milímetros. Las hojas son independientes.

a) Determine la media y la desviación estándar del espesor total de las dos hojas.

b) ¿Cuál es la probabilidad de que el espesor total sea mayor que 3.3 milímetros?

Solución: Sean las variables aleatorias

*X:* "espesor de la hoja 1" e *Y*: "espesor de la hoja 2"

Entonces $X \sim N(1.5, 0.1^2)$ ; $Y \sim N(1.5, 0.1^2)$ y $X$ e $Y$ independientes

a) Si definimos la v.a. *Z*: "espesor total de las dos hojas" , entonces $Z = X + Y$

Por lo tanto $Z \sim N(1.5+1.5, \quad 0.1^2 + 0.1^2)$ es decir $Z \sim N(3, \quad 0.02)$

En consecuencia $E(Z) = 3$ , $\sigma_Z = \sqrt{V(Z)} = \sqrt{0.02}$

b) Se pide calcular $P(Z > 3.3)$

$$P(Z > 3.3) = P\left(\frac{Z-3}{\sqrt{0.02}} > \frac{3.3-3}{\sqrt{0.02}}\right) = 1 - \Phi\left(\frac{3.3-3}{\sqrt{0.02}}\right) = 1 - \Phi(2.12132) = 1 - 0.983 = 0.017$$

2-Tengo tres mensajes que atender en el edificio administrativo. Sea $X_i$ : " el tiempo que toma el *i*-ésimo mensaje" ($i$ = 1, 2 ,3), y sea $X_4$ : " el tiempo total que utilizo para caminar hacia y desde el edificio y entre cada mensaje". Suponga que las $X_i$ son independientes, normalmente distribuidas, con las siguientes medias y desviaciones estándar:

$$\mu_1 = 15 \text{ min}, \quad \sigma_1 = 4, \quad \mu_2 = 5, \quad \sigma_2 = 1, \quad \mu_3 = 8, \quad \sigma_3 = 2, \quad \mu_4 = 12, \quad \sigma_4 = 3$$

Pienso salir de mi oficina precisamente a las 10.00 a.m. y deseo pegar una nota en mi puerta que dice "regreso a las *t* a.m." ¿A qué hora *t* debo escribir si deseo que la probabilidad de mi llegada después de *t* sea 0.01?

Solución: Definimos la v.a. *Z*: "tiempo transcurrido desde que salgo de mi oficina hasta que regreso", entonces $T = X_1 + X_2 + X_3 + X_4$

Por lo tanto $T \sim N\left(\sum_{i=1}^{4} \mu_i, \quad \sum_{i=1}^{4} \sigma_i^{\,2}\right)$, y se pide hallar *t* tal que $P(T > t) = 0.01$

$$\sum_{i=1}^{4} \mu_i = 15 + 5 + 8 + 12 = 50 \quad \text{y} \quad \sum_{i=1}^{4} \sigma_i^{\,2} = 4^2 + 1^2 + 2^2 + 3^2 = 30$$

Entonces $P(T > t) = 1 - \Phi\left(\frac{t-50}{\sqrt{30}}\right) = 0.01$ , es decir $\Phi\left(\frac{t-50}{\sqrt{30}}\right) = 0.99$

Buscando en la tabla de la normal $\frac{t-50}{\sqrt{30}} = 2.33 \quad \Rightarrow \quad t = 2.33 \times \sqrt{30} + 50 = 62.7619$

3- El ancho del marco de una puerta tiene una distribución normal con media 24 pulgadas y desviación estándar de 1/8 de pulgada. El ancho de la puerta tiene una distribución normal con media 23.875 de pulgadas y desviación estándar de 1/16 de pulgadas. Suponer independencia.

a) Determine la distribución, la media y la desviación estándar de la diferencia entre el ancho del marco y de la puerta.

b) ¿Cuál es la probabilidad de que la diferencia entre el ancho del marco y de la puerta sea mayor que ¼ de pulgada?.
c) ¿Cuál es la probabilidad de que la puerta no quepa en el marco?.

Solución: Sean las variables aleatorias
$X$: "ancho del marco de la puerta en pulgadas"
$Y$: "ancho de la puerta en pulgadas"
Entonces $X \sim N(24, \ (1/8)^2)$ , $Y \sim N(23.875, \ (1/16)^2)$ , $X$ e $Y$ independientes

a) Se pide la distribución de $X$-$Y$ , $E(X-Y)$ , $\sigma_{X-Y} = \sqrt{V(X-Y)}$

$$E(X-Y) = E(X) - E(Y) = 24 - 23.875 = 0.125$$

$$V(X-Y) = V(X) + V(Y) = \left(\frac{1}{8}\right)^2 + \left(\frac{1}{16}\right)^2 = \frac{5}{256} \quad \therefore \ \sigma_{X-Y} = \frac{\sqrt{5}}{16}$$

Por lo tanto $X - Y \sim N\left(0.125, \ \left(\frac{\sqrt{5}}{16}\right)^2\right)$

b) Se pide la probabilidad $P(X - Y > 1/4)$

$$P(X-Y > 1/4) = 1 - \Phi\left(\frac{0.25 - 0.125}{\frac{\sqrt{5}}{16}}\right) = 1 - \Phi\left(\frac{2\sqrt{5}}{5}\right) = 1 - \Phi(0.8944) = 1 - 0.8133 = 0.1867$$

c) Si la puerta no entra en el marco entonces se da el evento $\{X < Y\}$ o equivalentemente $\{X - Y < 0\}$, por lo tanto

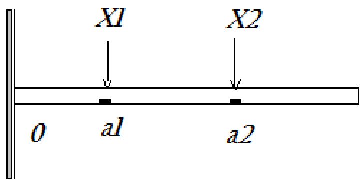
$$P(X-Y < 0) = \Phi\left(\frac{0 - 0.125}{\frac{\sqrt{5}}{16}}\right) = \Phi\left(-\frac{2\sqrt{5}}{5}\right) = 1 - \Phi\left(\frac{2\sqrt{5}}{5}\right) = 0.1867$$

4- Supongamos que las variables aleatorias $X$ e $Y$ denotan la longitud y el ancho en cm, respectivamente, de una pieza.
Supongamos además que $X$ e $Y$ son independientes y que $X \sim N(2 , 0.1^2)$ , $Y \sim N(5 , 0.2^2)$.
Entonces $Z = 2X + 2Y$ es una v.a. que representa el perímetro de la pieza.
Calcular la probabilidad de que el perímetro sea mayor que 14.5 cm.

Solución: tenemos que $Z \sim N\left(2\times 2 + 2\times 5, \ 2^2 \times 0.1^2 + 2^2 \times 0.2^2\right)$ , o sea $Z \sim N(14, \ 0.2)$
La probabilidad pedida es $P(Z > 14.5)$ , entonces

$$P(Z > 14.5) = 1 - \Phi\left(\frac{14.5 - 14}{\sqrt{0.2}}\right) = 1 - \Phi\left(\frac{\sqrt{5}}{2}\right) = 1 - \Phi(1.1180) = 1 - 0.8810 = 0.119$$

5- Si se aplican dos cargas aleatorias $X_1$ y $X_2$ a una viga voladiza como se muestra en la figura siguiente, el momento de flexión en 0 debido a las cargas es $a_1X_1 + a_2X_2$.




a) Suponga que $X_1$ y $X_2$ son v.a. independientes con medias 2 y 4 KLbs respectivamente, y desviaciones estándar 0.5 y 1.0 KLbs, respectivamente.

Si $a_1 = 5$ pies y $a_2 = 10$ pies, ¿cuál es el momento de flexión esperado y cuál es la desviación estándar del momento de flexión?

b) Si $X_1$ y $X_2$ están normalmente distribuidas, ¿cuál es la probabilidad de que el momento de flexión supere 75 KLbs?

Solución: Sea la v.a. $Z$: "momento de flexión en 0", entonces $Z = 5X_1 + 10X_2$

Por lo tanto

a) $E(Z) = 5E(X_1) + 10E(X_2) = 5\times 2 + 10\times 4 = 50$

$$V(Z) = 5^2 \times 0.5^2 + 10^2 \times 1^2 = 25\times 0.25 + 10\times 1 = \frac{65}{4} \quad \therefore \ \sigma_Z = \sqrt{\frac{65}{4}}$$

b) Si $X_1$ y $X_2$ están normalmente distribuidas, entonces $Z \sim N\left(50,\ \frac{65}{4}\right)$

Por lo tanto

$$P(Z > 75) = 1 - \Phi\left(\frac{75-50}{\sqrt{\frac{65}{4}}}\right) = 1 - \Phi\left(\frac{10\sqrt{65}}{13}\right) = 1 - \Phi(6.20) \approx 1 - 1 = 0$$

## Promedio de variables aleatorias normales independientes

Si $X_1, X_2, \ldots, X_n$ son $n$ variables aleatorias independientes donde $X_i \sim N(\mu, \sigma^2)$ para todo $i = 1,2,\ldots,n$ entonces la v.a. $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ tiene distribución normal con media $\mu$ y varianza $\frac{\sigma^2}{n}$

Dem.) Notar que $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ es un caso particular de combinación lineal de variables aleatorias donde $a_i = \frac{1}{n}$ para todo $i = 1,2,\ldots,n$

Además en este caso $\mu_i = \mu$ y $\sigma_i^2 = \sigma^2$ para todo $i = 1,2,\ldots,n$

Por lo tanto, $\overline{X}$ tiene distribución normal con esperanza $\sum_{i=1}^{n} \frac{1}{n}\mu_i = \sum_{i=1}^{n}\frac{1}{n}\mu = \frac{1}{n}n\mu = \mu$ y varianza

$$\sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \sigma_i^2 = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \sigma^2 = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$$

Es decir, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Observación: a $\overline{X}$ se lo llama ***promedio muestral*** o ***media muestral***

Ejemplos:

1) El diámetro interno de un anillo de pistón seleccionado al azar es una v.a. con distribución normal con media 12 cm y desviación estándar de 0.04 cm.

a) Si $\overline{X}$ es el diámetro promedio en una muestra de $n=16$ anillos, calcule $P(11.99 \le \overline{X} \le 12.01)$

b) ¿Qué tan probable es que el diámetro promedio exceda de 12.01 cuando $n=25$?

Solución:

a) Sean las variables aleatorias $X_i$ : "diámetro del anillo $i$" $i=1,2,\dots,16$

Entonces $X_i \sim N(12, 0.04^2)$ para cada $i$.

Por lo tanto $\overline{X} \sim N\left(12, \dfrac{0.04^2}{16}\right)$. Entonces

$$P(11.99 \le \overline{X} \le 12.01) = P\left(\frac{11.99-12}{\sqrt{\frac{0.04^2}{16}}} \le \frac{\overline{X}-12}{\sqrt{\frac{0.04^2}{16}}} \le \frac{12.01-12}{\sqrt{\frac{0.04^2}{16}}}\right) =$$

$$= \phi\left(\frac{12.01-12}{\sqrt{\frac{0.04^2}{16}}}\right) - \phi\left(\frac{11.99-12}{\sqrt{\frac{0.04^2}{16}}}\right) = \phi(1) - \phi(-1) = 2\phi(1) - 1 =$$

$$= 2 \times 0.8413 - 1 = 0.6826$$

b) En este caso $\overline{X} \sim N\left(12, \dfrac{0.04^2}{25}\right)$, entonces

$$P(\overline{X} > 12.01) = 1 - \phi\left(\frac{12.01-12}{\sqrt{\frac{0.04^2}{25}}}\right) = 1 - \phi(1.25) = 1 - 0.8944 = 0.1056$$

2) Una máquina embotelladora puede regularse de tal manera que llene un promedio de $\mu$ onzas por botella. Se ha observado que la cantidad de contenido que suministra la máquina presenta una distribución normal con $\sigma = 1$ onza. De la producción de la máquina un cierto día, se obtiene una muestra de 9 botellas llenas (todas fueron llenadas con las mismas posiciones del control operativo) y se miden las onzas del contenido de cada una.

a) Determinar la probabilidad de que la media muestral se encuentre a lo más a 0.3 onzas de la media real $\mu$ para tales posiciones de control

b) ¿Cuántas observaciones deben incluirse en la muestra si se desea que la media muestral esté a lo más a 0.3 onzas de $\mu$ con una probabilidad de 0.95?

Solución:

a) Sean las variables aleatorias $X_i$ : "contenido en onzas de la botella *i*" $i = 1,2,\dots,9$

Entonces $X_i \sim N(\mu,1)$ para cada *i*.

Por lo tanto $\overline{X} \sim N\left(\mu, \frac{1}{9}\right)$ . Se desea calcular

$$P(|\overline{X} - \mu| \le 0.3) = P(-0.3 \le \overline{X} - \mu \le 0.3) = P\left(-\frac{0.3}{\sigma/\sqrt{n}} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le \frac{0.3}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(-\frac{0.3}{\sigma/\sqrt{n}} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(-0.9 \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 0.9\right) = \Phi(0.9) - \Phi(-0.9) =$$

$$= 2\Phi(0.9) - 1 = 0.6318$$

b) Ahora se pretende que

$$P(|\overline{X} - \mu| \le 0.3) = P(-0.3 \le \overline{X} - \mu \le 0.3) = 0.95$$

Entonces

$$P(|\overline{X} - \mu| \le 0.3) = P\left(\frac{-0.3}{\sigma/\sqrt{n}} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(-0.3\sqrt{n} \le \frac{\overline{X} - \mu}{1/\sqrt{n}} \le 0.3\sqrt{n}\right) = 0.95$$

Mediante la tabla de la acumulada de la normal estándar se tiene que

$$P\left(-0.3\sqrt{n} \le \frac{\overline{X} - \mu}{1/\sqrt{n}} \le 0.3\sqrt{n}\right) = 2\Phi\left(0.3\sqrt{n}\right) - 1 = 0.95 \;\Rightarrow\; \Phi\left(0.3\sqrt{n}\right) = 0.975 \;\Rightarrow\; \left(0.3\sqrt{n}\right) = 1.96$$

O sea $\quad n \approx \left(\frac{1.96}{0.3}\right)^2 = 42.68$

Si tomamos $n = 43$, entonces $P(|\overline{X} - \mu| \le 0.3)$ será un poco mayor que 0.95

## 6.2 - <u>Teorema central del límite</u>

Acabamos de ver que la suma de un número finito *n* de variables aleatorias independientes que están normalmente distribuidas es una variable aleatoria también normalmente distribuida. Esta propiedad reproductiva no es exclusiva de la distribución normal. En efecto, por ejemplo, ya vimos que existen variables aleatorias discretas que la cumplen, es el caso de la Poisson y la Binomial.

En realidad, la propiedad que le da a la distribución normal el lugar privilegiado que ocupa entre todas las distribuciones es el hecho de que la suma de un número muy grande, rigurosamente un número infinito numerable, de variables aleatorias independientes con distribuciones ***arbitrarias*** (no necesariamente normales) es una variable aleatoria que tiene, aproximadamente, una distribución normal. Este es, esencialmente, el contenido del

Teorema central del límite (T.C.L.):
Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1,2,\ldots,n$, es decir ***independientes idénticamente distribuidas***

Sea la v.a. $S_n = \sum_{i=1}^{n} X_i$ y sea $Z_n = \dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$.

Entonces $\lim_{n\to\infty} P(Z_n \le z) = \Phi(z)$ , esto es $\lim_{n\to\infty} P\left(\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}} \le z\right) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$

Dem.) sin demostración

Observaciones:

1- Notar que $E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\mu$ y $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = n\sigma^2$

Por lo tanto $Z_n = \dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$ es la v.a. $S_n$ ***estandarizada***

2- Notar que $P\left(\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}} \le z\right) = P\left(\dfrac{\dfrac{S_n - n\mu}{n}}{\dfrac{\sqrt{n\sigma^2}}{n}} \le z\right) = P\left(\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$, por lo tanto también se puede enunciar el Teorema central del límite de la siguiente forma

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1,2,\ldots,n$, es decir ***independientes idénticamente distribuidas***

Sea la v.a. promedio muestral $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ y sea $Z_n = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$.

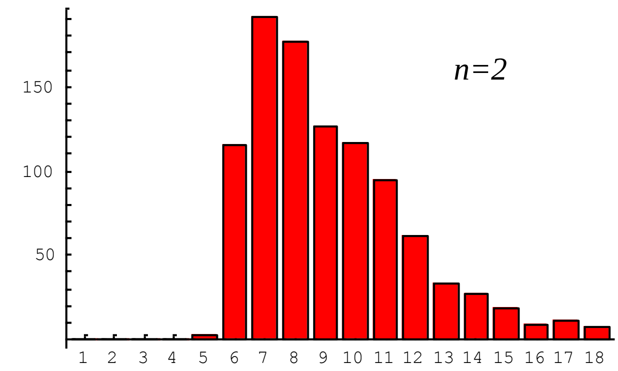
Entonces $\lim_{n\to\infty} P(Z_n \le z) = \Phi(z)$ , esto es $\lim_{n\to\infty} P\left(\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z\right) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$

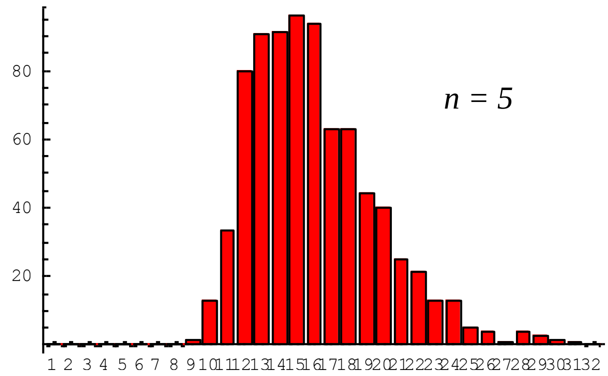
Donde $Z_n = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ es el ***promedio muestral estandarizado***

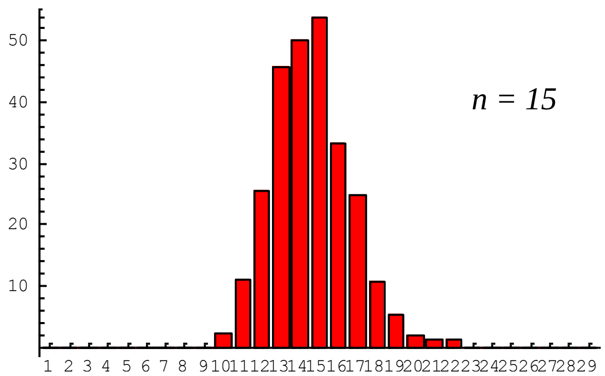
3- Aunque en muchos casos el T.C.L. funciona bien para valores de *n* pequeños , en particular donde la población es continua y simétrica, en otras situaciones se requieren valores de *n* mas grandes, dependiendo de la forma de la distribución de las $X_i$. En muchos casos de interés práctico, si $n \ge 30$, la aproximación normal será satisfactoria sin importar cómo sea la forma de la distribución de las $X_i$ . Si $n < 30$, el T.C.L. funciona si la distribución de las $X_i$ no está muy alejada de una distribución normal

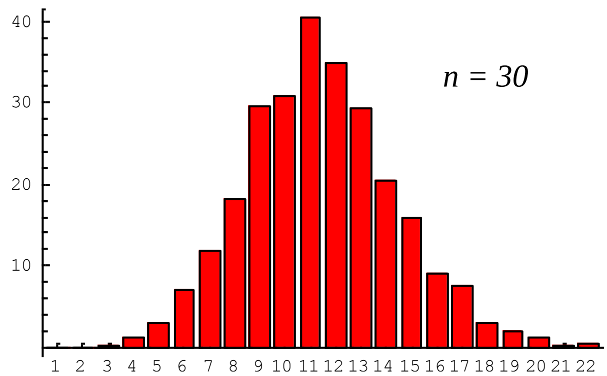
4- Para interpretar el significado del T.C.L., se generan (por computadora) *n* valores de una v.a. exponencial con parámetro $\lambda = 0.5$, y se calcula el promedio de esos *n* valores. Esto se repite 1000 veces, por lo tanto tenemos 1000 valores de la v.a. $\overline{X}$ .

Hacemos un ***histograma de frecuencias*** de $\overline{X}$, esto es, tomamos un intervalo $(a,b)$ donde "caen" todos los valores de $\overline{X}$, y lo subdividimos en intervalos mas chicos de igual longitud. La ***frecuencia de cada subintervalo*** es la cantidad de valores de $\overline{X}$ que caen en dicho subintervalo. Se grafican estas frecuencias obteniéndose los gráficos siguientes que se pueden considerar una aproximación a la verdadera distribución de $\overline{X}$.
Se observa que a medida que aumenta el valor de $n$ los gráficos se van haciendo más simétricos, pareciéndose a la gráfica de una distribución normal.




Ejemplos:
1- Supóngase que 30 instrumentos electrónicos $D_1$, $D_2$, ......,$D_{30}$, se usan de la manera siguiente: tan pronto como $D_1$ falla empieza a actuar $D_2$. Cuando $D_2$ falla empieza a actuar $D_3$, etc. Supóngase que el tiempo de falla de $D_i$ es una v.a. distribuida exponencialmente con parámetro $\lambda = 0.1$ por hora. Sea T el tiempo total de operación de los 30 instrumentos. ¿Cuál es la probabilidad de que T exceda 350 horas?

Solución:
Si $X_i$: "tiempo de falla del instrumento $D_i$" $i = 1,2,\dots,30$

Entonces $X_i \sim Exp(0.1)$ para $i = 1,2,\dots,30$

El tiempo total de operación de los 30 instrumentos es $T = \sum_{i=1}^{30} X_i$ , donde

$$E(T) = E\left(\sum_{i=1}^{30} X_i\right) = 30 \times E(X_i) = 30 \times \frac{1}{0.1} = 300$$

$$V(T)=V\left(\sum_{i=1}^{30}X_i\right)=30\times V(X_i)=30\times\frac{1}{0.1^2}=3000$$

Entonces por T.C.L. $\dfrac{T-300}{\sqrt{3000}}\sim N(0,1)$ ***aproximadamente*** pues $n=30$

La probabilidad pedida es

$$P(T>350)=P\left(\frac{T-300}{\sqrt{3000}}>\frac{350-300}{\sqrt{3000}}\right)\underset{\text{T.C.L.}}{\approx}1-\Phi\left(\frac{350-300}{\sqrt{3000}}\right)=1-\Phi(0.9128)=1-0.81859=0.18141$$

2- Suponga que el consumo de calorías por día de una determinada persona es una v.a. con media 3000 calorías y desviación estándar de 230 calorías. ¿Cuál es la probabilidad de que el promedio de consumo de calorías diario de dicha persona en el siguiente año (365 días) sea entre 2959 y 3050?

<u>Solución</u>:

Definimos las variables aleatorias

$X_i$: "cantidad de calorías que una persona consume en el día *i*" $i=1,2,\dots,365$

Se sabe que $E(X_i)=3000$ y $V(X_i)=230^2$

Si $\overline{X}=\dfrac{1}{365}\sum_{i=1}^{365}X_i$ entonces $E(\overline{X})=3000$ y $V(\overline{X})=\dfrac{\sigma^2}{n}=\dfrac{230^2}{365}$

La probabilidad pedida es

$$P\left(2959\le\overline{X}\le3050\right)=P\left(\frac{2959-3000}{230/\sqrt{365}}\le\frac{\overline{X}-3000}{230/\sqrt{365}}\le\frac{3050-3000}{230/\sqrt{365}}\right)\underset{\text{T.C.L.}}{\approx}$$

$$\approx\Phi\left(\frac{3050-3000}{230/\sqrt{365}}\right)-\Phi\left(\frac{2959-3000}{230/\sqrt{365}}\right)=\Phi(4.15)-\Phi(-3.40)\approx1-0=1$$

## Aplicaciones del Teorema central del límite

### Aproximación normal a la distribución binomial

El Teorema central del límite se puede utilizar para aproximar las probabilidades de algunas variables aleatorias discretas cuando es difícil calcular las probabilidades exactas para valores grandes de los parámetros.

Supongamos que *X* tiene una distribución binomial con parámetros *n* y *p*. Para calcular $P(X\le k)$ debemos hacer la suma $P(X\le k)=\sum_{i=0}^{k}P(X=i)$ o recurrir a las tablas de la F.d.a. , pero para valores de *n* grandes no existen tablas, por lo tanto habría que hacer el cálculo en forma directa y muchas veces es laborioso.

Como una opción podemos considerar a X como suma de variables aleatorias más simples, específicamente, si definimos

$$X_i = \begin{cases} 1 \quad si \quad en \quad la \quad í-ésima \quad repetición \quad de \quad \varepsilon \quad ocurre \quad éxito \\ \quad 0 \qquad caso \quad contrario \end{cases} \qquad i = 1,2,\ldots,n$$

entonces cada $X_i$ se la puede considerar $B(1, p)$, y además $X_1, X_2, \ldots, X_n$ son independientes

Podemos escribir $X = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$ y ***si n es grande*** entonces *X* tendrá ***aproximadamente*** una distribución normal con parámetros $np$ y $np(1-p)$, es decir

$$Z_n = \frac{X - n\mu}{\sqrt{n\sigma^2}} = \frac{X - n.p}{\sqrt{n.p(1-p)}} \approx N(0,1) \quad \text{si } n \text{ es lo suficientemente grande}$$

Observaciones:
1- La aproximación normal a la distribución binomial funciona bien aun cuando *n* no sea muy grande si *p* no está demasiado cerca de cero o de uno. En particular la aproximación normal a la binomial es buena si *n* es grande , $np > 5$ y $n(1-p) > 5$, ***pero es más efectivo aplicar esta aproximación cuando*** $np > 10$ ***y*** $n(1-p) > 10$

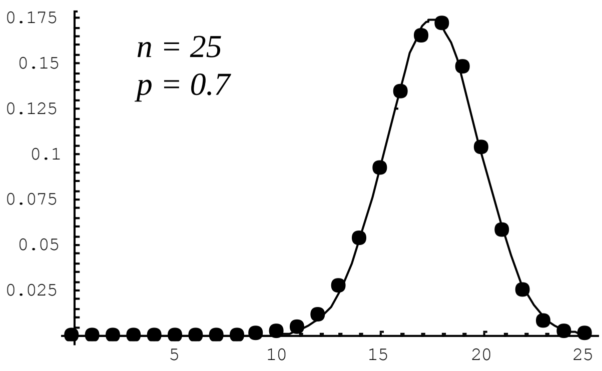
2- Corrección por continuidad.
Acabamos de ver que si $X \sim B(n,p)$ entonces, para *n* suficientemente grande, podemos considerar que aproximadamente es $X \sim N[n.p, n.p(1-p)]$. El problema que surge de inmediato si deseo calcular, por ejemplo, la probabilidad de que $X = k$ (con *k* alguno de los valores posibles 0,1,2,…,*n*) es que la binomial es una distribución discreta y tiene sentido calcular probabilidades como $P(X = k)$ mientras que la normal es una distribución continua y, en consecuencia, $P(X = k) = 0$ puesto que para una variable aleatoria continua la probabilidad de que ésta tome un valor aislado es cero. Esto se resuelve si se considera $P(X = k) \approx P\left(k - \frac{1}{2} \le X \le k + \frac{1}{2}\right)$

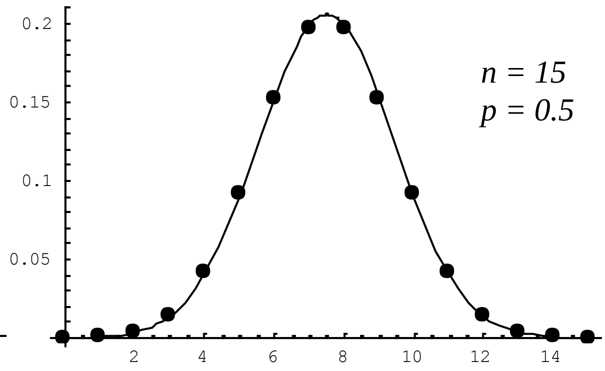
También se puede usar esta corrección para mejorar la aproximación en otros casos, específicamente en lugar de $P(X \le k)$ calculamos

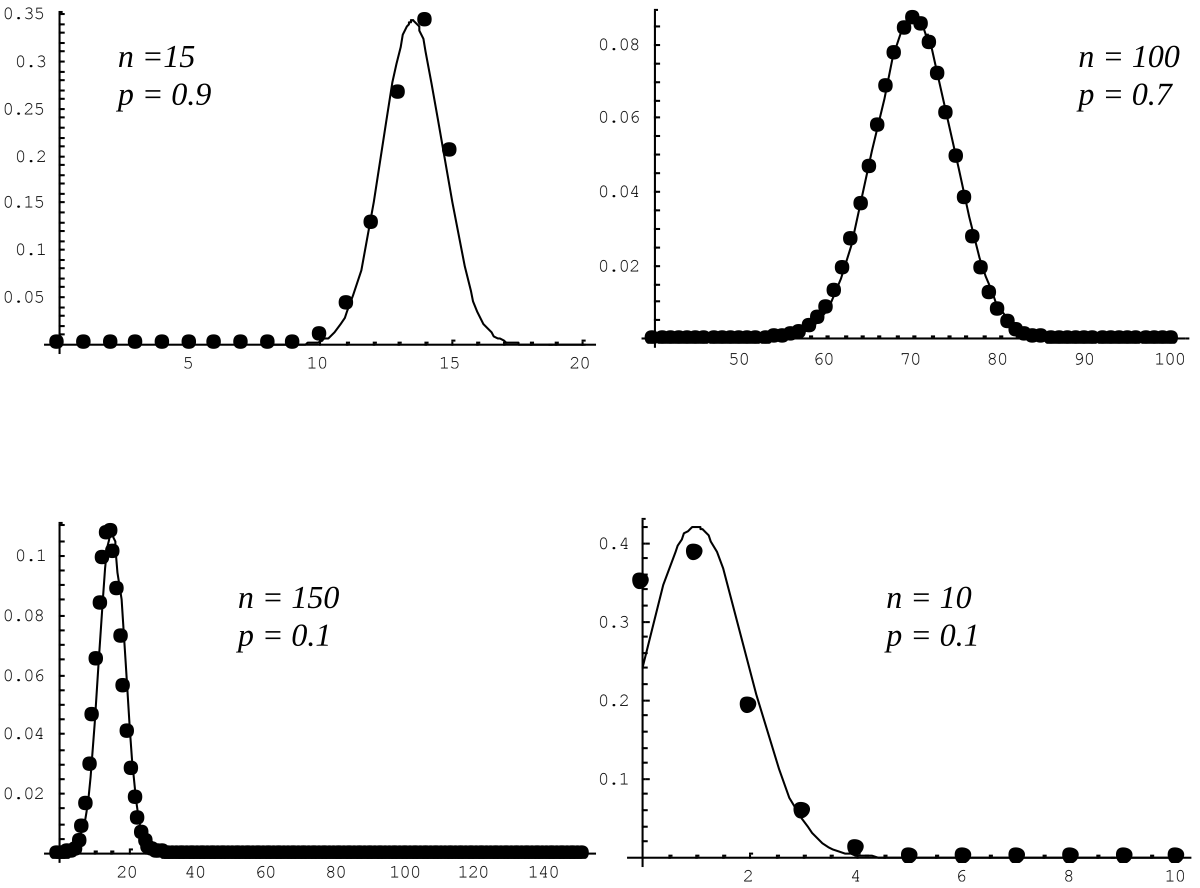
$$P(X \le k) \approx P\left(X \le k + \frac{1}{2}\right)$$

Y en lugar de $P(X \ge k) \approx P\left(X \ge k - \frac{1}{2}\right)$

En los gráficos siguientes se muestra para diferentes valores de n y p cómo aproxima la distribución $N(np,\ np(1-p))$ a la distribución $B(n, p)$






<u>Ejemplos</u>:

1- Sea $X \sim B(25,0.4)$. Hallar las probabilidades exactas de que $X \le 8$ y $X = 8$ y comparar estos resultados con los valores correspondientes encontrados por la aproximación normal.

<u>Solución</u>:

De la tabla de la $F.d.a.$ de la binomial encontramos $P(X \le 8) = 0.274$

Y $P(X = 8) = P(X \le 8) - P(X \le 7) = 0.274 - 0.154 = 0.120$

Ahora usamos la aproximación normal

$$P(X \le 8) \approx P(X \le 8.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \le \frac{8.5 - 10}{\sqrt{25 \times 0.4 \times 0.6}}\right) \approx \Phi(-0.61) = 0.2709$$

corrección por continuidad

Observar que el valor aproximado está muy cercano al valor exacto para $P(X \le 8) = 0.274$

$$P(X = 8) \approx P(7.5 \le X \le 8.5) = P\left(\frac{7.5-10}{\sqrt{6}} \le \frac{X-10}{\sqrt{6}} \le \frac{8.5-10}{\sqrt{6}}\right) = P\left(-1.02 \le \frac{X-10}{\sqrt{6}} \le -0.61\right) =$$

$$= 0.2709 - 0.1593 = 0.1170$$

Nuevamente este valor aproximado está muy cerca del valor real de $P(X = 8) = 0.120$

2- Suponga que el 10% de todos los ejes de acero producidos por cierto proceso están fuera de especificaciones, pero se pueden volver a trabajar (en lugar de tener que enviarlos a la chatarra). Considere una muestra aleatoria de 200 ejes y denote por *X* el número entre ellos que estén fuera de especificaciones y se puedan volver a trabajar. ¿Cuál es la probabilidad (aproximada) de que *X* sea

a) a lo sumo 30?
b) menos de 30?
c) entre 15 y 25 (inclusive)?

Solución:
Sea la v.a. *X*: "número de ejes fuera de especificaciones"
Entonces $X \sim B(200,\ 0.1)$, además $np = 200 \times 0.1 = 20 > 5$ y $n(1-p) = 200 \times (1-0.1) = 180 > 5$

Por lo tanto podemos aplicar la aproximación normal a la binomial
a) la probabilidad pedida es $P(X \le 30)$

$$P(X \le 30) \approx P(X \le 30.5) = P\left( \frac{X - np}{\sqrt{np(1-p)}} \le \frac{30.5-20}{\sqrt{18}} \right) \approx \Phi\left( \frac{30.5-20}{\sqrt{18}} \right) = \Phi(2.474) = 0.993244$$

b) La probabilidad pedida es $P(X < 30)$
Al ser *X* una v.a. ***discreta*** con distribución binomial $P(X < 30) = P(X \le 29)$

$$P(X \le 29) \approx P(X \le 29.5) \approx \Phi\left( \frac{29.5-20}{\sqrt{18}} \right) = \Phi(2.2391) = 0.98745$$

c)

$$P(15 \le X \le 25) \approx P(14.5 \le X \le 25.5) \approx \Phi\left( \frac{25.5-20}{\sqrt{18}} \right) - \Phi\left( \frac{14.5-20}{\sqrt{18}} \right) =$$
$$= \Phi(1.2963) - \Phi(-1.2963) = 2\Phi(1.2963) - 1 = 2 \times 0.90147 - 1 = 0.80294$$

3- El gerente de un supermercado desea recabar información sobre la proporción de clientes a los que no les agrada una nueva política respecto de la aceptación de cheques. ¿Cuántos clientes tendría que incluir en una muestra si desea que la fracción de la muestra se desvíe a lo mas en 0.15 de la verdadera fracción, con probabilidad de 0.98?.

Solución:
Sea *X*: "número de clientes a los que no les agrada la nueva política de aceptación de cheques"
Entonces $X \sim B(n,\ p)$ donde *p* es desconocido y es la ***verdadera proporción*** de clientes a los que no les agrada la nueva política de aceptación de cheques. El gerente tomará una muestra de *n* clientes para ***"estimar"*** *p* con $\overline{X} = \frac{X}{n}$ ya que $\overline{X} = \frac{X}{n}$ es la proporción de clientes a los que no les agrada la nueva política de aceptación de cheques ***en la muestra de n clientes***. Si no se toman a ***todos los clientes,*** entonces $\overline{X} = \frac{X}{n}$ ***no será igual a p***.

La pregunta es cuál debe ser *n* para que $\overline{X} = \frac{X}{n}$ se aleje del verdadero *p* en menos de 0.15 con probabilidad 0.98 por lo menos, o sea para que $P(|\overline{X} - p| \le 0.15) \ge 0.98$
Entonces planteamos

$$P(|\overline{X} - p| \le 0.15) = P(-0.15 \le \overline{X} - p \le 0.15) = P\left( \frac{-0.15n}{\sqrt{np(1-p)}} \le \frac{X - np}{\sqrt{np(1-p)}} \le \frac{0.15n}{\sqrt{np(1-p)}} \right) \approx$$

$$\approx \Phi\left(\frac{0.15n}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-0.15n}{\sqrt{np(1-p)}}\right) = 2\Phi\left(\frac{0.15n}{\sqrt{np(1-p)}}\right) - 1 \geq 0.98$$

Por lo tanto $\Phi\left(\frac{0.15n}{\sqrt{np(1-p)}}\right) \geq \frac{0.98+1}{2} = 0.99$

Además $\frac{0.15n}{\sqrt{np(1-p)}} = \frac{0.15\sqrt{n}}{\sqrt{p(1-p)}} \geq \frac{0.15\sqrt{n}}{\sqrt{0.5(1-0.5)}} = 0.3\sqrt{n}$

Entonces debe cumplirse que $0.3\sqrt{n} \geq 2.33$ o sea $n \geq \left(\frac{2.33}{0.3}\right)^2 = 60.3211$

O sea ***se debe tomar una muestra de al menos 61 clientes***

## Aproximación normal a la distribución Poisson

Se puede probar aplicando Teorema central del límite que

| |
|---|
| Si $X \sim P(\lambda)$ entonces para $\lambda$ suficientemente grande $\frac{X-\lambda}{\sqrt{\lambda}}$ tiene aproximadamente distribución $N(0,1)$ |

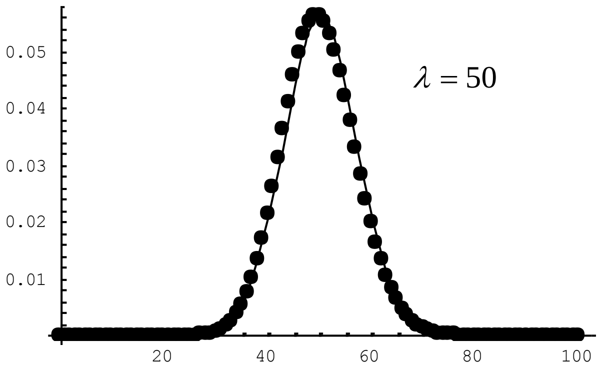
Es decir para $\lambda$ suficientemente grande $\frac{X-\lambda}{\sqrt{\lambda}} \approx N(0,1)$
En la práctica si $\lambda \geq 30$ la aproximación es buena.

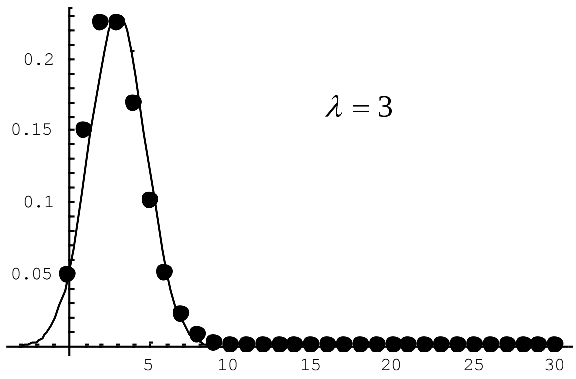
Observación: la demostración es sencilla si $\lambda$ es igual a un número natural *n* pues, si consideramos las variables aleatorias $X_i \sim P(1)$ con $i = 1,2,\ldots,n$ independientes, entonces ya sabemos que $\sum_{i=1}^{n} X_i \sim P\left(\sum_{i=1}^{n} 1\right)$, es decir $\sum_{i=1}^{n} X_i \sim P(n)$

Pero además por T.C.L. si *n* es grande $\sum_{i=1}^{n} X_i$ tiene aproximadamente distribución normal con parámetros $n\mu = n \times 1 = n$ y $n\sigma^2 = n \times 1 = n$

O sea la distribución de $\sum_{i=1}^{n} X_i$ que es exactamente Poisson con parámetro *n*, se puede aproximar con una $N(n,n)$, por lo tanto $\frac{X-n}{\sqrt{n}} \approx N(0,1)$ aproximadamente para valores de *n* suficientemente grandes

En los gráficos siguientes se muestra para diferentes valores de $\lambda$ cómo aproxima la distribución $N(\lambda,\ \lambda)$ a la distribución $P(\lambda)$




Ejemplo:

El número de infracciones por estacionamiento en cierta ciudad en cualquier día hábil tiene una distribución de Poisson con parámetro λ = 50. ¿Cuál es la probabilidad aproximada de que:

a) más de 35 y a lo sumo 70 infracciones se expidan en un día en particular?

b) el número total de infracciones expedidas durante una semana de 5 días sea más 225 y a lo sumo 275?

Solución:

Sea $X$: "número de infracciones por estacionamiento en cierta ciudad en cualquier día hábil"

Entonces $X \sim P(\lambda)$ donde $\lambda = 50$

Como $\lambda = 50$ entonces $\dfrac{X-50}{\sqrt{50}} \approx N(0,1)$ (aproximadamente)

a) la probabilidad pedida es

$$P(35 < X \le 70) \approx \Phi\left(\frac{70-50}{\sqrt{50}}\right) - \Phi\left(\frac{35-50}{\sqrt{50}}\right) = \Phi(2.8284) - \Phi(-2.12132) =$$

$$= 0.997599 - 0.017 = 0.9805$$

b) Sea $Y$: "número total de infracciones expedidas durante una semana de 5 días"

Entonces $Y \sim P(\lambda)$ donde $\lambda = 50 \times 5 = 250$

La probabilidad pedida es

$$P(225 < Y \le 275) \approx \Phi\left(\frac{275-250}{\sqrt{250}}\right) - \Phi\left(\frac{225-250}{\sqrt{250}}\right) = \Phi(1.5811) - \Phi(-1.5811) =$$

$$= 2\Phi(1.5811) - 1 = 2 \times 0.94295 - 1 = 0.8859$$

# 8- Estimación puntual

## 8. 1 – Introducción

Supongamos la siguiente situación: en una fábrica se producen artículos, el interés está en la producción de un día, específicamente, de todos los artículos producidos en un día nos interesa una característica determinada, si el artículo es o no defectuoso. Sea $p$ la proporción de artículos defectuosos en la ***población***, es decir en la producción de un día.
Tomamos una ***muestra*** de 25 artículos, podemos definir la v.a. $X$: "número de artículos defectuosos en la muestra", y podemos asumir que $X \sim B(25, p)$.
En ***Probabilidades*** se ***conocían todos los datos sobre la v.a. X***, es decir conocíamos $p$. De esa forma podíamos responder preguntas como: ¿cuál es la probabilidad que entre los 25 artículos halla 5 defectuosos?. Si, por ejemplo, $p = 0.1$ entonces calculábamos $P(X = 5)$ donde $X \sim B(25,\ 0.1)$.
En ***Estadística desconocemos las características de X*** total o parcialmente, y a partir de la muestra de 25 artículos tratamos de ***inferir*** información sobre la distribución de $X$, o dicho de otra forma tratamos de inferir información sobre la ***población***.
Por ejemplo, en estadística sabremos que $X$ tiene distribución binomial pero ***desconocemos p***, y a partir de la muestra de 25 artículos trataremos de hallar información sobre $p$.
En Estadística nos haremos preguntas tales como: si en la muestra de 25 artículos se encontraron 5 defectuosos, ¿ese hecho me permite inferir que el verdadero $p$ es 0.1?.
El campo de la ***inferencia estadística*** está formado por los métodos utilizados para tomar decisiones o para obtener conclusiones sobre el o los parámetros de una población. Estos métodos utilizan la información contenida en una ***muestra*** de la población para obtener conclusiones.
La inferencia estadística puede dividirse en dos grandes áreas: ***estimación de parámetros*** y ***pruebas de hipótesis***.

## 8.2 – Muestreo aleatorio

En muchos problemas estadísticos es necesario utilizar una muestra de observaciones tomadas de la población de interés con objeto de obtener conclusiones sobre ella. A continuación se presenta la definición de algunos términos

| |
|---|
| Una ***población*** está formada por la totalidad de las observaciones en las cuales se tiene cierto interés |

En muchos problemas de inferencia estadística es poco práctico o imposible, observar toda la población, en ese caso se toma una parte o subconjunto de la población

| |
|---|
| Una ***muestra*** es un subconjunto de observaciones seleccionada de una población |

Para que las inferencias sean válidas, la muestra debe ser representativa de la población. Se selecciona una muestra aleatoria como el resultado de un mecanismo aleatorio. En consecuencia, la selección de una muestra es un experimento aleatorio, y cada observación de la muestra es el valor observado de una variable aleatoria. Las observaciones en la población determinan la distribución de probabilidad de la variable aleatoria.
Para definir muestra aleatoria, sea $X$ la v.a. que representa el resultado de tomar una observación de la población. Sea $f(x)$ la *f.d.p.* de la v.a. $X$. supongamos que cada observación en la muestra se ob-

tiene de manera independiente, bajo las mismas condiciones. Es decir, las observaciones de la muestra se obtienen al observar *X* de manera independiente bajo condiciones que no cambian, digamos *n* veces.
Sea $X_i$ la variable aleatoria que representa la *i*-ésima observación. Entonces $X_1, X_2, ..., X_n$ constituyen una muestra aleatoria, donde los valores numéricos obtenidos son $x_1, x_2, ..., x_n$. Las variables aleatorias en una muestra aleatoria son independientes, con la misma distribución de probabilidad *f(x)* debido a que cada observación se obtiene bajo las mismas condiciones. Es decir las funciones de densidad marginales de $X_1, X_2, ..., X_n$ son todas iguales a *f(x)* y por independencia, la distribución de probabilidad conjunta de la muestra aleatoria es el producto de las marginales $f(x_1)f(x_2)...f(x_n)$

Las variables aleatorias $(X_1, X_2, ..., X_n)$ constituyen una ***muestra aleatoria*** de tamaño *n* de una v.a. $X$ si $X_1, X_2, ..., X_n$ son independientes idénticamente distribuidas

El propósito de tomar una muestra aleatoria es obtener información sobre los parámetros desconocidos de la población. Por ejemplo, se desea alcanzar una conclusión acerca de la proporción de artículos defectuosos en la producción diaria de una fábrica. Sea *p* la proporción de artículos defectuosos en la población, para hacer una inferencia con respecto a *p*, se selecciona una muestra aleatoria (de un tamaño apropiado) y se utiliza la proporción observada de artículos defectuosos en la muestra para estimar *p*.
La proporción de la muestra $\hat{p}$ se calcula dividiendo el número de artículos defectuosos en la muestra por el número total de artículos de la muestra. Entonces $\hat{p}$ es una función de los valores observados en la muestra aleatoria. Como es posible obtener muchas muestras aleatorias de una población, el valor de $\hat{p}$ cambiará de una a otra. Es decir $\hat{p}$ ***es una variable aleatoria***. Esta variable aleatoria se conoce como ***estadístico.***

Un ***estadístico*** es cualquier función de la muestra aleatoria

## Estadísticos usuales

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una v.a. $X$ donde $E(X) = \mu$ y $V(X) = \sigma^2$
Si desconocemos $\mu$ un estadístico que se utiliza para estimar ese parámetro es la ***media o promedio muestral*** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Análogamente si se desconoce $\sigma^2$ un estadístico usado para tener alguna información sobre ese parámetro es la ***varianza muestral*** que se define como $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

Otro estadístico es la ***desviación estándar muestral*** $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$

Como un estadístico es una variable aleatoria, éste tiene una distribución de probabilidad, esperanza y varianza.
Una aplicación de los estadísticos es obtener ***estimaciones puntuales*** de los parámetros desconocidos de una distribución. Por ejemplo como se dijo antes se suelen estimar la media y la varianza de una población.

Cuando un estadístico se utiliza para estimar un parámetro desconocido se lo llama ***estimador puntual***. Es habitual simbolizar en forma genérica a un parámetro con la letra $\theta$ y al estadístico que se utiliza como estimador puntual de $\theta$, simbolizarlo con $\hat{\Theta}$.
Por lo tanto $\hat{\Theta}$ es una función de la muestra aleatoria: $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$
Al medir la muestra aleatoria se obtienen $x_1, x_2, \ldots, x_n$, y entonces ***el valor que toma*** $\hat{\Theta}$ es $\hat{\theta} = h(x_1, x_2, \ldots, x_n)$ y se denomina ***estimación puntual de*** $\theta$
El objetivo de la estimación puntual es seleccionar un número, a partir de los valores de la muestra, que sea el valor más probable de $\theta$.
Por ejemplo, supongamos que $X_1, X_2, X_3, X_4$ es una muestra aleatoria de una v.a. $X$. Sabemos que $X$ tiene distribución normal pero desconocemos $\mu$.
Tomamos como ***estimador*** de $\mu$ al promedio muestral $\overline{X}$, es decir $\hat{\mu} = \overline{X}$
Tomamos la muestra (medimos $X_1, X_2, X_3, X_4$) y obtenemos $x_1 = 24,\ x_2 = 30,\ x_3 = 27,\ x_4 = 32$
Entonces la ***estimación puntual*** de $\mu$ es $\overline{x} = \dfrac{24+30+27+32}{4} = 28.25$
Si la varianza $\sigma^2$ de $X$ también es desconocida, un estimador puntual usual de $\sigma^2$ es la varianza muestral, es decir $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$, para la muestra dada la estimación de $\sigma^2$ es 12.25.
Otro parámetro que a menudo es necesario estimar es la proporción $p$ de objetos de una población que cumplen una determinada característica.
En este caso el estimador puntual de $p$ sería $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ donde

$$X_i = \begin{cases} 1 \text{ si la í-ésima observación tiene la característica de interés} \\ 0 \quad \text{caso contrario} \end{cases} \qquad i = 1, 2, \ldots, n$$

Por lo tanto $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ es la ***proporción de objetos en la muestra*** cumplen la característica de interés

Puede ocurrir que se tenga más de un estimador para un parámetro, por ejemplo para estimar la media muestral se pueden considerar el promedio muestral, o también la semisuma entre $X_1$ y $X_n$, es decir $\hat{\mu} = \dfrac{X_1 + X_n}{2}$. En estos casos necesitamos de algún ***criterio para decidir cuál es mejor estimador*** de $\mu$.

## 8.3 – Criterios para evaluar estimadores puntuales

Lo que se desea de un estimador puntual es ***que tome valores "próximos" al verdadero parámetro.***

> Se dice que el estimador puntual $\hat{\Theta}$ es un ***estimador insesgado*** del parámetro $\theta$ si $E(\hat{\Theta}) = \theta$ cualquiera sea el valor verdadero de $\theta$

Podemos exigir que el estimador $\hat{\Theta}$ tenga una distribución cuya media sea $\theta$.

> La diferencia $E(\hat{\Theta}) - \theta$ se conoce como ***sesgo de estimador*** $\hat{\Theta}$**.** Anotamos $b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$

***Notar que si un estimador es insesgado entonces su sesgo es cero***

Ejemplos:

1- Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una v.a. $X$ donde $E(X) = \mu$ y $V(X) = \sigma^2$

Si desconocemos $\mu$ un estadístico que se utiliza usualmente para estimar este parámetro es la ***media o promedio muestral*** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Veamos si es un estimador insesgado de $\mu$. Debemos ver si $E(\overline{X}) = \mu$.

Usamos las propiedades de la esperanza, particularmente la propiedad de linealidad.

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i).$$

Pero, tratándose de las componentes de una muestra aleatoria es:

$E(X_i) = E(X) = \mu \quad \forall i = 1,2,\dots,n$. Luego:

$$E(\overline{X}) = \frac{1}{n} n\mu = \mu.$$

2- Sea $X$ una variable aleatoria asociada con alguna característica de los individuos de una población y sean $E(X) = \mu$ y $V(X) = \sigma^2$ . Sea $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ la varianza muestral (con $\overline{X} = \left(\sum_{i=1}^{n} X_i\right)/n$ la esperanza muestral) para una muestra aleatoria de tamaño $n$, $(X_1, X_2, ..., X_n)$.

Entonces $E(S^2) = \sigma^2$ es decir $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ es un estimador insesgado de $V(X) = \sigma^2$ pues:

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right).$$

Reescribiremos la suma de una forma más conveniente. Sumamos y restamos $\mu$ y desarrollamos el cuadrado:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n}\left(X_i - \mu + \mu - \overline{X}\right)^2 = \sum_{i=1}^{n}\left([X_i - \mu] + [\mu - \overline{X}]\right)^2 =$$

$$= \sum_{i=1}^{n}\left\{ [X_i - \mu]^2 + 2[X_i - \mu][\mu - \overline{X}] + [\mu - \overline{X}]^2 \right\} = \sum_{i=1}^{n}[X_i - \mu]^2 + 2[\mu - \overline{X}] \sum_{i=1}^{n}[X_i - \mu] + n[\mu - \overline{X}]^2 =$$

$$= \sum_{i=1}^{n}[X_i - \mu]^2 + 2[\mu - \overline{X}]n[\overline{X} - \mu] + n[\mu - \overline{X}]^2 = \sum_{i=1}^{n}[X_i - \mu]^2 - 2n[\mu - \overline{X}]^2 + n[\mu - \overline{X}]^2.$$

Esto es:

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}[X_i-\mu]^2-n\left[\mu-\overline{X}\right]^2$$

Entonces:

$$E\left(S^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}[X_i-\mu]^2-n\left[\mu-\overline{X}\right]^2\right)=$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E[X_i-\mu]^2-nE\left[\overline{X}-\mu\right]^2\right)=$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nE\left[\overline{X}-E\left(\overline{X}\right)\right]^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nV\left(\overline{X}\right)\right)=\frac{1}{n-1}\left(n\sigma^2-n\frac{\sigma^2}{n}\right),$$

donde en la última igualdad tuvimos en cuenta que $V(X_i)=V(X)=\sigma^2 \quad \forall i=1,2,\ldots,n$ y que $V\left(\overline{X}\right)=\frac{\sigma^2}{n}$. Luego llegamos a lo que se deseaba demostrar: $E\left(S^2\right)=\sigma^2$.

3- Supongamos que tomamos como estimador de $\sigma^2$ a $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

Entonces notar que podemos escribir $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{n-1}{n}\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}=\frac{n-1}{n}S^2$

Por lo tanto $E\left(\hat{\sigma}^2\right)=E\left(\frac{n-1}{n}S^2\right)=\frac{n-1}{n}E\left(S^2\right)=\frac{n-1}{n}\sigma^2\neq\sigma^2$

Es decir $\hat{\sigma}^2$ no es un estimador insesgado de $\sigma^2$, ***es sesgado***, y su sesgo es

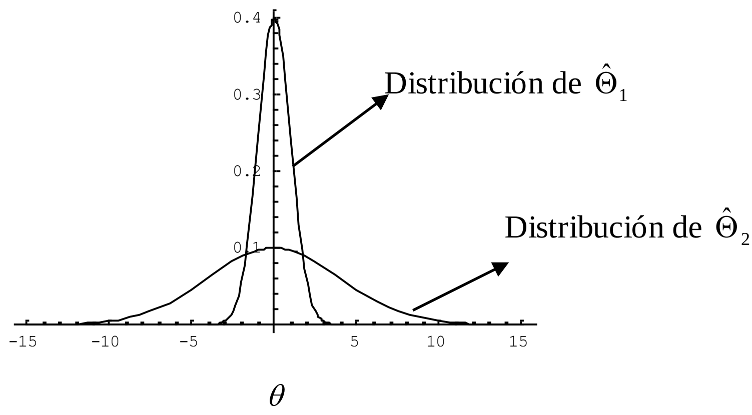
$$b\left(\hat{\sigma}^2\right)=E\left(\hat{\sigma}^2\right)-\sigma^2=\left(\frac{n-1}{n}\right)\sigma^2-\sigma^2=-\frac{1}{n}\sigma^2$$

Como el sesgo es negativo el estimador tiende a subestimar el valor de verdadero parámetro

En ocasiones hay más de un estimador insesgado de un parámetro $\theta$
Por lo tanto necesitamos un método para seleccionar un estimador entre varios estimadores insesgados.

**<u>Varianza y error cuadrático medio de un estimador puntual</u>**

Supongamos que $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores insegados de un parámetro $\theta$. Esto indica que la distribución de cada estimador está centrada en el verdadero parámetro $\theta$. Sin embargo las varianzas de estas distribuciones pueden ser diferentes. La figura siguiente ilustra este hecho.




Como $\hat{\Theta}_1$ tiene menor varianza que $\hat{\Theta}_2$, entonces es más probable que el estimador $\hat{\Theta}_1$ produzca una estimación más cercana al verdadero valor de $\theta$. Por lo tanto si tenemos dos estimadores insesgados se seleccionará aquel que tenga menor varianza.

Ejemplo: Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una v.a. $X$ donde $E(X) = \mu$ y $V(X) = \sigma^2$
Suponemos $\mu$ desconocido.

Estimamos al parámetro $\mu$ con la ***media o promedio muestral*** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Sabemos que es un estimador insesgado de $\mu$. Anotamos $\hat{\mu}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Supongamos que tomamos otro estimador para $\mu$, lo anotamos $\hat{\mu}_2 = \frac{X_1 + X_n}{2}$

Entonces como

$$E(\hat{\mu}_2) = E\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2}\left(E(X_1) + E(X_2)\right) = \frac{1}{2}(\mu + \mu) = \frac{1}{2}2\mu = \mu \ ,$$

$\hat{\mu}_2 = \frac{X_1 + X_n}{2}$ es también un estimador insesgado de $\mu$

¿Cuál de los dos estimadores es mejor?

Calculamos la varianza de cada uno utilizando las propiedades de la varianza.

Ya sabemos cuál es la varianza de $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (se la halló para T.C.L.):

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i),$$

donde en la última igualdad hemos tenido en cuenta que, por tratarse de una muestra aleatoria, las $X_i$ con $i$=1,2,…,$n$ son variables aleatorias independientes y, en consecuencia, la varianza de la suma de ellas es la suma de las varianzas. Si tenemos en cuenta que además todas tienen la misma distribución que $X$ y por lo tanto la misma varianza:

$V(X_i) = V(X) = \sigma^2 \quad \forall i = 1,2,\ldots,n$, tenemos

$$V(\overline{X}) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Análogamente calculamos la varianza de $\hat{\mu}_2 = \frac{X_1 + X_n}{2}$:

$$V(\hat{\mu}_2) = V\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{4}(V(X_1) + V(X_2)) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

Vemos que si $n > 2$ entonces $V(\hat{\mu}_1) < V(\hat{\mu}_2)$. Por lo tanto si $n > 2$ es mejor estimador $\hat{\mu}_1$

Supongamos ahora que $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de un parámetro $\theta$ y ***alguno de ellos no es insesgado***.
A veces es necesario utilizar un estimador sesgado. En esos casos puede ser importante el ***error cuadrático medio*** del estimador.

El ***error cuadrático medio*** de un estimador $\hat{\Theta}$ de un parámetro $\theta$ está definido como

$$ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - \theta)^2\right]$$

El error cuadrático medio puede escribirse de la siguiente forma:

$$ECM(\hat{\Theta}) = V(\hat{\Theta}) + (b(\hat{\Theta}))^2$$

Dem.) Por definición $ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - \theta)^2\right]$. Sumamos y restamos el número $E(\hat{\Theta})$:

$ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - E(\hat{\Theta}) + E(\hat{\Theta}) - \theta)^2\right]$ , y desarrollamos el cuadrado:

$$ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - E(\hat{\Theta}) + E(\hat{\Theta}) - \theta)^2\right] = E\left[(\hat{\Theta} - E(\hat{\Theta}))^2 + (E(\hat{\Theta}) - \theta)^2 + 2(\hat{\Theta} - E(\hat{\Theta}))(E(\hat{\Theta}) - \theta)\right] =$$

Aplicamos propiedades de la esperanza:

$$= \underbrace{E\left[(\hat{\Theta} - E(\hat{\Theta}))^2\right]}_{V(\hat{\Theta})} + \underbrace{(E(\hat{\Theta}) - \theta)^2}_{b(\hat{\Theta})^2} + 2(E(\hat{\Theta}) - \theta)\underbrace{E(\hat{\Theta} - E(\hat{\Theta}))}_{0} = V(\hat{\Theta}) + (b(\hat{\Theta}))^2$$

El error cuadrático medio es un criterio importante para comparar estimadores.

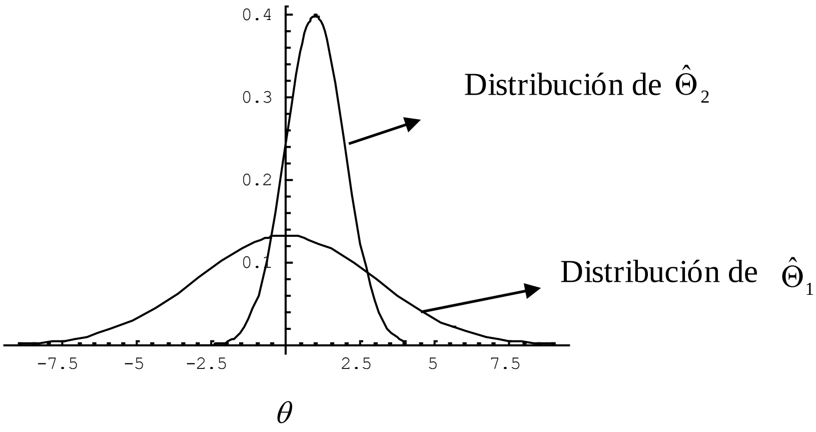
Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de un parámetro $\theta$.
La eficiencia relativa de $\hat{\Theta}_2$ con respecto a $\hat{\Theta}_1$ se define como $\dfrac{ECM(\hat{\Theta}_1)}{ECM(\hat{\Theta}_2)}$

Si la eficiencia relativa es menor que 1 entonces $\hat{\Theta}_1$ tiene menor error cuadrático medio que $\hat{\Theta}_2$
***Por lo tanto*** $\hat{\Theta}_1$ ***es más eficiente que*** $\hat{\Theta}_2$

Observaciones:

1- Si $\hat{\Theta}$ es un estimador insesgado de $\theta$, entonces $ECM(\hat{\Theta}) = V(\hat{\Theta})$

2- A veces es preferible utilizar estimadores sesgados que estimadores insesgados, si es que tienen un error cuadrático medio menor.
En el error cuadrático medio se consideran tanto la varianza como el sesgo del estimador.
Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de un parámetro $\theta$, tales que $E(\hat{\Theta}_1)=\theta$; $E(\hat{\Theta}_2)\neq\theta$ y $V(\hat{\Theta}_2)<V(\hat{\Theta}_1)$, habría que calcular el error cuadrático medio de cada uno, y tomar el que tenga menor error cuadrático medio. Pues puede ocurrir que $\hat{\Theta}_2$, aunque sea sesgado, al tener menor varianza tome valores mas cercanos al verdadero parámetro que $\hat{\Theta}_1$




Ejemplo:
Supóngase que $\hat{\Theta}_1$ ,$\hat{\Theta}_2$ y $\hat{\Theta}_3$ son dos estimadores de un parámetro $\theta$, y que $E(\hat{\Theta}_1)=E(\hat{\Theta}_2)=\theta$; $E(\hat{\Theta}_3)\neq\theta$, $V(\hat{\theta}_1)=10$, $V(\hat{\Theta}_2)=6$ y $E\left[\left(\hat{\Theta}_3-\theta\right)^2\right]=4$. Haga una comparación de estos estimadores. ¿Cuál prefiere y por qué?

Solución: Calculamos el error cuadrático medio de cada estimador
$ECM(\hat{\Theta}_1)=V(\hat{\Theta}_1)=10$ pues $\hat{\Theta}_1$ es insesgado
$ECM(\hat{\Theta}_2)=V(\hat{\Theta}_2)=6$ pues $\hat{\Theta}_2$ es insesgado
$ECM(\hat{\Theta}_3)=E\left[\left(\hat{\Theta}_3-\theta\right)^2\right]=4$ es dato

En consecuencia $\hat{\Theta}_3$ es el mejor estimador de los tres dados porque tiene menor error cuadrático medio.

## Consistencia de estimadores puntuales

Sea $\hat{\Theta}_n$ un estimador del parámetro $\theta$, basado en una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ de tamaño *n*. Se dice que $\hat{\Theta}_n$ es un estimador consistente de $\theta$ si

$$\lim_{n\to\infty} P\left(\left|\hat{\Theta}_n-\theta\right|\geq\varepsilon\right)=0 \quad \text{para todo } \varepsilon>0$$

Observación:

Este tipo de convergencia, que involucra a una sucesión de variables aleatorias, se llama ***convergencia en probabilidad*** y es la misma que consideramos en relación a la ***ley de los grandes números***

Suele escribirse también $\hat{\Theta}_n \xrightarrow{P} \theta$.

Este tipo de convergencia debe distinguirse de la considerada en relación al teorema central del límite. En este último caso teníamos una sucesión de distribuciones: $F_{Z_n}(z) = P(Z_n \le z)$ y se considera el límite $\lim_{n\to\infty} F_{Z_n}(z) = \lim_{n\to\infty} P(Z_n \le z) = \Phi(z)$.

Se habla, entonces, de ***convergencia en distribución*** y suele indicarse $Z_n \xrightarrow{d} Z \sim N(0,1)$.

**<u>Teorema</u>**. Sea $\hat{\Theta}_n$ un estimador del parámetro $\theta$ basado en una muestra aleatoria $(X_1, X_2, ..., X_n)$.
Si $\lim_{n\to\infty} E(\hat{\Theta}_n) = \theta$ y $\lim_{n\to\infty} V(\hat{\Theta}_n) = 0$, entonces $\hat{\Theta}_n$ es un estimador consistente de $\theta$.

Dem.)

Utilizamos la desigualdad de Chebyshev $\forall \varepsilon > 0$:

$$P\left(\left|\hat{\Theta}_n - \theta\right| \ge \varepsilon\right) \le \frac{E\left(\hat{\Theta}_n - \theta\right)^2}{\varepsilon^2} = \frac{1}{\varepsilon^2} ECM\left(\hat{\Theta}_n\right) = \frac{1}{\varepsilon^2}\left[V\left(\hat{\Theta}_n\right) + b\left(\hat{\Theta}_n\right)^2\right]$$

Entonces, al tomar el límite $\lim_{n\to\infty}$ y teniendo presente que $\lim_{n\to\infty} E(\hat{\Theta}_n) = \theta$ y $\lim_{n\to\infty} V(\hat{\Theta}_n) = 0$, vemos que $\lim_{n\to\infty} P\left(\left|\hat{\Theta}_n - \theta\right| \ge \varepsilon\right) = 0 \quad \forall \varepsilon > 0$, es decir $\hat{\Theta}_n$ es un estimador convergente de $\theta$.

<u>Ejemplo</u>:

Sea $X$ una variable aleatoria que describe alguna característica numérica de los individuos de una población y sean $\mu = E(X)$ y $\sigma^2 = V(X)$ la esperanza poblacional y la varianza poblacional, respectivamente. Sea $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la esperanza muestral basada en una muestra aleatoria $(X_1, X_2, ..., X_n)$. Entonces $\overline{X}$ es un estimador consistente de la esperanza poblacional $\mu = E(X)$.

Sabemos que

a) $E(\overline{X}) = \mu = E(X) \qquad \forall n$

b) $V(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{V(X)}{n} \qquad \forall n$

La propiedad a) ya me dice que $\overline{X}$ es un estimador insesgado de $\mu = E(X)$.

Por otra parte si a) vale para todo *n*, también vale en particular en el límite $n \to \infty$:

$\lim_{n\to\infty} E(\overline{X}) = \mu = E(X)$.

Además, de b) deducimos inmediatamente que

$\lim_{n\to\infty} V(\overline{X}) = 0$.

Por lo tanto vemos que $\overline{X}$ es un estimador consistente de $\mu = E(X)$.

## 8.4 – <u>Métodos de estimación puntual</u>

Los criterios anteriores establecen propiedades que es deseable que sean verificadas por los estimadores. Entre dos estimadores posibles para un dado parámetro poblacional es razonable elegir aquél que cumple la mayor cantidad de criterios o alguno en particular que se considera importante para el problema que se esté analizando. Sin embargo estos criterios no nos enseñan por sí mismos a construir los estimadores. Existen una serie de métodos para construir estimadores los cuales en general se basan en principios básicos de razonabilidad. Entre éstos podemos mencionar:

- ***Método de los momentos***

- ***Método de máxima verosimilitud***

## Método de los momentos

Se puede probar usando la desigualdad de Chebyshev el siguiente resultado:

Ley débil de los grandes números:
Sean $(X_1, X_2, \ldots, X_n)$ $n$ variables aleatorias independientes todas las cuales tienen la misma esperanza $\mu = E(X)$ y varianza $\sigma^2 = V(X)$. Sea $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ . Entonces

$$\lim_{n\to\infty} P\left(\left|\overline{X} - \mu\right| \geq \varepsilon\right) = 0$$

Decimos que $\overline{X}$ converge a $\mu$ en probabilidad y lo indicamos: $\overline{X} \xrightarrow{p} \mu$ .

Definimos los ***momentos de orden k de una variable aleatoria*** como:

$$\mu_k = E\left(X^k\right) = \sum_{x_i \in R_X} x_i^k\, p(x_i) \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es discreta}$$

$$\mu_k = E\left(X^k\right) = \int_{-\infty}^{+\infty} x^k f(x)dx \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es continua,}$$

y definimos los correspondientes momentos ***muestrales de orden k*** como:

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad (k = 0,1,2,\ldots),$$

Entonces la ley débil de los grandes números se puede generalizar:

$$\lim_{n\to\infty} P\left(\left|M_k - \mu_k\right| \geq \varepsilon\right) = 0 \qquad (k = 0,1,2,\ldots).$$

De acuerdo con esto parece razonable estimar los momentos poblacionales de orden $k$ mediante los momentos muestrales de orden $k$: $\mu_k \sim M_k \quad (k = 0,1,2,\ldots)$.

Supongamos, entonces, una variable aleatoria $X$ y supongamos que la distribución de $X$ depende de $r$ parámetros $\theta_1, \theta_2, \ldots, \theta_r$, esto es la *fdp* poblacional es $p(x_i, \theta_1, \theta_2, \ldots, \theta_r)$ si $X$ es discreta o $f(x, \theta_1, \theta_2, \ldots, \theta_r)$ si es continua. Sean $\mu_1, \mu_2, \ldots, \mu_r$ los primeros $r$ momentos poblacionales:

$$\mu_k = E(X^k) = \sum_{x_i \in R_X} x_i^k \, p(x_i, \theta_1, \theta_2, \ldots, \theta_r) \qquad (k = 1,2,\ldots,r) \qquad \text{Si } X \text{ es discreta}$$

$$\mu_k = E(X^k) = \int_{-\infty}^{+\infty} x^k f(x, \theta_1, \theta_2, \ldots, \theta_r) dx \qquad (k = 1,2,\ldots,r) \qquad \text{Si } X \text{ es continua,}$$

y sean

$M_k = \dfrac{1}{n} \sum_{i=1}^{n} X_i^k \quad (k = 1,2,\ldots r,)$ los $r$ primeros momentos maestrales para una muestra de tamaño $n$ $(X_1, X_2, \ldots, X_n)$. Entonces ***el método de los momentos consiste en plantear el sistema de ecuaciones***:

$$\begin{cases} \mu_1 = M_1 \\ \mu_2 = M_2 \\ \vdots \quad \vdots \; \vdots \\ \mu_r = M_r \end{cases}$$

Es decir

$$\begin{cases} \sum_{x_i \in R_X} x_i \, p(x_i, \theta_1, \theta_2, \ldots, \theta_r) = \dfrac{1}{n} \sum_{i=1}^{n} X_i^1 \\ \sum_{x_i \in R_X} x_i^2 \, p(x_i, \theta_1, \theta_2, \ldots, \theta_r) = \dfrac{1}{n} \sum_{i=1}^{n} X_i^2 \\ \qquad \vdots \qquad\qquad \vdots \quad \vdots \\ \sum_{x_i \in R_X} x_i^r \, p(x_i, \theta_1, \theta_2, \ldots, \theta_r) = \dfrac{1}{n} \sum_{i=1}^{n} X_i^r \end{cases} \qquad \text{Si } X \text{ es discreta,}$$

o

$$\begin{cases} \int_{-\infty}^{+\infty} x f(x, \theta_1, \theta_2, \ldots, \theta_r) dx = \dfrac{1}{n} \sum_{i=1}^{n} X_i^1 \\ \int_{-\infty}^{+\infty} x^2 f(x, \theta_1, \theta_2, \ldots, \theta_r) dx = \dfrac{1}{n} \sum_{i=1}^{n} X_i^2 \\ \qquad \vdots \qquad\qquad \vdots \quad \vdots \\ \int_{-\infty}^{+\infty} x^r f(x, \theta_1, \theta_2, \ldots, \theta_r) dx = \dfrac{1}{n} \sum_{i=1}^{n} X_i^r \end{cases} \qquad \text{Si } X \text{ es continua.}$$

Resolviendo estos sistema de ecuaciones para los parámetros desconocidos $\theta_1,\theta_2,...,\theta_r$ en función de la muestra aleatoria $(X_1,X_2,...,X_n)$ obtenemos los estimadores:

$$\begin{cases} \hat{\Theta}_1 = H_1(X_1,X_2,...,X_n) \\ \hat{\Theta}_2 = H_2(X_1,X_2,...,X_n) \\ \vdots \\ \hat{\Theta}_r = H_r(X_1,X_2,...,X_n) \end{cases}$$

Observación:
En la forma que presentamos aquí el método necesitamos conocer la forma de la *fdp* poblacional, por lo tanto estamos frente a un caso de estimación puntual ***paramétrica***.

Ejemplos:
1- Sea *X* una variable aleatoria. Supongamos que *X* tiene ***distribución gama con parámetros*** $\sigma$ ***y*** $\lambda$ : $X \sim \Gamma(\sigma,\lambda)$, es decir su *fdp* está dada por:

$$f(x)=\begin{cases} \dfrac{1}{\sigma\Gamma(\lambda)}\left(\dfrac{x}{\sigma}\right)^{\lambda-1} e^{-\frac{x}{\sigma}} & x>0 \\ 0 & demás \quad valores \end{cases}$$

con $\sigma>0$; $\lambda>0$ y $\Gamma(\lambda)=\int_0^\infty x^{\lambda-1}e^{-x}dx$.

Sea $(X_1,X_2,...,X_n)$ una muestra aleatoria de tamaño *n*. Deseamos calcular los estimadores de $\sigma$ y $\lambda$ dados por el método de los momentos.

Solución:
Como tenemos dos parámetros desconocidos a estimar, planteamos el sistema de ecuaciones:

$$\begin{cases} \mu_1 = M_1 \\ \mu_2 = M_2 \end{cases}$$

Se puede probar que

$$\mu_1 = \lambda.\sigma$$
$$\mu_2 = \lambda^2.\sigma^2 + \lambda.\sigma^2$$

Tenemos, entonces, el sistema de ecuaciones

$$\begin{cases} \lambda.\sigma = \dfrac{1}{n}\sum_{i=1}^{n} X_i \\ \lambda^2.\sigma^2 + \lambda.\sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases} \Rightarrow \begin{cases} \lambda.\sigma = \overline{X} \\ \lambda^2.\sigma^2 + \lambda.\sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

Reemplazando en la segunda ecuación: $\overline{X}^2 + \sigma\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} \quad \Rightarrow \quad \sigma = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} - \overline{X}^2}{\overline{X}}$

Y despejando $\lambda$ de la primera ecuación y reemplazando la expresión hallada para $\sigma$

$$\begin{cases} \hat{\lambda} = \dfrac{n\overline{X}^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \\ \hat{\sigma} = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n\overline{X}} \end{cases}$$

2- Sea $\left(X_1, X_2, \ldots, X_n\right)$ una muestra aleatoria de tamaño *n* de una v.a. *X* donde $X \sim U[0,\theta]$, $\theta$ desconocido. Hallar el estimador de $\theta$ por el método de los momentos.

Solución:
Planteamos la ecuación: $\mu_1 = M_1$

Sabemos que $\mu_1 = E(X) = \frac{0+\theta}{2} = \frac{\theta}{2}$. Entonces $\frac{\theta}{2} = \overline{X} \quad \Rightarrow \quad \hat{\Theta} = 2\overline{X}$

Observación: notar que el estimador $\hat{\Theta} = 2\overline{X}$ es un estimador consistente de $\theta$, pues

$E\left(\hat{\Theta}\right) = E\left(2\overline{X}\right) = 2E\left(\overline{X}\right) = 2\frac{\theta}{2} = \theta$ y $V\left(\hat{\Theta}\right) = V\left(2\overline{X}\right) = 4V\left(\overline{X}\right) = 4\frac{(\theta-0)^2}{12n} = \frac{\theta^2}{3n} \xrightarrow[n\to\infty]{} 0$

3- Sea $\left(X_1, X_2, \ldots, X_n\right)$ una muestra aleatoria de una v.a. $X \sim N(\mu, \sigma^2)$.
Encuentra los estimadores de $\mu$ y $\sigma$ por el método de momentos.

Solución:
Planteamos las ecuaciones

$$\begin{cases} \mu_1 = M_1 \\ \mu_2 = M_2 \end{cases} \quad \Rightarrow \quad \begin{cases} \mu = \overline{X} \\ E\left(X^2\right) = \dfrac{1}{n}\sum_{i=1}^{n} X_i^{\,2} \end{cases}$$

pero en general es válido que $V(X) = E(X^2) - \mu^2 \quad \Rightarrow \quad E(X^2) = V(X) + \mu$

Entonces las ecuaciones quedan

$$\begin{cases} \mu = \overline{X} \\ \sigma^2 + \mu^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^{\,2} \end{cases} \quad \Rightarrow \quad \begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^{\,2} - \overline{X}^2 \end{cases}$$

4- Sea $\left(X_1, X_2, \ldots, X_n\right)$ una muestra aleatoria de una v.a. $X \sim N(0, \sigma^2)$.
Hallar un estimador por el método de los momentos de $\sigma^2$

Solución: en este caso no es conveniente plantear $\mu_1 = M_1$ pues quedaría

la ecuación $0 = \overline{X}$ que no conduce a nada.
Entonces podemos plantear $\mu_2 = M_2$ es decir

$$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2} \quad \Rightarrow \quad \sigma^2 + 0 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2} \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2}$$

Observación: si $\hat{\Theta}$ es un estimador por el método de los momentos de un parámetro $\theta$, el estimador de los momentos de $g(\theta)$ es $g(\hat{\Theta})$, si $g(x)$ es una función inyectiva.
Por ejemplo, en el ejemplo anterior un estimador de $\sigma$ por el método de los momentos sería $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2}}$ . Notar que $g(x) = \sqrt{x}$ es inyectiva para los reales positivos.

## Método de máxima verosimilitud

Uno de los mejores métodos para obtener un estimador puntual de un parámetro es el método de máxima verosimilitud.

Supongamos que *X* es una v.a. discreta con función de distribución de probabilidad $p(x,\theta)$, donde $\theta$ es un parámetro desconocido. Sean $x_1, x_2,...,x_n$ los valores observados de una muestra aleatoria de tamaño *n*.
Se define la ***función de verosimilitud*** como la función de distribución conjunta de las observaciones:

$$L(x_1, x_2,...,x_n, \theta) = P(X_1 = x_1)P(X_2 = x_2)..P(X_n = x_n) = p(x_1,\theta).p(x_2,\theta).....p(x_n,\theta)$$

Notar que la función de verosimilitud es una función de $\theta$.
El estimador de máxima verosimilitud de $\theta$ es aquel valor de $\theta$ que maximiza la función de verosimilitud

***La interpretación del método sería: el estimador de máxima verosimilitud es aquel valor del parámetro que maximiza la probabilidad de ocurrencia de los valores muestrales***

La adaptación para el caso en que X es una v.a. continua sería la siguiente

Supongamos que *X* es una v.a. continua con función de densidad de probabilidad $f(x,\theta)$, donde $\theta$ es un parámetro desconocido. Sean $x_1, x_2,...,x_n$ los valores observados de una muestra aleatoria de tamaño *n*.
Se define la ***función de verosimilitud*** como la función de distribución conjunta de las observaciones:

$$L(x_1, x_2,...,x_n, \theta) = f(x_1,\theta).f(x_2,\theta).....f(x_n,\theta)$$

La función de verosimilitud es una función de $\theta$.
El estimador de máxima verosimilitud de $\theta$ es aquel valor de $\theta$ que maximiza la función de verosimilitud

Notación: abreviamos estimador de máxima verosimilitud con EMV

Ejemplos:

1- Sea $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de una v.a. $X \sim B(1, p)$

Por ejemplo, se eligen al azar $n$ objetos de una línea de producción, y cada uno se clasifica como defectuoso (en cuyo caso $x_i = 1$) o no defectuoso (en cuyo caso $x_i = 0$).

Entonces $p = P(X_i = 1)$, es decir es la verdadera proporción de objetos defectuosos en la producción total.

Queremos hallar el EMV de $p$

Solución:

Si $X \sim B(1, p)$ entonces $P(X = k) = \binom{1}{k} p^k (1-p)^{1-k} \quad k = 0,1$

Planteamos la función de verosimilitud

$$L(x_1, x_2, .., x_n; p) = p(x_1; p)p(x_2; p)..p(x_n; p) = \left[p^{x_1}(1-p)^{1-x_1}\right]\left[p^{x_2}(1-p)^{1-x_2}\right]..\left[p^{x_n}(1-p)^{1-x_n}\right]$$

Esto puede escribirse:

$$L(x_1, x_2, \ldots, x_n; p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Para maximizar la función de verosimilitud y ***facilitar los cálculos*** tomamos el logaritmo natural de $L$

Pues maximizar $L$ es equivalente a maximizar *ln(L)* y al tomar logaritmos transformamos productos en sumas.

Entonces

$$\ln(L(x_1, x_2, \ldots, x_n; p)) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

Y ahora podemos maximizar la función derivando e igualando a cero

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; p)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$$

de donde despejando $p$

$$p = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \text{la proporción de defectuosos en la muestra}$$

Por lo tanto se toma como estimador a $\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

2- El tiempo de fallar $T$ de una componente tiene una distribución exponencial con parámetro $\lambda$: $T \sim Exp(\lambda)$, es decir la *fdp* es

$$f(t; \lambda) = \begin{cases} \lambda e^{-\lambda t} & 0 \le t < \infty \\ 0 & demás \quad valores \end{cases}$$

Recordemos que la esperanza y varianza son:

$E(T) = 1/\lambda$ y $V(T) = 1/\lambda^2$ , respectivamente.

Se desea calcular el estimador de máxima verosimilitud del parámetro $\lambda$ para una muestra de tamaño *n*.

Solución:
La función de probabilidad es:
$L(t_1, t_2, \ldots, t_n; \lambda) = f(t_1; \lambda) f(t_2; \lambda) \ldots f(t_n; \lambda) = \left[\lambda e^{-\lambda t_1}\right] \times \left[\lambda e^{-\lambda t_2}\right] \times \ldots \times \left[\lambda e^{-\lambda t_n}\right]$,

que puede escribirse:

$$L(t_1, t_2, \ldots, t_n; \lambda) = (\lambda)^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

Nuevamente tomamos logaritmo natural

$$\ln L(t_1, t_2, \ldots, t_n; \sigma) = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

$$\frac{\partial \ln L(t_1, t_2, \ldots, t_n; \lambda)}{\partial \lambda} = n\frac{1}{\lambda} - \sum_{i=1}^{n} T_i = 0$$

de donde podemos despejar $\lambda$ :

$\lambda = \dfrac{n}{\sum_{i=1}^{n} t_i} = \bar{t}$ , entonces el estimador de $\lambda$ es $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} T_i}$

El método de máxima verosimilitud presenta, algunas veces, dificultades para maximizar la función de verosimilitud debido a que la ecuación obtenida a partir de $\dfrac{d}{d\theta} L(\theta) = 0$ no resulta fácil de resolver. O también puede ocurrir que los métodos de cálculo para maximizar $L(\theta)$ no son aplicables.

Por ejemplo:
Sea $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de tamaño *n* de una v.a. *X* donde $X \sim U[0, \theta]$, $\theta$ desconocido. Hallar el estimador de $\theta$ por el método máxima verosimilitud.
Solución:
La *f.d.p.* de *X* es

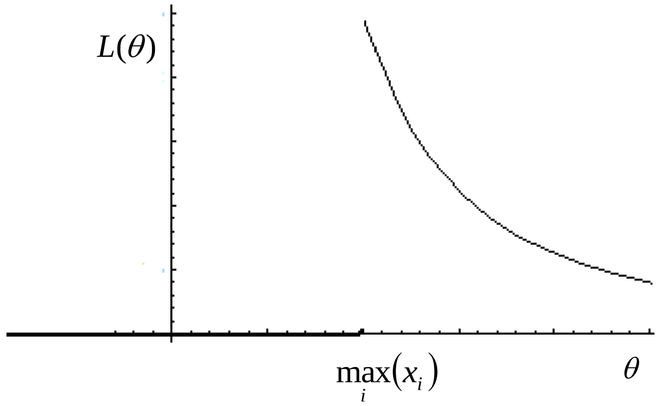
$$f(x) = \begin{cases} \dfrac{1}{\theta} & si \quad 0 < x < \theta \\ 0 & caso \quad contrario \end{cases}$$

Planteamos la función de verosimilitud

$$L(x_1, x_2, \ldots x_n, \theta) = \begin{cases} \dfrac{1}{\theta^n} & si \quad 0 < x_i < \theta \quad \forall i \\ 0 & caso \quad contrario \end{cases} = \begin{cases} \dfrac{1}{\theta^n} & si \quad \max_i (x_i) < \theta \\ 0 & caso \quad contrario \end{cases}$$

Si derivamos con respecto a $\theta$ obtenemos $\frac{d}{d\theta}\theta^{-n} = -\frac{n}{\theta^{n+1}}$ que es siempre menor que cero. Por lo tanto la función de verosimilitud es una función decreciente para todos los $\theta > \max_i(x_i)$

Si hacemos un gráfico de la función de verosimilitud




Vemos que donde la función tiene el máximo hay una discontinuidad no evitable.

Por lo tanto $\hat{\Theta} = \max_i(x_i)$

El método de máxima verosimilitud puede emplearse en el caso donde hay más de un parámetro desconocido para estimar. En ese caso la función de verosimilitud es una función de varias variables. Específicamente si tenemos para estimar *k* parámetros $\theta_1, \theta_2, ..\theta_k$ , entonces la función de verosimilitud es una función de *k* variables $L(x_1, x_2, ..., x_n, \theta_1, \theta_2, ..\theta_k)$ y los estimadores de máxima verosimilitud $\hat{\Theta}_1, \hat{\Theta}_2, ..\hat{\Theta}_k$ se obtienen al plantear ( si existen las derivadas parciales) y resolver el sistema de *k* ecuaciones con *k* incógnitas $\theta_1, \theta_2, ..\theta_k$

$$\frac{d}{d\theta_i} L(x_1, x_2, ..., x_n, \theta_1, \theta_2, ..\theta_k) = 0 \qquad i = 1,2,..k$$

<u>Ejemplo</u>:

La variable aleatoria *X* tiene distribución $N(\mu, \sigma^2)$ con $\mu$ y $\sigma^2$ ambos parámetros desconocidos para los cuales se desea encontrar los estimadores máxima verosimilitud. La *fdp* es

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty,$$

La función de verosimilitud para una muestra aleatoria de tamaño *n* es

$$L(x_1, x_2, ..., x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2} =$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Luego

$$\ln L\left(x_1, x_2,\ldots,x_n;\mu,\sigma^2\right)=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

y el sistema de ecuaciones de verosimilitud queda:

$$\begin{cases}\dfrac{\partial \ln L\left(x_1, x_2,\ldots,x_n;\mu,\sigma^2\right)}{\partial\mu}=\displaystyle\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)=0\\[2ex] \dfrac{\partial \ln L\left(x_1, x_2,\ldots,x_n;\mu,\sigma^2\right)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\frac{\left(x_i-\mu\right)^2}{\sigma^4}=0\end{cases}$$

Resolvemos con respecto a $\mu$ y $\sigma^2$ :

$$\begin{cases}\mu=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i=\overline{x}\\[2ex] \sigma^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(x_i-\mu\right)^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\end{cases}$$

Entonces los estimadores máxima verosimilitud de $\mu$ y $\sigma^2$ son

$$\begin{cases}\hat{\mu}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i=\overline{X}\\[2ex] \hat{\sigma}^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\end{cases}$$

## Propiedades de los estimadores máxima verosimilitud

1- Los EMV pueden ser ***sesgados***, pero en general si $\hat{\Theta}$ es el EMV de un parámetro $\theta$ basado en una muestra de tamaño *n*, entonces $\lim_{n\to\infty} E(\hat{\Theta})=\theta$, es decir son ***asintóticamente insesgados***

2- Bajo condiciones bastantes generales se puede probar que los EMV son ***consistentes***

3- Bajo condiciones bastantes generales se puede probar que los EMV ***asintóticamente tienen varianza mínima***

4-Los EMV cumplen la ***propiedad de invarianza*** es decir:
si $\hat{\Theta}$ es un EMV de un parámetro $\theta$, el EMV de $g(\theta)$ es $g\left(\hat{\Theta}\right)$, si $g(x)$ es una función inyectiva.

Ejemplos:

1- Si consideramos nuevamente la situación considerada en el Ejemplo 2, donde teníamos una v.a. *T* cuya distribución es una exponencial: $T \sim Exp(\lambda)$, entonces, si queremos el EMV de la varianza poblacional, podemos calcularlo recordando que $V(T)=1/\lambda^2$, es decir, $V(T)=g(\lambda)=1/\lambda^2$. Vimos que $\hat{\lambda}=\dfrac{n}{\sum_{i=1}^{n}T_i}=\dfrac{1}{\overline{T}}$. Por lo tanto el EMV de la varianza es $\hat{\sigma}^2=\dfrac{1}{\hat{\lambda}^2}$.

2- Sea $X_1, X_2,\ldots\ldots,X_n$ una muestra aleatoria de una v.a. $B(1,p)$. Un EMV de *p* es $\hat{p}=\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$

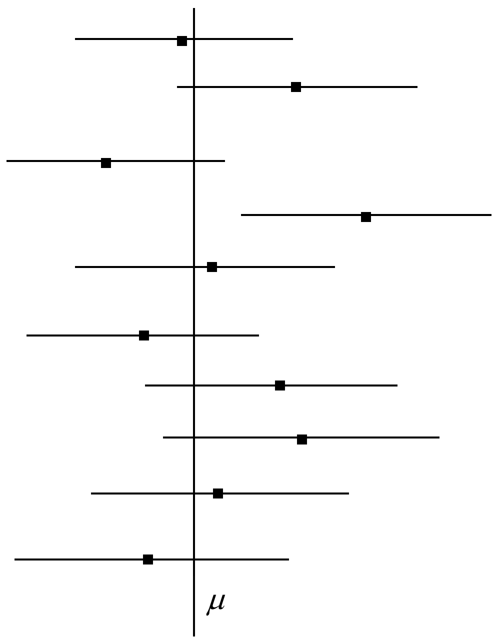
Se selecciona una muestra aleatoria de $n$ cascos para ciclistas fabricados por cierta compañía.
Sea $X$ : " el número entre los $n$ que tienen defectos" , y $p$ = P(el casco tiene defecto).
Supongamos que solo se observa $X$ ( el número de cascos con defectos). Si $n = 20$ y $x = 3$, es la estimación de $p$ es $\hat{p} = \dfrac{3}{20}$

El E.M.V. de la probabilidad $(1-p)^5$, de que ninguno de los siguientes cinco cascos que se examinen tenga defectos será $(1-\hat{p})^5$ y su estimación en este caso $\left(1-\dfrac{3}{20}\right)^5$

# *9- Intervalos de confianza*

## 9.1 – Introducción

Se ha visto como construir a partir de una muestra aleatoria un estimador puntual de un parámetro desconocido. En esos casos necesitábamos dar algunas características del estimador, como por ejemplo si era insesgado o su varianza.
A veces resulta más conveniente dar un ***intervalo de valores posibles*** del parámetro desconocido, de manera tal que dicho intervalo contenga al verdadero parámetro con determinada probabilidad.
Específicamente, a partir de una muestra aleatoria se construye un intervalo $\left(\hat{\Theta}_1, \hat{\Theta}_2\right)$ donde los extremos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estadísticos, tal que $P\left(\theta \in \left(\hat{\Theta}_1, \hat{\Theta}_2\right)\right) = 1-\alpha$ donde $\theta$ es el parámetro desconocido a estimar y $\alpha$ es un valor real entre cero y uno dado de antemano. Por ejemplo si $\alpha = 0.05$, se quiere construir un intervalo $\left(\hat{\Theta}_1, \hat{\Theta}_2\right)$ tal que $P\left(\theta \in \left(\hat{\Theta}_1, \hat{\Theta}_2\right)\right) = 0.95$, o escrito de otra forma $P\left(\hat{\Theta}_1 \le \theta \le \hat{\Theta}_2\right) = 0.95$
Esta probabilidad tiene el siguiente significado: como $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son estadísticos, los valores que ellos toman varían con los valores de la muestra, es decir si $x_1, x_2, ..., x_n$ son los valores medidos de la muestra entonces el estadístico $\hat{\Theta}_1$ tomará el valor $\theta_1$ y el estadístico $\hat{\Theta}_2$ tomará el valor $\theta_2$. Si medimos nuevamente la muestra obtendremos ahora valores $x_1', x_2', ..., x_n'$ y por lo tanto $\hat{\Theta}_1$ tomará el valor $\theta_1'$ y el estadístico $\hat{\Theta}_2$ tomará el valor $\theta_2'$, diferentes en general de los anteriores. Esto significa que si medimos la muestra 100 veces obtendremos 100 valores diferentes para $\hat{\Theta}_1$ y $\hat{\Theta}_2$ y por lo tanto obtendremos 100 intervalos distintos, de los cuales aproximadamente 5 de ellos no contendrán al verdadero parámetro.
Al valor $1-\alpha$ se lo llama ***nivel de confianza*** del intervalo. También se suele definir como nivel de confianza al $(1-\alpha)100\%$
La construcción repetida de un intervalo de confianza para $\mu$ se ilustra en la siguiente figura



**9.2 – <u>Intervalo de confianza para la media de una distribución normal, varianza conocida</u>.**

El método general para construir intervalos de confianza es el siguiente llamado ***método del pivote:***

Supongamos el siguiente caso particular, sea $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de tamaño *n* de una v.a. *X* donde $X \sim N(\mu, \sigma^2)$, $\sigma^2$ conocido, se quiere construir un intervalo de confianza para $\mu$ de nivel $1-\alpha$. Supongamos $\alpha = 0.05$.

1- tomamos un estimador puntual de $\mu$, sabemos que $\hat{\mu} = \overline{X}$ es un estimador con buenas propiedades.

2- a partir de $\hat{\mu} = \overline{X}$ construimos el estadístico $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$. Notar que *Z* ***(pivote)*** contiene al verdadero parámetro $\mu$ y que bajo las condiciones dadas $Z \sim N(0,1)$

3- como conocemos la distribución de *Z*, podemos plantear: hallar un número $z$ tal que $P(-z \le Z \le z) = 0.95$

Por la simetría de la distribución normal estándar podemos escribir

$P(-z \le Z \le z) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = 0.95 \quad \Rightarrow \quad \Phi(z) = 0.975 \quad \Rightarrow \quad z = 1.96$

Por lo tanto $P(-1.96 \le Z \le 1.96) = P\left(-1.96 \le \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 1.96\right) = 0.95$

Despejamos $\mu$:

$$P\left(-1.96 \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 1.96\right) = P\left(-1.96\frac{\sigma}{\sqrt{n}} \le \overline{X} - \mu \le 1.96\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-1.96\frac{\sigma}{\sqrt{n}} - \overline{X} \le -\mu \le 1.96\frac{\sigma}{\sqrt{n}} - \overline{X}\right) = P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Entonces

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = P\left(\mu \in \left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} ; \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)\right) = 0.95$$

Es decir el intervalo de confianza para $\mu$ es $\left(\overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}} ; \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}\right)$ y tiene nivel de confianza 0.95 o 95%.

Aquí $\hat{\Theta}_1 = \overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}}$ y $\hat{\Theta}_2 = \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}$

Repetimos el procedimiento anterior y construimos un intervalo de confianza para $\mu$ con nivel de confianza $1-\alpha$

1-Partimos de la esperanza muestral $\overline{X} = \frac{1}{n}\sum_{1=1}^{n} X_i$ para una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ de tamaño *n*. Sabemos que es un estimador insesgado y consistente de $\mu$.

2-Construimos el estadístico

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

La variable aleatoria *Z* cumple las condiciones necesarias de un pivote

Para construir un intervalo de confianza al nivel de confianza 1-$\alpha$ partiendo del pivote *Z*, comenzamos por plantear la ecuación

$$P(-z \leq Z \leq z) = 1\text{-}\alpha\,,$$

donde la incógnita es el número real *z*.

Si reemplazamos la v.a. *Z* por su expresión tenemos:

$$P\left(-z \leq \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \leq z\right) = P\left(-z\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq z\frac{\sigma}{\sqrt{n}}\right) = P\left(-\overline{X} - z\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\overline{X} + z\frac{\sigma}{\sqrt{n}}\right) = 1\text{-}\alpha$$

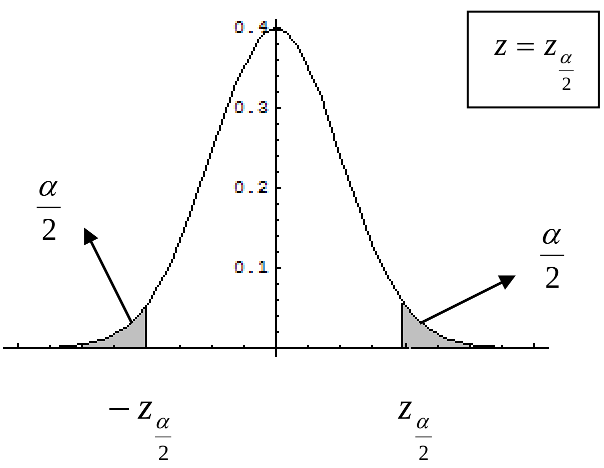
Multiplicando todos los miembros de la desigualdad por -1 (el orden de los miembros se invierte) llegamos a:

$$P\left(\overline{X} - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z\frac{\sigma}{\sqrt{n}}\right) = 1\text{-}\alpha$$

Evidentemente, si definimos

$\begin{cases} \hat{\Theta}_1 = \overline{X} - z\dfrac{\sigma}{\sqrt{n}} \\ \hat{\Theta}_2 = \overline{X} + z\dfrac{\sigma}{\sqrt{n}} \end{cases}$, hemos construido dos estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ tales que $P\left(\hat{\Theta}_1 \leq \mu \leq \hat{\Theta}_2\right) = 1\text{-}\alpha$,

es decir hemos construido el intervalo de confianza bilateral deseado $\left[\hat{\Theta}_1, \hat{\Theta}_2\right]$. Todos los elementos que forman los estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son conocidos ya que el número *z* verifica la ecuación anterior, es decir (ver figura):



$P(-z \le Z \le z) = \Phi(z) - \Phi(-z)$=1- $\alpha$   donde $\Phi(z)$ es la *Fda* para la v.a. $Z \sim N(0,1)$

Recordando que $\Phi(-z) = 1 - \Phi(z)$, esta ecuación queda:
$\Phi(z) - \Phi(-z) = 2\Phi(z) - 1$=1- $\alpha$,   o bien (ver figura anterior),

$\Phi(z) = 1 - \dfrac{\alpha}{2}$   o   de otra forma $P(Z > z) = \dfrac{\alpha}{2}$.

Al valor de $z$ que verifica esta ecuación se lo suele indicar $z_{\frac{\alpha}{2}}$. En consecuencia, el intervalo de confianza bilateral al nivel de significación 1- $\alpha$ queda:

$$\left[\hat{\Theta}_1, \hat{\Theta}_2\right] = \left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

En consecuencia:

Si $(X_1, X_2, ..., X_n)$ una muestra aleatoria de tamaño $n$ de una v.a. $X$ donde $X \sim N(\mu, \sigma^2)$, $\sigma^2$ conocido, un intervalo de confianza para $\mu$ de nivel $1-\alpha$ es

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \qquad (8.1)$$

Ejemplo:
Un ingeniero civil analiza la resistencia a la compresión del concreto. La resistencia está distribuida aproximadamente de manera normal, con varianza 1000 (psi)$^2$. Al tomar una muestra aleatoria de 12 especímenes, se tiene que $\bar{x} = 3250$ psi.

a) Construya un intervalo de confianza del 95% para la resistencia a la compresión promedio.
b) Construya un intervalo de confianza del 99% para la resistencia a la compresión promedio. Compare el ancho de este intervalo de confianza con el ancho encontrado en el inciso a).

Solución:
La v. a. de interés es $X_i$: "resistencia a la compresión del concreto en un espécimen $i$"
Tenemos una muestra de $n = 12$ especímenes.
Asumimos que $X_i \sim N(\mu, \sigma^2)$ para $i = 1,2,3,\ldots,12$   con $\sigma^2 = 1000$

a) Queremos un intervalo de confianza para $\mu$ de nivel 95%. Por lo tanto $\alpha = 0.05$

El intervalo a utilizar es $\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$.

Buscamos en la tabla de la normal estándar el valor de $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

Reemplazando:

$$\left[3250 - 1.96 \times \frac{\sqrt{1000}}{\sqrt{12}}, 3250 + 1.96 \times \frac{\sqrt{1000}}{\sqrt{12}}\right] = \left[3232.10773, \;\; 3267.89227\right]$$

b) repetimos lo anterior pero ahora $\alpha = 0.01$

El intervalo a utilizar es $\left[ \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$.

Buscamos en la tabla de la normal estándar el valor de $z_{\frac{\alpha}{2}} = z_{0.005} = 2.58$

Reemplazando:

$$\left[ 3250 - 2.58 \times \frac{\sqrt{1000}}{\sqrt{12}}, 3250 + 2.58 \times \frac{\sqrt{1000}}{\sqrt{12}} \right] = \left[ 3226.44793, \quad 3273.55207 \right]$$

La longitud del intervalo encontrado en a) es: 35.78454
La longitud del intervalo encontrado en b) es: 47.10414
Notar que la seguridad de que el verdadero parámetro se encuentre en el intervalo hallado es mayor en el intervalo b) que en el a), pero la longitud del intervalo b) es mayor que la del intervalo a).
Al aumentar el nivel de confianza se perdió ***precisión en la estimación***, ya que a menor longitud hay mayor precisión en la estimación.

En general la longitud del intervalo es $L = 2z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Notar que:

a) si $n$ y $\sigma$ están fijos, a medida que $\alpha$ disminuye tenemos que $z_{\frac{\alpha}{2}}$ aumenta, por lo tanto $L$ aumenta.
b) si $\alpha$ y $\sigma$ están fijos, entonces a medida que $n$ aumenta tenemos que $L$ disminuye.

Podemos plantearnos la siguiente pregunta relacionada con el ejemplo anterior: ¿qué tamaño $n$ de muestra se necesita para que el intervalo tenga nivel de confianza 95% y longitud la mitad de la longitud del intervalo hallado en a)?
Solución: el intervalo hallado en a) tiene longitud 35.78454, y queremos que el nuevo intervalo tenga longitud 17.89227 aproximadamente. Planteamos:

$$L = 2\, z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 17.89227 \qquad \Rightarrow \qquad 2 \times 1.96 \times \frac{\sqrt{1000}}{\sqrt{n}} \leq 17.89227$$

Despejando $n$ :

$$\left( 2 \times 1.96 \times \frac{\sqrt{1000}}{17.89227} \right)^2 \leq n \qquad \Rightarrow \qquad n \geq 48$$

O sea, hay que tomar por lo menos 84 especímenes para que el intervalo tenga la longitud pedida.

En general, si queremos hallar $n$ tal que $L = 2z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq l$ , donde $l$ es un valor dado, entonces despejando $n$

$$n \geq \left( \frac{2z_{\frac{\alpha}{2}}\sigma}{l} \right)^2$$

Si estimamos puntualmente al parámetro $\mu$ con $\overline{X}$ estamos cometiendo un error en la estimación menor o igual a $\dfrac{L}{2} = z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$, que se conoce como ***precisión del estimador***

<u>Ejemplo</u>: Se estima que el tiempo de reacción a un estímulo de cierto dispositivo electrónico está distribuido normalmente con desviación estándar de 0.05 segundos. ¿Cuál es el número de mediciones temporales que deberá hacerse para que la confianza de que el error de la estimación de la esperanza no exceda de 0.01 sea del 95%?

Nos piden calcular $n$ tal que $\dfrac{L}{2} = z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}} < 0.01$ con $\alpha = 0.05$.

Por lo tanto $n \geq \left( z_{0.025} \dfrac{0.05}{0.01} \right)^2$.

Además $z_{0,025}$=1.96. Entonces $n \geq \left( z_{0.975} \dfrac{0.05}{0.01} \right)^2 = (1.96 \times 5)^2 = 96.04$.

O sea hay que tomar por lo menos 97 mediciones temporales.

> Para muestras tomadas de una población normal, o para muestras de tamaño $n \geq 30$, de una población cualquiera, el intervalo de confianza dado anteriormente en (8.1), proporciona buenos resultados.
> En el caso de que la población de la que se extrae la muestra no sea normal pero $n \geq 30$, el nivel de confianza del intervalo (8.1) es ***aproximadamente*** $1-\alpha$ .
> Pero para muestras pequeñas tomadas de poblaciones que no son normales no se puede garantizar que el nivel de confianza sea $1-\alpha$ si se utiliza (8.1).

<u>Ejemplo</u>:
Supongamos que $X$ representa la duración de una pieza de equipo y que se probaron 100 de esas piezas dando una duración promedio de 501.2 horas. Se sabe que la desviación estándar poblacional es $\sigma$=4 horas. Se desea tener un intervalo del 95% de confianza para la esperanza poblacional $E(X) = \mu$.
<u>Solución</u>:
En este caso, si bien no conocemos cuál es la distribución de $X$ tenemos que el tamaño de la muestra es $n = 100 > 30$ (muestra grande) por lo tanto el intervalo buscado es

$$\left[ \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Puesto que 1-$\alpha$=0.95 $\rightarrow \alpha = 1 - 0.95 = 0.05 \rightarrow \dfrac{\alpha}{2} = 0.025$

De la tabla de la normal estandarizada obtenemos $z_{0,025}$=1.96. Entonces reemplazando:

$$\left[ \overline{X} - 1.96 \frac{4}{\sqrt{100}}, \overline{X} + 1.96 \frac{4}{\sqrt{100}} \right]$$

Para el valor particular $\bar{x}$ =501.2 tenemos el intervalo

$$\left[\overline{x}-1.96\frac{4}{\sqrt{100}}, \overline{x}+1.96\frac{4}{\sqrt{n}}\right]=\left[501.2-1.96\frac{4}{10}, 501.2+1.96\frac{4}{10}\right]=\left[500.4,\ \ 502.0\right].$$

Al establecer que $\left[500.4,\ \ 502.0\right]$ es un intervalo al 95% de confianza de $\mu$ estamos diciendo que la probabilidad de que el intervalo $\left[500.4,\ \ 502.0\right]$ contenga a $\mu$ es 0.95. O, en otras palabras, la probabilidad de que la muestra aleatoria $(X_1, X_2, \ldots, X_n)$ tome valores tales que el intervalo aleatorio $\left[\overline{X}-1.96\frac{4}{\sqrt{100}}, \overline{X}+1.96\frac{4}{\sqrt{100}}\right]$ defina un intervalo numérico que contenga al parámetro fijo desconocido $\mu$ es 0.95.

## 9.3 - Intervalo de confianza para la media de una distribución normal, varianza desconocida

Nuevamente como se trata de encontrar un intervalo de confianza para $\mu$ nos basamos en la esperanza muestral $\overline{X}=\frac{1}{n}\sum_{1=1}^{n}X_i$ que sabemos es un buen estimador de $\mu$. Pero ahora no podemos usar como pivote a

$$Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

porque desconocemos $\sigma$ y una condición para ser pivote es que, excepto por el parámetro a estimar ( en este caso $\mu$ ), todos los parámetros que aparecen en él deben ser conocidos. Entonces proponemos como pivote una variable aleatoria definida en forma parecida a *Z* pero reemplazando $\sigma$ por un estimador adecuado.
Ya vimos que la varianza muestral definida

$$S^2=\frac{1}{n-1}\sum_{1=1}^{n}\left(X_i-\overline{X}\right)^2,$$

donde $\overline{X}$ es la esperanza muestral, es un estimador insesgado de la varianza poblacional $V(X)$, es decir, $E(S^2)=V(X)=\sigma^2\ \ \forall n$. Entonces estimamos $\sigma$ con $S$ y proponemos como pivote a la variable aleatoria

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}.$$

Pero para poder usar a *T* como pivote debemos conocer su distribución.
Se puede probar que la distribución de *T* es una distribución llamada ***Student con parámetro n-1***.

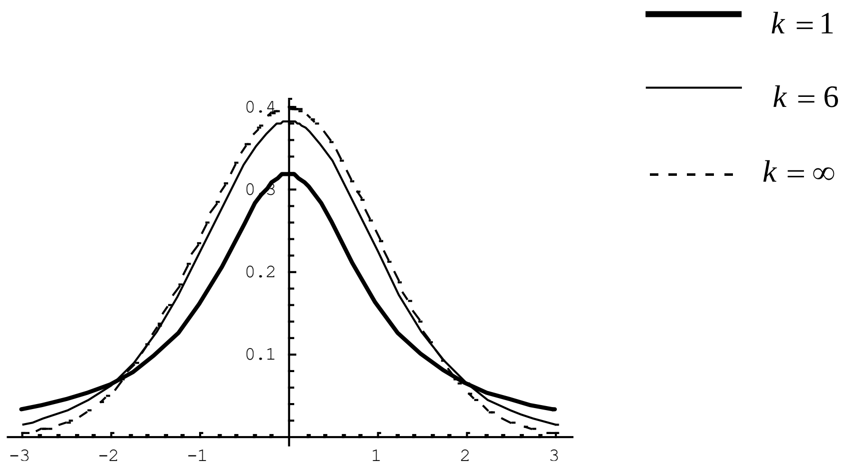
<u>Nota</u>: Una v.a. continua tiene distribución ***Student con k grados de libertad***, si su *f.d.p.* es de la forma

$$f(x)=\frac{\Gamma\left[\frac{(k+1)}{2}\right]}{\sqrt{\pi k}\;\Gamma\left(\frac{k}{2}\right)}\frac{1}{\left[\left(\frac{x^2}{k}\right)+1\right]^{\frac{k+1}{2}}}\qquad -\infty<x<\infty$$

<u>Notación</u>: $T\sim t_k$

La gráfica de la *f.d.p.* de la distribución Student tiene forma de campana como la normal, pero tiende a cero más lentamente. Se puede probar que cuando $k\to\infty$ la *fdp* de la Student tiende a la *fdp* de la $N(0,\ 1)$.

En la figura siguiente se grafica *f(x)* para diferentes valores de *k*




Anotaremos $t_{\alpha,k}$ al cuantil de la Student con *k* grados de libertad que deja bajo la *fdp* a derecha un área de $\alpha$, y a su izquierda un área de $1-\alpha$.

Luego, para construir el intervalo de confianza buscado a partir del pivote *T* procedemos como en los casos anteriores:

Comenzamos por plantear la ecuación

$$P(-t\le T\le t)=1\text{-}\alpha,$$

donde la incógnita es el número real *t*.

Si reemplazamos la v.a. *T* por su expresión, tenemos sucesivamente (multiplicando por $S/\sqrt{n}$ y restando $\overline{X}$):

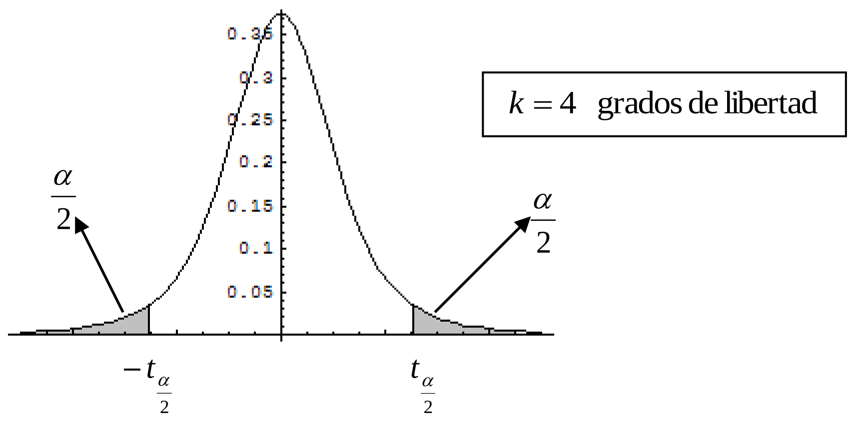
$$P\left(-t\le\frac{\overline{X}-\mu}{S/\sqrt{n}}\le t\right)=P\left(-t\frac{S}{\sqrt{n}}\le\overline{X}-\mu\le t\frac{S}{\sqrt{n}}\right)=P\left(-\overline{X}-t\frac{S}{\sqrt{n}}\le-\mu\le-\overline{X}+t\frac{S}{\sqrt{n}}\right)=1\text{-}\alpha$$

Multiplicando todos los miembros de la desigualdad por -1 (el orden de los miembros se invierte) llegamos a:

$$P\left(\overline{X}-t\frac{S}{\sqrt{n}}\leq\mu\leq\overline{X}+t\frac{S}{\sqrt{n}}\right)=1\text{-}\alpha$$

Evidentemente, si definimos

$\begin{cases}\hat{\Theta}_1=\overline{X}-t\dfrac{S}{\sqrt{n}}\\ \hat{\Theta}_2=\overline{X}+t\dfrac{S}{\sqrt{n}}\end{cases}$, hemos construido dos estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ tales que $P\left(\hat{\Theta}_1\leq\mu\leq\hat{\Theta}_2\right)=1\text{-}\alpha$, veamos quien es el número *t* que verifica la ecuación, es decir (ver figura):




$P(-t\leq T\leq t)=F(t)-F(-t)=1\text{-}\alpha$ donde $F(t)$ es la *Fda* para la v.a. $T\sim t_{n-1}$.

Por la simetría de la distribución t de Student se deduce fácilmente de la figura anterior que $F(-t)=1-F(t)$, entonces:

$F(t)-F(-t)=2F(t)-1=1\text{-}\alpha$, o bien (ver figura anterior),

$F(t)=1-\dfrac{\alpha}{2}$.

Al valor de *t* que verifica esta ecuación se lo suele indicar $t_{\frac{\alpha}{2},n-1}$. En consecuencia, el intervalo de confianza bilateral al nivel de significación $1\text{-}\alpha$ queda:

$$\left[\overline{X}-t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}},\overline{X}+t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right]\quad\text{con}\quad F\left(t_{\frac{\alpha}{2},n-1}\right)=1-\frac{\alpha}{2}.$$

En consecuencia:

Si $(X_1,X_2,...,X_n)$ una muestra aleatoria de tamaño *n* de una v.a. *X* donde $X\sim N(\mu,\sigma^2)$, $\sigma^2$ desconocido, un intervalo de confianza para $\mu$ de nivel $1-\alpha$ es

$$\left[\overline{X}-t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\overline{X}+t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]\qquad(8.2)$$

Ejemplo:
Se hicieron 10 mediciones sobre la resistencia de cierto tipo de alambre que dieron valores $x_1, x_2, \dots, x_{10}$ tales que $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 10.48$ ohms y $S = \sqrt{\frac{1}{9}\sum_{i=!}^{10}(x_i - \bar{x})^2}$ = 1.36 ohms. Supóngase que $X \sim N(\mu, \sigma^2)$.
Se desea obtener un intervalo de confianza para la esperanza poblacional $\mu$ al 90 %.

Tenemos que $1-\alpha = 0.90 \rightarrow \alpha = 0.1 \rightarrow \alpha/2 = 0.05$
De la Tabla de la $t$ de Student tenemos que $t_{0.05,9} = 1.8331$. Entonces el intervalo de confianza buscado es:

$$\left[\overline{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right] = \left[10.48 - 1.8331\frac{1.36}{\sqrt{10}}, 10.48 + 1.8331\frac{1.36}{\sqrt{10}}\right]$$

Esto es: $[9.69,\ 11.27]$.

Si $\sigma^2$ es desconocido y el tamaño de la muestra grande, entonces se puede probar que al reemplazar $\sigma$ por $S$, el estadístico

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1) \quad \textbf{\textit{aproximadamente}}$$

y puedo construir el intervalo para $\mu$ como antes:

$$\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$ , pero su nivel es aproximadamente $1-\alpha$

## 9.4 – <u>Intervalo de confianza para la diferencia de dos medias, varianzas conocidas</u>

Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son conocidas.

Sean además
$(X_{11}, X_{12}, \dots, X_{1n_1})$ una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, \dots, X_{2n_2})$ una muestra aleatoria de tamaño $n_2$ de $X_2$.

Deseamos construir un intervalo al nivel de confianza $1-\alpha$ para la diferencia de esperanzas $\mu_1 - \mu_2$.
Ya sabemos cuál es la distribución del promedio de variables aleatorias normales independientes:

$$\begin{cases} \overline{X}_1 = \dfrac{1}{n_1}\displaystyle\sum_{i=1}^{n_1} X_{1i} \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \overline{X}_2 = \dfrac{1}{n_2}\displaystyle\sum_{i=1}^{n_2} X_{2i} \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n_2}\right) \end{cases}$$

Consideremos ahora la diferencia $\overline{Y} = \overline{X}_1 - \overline{X}_2$. Si $\overline{X}_1$ y $\overline{X}_2$ tienen distribución normal y son independientes, su diferencia también es normal, con esperanza igual a la diferencia de las esperanzas y la varianza es la suma de las varianzas:

$$\overline{X}_1 - \overline{X}_2 \sim \mathrm{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Por lo tanto

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1),$$ es decir, tiene distribución normal estandarizada.

La v.a. $Z$ cumple con toda las condiciones para servir de pivote y construiremos nuestro intervalo en forma análoga a cómo hicimos en los casos anteriores:
Comenzamos por plantear la ecuación

$$P(-z \le Z \le z) = 1\text{-}\alpha,$$

donde la incógnita es el número real $z$.

Reemplazamos la v.a. $Z$ por su expresión y tenemos sucesivamente (multiplicando por $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ y restando $\overline{X}_1 - \overline{X}_2$):

$$P\left(-z \le \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \le z\right) = P\left(-z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2) \le z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) =$$

$$= P\left(-\left(\overline{X}_1 - \overline{X}_2\right) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le -(\mu_1 - \mu_2) \le -\left(\overline{X}_1 - \overline{X}_2\right) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Multiplicando todos los miembros de la desigualdad por -1 (el orden de los miembros se invierte) llegamos a:

$$P\left(\overline{X}_1 - \overline{X}_2 - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le (\mu_1 - \mu_2) \le \overline{X}_1 - \overline{X}_2 + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Evidentemente, si definimos

$$\begin{cases} \hat{\Theta}_1 = \overline{X}_1 - \overline{X}_2 - z\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \\ \hat{\Theta}_2 = \overline{X}_1 - \overline{X}_2 + z\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \end{cases},$$

habremos construido dos estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ tales que $P\left(\hat{\Theta}_1 \le (\mu_1 - \mu_2) \le \hat{\Theta}_2\right) = 1\text{-}\alpha$, es decir habremos construido el intervalo de confianza bilateral deseado $\left[\hat{A}_1, \hat{A}_2\right]$. Todos los elementos que forman los estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son conocidos ya que el número $z$ verifica la ecuación anterior, es decir:

$P(-z \le Z \le z) = \Phi(z) - \Phi(-z) = 1\text{-}\alpha$ donde $\Phi(z)$ es la *Fda* para la v.a. $Z \sim N(0,1)$

o bien, según vimos:

$\Phi(z) = 1 - \dfrac{\alpha}{2}$ que anotamos $z_{\frac{\alpha}{2}}$

En consecuencia, el intervalo de confianza bilateral al nivel de significación 1-$\alpha$ queda:

$$\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Por lo tanto

> Si $X_1$ y $X_2$ son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^{\ 2})$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^{\ 2})$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son conocidas. Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel $1-\alpha$ es
>
> $$\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \overline{X}_1 - \vec{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \qquad (8.3)$$

Ejemplo:

Se utilizan dos máquinas para llenar botellas de plástico con detergente para máquinas lavaplatos. Se sabe que las desviaciones estándar de volumen de llenado son $\sigma_1 = 0.10$ onzas de líquido y $\sigma_2 = 0.15$ onzas de líquido para las dos máquinas respectivamente. Se toman dos muestras aleatorias, $n_1 = 12$ botellas de la máquina 1 y $n_2 = 10$ botellas de la máquina 2. Los volúmenes promedio de llenado son $\bar{x}_1 = 30.87$ onzas de líquido y $\bar{x}_2 = 30.68$ onzas de líquido.

Asumiendo que ambas muestras provienen de distribuciones normales

Construya un intervalo de confianza de nivel 90% para la diferencia entre las medias del volumen de llenado.

Solución:

Como $1-\alpha = 0.90$ entonces $\alpha = 0.10$

Por lo tanto $z_{\frac{\alpha}{2}} = z_{0.05} = 1.65$

El intervalo será $\left[ (30.87 - 30.68) - 1.65\sqrt{\frac{0.10^2}{12} + \frac{0.15^2}{10}} ; (30.87 - 30.68) + 1.65\sqrt{\frac{0.10^2}{12} + \frac{0.15^2}{10}} \right]$

O sea $\left[ 0.09837; \quad 0.281620 \right]$

Si se conocen las desviaciones estándar y los tamaños de las muestras son iguales (es decir $n_1 = n_2 = n$), entonces puede determinarse el tamaño requerido de la muestra de manera tal que la longitud del intervalo sea menor que $l$

$$L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} \le l \qquad \Rightarrow \qquad n \ge \left( \frac{2z_{\frac{\alpha}{2}}}{l} \right)^2 \left( \sigma_1^2 + \sigma_2^2 \right)$$

Si $\sigma_1$ y $\sigma_2$ son ***desconocidos***, $n_1 \ge 30$ y $n_2 \ge 30$, entonces se puede probar que al reemplazar $\sigma_1$ por *S1* y $\sigma_2$ por *S2*, el estadístico

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0,1) \quad \textit{aproximadamente}$$

y puedo construir el intervalo para $\mu_1 - \mu_2$ como antes:

$$\left[ \overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \quad \overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right], \qquad (8.4)$$

pero su nivel es aproximadamente $1 - \alpha$

Para muestras tomadas de dos poblaciones normales, o para muestras de tamaño $n_1 \ge 30$ y $n_2 \ge 30$, de dos poblaciones cualesquiera, el intervalo de confianza dado anteriormente en (8.3), proporciona buenos resultados.
En el caso de que la población de la que se extrae la muestra no sea normal pero $n_1 \ge 30$ y $n_2 \ge 30$, el nivel de confianza del intervalo (8.3) es ***aproximadamente*** $1 - \alpha$ .

Ejemplo:
De una muestra de 150 lámparas del fabricante A se obtuvo una vida media de 1400 hs y una desviación típica de 120 hs. Mientras que de una muestra de 100 lámparas del fabricante B se obtuvo una vida media de 1200 hs. y una desviación típica de 80 hs.
Halla los límites de confianza del 95% para la diferencia las vidas medias de las poblaciones A y B.

Solución:

Sean las variables aleatorias:
$X_1$ : "duración en horas de una lámpara del fabricante A"
$X_2$ : "duración en horas de una lámpara del fabricante B"
No se dice cuál es la distribución de estas variables, pero como $n_1 = 150$ y $n_2 = 100$ podemos usar el intervalo dado en (8.4)

Tenemos que $\bar{x}_1 = 1400$, $\bar{x}_2 = 1200$, $s_1 = 120$ y $s_2 = 80$.
Además $1 - \alpha = 0.95 \rightarrow z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$
Entonces el intervalo es

$$\left[1400 - 1200 - 1.96\sqrt{\frac{120^2}{150} + \frac{80^2}{100}}\ ;\ 1400 - 1200 + 1.96\sqrt{\frac{120^2}{150} + \frac{80^2}{100}}\right] = \left[175.2077;\ \ 224.7922\right]$$

Observación: como este intervalo no contiene al cero, podemos inferir que hay diferencia entre las medias con probabilidad 0.95, es más, podemos inferir que la media del tiempo de duración de las lámparas del fabricante A es mayor que la media del tiempo de duración de las lámparas del fabricante B con probabilidad 0.95 .

## 9.5 – Intervalo de confianza para la diferencia de dos medias, varianzas desconocidas

Nuevamente supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son ***desconocidas*** .

Sean además
$(X_{11}, X_{12}, \ldots, X_{1n_1})$ una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, \ldots, X_{2n_2})$ una muestra aleatoria de tamaño $n_2$ de $X_2$.
Pero ahora $n_1$ ***o*** $n_2$ ***no son mayores que 30***
Supongamos que es razonable suponer que las varianzas desconocidas son iguales, es decir $\sigma_1 = \sigma_2 = \sigma$
Deseamos construir un intervalo al nivel de confianza $1 - \alpha$ para la diferencia de esperanzas $\mu_1 - \mu_2$

Sean $\bar{X}_1$ y $\bar{X}_2$ las medias muestrales y $S_1^2$ y $S_2^2$ las varianzas muestrales. Como $S_1^2$ y $S_2^2$ son los estimadores de la varianza común $\sigma^2$, entonces construimos un ***estimador combinado*** de $\sigma^2$. Este estimador es

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Se puede comprobar que es un estimador insesgado de $\sigma^2$.
Se puede probar que el estadístico

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ tiene distribución Student con $n_1 + n_2 - 2$ grados de libertad

Por lo tanto se plantea la ecuación

$$P\left(-t_{\frac{\alpha}{2},n_1+n_2-2} \le T \le t_{\frac{\alpha}{2},n_1+n_2-2}\right) = 1-\alpha$$

o

$$P\left(-t_{\frac{\alpha}{2},n_1+n_2-2} \le \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\frac{\alpha}{2},n_1+n_2-2}\right) = 1-\alpha$$

Despejamos $\mu_1 - \mu_2$ y queda la expresión

$$P\left(\bar{X}_1 - \bar{X}_2 - t_{\frac{\alpha}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{X}_1 - \bar{X}_2 + t_{\frac{\alpha}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1-\alpha$$

Entonces

> Si $X_1$ y $X_2$ son dos variables aleatorias ***independientes*** normalmente distribuidas:
> $X_1 \sim N(\mu_1, \sigma_1^2)$ , $X_2 \sim N(\mu_2, \sigma_2^2)$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son ***desconocidas e iguales***, es decir $\sigma_1 = \sigma_2 = \sigma$
> Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel $1-\alpha$ es
> $$\left[\bar{X}_1 - \bar{X}_2 - t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\ ;\ \bar{X}_1 - \bar{X}_2 + t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] \qquad (8.5)$$

<u>Ejemplo</u>:
Se piensa que la concentración del ingrediente activo de un detergente líquido para ropa, es afectada por el tipo de catalizador utilizado en el proceso de fabricación. Se realizan 10 observaciones con cada catalizador, y se obtienen los datos siguientes:
Catalizador 1: 57.9, 66.2, 65.4, 65.4, 65.2, 62.6, 67.6, 63.7, 67.2, 71.0
Catalizador 2: 66.4, 71.7, 70.3, 69.3, 64.8, 69.6, 68.6, 69.4, 65.3, 68.8
a) Encuentre un intervalo de confianza del 95% para la diferencia entre las medias de las concentraciones activas para los dos catalizadores. Asumir que ambas muestras fueron extraídas de poblaciones normales con varianzas iguales.
b) ¿Existe alguna evidencia que indique que las concentraciones activas medias dependen del catalizador utilizado?
<u>Solución</u>:
Sean las variables aleatorias

$X_1$ : " concentración del ingrediente activo con catalizador 1"
$X_2$ : " concentración del ingrediente activo con catalizador 2"
Asumimos que ambas variables tienen distribución normal con varianzas iguales
Estamos e3n las condiciones para usar (8.5)
Tenemos que $\bar{x}_1 = 65.22$ , $\bar{x}_2 = 68.42$ , $s_1 = 3.444$, $s_2 = 2.224$ , $n_1 = n_2 = 10$
Calculamos $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \dfrac{9\times 3.444^2 + 9\times 2.224^2}{10+10-2} = 8.4036$
Por lo tanto $S_p = \sqrt{8.4036} = 2.89890$
Buscamos en la tabla de la Student $t_{\frac{\alpha}{2},n_1+n_2-2} = t_{0.025,18} = 2.101$

Entonces el intervalo es

$$\left[65.22 - 68.42 - 2.101\times 2.89890\sqrt{\frac{1}{10}+\frac{1}{10}};65.22 - 68.42 + 2.101\times 2.89890\sqrt{\frac{1}{10}+\frac{1}{10}}\right] =$$
$$= [-5.9237;-0.476301]$$

b) Existe alguna evidencia que indique que las concentraciones activas medias dependen del catalizador utilizado, pues el 0 no pertenece al intervalo.

En muchas ocasiones ***no es razonable suponer que las varianzas son iguales***. Si no podemos garantizar que las varianzas son iguales, para construir un intervalo de confianza de nivel $1-\alpha$ para $\mu_1 - \mu_2$ utilizamos es estadístico

$$T^* = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Se puede probar que $T^*$ tiene ***aproximadamente*** una distribución Student con $\nu$ grados de libertad donde

$$\nu = \frac{\left( S_1^2/n_1 + S_2^2/n_2 \right)^2}{\frac{\left( S_1^2/n_1 \right)^2}{n_1-1} + \frac{\left( S_2^2/n_2 \right)^2}{n_2-1}}$$

si $\nu$ no es entero, ***se toma el entero más próximo a*** $\nu$

Por lo tanto planteamos la ecuación

$$P\left(-t_{\frac{\alpha}{2},\nu} \le T^* \le t_{\frac{\alpha}{2},\nu}\right) = 1-\alpha$$

Y despejando $\mu_1 - \mu_2$ el intervalo es

$$\left[\overline{X}_1-\overline{X}_2-t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\quad \overline{X}_1-\overline{X}_2+t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

Entonces

Si $X_1$ y $X_2$ son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim N(\mu_1,\sigma_1^{\,2})$, $X_2 \sim N(\mu_2,\sigma_2^{\,2})$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son ***desconocidas*** y ***distintas***

Un intervalo de confianza para la diferencia $\mu_1-\mu_2$ de nivel ***aproximadamente*** $1-\alpha$ es

$$\left[\overline{X}_1-\overline{X}_2-t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\quad \overline{X}_1-\overline{X}_2+t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right] \qquad (8.6)$$

Donde

$$\nu=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

<u>Ejemplo</u>:

Una muestra de 6 soldaduras de un tipo tenía promedio de prueba final de resistencia de 83.2 ksi y desviación estándar de 5.2. Y una muestra de 10 soldaduras de otro tipo tenía resistencia promedio de 71.3 ksi y desviación estándar de 3.1. supongamos que ambos conjuntos de soldaduras son muestras aleatorias de poblaciones normales. Se desea encontrar un intervalo de confianza de 95% para la diferencia entre las medias de las resistencias de los dos tipos de soldaduras.

<u>Solución</u>:

Ambos tamaños muestrales son pequeños y las muestras provienen de poblaciones normales. No podemos asumir igualdad de varianzas. Entonces aplicamos (8.6)

Tenemos que $\bar{x}_1=83.2$, $\bar{x}_2=71.3$, $s_1=5.2$, $s_2=3.1$, $n_1=6$; $n_2=10$

Como $1-\alpha=0.95$ entonces $\dfrac{\alpha}{2}=0.025$

Además $\nu=\dfrac{\left(S_1^2/n_1+S_2^2/n_2\right)^2}{\dfrac{\left(S_1^1/n_1\right)^2}{n_1-1}+\dfrac{\left(S_2^2/n_2\right)^2}{n_2-1}}=\dfrac{\left(\dfrac{5.2^2}{6}+\dfrac{3.1^2}{10}\right)^2}{\dfrac{\left(5.2/6\right)^2}{5}+\dfrac{\left(3.1/10\right)^2}{9}}=7.18\approx 7$

Entonces buscamos en la tabla de la Student $t_{0.025,7}=2.365$

Por lo tanto el intervalo es

$$\left[\overline{X}_1-\overline{X}_2-t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\quad \overline{X}_1-\overline{X}_2+t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]=$$

$$=\left[83.2-71.3-2.365\sqrt{\frac{5.2^2}{6}+\frac{3.1^2}{10}};\ 83.2-71.3+2.365\sqrt{\frac{5.2^2}{6}+\frac{3.1^2}{10}}\right]=\left[6.37,\ \ 17.43\right]$$

## 9.6 – <u>Intervalo de confianza para $\mu_1-\mu_2$ para datos pareados</u>

Hasta ahora se obtuvieron intervalos de confianza para la diferencia de medias donde se tomaban dos muestras aleatorias independientes de dos poblaciones de interés. En ese caso se tomaban $n_1$ observaciones de una población y $n_2$ observaciones de la otra población.
En muchas situaciones experimentales, existen solo $n$ unidades experimentales diferentes y los datos están ***recopilados por pares***, esto es cada unidad experimental está formada por ***dos observaciones***.
Por ejemplo, supongamos que se mide el tiempo en segundos que un individuo tarda en hacer una maniobra de estacionamiento con dos automóviles diferentes en cuanto al tamaño de la llanta y la relación de vueltas del volante. Notar que cada individuo es la unidad experimental y de esa unidad experimental se toman dos observaciones que ***no serán independientes***. Se desea obtener un intervalo de confianza para la diferencia entre el tiempo medio para estacionar los dos automóviles.
En general, supongamos que tenemos los siguientes datos $(X_{11},X_{21});(X_{12},X_{22});\ldots;(X_{1n_1},X_{2n})$.
Las variables aleatorias $X_1$ y $X_2$ tienen medias $\mu_1$ y $\mu_2$ respectivamente.
Sea $D_j=X_{1j}-X_{2j}$ con $j=1,2,\ldots,n$.
Entonces

$$E(D_j)=E(X_{1j}-X_{2j})=E(X_{1j})-E(X_{2j})=\mu_1-\mu_2$$

y

$$V(D_j)=V(X_{1j}-X_{2j})=V(X_{1j})+V(X_{2j})-2Cov(X_{1j},X_{2j})=\sigma_1^2+\sigma_2^2-2Cov(X_1,X_2)$$

Estimamos $E(D_j)=\mu_1-\mu_2$ con $\overline{D}=\frac{1}{n}\sum_{j=1}^{n}D_j=\frac{1}{n}\sum_{j=1}^{n}(X_{1j}-X_{2j})=\overline{X}_1-\overline{X}_2$

En lugar de tratar de estimar la covarianza, estimamos la $V(D_j)$ con $S_D^2=\frac{1}{n-1}\sum_{j=1}^{n}(D_j-\overline{D})^2$

Anotamos $\mu_D=\mu_1-\mu_2$ y $\sigma^2{}_D=V(D_j)$

Asumimos que $D_j\sim \mathrm{N}(\mu_D,\sigma_D{}^2)$ con $j=1,2,\ldots,n$

Las variables aleatorias en pares diferentes son independientes, no lo son dentro de un mismo par.
Para construir el intervalo de confianza notar que

$$T = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \sim t_{n-1}$$

entonces al plantear la ecuación $P(-t \le T \le t) = 1\text{-}\alpha$, deducimos que $t = t_{\frac{\alpha}{2},n-1}$

Por lo tanto el intervalo de confianza para $\mu_D = \mu_1 - \mu_2$ de nivel $1-\alpha$ se obtendrá al sustituir $T$ en la ecuación anterior y despejar $\mu_D = \mu_1 - \mu_2$

El intervalo resultante es

$$\left[\overline{D} - t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}};\ \overline{D} + t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}}\right]$$

Entonces

> Cuando las observaciones se dan de a pares $(X_{11}, X_{21}); (X_{12}, X_{22}); \dots; (X_{1n_1}, X_{2n})$, y las diferencias
> $D_j = X_{1j} - X_{2j}$ son tales que $D_j \sim \mathrm{N}(\mu_D, \sigma_D{}^2)$ para $j = 1,2,\dots,n$, un intervalo de confianza de nivel $1-\alpha$ para $\mu_D = \mu_1 - \mu_2$ es
>
> $$\left[\overline{D} - t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}};\ \overline{D} + t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}}\right] \qquad (8.7)$$

<u>Ejemplo</u>:

Consideramos el ejemplo planteado al comienzo. Deseamos un intervalo de nivel 0.90

Sean las variables aleatorias

$X_{1j}$: "tiempo en segundos que tarda el individuo *j* en estacionar automóvil 1" con $j = 1,2,\dots,n$

$X_{2j}$: "tiempo en segundos que tarda el individuo *j* en estacionar automóvil 2" con $j = 1,2,\dots,n$

Medimos estas variables de manera que tenemos las siguientes observaciones

| sujeto | ***Automóvil 1*** ***(observación $x_{1j}$)*** | ***Automóvil 2*** ***(observación $x_{2j}$)*** | ***diferencia*** $D_j$ |
|---|---|---|---|
| 1 | 37.0 | 17.8 | 19.2 |
| 2 | 25.8 | 20.2 | 5.6 |
| 3 | 16.2 | 16.8 | -0.6 |
| 4 | 24.2 | 41.4 | -17.2 |
| 5 | 22.0 | 21.4 | 0.6 |
| 6 | 33.4 | 38.4 | -5.0 |
| 7 | 23.8 | 16.8 | 7.0 |
| 8 | 58.2 | 32.2 | 26.0 |
| 9 | 33.6 | 27.8 | 5.8 |
| 10 | 24.4 | 23.2 | 1.2 |
| 11 | 23.4 | 29.6 | -6.2 |
| 12 | 21.2 | 20.6 | 0.6 |
| 13 | 36.2 | 32.2 | 4.0 |
| 14 | 29.8 | 53.8 | -24.0 |

A partir de la columna de diferencias observadas se calcula $\overline{D} = 1.21$ y $S_D = 12.68$

Además $t_{\frac{\alpha}{2},n-1} = t_{0.05,13} = 1.771$, entonces el intervalo para la diferencia $\mu_D = \mu_1 - \mu_2$ de nivel 0.90 es

$$\left[1.21 - 1.771 \times \frac{12.68}{\sqrt{14}};\ \ 1.21 + 1.771 \times \frac{12.68}{\sqrt{14}}\right] = \left[-4.79;\ \ 7.21\right]$$

## 9.7 – Intervalo de confianza para la varianza de una distribución normal

Supongamos que se quiere hallar un intervalo de confianza para la varianza $\sigma^2$ de una distribución normal.
Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria de una v.a. *X*, donde $X \sim N(\mu, \sigma^2)$.

Tomamos como estimador puntual de $\sigma^2$ a $S^2 = \frac{1}{n-1}\sum_{1=1}^{n}\left(X_i - \overline{X}\right)^2$

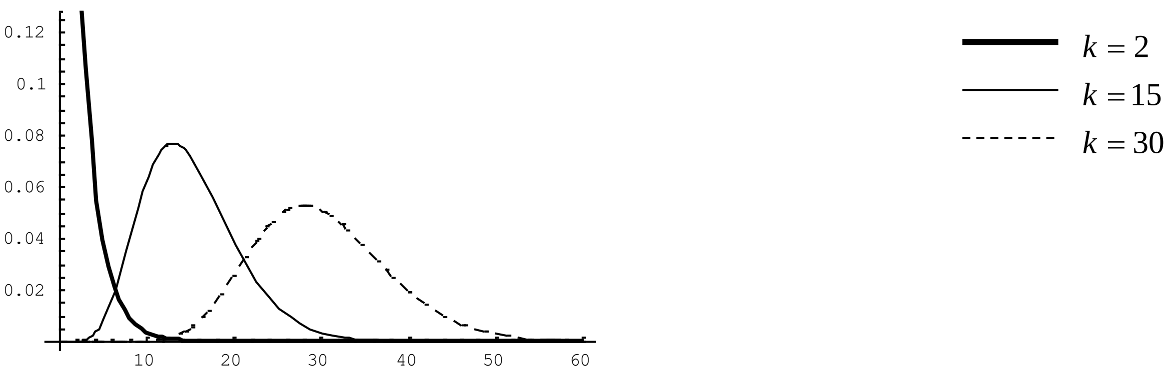
Luego a partir de este estimador puntual construimos el estadístico $X = \frac{(n-1)S^2}{\sigma^2}$

Este estadístico contiene al parámetro desconocido a estimar $\sigma^2$ y tiene una distribución conocida, se puede probar que *X* tiene una distribución llamada ***ji-cuadrado con n-1 grados de libertad***

Observación: Si X es una v.a. continua se dice que tiene distribución ***ji-cuadrado con k grados de libertad*** si su f.d.p. es

$$f(x) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{(k/2)-1} e^{-x/2} \qquad x > 0$$

Notación: X~ $\chi_k^2$
La distribución ji-cuadrdo es asimétrica. En la figura siguiente se grafica la densidad para diferentes valores de *k*



Anotaremos $\chi^2{}_{\alpha,k}$ al cuantil de la ji-cuadrado con *k* grados de libertad que deja bajo la *fdp* a derecha un área de $\alpha$, y a su izquierda un área de $1-\alpha$.

<u>Propiedades</u>:

1- Se puede probar que si $X_1, X_2, ..., X_n$ son variables aleatorias independientes con distribución $N(0,1)$ entonces $Z = X_1{}^2 + X_2{}^2 + ... + X_n{}^2$ tiene distribución ji-cuadrado con *n* grados de libertad.

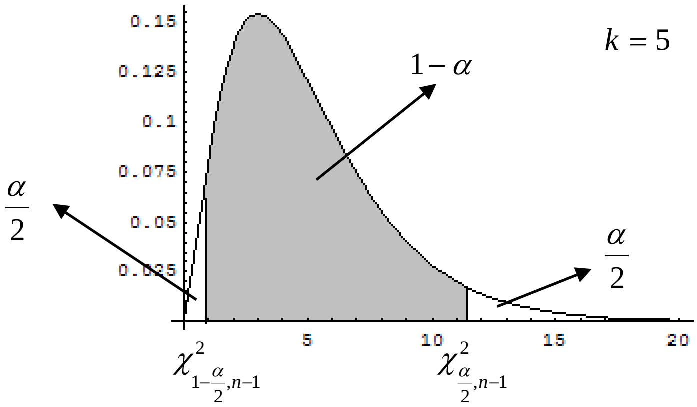
2- Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes tal que $X_i$ tiene distribución ji-cuadrado con $k_i$ grados de libertad, entonces $Z = X_1 + X_2 + ... + X_n$ tiene distribución ji-cuadrado con *k* grados de libertad donde $k = k_1 + k_2 + ... + k_n$

3- Si $X \sim \chi_k^2$ entonces ***para k grande*** $\sqrt{2X} \sim N\left(\sqrt{2k-1},\ 1\right)$ aproximadamente.

Para desarrollar el intervalo de confianza planteamos hallar dos números *a* y *b* tales que

$$P(a \le X \le b) = 1-\alpha \qquad \text{es decir} \qquad P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) = 1-\alpha$$

Se puede probar que la mejor elección de *a* y *b* es: $a = \chi^2_{1-\frac{\alpha}{2},n-1}$ y $b = \chi^2_{\frac{\alpha}{2},n-1}$




Por lo tanto

$$P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

y despejando $\sigma^2$ se llega a

$$P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right) = 1-\alpha$$

Entonces

Si $(X_1, X_2, ..., X_n)$ es una muestra aleatoria de una v.a. *X*, donde $X \sim N(\mu, \sigma^2)$, un intervalo de confianza para $\sigma^2$ de nivel $1-\alpha$ es

$$\left( \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} ; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right) \qquad (8.8)$$

Observación: un intervalo de confianza para $\sigma$ de nivel $1-\alpha$, es $\left( \sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}} ; \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}} \right)$

Ejemplo:
Un fabricante de detergente líquido está interesado en la uniformidad de la máquina utilizada para llenar las botellas. De manera específica, es deseable que la desviación estándar $\sigma$ del proceso de llenado sea menor que 0.15 onzas de líquido; de otro modo, existe un porcentaje mayor del deseable de botellas con un contenido menor de detergente. Supongamos que la distribución del volumen de llenado es aproximadamente normal. Al tomar una muestra aleatoria de 20 botellas, se obtiene una varianza muestral $S^2 = 0.0153$. Hallar un intervalo de confianza de nivel 0.95 para la verdadera varianza del volumen de llenado.

Solución:
La v.a. de interés es *X*: " volumen de llenado de una botella"
Se asume que $X \sim N(\mu, \sigma^2)$ con $\sigma$ desconocido.
Estamos en las condiciones para aplicar (8.8)

Tenemos que $1-\alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \chi^2_{1-\frac{\alpha}{2}, n-1} = \chi^2_{0.975,19} = 8.91$ y $\chi^2_{\frac{\alpha}{2}, n-1} = \chi^2_{0.025,19} = 32.85$

Además $S^2 = 0.0153$

Por lo tanto el intervalo es

$$\left( \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} ; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right) = \left( \frac{(20-1) \times 0.0153}{32.85} ; \frac{(20-1) \times 0.0153}{8.91} \right) = (0.00884;\ 0.0326)$$

Y un intervalo para $\sigma$ es $\left(\sqrt{0.00884};\ \sqrt{0.0326}\right) = (0.09;\ 0.1805)$

Por lo tanto con un nivel de 0.95 los datos ***no apoyan la afirmación que*** $\sigma < 0.15$

## 9.8 – Intervalo de confianza para el cociente de varianzas de dos distribuciones normales

Supongamos que se tienen dos poblaciones normales e independientes con varianzas desconocidas $\sigma_1^2$ y $\sigma_2^2$ respectivamente. Se desea encontrar un intervalo de nivel $1-\alpha$ para el cociente de las dos varianzas $\frac{\sigma_1^2}{\sigma_2^2}$.

Se toma una muestra aleatoria de tamaño $n_1$ de una de las poblaciones y una muestra de tamaño $n_2$ de la otra población. Sean $S_1^2$ y $S_2^2$ las dos varianzas muestrales.
Consideramos el estadístico

$$F = \frac{S_2^2 / \sigma_2^2}{S_1^2 / \sigma_1^2}$$

Notar que *F* contiene al parámetro de interés $\frac{\sigma_1^2}{\sigma_2^2}$, pues $F = \frac{S_2^2 \times \sigma_1^2}{S_1^2 \times \sigma_2^2}$

Se puede probar que *F* tiene una distribución llamada Fisher con $n_2 - 1$ y $n_1 - 1$ grados de libertad.

<u>Observación</u>:
Sea *X* una variable aleatoria continua, se dice que tiene distribución Fisher con *u* grados de libertad en el numerador y *v* grados de libertad en el denominador si su *fdp* es de la forma

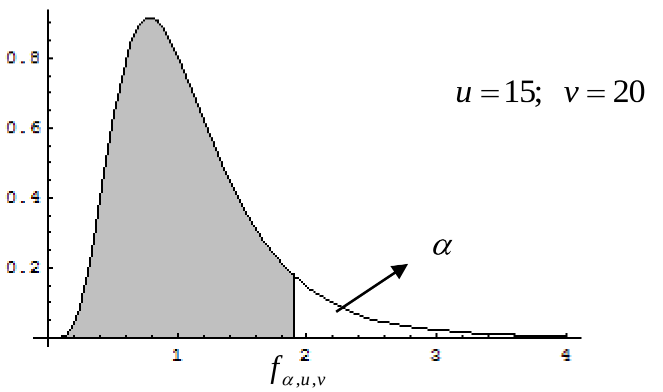
$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{\frac{u}{2}} x^{\frac{u}{2}-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\left(\frac{u}{v}\right)x+1\right]^{\frac{u+v}{2}}} \qquad 0 < x < \infty$$

En particular si *W* e *Y* son variables aleatorias independientes ji-cuadrado con *u* y *v* grados de libertad respectivamente, entonces el cociente

$$F = \frac{W/u}{Y/v}$$

Tiene una distribución Fisher con *u* grados de libertad en el numerador y *v* grados de libertad en el denominador.
<u>Notación</u>: $F \sim F_{u,v}$

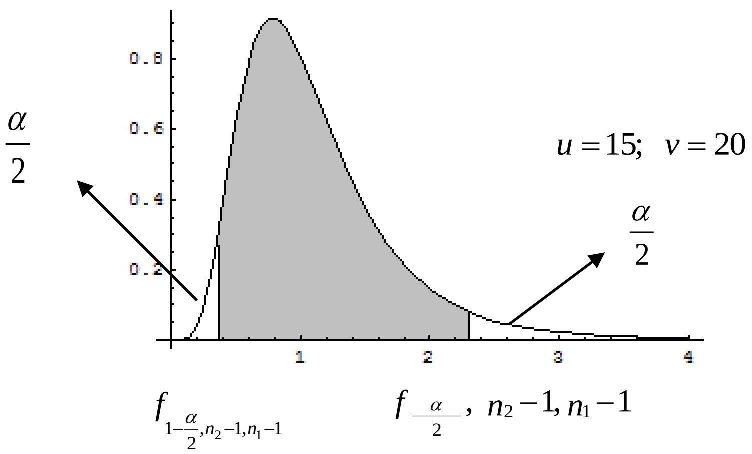
La gráfica de una distribución Fisher es similar a la de una ji-cuadrado, es asimétrica. Anotamos $f_{\alpha,u,v}$ al cuantil que deja a su derecha un área de $\alpha$ bajo la curva de densidad.



Existe la siguiente relación entre los cuantiles de una $F_{u,v}$ y de una $F_{v,u}$

$$\boxed{f_{1-\alpha,u,v} = \frac{1}{f_{\alpha,v,u}}}$$

Planteamos la siguiente ecuación $P(a \le F \le b) = 1-\alpha$ y se pede probar que la mejor elección de $a$ y $b$ es : $a = f_{1-\frac{\alpha}{2},n_2-1,n_1-1}$ y $b = f_{\frac{\alpha}{2},n_2-1,n_1-1}$




Entonces $P\left( f_{1-\frac{\alpha}{2},n_2-1,n_1-1} \le \dfrac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \le f_{\frac{\alpha}{2},n_2-1,n_1-1} \right) = 1-\alpha$

Despejando el cociente $\dfrac{\sigma_1^2}{\sigma_2^2}$ queda :

$$P\left( \frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2},n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2},n_2-1,n_1-1} \right) = 1-\alpha$$

Por lo tanto

Si se tienen dos poblaciones normales e independientes con varianzas desconocidas $\sigma_1^2$ y $\sigma_2^2$ respectivamente, entonces un intervalo de nivel $1-\alpha$ para el cociente de las dos varianzas $\dfrac{\sigma_1^2}{\sigma_2^2}$ es

$$\left[ \frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2},n_2-1,n_1-1} ; \ \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2},n_2-1,n_1-1} \right] \quad (8.9)$$

<u>Ejemplo</u>:
Una compañía fabrica propulsores para uso en motores de turbina. Una de las operaciones consiste en esmerilar el terminado de una superficie particular con una aleación de titanio. Pueden emplear-

se dos procesos de esmerilado, y ambos pueden producir partes que tienen la misma rugosidad superficial promedio. Interesaría seleccionar el proceso que tenga la menor variabilidad en la rugosidad de la superficie. Para esto se toma una muestra de 12 partes del primer proceso, la cual tiene una desviación estándar muestral $S_1 = 5.1$ micropulgadas, y una muestra aleatoria de 15 partes del segundo proceso, la cual tiene una desviación estándar muestral $S_2 = 4.7$ micropulgadas. Se desea encontrar un intervalo de confianza de nivel 90% para el cociente de las dos varianzas.
Suponer que los dos procesos son independientes y que la rugosidad de la superficie está distribuida de manera normal.

<u>Solución</u>:
Estamos en las condiciones para aplicar (8.9)
Buscamos en la tabla de la Fisher $f_{1-\frac{\alpha}{2},n_2-1,n_1-1} = f_{0.95,14,11} = \dfrac{1}{f_{0.05,11,14}} = \dfrac{1}{2.58} = 0.39$

y $f_{\frac{\alpha}{2},n_2-1,n_1-1} = f_{0.05,14,11} = 2.74$

Entonces el intervalo es

$$\left[\frac{5.1^2}{4.7^2}0.39;\ \frac{5.1^2}{4.7^2}2.74\right] = [0.46;\ 3.23]$$

Como este intervalo incluye al 1, no podemos afirmar que las desviaciones estándar de los dos procesos sean diferentes con una confianza de 90%.

## 9.9 – <u>Intervalo de confianza para una proporción</u>

Sea una población de tamaño *N* (eventualmente puede ser infinito) de cuyos individuos nos interesa cierta propiedad *A*. Supongamos que la probabilidad de que un individuo de la población verifique *A* es $p = P(A)$ .El significado del parámetro *p* es, en consecuencia, el de proporción de individuos de la población que verifican la propiedad *A*. Podemos definir una variable
aleatoria $X_i$ que mide a los individuos de la población la ocurrencia o no de la propiedad $A$.
La variable aleatoria tendrá la distribución:

$$p(x) = \begin{cases} p(1) = P(X_i = 1) = p \\ p(0) = P(X_i = 0) = 1 - p, \end{cases}$$

es decir, $X_i$ es una v.a. que toma sólo dos valores: 1 (si el individuo verifica *A*) con probabilidad *p* y 0 (cuando no verifica *A*) con probabilidad 1-*p*. Esto es equivalente a decir que $X_i$ tiene una distribución binomial con parámetros 1 y *p*: $X_i \sim B(1,p)$.

Supongamos que consideramos una muestra aleatoria $(X_1, X_2 ..., X_n)$ de tamaño $n$. Si formamos el estadístico $X = X_1 + X_2 + ... + X_n$ , es evidente que esta v.a. mide el número de individuos de la muestra de tamaño *n* que verifican la propiedad *A*. Por lo tanto por su significado *X* es una v.a. cuya distribución es binomial con parámetros *n* y *p*: $X \sim B(n,p)$. De acuerdo con esto, la variable

aleatoria $\hat{P}$ definida: $\hat{P} = \frac{X}{n}$ representa la proporción de individuos de la muestra que verifican la propiedad *A*.
Observemos que siendo $X_i \sim B(1,p)$ es $E(X_i) = p$. Y, dado que $X \sim B(n,p)$, también es $E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$, es decir $\hat{P}$ es un estimador insesgado de $p$. Esto es de esperar pues $\hat{P} = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Pero además, es fácil ver que $\hat{P}$ es estimador consistente de $p$. En efecto, tenemos que $E(\hat{P}) = p$, pero también es

$$V(\hat{P}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Deseamos construir un intervalo de confianza de *p*. Es razonable basarnos en el estimador insesgado $\hat{P}$. Consideramos como pivote a la variable aleatoria

$Z = \dfrac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ cuya distribución es, para *n* suficientemente grande, aproximadamente $N(0,1)$. En efecto:

Siendo $\hat{P} = \frac{X_1}{n} + \frac{X_2}{n} + \ldots + \frac{X_n}{n}$, es $E(\hat{P}) = \sum_{i=1}^{n} E\left(\frac{X_i}{n}\right) = p$ y $V(\hat{P}) = \sum_{i=1}^{n} V\left(\frac{X_i}{n}\right) = \frac{p(1-p)}{n}$

Por lo tanto:

$$Z = \frac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \underset{n\ grande}{\sim} N(0,1),$$

El pivote puede ponerse en una forma más conveniente si tenemos en cuenta que, según vimos recién, $\hat{P}$ es estimador consistente de *p* y en consecuencia, en el denominador reemplazamos el parámetro desconocido *p* por su estimador $\hat{P}$, y se puede probar que :

$$Z = \frac{\hat{P} - p}{\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}} \approx N(0,1). \quad \text{aproximadamente si } n \text{ es grande}$$

Partiendo de este pivote podemos seguir los mismos pasos de los casos anteriores para llegar al siguiente intervalo de confianza al nivel $1-\alpha$ de *p*:

$$\left[\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right] \quad \text{con} \quad \Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

Entonces

Si $\hat{P}$ es la proporción de observaciones de una muestra aleatoria de tamaño $n$ que verifican una propiedad de interés, entonces un intervalo de confianza para la proporción p de la población que cumple dicha propiedad de nivel aproximadamente $1-\alpha$ es

$$\left[\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}},\ \ \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}}\right] \qquad (8.10)$$

Observaciones:

1- Este procedimiento depende de la aproximación normal a la distribución binomial. Por lo tanto el intervalo (8.10) se puede utilizar si $n\hat{P} > 10$ y $n(1-\hat{P}) > 10$, es decir, ***la muestra debe contener un mínimo de diez éxitos y diez fracasos.***

2- La longitud del intervalo es $L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}}$, pero esta expresión está en función de $\hat{P}$

Si nos interesa hallar un valor de $n$ de manera tal que la longitud $L$ sea menor que un valor determinado, podemos hacer dos cosas:

a) tomar una muestra preliminar, con ella estimar $p$ con $\hat{P}$ y de la expresión anterior despejar $n$, lo que lleva a

$$L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}} \le l \quad \Rightarrow \quad n \ge \left(\frac{2z_{\frac{\alpha}{2}}}{l}\right)^2 \hat{P}\left(1-\hat{P}\right)$$

b) si no tomamos una muestra preliminar, entonces acotamos $\hat{P}\left(1-\hat{P}\right) \le 0.5 \times \left(1-0.5\right)$, entonces

$$L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}} \le 2z_{\frac{\alpha}{2}}\sqrt{\frac{0.5\left(1-0.5\right)}{n}} \le l \quad \Rightarrow \quad n \ge \left(\frac{z_{\frac{\alpha}{2}}}{l}\right)^2$$

Ejemplo:

Un fabricante de componentes compra un lote de dispositivos de segunda mano y desea saber la proporción de la población que están fallados. Con ese fin experimenta con 140 dispositivos elegidos al azar y encuentra que 35 de ellos están fallados.

a) Calcular un intervalo de confianza del 99% para la proporción poblacional $p$.

b) ¿De qué tamaño deberá extraerse la muestra a fin de que la proporción muestral no difiera de la proporción poblacional en más de 0.03 con un 95% de confianza?

Solución:

a) El tamaño de la muestra es $n = 140$ (muestra grande)

La proporción muestral es $\hat{P} = \frac{35}{140} = 0.25$

El nivel de confianza es $1-\alpha = 0.99 \rightarrow \alpha = 0.01 \rightarrow \frac{\alpha}{2} = 0.005$.

De la tabla de la normal estandarizada vemos que $z_{0.005} = 2.58$. Entonces el intervalo buscado es:

$$\left[0.25 - 2.58\sqrt{\frac{0.25(1-0.25)}{140}},\ \ 0.25 + 2.58\sqrt{\frac{0.25(1-0.25)}{140}}\right] = [0.15558,\ 0.34441]$$

b) Buscamos el tamaño *n* de la muestra tal que con un 95% de confianza la proporción muestral $\hat{P}$ esté a una distancia 0.03 de la proporción poblacional *p*, es decir buscamos *n* tal que

$\frac{L}{2} \le 0.03$ , por lo tanto como $\alpha = 0.05 \ \rightarrow\ \frac{\alpha}{2} = 0.025$ ***si tomamos la muestra anterior como preliminar*** :

$$n \ge \left(\frac{2z_{\frac{\alpha}{2}}}{l}\right)^2 \hat{P}(1-\hat{P}) = \left(\frac{2\times 1.96}{2\times 0.03}\right)^2 0.25(1-0.25) = 800.3333$$

Por lo tanto hay que tomar una muestra de tamaño por lo menos 801. Como ya se tomó una muestra de tamaño 140, hay que tomar otra adicional de tamaño $801 - 140 = 661$
Supongamos que no tomamos una muestra inicial, entonces directamente planteamos

$$n \ge \left(\frac{z_{\frac{\alpha}{2}}}{l}\right)^2 = \left(\frac{1.96}{2\times 0.03}\right)^2 = 1067.1111$$

Entonces hay que tomar una muestra de tamaño 1068 por lo menos.

## 9.10 – <u>Intervalo de confianza para la diferencia entre dos proporciones</u>

Supongamos que existen dos proporciones de interés $p_1$ y $p_2$ y es necesario obtener un intervalo de confianza de nivel $1-\alpha$ para la diferencia $p_1 - p_2$.
Supongamos que se toman dos muestras independientes de tamaños $n_1$ y $n_2$ respectivamente de dos poblaciones.
Sean las variables aleatorias
$X_1$ : “número de observaciones en la primera muestra que tienen la propiedad de interés”
$X_2$ : “número de observaciones en la segunda muestra que tienen la propiedad de interés”
Entonces $X_1$ y $X_2$ son variables aleatorias independientes y $X_1 \sim B(n_1, p_1)$ ; $X_2 \sim B(n_2, p_2)$
Además $\hat{P}_1 = \frac{X_1}{n_1}$ y $\hat{P}_2 = \frac{X_2}{n_2}$ son estimadores puntuales de $p_1$ y $p_2$ respectivamente.
Vemos que $E(\hat{P}_1 - \hat{P}_2) = p_1 - p_2$ y $V(\hat{P}_1 - \hat{P}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$
Aplicando la aproximación normal a la binomial podemos decir que

$Z = \dfrac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \approx N(0,1)$ , y como en el caso de intervalo para una proporción estimamos $\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ con $\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$ y entonces

$Z = \dfrac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2}}} \approx N(0,1)$ aproximadamente.

Planteamos la ecuación $P(-z \le Z \le z) \approx \Phi(z) - \Phi(-z) = 1\text{-}\alpha$, lo que lleva a $z = z_{\frac{\alpha}{2}}$, y con una deducción análoga a las anteriores se llega al intervalo

$$\left[ \hat{P}_1 - \hat{P}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}};\ \hat{P}_1 - \hat{P}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}} \right]$$

Entonces

Si $\hat{P}_1$ y $\hat{P}_2$ son las proporciones muestrales de una observación de dos muestras aleatorias independientes de tamaños $n_1$ y $n_2$ respectivamente que verifican la propiedad de interés, entonces un intervalo de confianza de nivel $1-\alpha$ aproximadamente es

$$\left[ \hat{P}_1 - \hat{P}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}};\ \hat{P}_1 - \hat{P}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}} \right] \quad (8.11)$$

Ejemplo:
Se lleva a cabo un estudio para determinar la efectividad de una nueva vacuna contra la gripe. Se administra la vacuna a una muestra aleatoria de 3000 sujetos, y de ese grupo 13 contraen gripe. Como grupo de control se seleccionan al azar 2500 sujetos, a los cuales no se les administra la vacuna, y de ese grupo 170 contraen gripe. Construya un intervalo de confianza de nivel 0.95 para la diferencia entre las verdaderas proporciones de individuos que contraen gripe.

Solución:
Sean las variables aleatorias

$X_1$ : "número de personas que contraen gripe del grupo que recibió la vacuna"

$X_2$ : "número de personas que contraen gripe del grupo que no recibió la vacuna"

Entonces $X_1 \sim B(n_1,p_1)$ ; $X_2 \sim B(n_2,p_2)$ donde $n_1 = 3000$ ; $n_2 = 2500$

Además $\hat{P}_1 = \dfrac{13}{3000}$ ; $\hat{P}_2 = \dfrac{170}{2500}$

Y $1-\alpha = 0.95 \quad \rightarrow \quad z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

Entonces

$$\left[\hat{P}_1-\hat{P}_2-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1\left(1-\hat{P}_1\right)}{n_1}+\frac{\hat{P}_2\left(1-\hat{P}_2\right)}{n_2}};\ \ \hat{P}_1-\hat{P}_2+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1\left(1-\hat{P}_1\right)}{n_1}+\frac{\hat{P}_2\left(1-\hat{P}_2\right)}{n_2}}\right]=$$

$$=\left[\frac{13}{3000}-\frac{170}{2500}-1.96\sqrt{\frac{\frac{13}{3000}\left(1-\frac{13}{3000}\right)}{3000}+\frac{\frac{170}{2500}\left(1-\frac{170}{2500}\right)}{2500}};\right.$$

$$\left.\frac{13}{3000}-\frac{170}{2500}+1.96\sqrt{\frac{\frac{13}{3000}\left(1-\frac{13}{3000}\right)}{3000}+\frac{\frac{170}{2500}\left(1-\frac{170}{2500}\right)}{2500}}\right]=\left[-0.0738112;\ \ -0.0535222\right]$$

# *9- Test o prueba de hipótesis*

## 9.1 – Introducción

Hasta ahora hemos estudiado el problema de estimar un parámetro desconocido a partir de una muestra aleatoria.
En muchos problemas se requiere tomar una decisión entre aceptar o rechazar una proposición sobre algún parámetro. Esta proposición recibe el nombre de ***hipótesis estadística***, y el procedimiento de toma de decisión sobre la hipótesis se conoce como ***prueba o test de hipótesis***.
Como se emplean distribuciones de probabilidad para representar poblaciones, también podemos decir que una hipótesis estadística es una proposición sobre la distribución de probabilidad de una variable aleatoria, donde la hipótesis involucra a uno más parámetros de esta distribución.
Por ejemplo, supongamos que cierto tipo de motor de automóvil emite una media de 100 mg de óxidos de nitrógeno ($NO_x$) por segundo con 100 caballos de fuerza. Se ha propuesto una modificación al diseño del motor para reducir las emisiones de $NO_x$. El nuevo diseño se producirá si se demuestra que la media de su tasa de emisiones es menor de 100 mg/s. Se construye y se prueba una muestra de 50 motores modificados. La media muestral de emisiones de $NO_x$ es de 92 mg/s, y la desviación estándar muestral es de 21 mg/s.
La variable aleatoria de interés en este caso es *X*: "tasa de emisión de un motor modificado tomado al azar".
La preocupación de los fabricantes consiste en que los motores modificados no puedan reducir todas la emisiones; es decir que la media poblacional pudiera ser 100 o mayor que 100.
Entonces, la pregunta es: ¿es factible que esta muestra pueda provenir de una v.a. con media 100 o mayor?
Éste es el tipo de preguntas que las pruebas de hipótesis están diseñadas para responder. Veremos cómo construir una prueba de hipótesis, pero podemos decir que en general se basa en construir a partir de la muestra aleatoria un ***estadístico***, y según el valor que tome este ***estadístico de prueba*** se aceptará o se rechazará la hipótesis.
Se ha observado una muestra con media $\overline{X} = 92$.
Hay dos interpretaciones posibles de esta observación:

1- La media poblacional es realmente mayor o igual que 100, y la media muestral es menor que 100 debido a la ***variabilidad*** propia de la variable aleatoria $\overline{X}$
2- La media poblacional es en realidad menor que 100, y la media muestral refleja este hecho.

Estas dos explicaciones tienen nombres: la primera se llama ***hipótesis nula***; la segunda es la ***hipótesis alternativa***.
En la mayoría de las situaciones la hipótesis nula dice que el efecto que indica la muestra es atribuible solamente a la variación aleatoria del estadístico de prueba.
La hipótesis alternativa establece que el efecto que indica la muestra es verdadero.
Para hacer las cosas más precisas, todo se expresa mediante símbolos. La hipótesis nula se denota por $H_0$, la hipótesis alternativa se denota con $H_1$. Como es usual la media poblacional se anota $\mu$. Por lo tanto se tiene

$$H_0 : \mu \geq 100 \quad \text{contra} \quad H_1 : \mu < 100 \quad \text{(hipótesis alternativa unilateral)}$$

Esencialmente, para realizar una prueba de hipótesis se pone la hipótesis nula en juicio. Se asume que $H_0$ es verdadera, de la misma manera como se empieza en un juicio bajo el supuesto de que un acusado es inocente. La muestra aleatoria proporciona la evidencia.

Las hipótesis son siempre proposiciones ***sobre los parámetros de la población o distribución bajo estudio***, no proposiciones sobre la muestra.
Otros tipos de hipótesis que podrían formularse son

$$H_0: \mu \leq 100 \quad \text{contra} \quad H_1: \mu > 100 \qquad \text{(hipótesis alternativa unilateral)}$$

o

$$H_0: \mu = 100 \quad \text{contra} \quad H_1: \mu \neq 100 \qquad \text{(hipótesis alternativa bilateral)}$$

En el ejemplo tenemos $X_1, X_2, ..., X_{50}$ muestra aleatoria de la v.a. $X$ definida anteriormente.
Como estamos haciendo una hipótesis sobre la media poblacional es razonable tomar como estadístico de prueba a $\overline{X}$. El valor observado de la media muestral es $\overline{X} = 92$.
Si el valor de $\overline{X}$ es muy "menor" que 100 entonces se considera que hay evidencia en contra $H_0$ y se la rechaza, aceptando la hipótesis alternativa.
Si el valor de $\overline{X}$ no es "muy menor" que 100 entonces se considera que no hay evidencia en contra $H_0$ y se rechaza la hipótesis alternativa.
Ya veremos como construir una ***regla de decisión***, supongamos ahora que tenemos la siguiente regla:

$$\begin{cases} se \quad rechaza \ H_0 \quad si \quad \overline{X} < 95 \\ se \quad acepta \ H_0 \quad si \quad \overline{X} \geq 95 \end{cases}$$

El intervalo $\left(95, \ \infty\right)$ es la ***zona de aceptación.***

La región $\left(-\infty; \ 95\right)$ es la ***zona de rechazo o región crítica.***

Mientras que 95 es el ***punto crítico***.

Como estamos tomando una decisión basados en el valor de un estadístico podemos cometer dos tipos de errores: rechazar $H_0$ cuando ésta es verdadera, es decir el estadístico toma valores en la zona de rechazo cuando $H_0$ es verdadera; o aceptar $H_0$ cuando ésta es falsa, es decir que el estadístico tome valores en la zona de aceptación cuando $H_0$ es falsa.
El primero se conoce como ***error de tipo I***, y el segundo como ***error de tipo II***.
Debido a que la decisión se basa en variables aleatorias es posible asociar probabilidades a los errores de tipo I y II, específicamente anotamos

$$\alpha = P(error \quad de \quad tipo \quad I)$$

$$\beta = P(error \quad de \quad tipo \quad II)$$

A $\alpha = P(error \quad de \quad tipo \quad I)$ se lo conoce como ***nivel de significancia del test***.

Para calcular estas probabilidades debemos conocer la distribución del estadístico de prueba en el caso de ser $H_0$ verdadera, es decir debemos conocer la distribución del estadístico de prueba "bajo $H_0$".

En el ejemplo anterior la muestra es grande, ya sabemos que por T.C.L. el estadístico

$$Z = \frac{\overline{X} - 100}{s/\sqrt{n}} \approx N(0,1) \quad \textbf{\textit{si}} \ \ H_0 \ \textbf{\textit{es verdadera,}} \ \text{ o sea } \quad Z = \frac{\overline{X} - 100}{21/\sqrt{50}} \approx N(0,1)$$

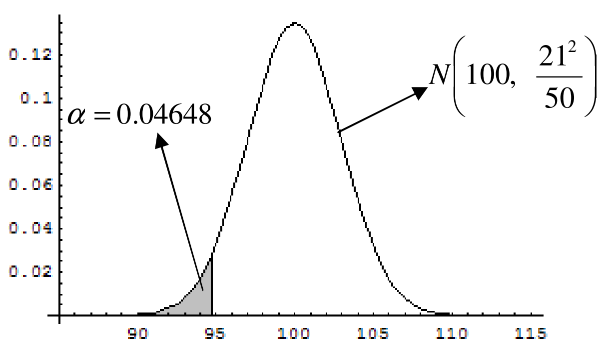
Entonces para calcular $\alpha$ planteamos:

$$\alpha = P(error \ \ de \ \ tipo \ \ I) = P\left( rechazar \ \ H_0 / H_0 \ \ es \ \ V \right) = P\left(\overline{X} < 95 / \mu = 100\right) =$$

$$= P\left( \frac{\overline{X} - 100}{21/\sqrt{50}} < \frac{95 - 100}{21/\sqrt{50}} \right) \approx \Phi\left( \frac{95 - 100}{21/\sqrt{50}} \right) = \Phi(-1.6835) = 1 - 0.95352 = 0.04648$$

Esto significa que el 4.64% de las muestras aleatorias conducirán al rechazo de la hipótesis $H_0 : \mu \geq 100$ cuando el verdadero $\mu$ sea mayor o igual que 100.

En este caso el gráfico de la zona de rechazo es




Del gráfico anterior vemos que podemos reducir $\alpha$ al aumentar la zona de aceptación. Por ejemplo supongamos que ahora la regla de decisión es

$$\begin{cases} se \ \ rechaza \ H_0 \ \ si \ \ \ \overline{X} < 93 \\ se \ \ acepta \ \ H_0 \ \ si \ \ \overline{X} \geq 93 \end{cases}$$

Entonces $\alpha = P(error \ \ de \ \ tipo \ \ I) = P\left( rechazar \ \ H_0 / H_0 \ \ es \ \ V \right) = P\left(\overline{X} < 93 / \mu = 100\right) =$

$$= P\left( \frac{\overline{X} - 100}{21/\sqrt{50}} < \frac{93 - 100}{21/\sqrt{50}} \right) \approx \Phi\left( \frac{93 - 100}{21/\sqrt{50}} \right) = \Phi(-2.357) = 1 - 0.99061 = 0.00939$$

También se puede reducir $\alpha$ aumentando el tamaño de la muestra. Supongamos que $n=85$, entonces

$$\alpha = P(error \;\; de \;\; tipo \;\; I) = P\left( rechazar \;\; H_0 / H_0 \;\; es \;\; V \right) = P\left(\overline{X} < 95 / \mu = 100\right) =$$

$$= P\left( \frac{\overline{X}-100}{21/\sqrt{85}} < \frac{95-100}{21/\sqrt{85}} \right) \approx \Phi\left( \frac{95-100}{21/\sqrt{85}} \right) = \Phi(-2.195) = 1-0.98574 = 0.01426$$

También es importante examinar la probabilidad de cometer error de tipo II, esto es

$$\beta = P(error \;\; de \;\; tipo \;\; II) = P(aceptar\, H_0 / H_0 \; es \; falsa)$$

Pero en este caso para llegar a un valor numérico necesitamos tener una alternativa específica pues en nuestro ejemplo:

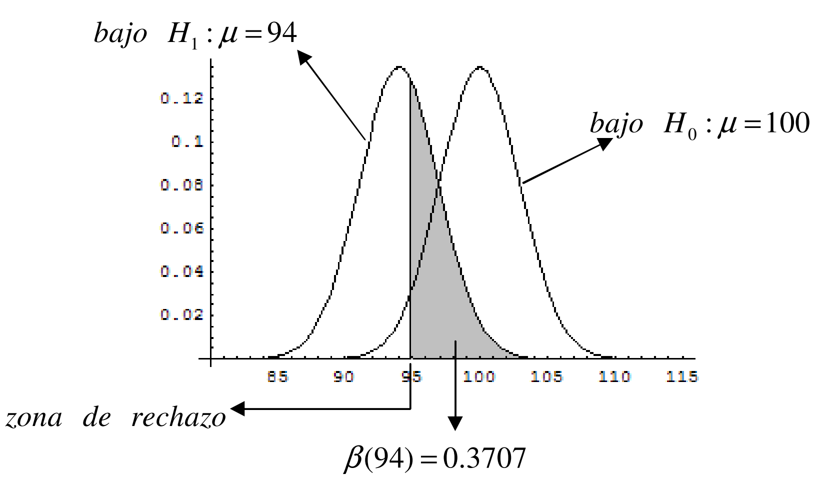
$$\beta = P(error \;\; de \;\; tipo \;\; II) = P(aceptar\, H_0 / H_0 \; es \; falsa) = P\left(\overline{X} \geq 95 / \mu \neq 100\right) =$$

$$= P\left( \frac{\overline{X}-\mu}{21/\sqrt{50}} \geq \frac{95-\mu}{21/\sqrt{50}} \right) = 1-\Phi\left( \frac{95-\mu}{21/\sqrt{50}} \right) = \beta(\mu)$$

Donde anotamos con $\mu$ a la verdadera media poblacional ***desconocida***.
Podemos entonces calcular $\beta$ para un valor particular de $\mu$, por ejemplo nos puede interesar como se comporta el test cuando la verdadera media es $\mu = 94$, entonces

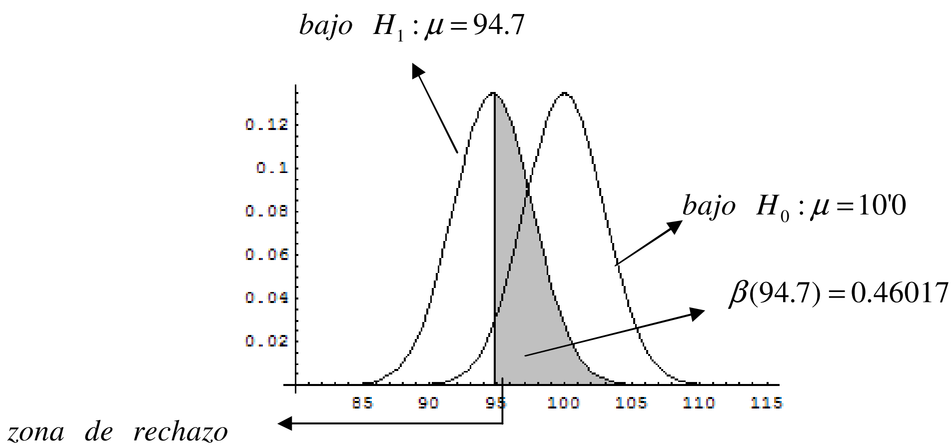
$$\beta(94) = P\left( \frac{\overline{X}-94}{21/\sqrt{50}} \geq \frac{95-94}{21/\sqrt{50}} \right) = 1-\Phi\left( \frac{95-94}{21/\sqrt{50}} \right) = 1-\Phi(0.3367) = 1-0.62930 = 0.3707$$

Gráficamente:



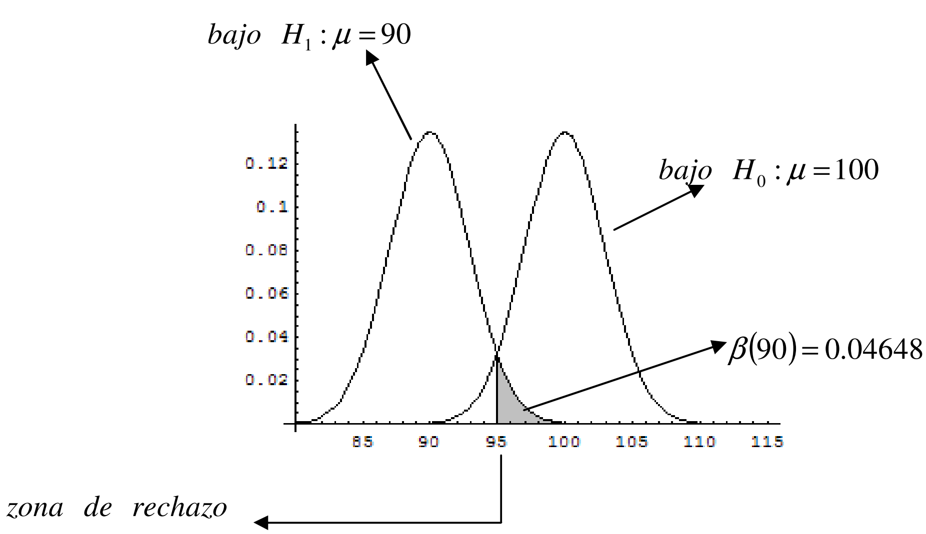
La probabilidad $\beta$ de cometer error de tipo II ***crece*** a medida que el valor verdadero de $\mu$ se acerca al valor hipotético. Por ejemplo si el verdadero valor de $\mu$ fuera 94.7 entonces

$$\beta(94.7) = P\left( \frac{\overline{X} - 94.7}{21/\sqrt{50}} \geq \frac{95 - 94.7}{21/\sqrt{50}} \right) = 1 - \Phi\left( \frac{95 - 94.7}{21/\sqrt{50}} \right) = 1 - \Phi(0.101015) = 1 - 0.53983 = 0.46017$$




Además, la probabilidad $\beta$ de cometer error de tipo II ***disminuye*** a medida que el valor verdadero de $\mu$ se aleja del valor hipotético. Por ejemplo si el verdadero valor de $\mu$ fuera 90 entonces

$$\beta(90) = P\left( \frac{\overline{X} - 90}{21/\sqrt{50}} \geq \frac{95 - 90}{21/\sqrt{50}} \right) = 1 - \Phi\left( \frac{95 - 90}{21/\sqrt{50}} \right) = 1 - \Phi(1.6835) = 1 - 0.95352 = 0.04648$$



También se puede reducir la probabilidad de cometer error de tipo II con el tamaño de la muestra. Por ejemplo si $n = 85$ entonces y $\mu = 94$

$$\beta(94) = P\left(\frac{\overline{X} - 94}{21/\sqrt{85}} \geq \frac{95-94}{21/\sqrt{85}}\right) = 1 - \Phi\left(\frac{95-94}{21/\sqrt{85}}\right) = 1 - \Phi(0.4390) = 1 - 0.67003 = 0.32997$$

Lo que se ha visto en los ejemplos anteriores se puede generalizar. Podemos recalcar los siguientes puntos importantes:

1- El tamaño de la región crítica, y en consecuencia la probabilidad $\alpha$ de cometer error de tipo I, siempre pueden reducirse mediante una selección apropiada de los valores críticos.
2- Los errores tipo I y II están relacionados. Una disminución en la probabilidad en un tipo de error siempre da como resultado un aumento en la probabilidad del otro, siempre que el tamaño de la muestra no cambie.
3- En general, un aumento en el tamaño de la muestra reduce tanto a $\alpha$ como a $\beta$, siempre que los valores críticos se mantengan constantes.
4- Cuando la hipótesis nula es falsa, $\beta$ aumenta a medida que el valor verdadero del parámetro tiende al valor hipotético propuesto por la hipótesis nula. El valor de $\beta$ disminuye a medida que aumenta la deferencia entre el verdadero valor medio y el propuesto.

En general el investigador controla la probabilidad $\alpha$ del error de tipo I cuando selecciona los valores críticos. Por lo tanto el rechazo de la hipótesis nula de manera errónea se puede fijar de antemano. Eso hace que rechazar la hipótesis nula sea una ***conclusión fuerte***.
La probabilidad $\beta$ de error de tipo II no es constante, sino que depende del valor verdadero del parámetro. También depende $\beta$ del tamaño de la muestra que se haya seleccionado. Como $\beta$ está en función del tamaño de la muestra y del valor verdadero del parámetro, la decisión de aceptar la hipótesis nula se la considera una ***conclusión débil***, a menos que se sepa que $\beta$ es aceptablemente pequeño. Por lo tanto ***cuando se acepta $H_0$ en realidad se es incapaz de rechazar $H_0$. No se puede rechazar $H_0$ pues no hay evidencia en contra $H_0$.***
Un concepto importante es el siguiente:
La ***potencia*** de un test es la probabilidad de rechazar la hipótesis nula. La simbolizamos $\pi(\mu)$. Para los ***valores de $\mu$ tal que la alternativa es verdadera*** se tiene

$$\pi(\mu) = P\left(rechazar \quad H_0 / H_0 \quad es \quad falsa\right) = 1 - \beta(\mu)$$

Las pruebas estadísticas se comparan mediante la comparación de sus propiedades de potencia.
La potencia es una medida de la ***sensibilidad*** del test, donde por sensibilidad se entiende la capacidad de una prueba para detectar diferencias.
En el ejemplo anterior, la sensibilidad de la prueba para detectar la diferencia entre una tasa de emisión media de 100 y otra de 94 es $\pi(94) = 1 - \beta(94) = 1 - 0.3707 = 0.6293$. Es decir si el valor verdadero de la tasa de emisión media es 94, la prueba rechazará de manera correcta $H_0$ y detectará esta diferencia el 62.93% de las veces. Si el investigador piensa que este valor es bajo entonces el investigador puede aumentar $\alpha$ o el tamaño de la muestra.

## 9.2 – Prueba de hipótesis sobre la media, varianza conocida

Veamos ahora cómo construir una regla de decisión sobre la media de una población.
***Supongamos que la variable aleatoria de interés X tiene una media $\mu$ y una varianza $\sigma^2$ conocida. Asumimos que X tiene distribución normal, es decir $X \sim N(\mu, \sigma^2)$.***
Nuevamente, como en el ejemplo introductorio, es razonable tomar como estadístico de prueba al promedio muestral $\overline{X}$. Bajo las suposiciones hechas tenemos que $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Supongamos que tenemos las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu \neq \mu_0$$

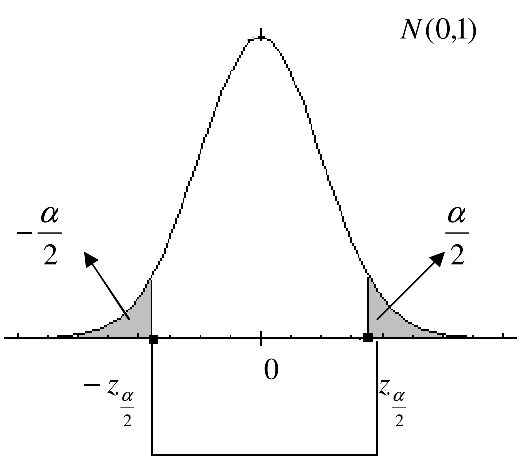
Donde $\mu_0$ es una constante específica. Se toma una muestra aleatoria $X_1, X_2, ..., X_n$ de la población.

Si $H_0 : \mu = \mu_0$ es verdadera, entonces $\overline{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$, por lo tanto el estadístico

$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ tiene distribución $N(0,1)$ si $H_0 : \mu = \mu_0$ es verdadera

Tomamos a $Z$ como ***estadístico de prueba***

Si $H_0 : \mu = \mu_0$ es verdadera entonces $P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$




Es evidente que una muestra que produce un valor del estadístico de prueba que cae en las colas de la distribución de $Z$ será inusual si $H_0 : \mu = \mu_0$ es verdadera, por lo tanto esto es un indicador que $H_0$ es falsa.
Entonces la regla de decisión es:

$$\begin{cases} rechazar \;\; H_0 \;\; si \;\; |Z| > z_{\frac{\alpha}{2}} \\ aceptar \;\; H_0 \;\; si \;\; |Z| \leq z_{\frac{\alpha}{2}} \end{cases}$$

Notar que la probabilidad que la estadística de prueba tome un valor que caiga en la zona de rechazo si $H_0$ es verdadera es igual a $\alpha$ , es decir la probabilidad de cometer error de tipo I es $\alpha$ pues

$$P(error \;\; de \;\; tipo \;\; I) = P\left( rechazar \;\; H_0 / H_0 \;\; es \;\; V \right) = P\left( \left| \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \right| > z_{\frac{\alpha}{2}} \Big/ \mu = \mu_0 \right) =$$

$$= P\left( \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_{\frac{\alpha}{2}} \right) + P\left( \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_{\frac{\alpha}{2}} \right) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

Ejemplo:
El porcentaje deseado de $SiO_2$ en cierto tipo de cemento aluminoso es 5.5. Para probar si el verdadero promedio de porcentaje es 5.5 para una planta de producción en particular, se analizaron 16 muestras obtenidas de manera independiente. Supongamos que el porcentaje de $SiO_2$ en una muestra está normalmente distribuido con $\sigma = 0.3$ , y que $\overline{x} = 5.25$.
¿Indica esto de manera concluyente que el verdadero promedio de porcentaje difiere de 5.5?. Utilice $\alpha = 0.01$

Solución:
La v.a. de interés es $X$: "porcentaje de $SiO_2$ en cierto tipo de cemento aluminoso"
Asumimos que $X \sim N(\mu, 3^2)$
Podemos plantear las hipótesis

$$H_0 : \mu = 5.5 \quad \text{contra} \quad H_1 : \mu \neq 5.5$$

Tenemos una muestra de tamaño $n = 16$ que dio un promedio muestral $\overline{x} = 5.25$
Como $\alpha = 0.01$ entonces $z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$

Por lo tanto la regla de decisión es
$$\begin{cases} rechazar \;\; H_0 \;\; si \;\; \left| \dfrac{\overline{X} - 5.5}{3 / \sqrt{16}} \right| > 2.575 \\[2ex] aceptar \;\; H_0 \;\; si \;\; \left| \dfrac{\overline{X} - 5.5}{3 / \sqrt{16}} \right| \leq 2.575 \end{cases}$$

El estadístico $\left| \dfrac{\overline{X} - 5.5}{3 / \sqrt{16}} \right|$ toma el valor $z_0 = \left| \dfrac{5.25 - 5.5}{3 / \sqrt{16}} \right| = 0.333333$

Como $z_0 = 0.333333 < 2.575 = z_{\frac{0.01}{2}}$ se acepta $H_0$

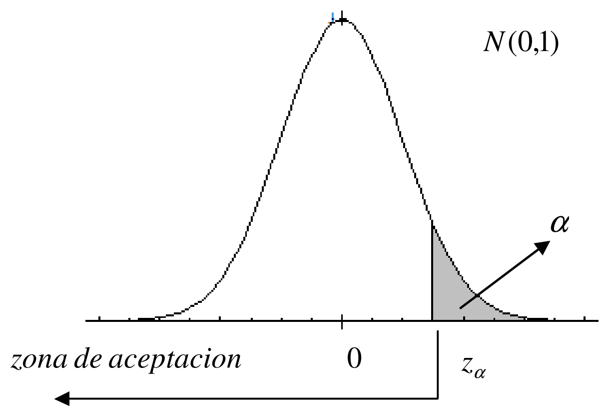
También podemos desarrollar tests o pruebas de hipótesis para el caso de que la hipótesis alternativa es unilateral.

Supongamos las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu > \mu_0$$

En este caso la región crítica debe colocarse en la cola superior de la distribución normal estándar y el rechazo de $H_0$ se hará cuando el valor calculado de $z_0$ sea muy grande, esto es la regla de decisión será

$$\begin{cases} rechazar \quad H_0 \quad si \quad \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha \\ aceptar \quad H_0 \quad si \quad \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \le z_\alpha \end{cases}$$

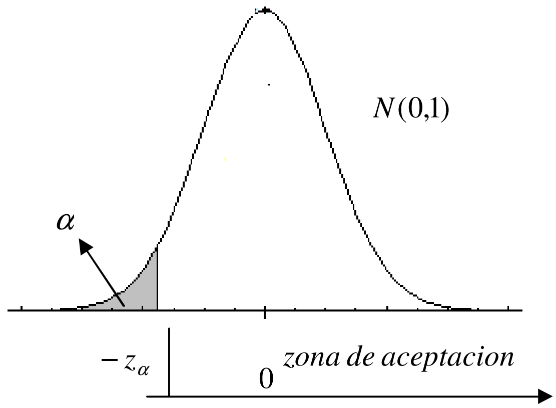



De manera similar para las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu < \mu_0$$

se calcula el valor del estadístico de prueba $z_0$ y se rechaza $H_0$ si el valor de $z_0$ es muy pequeño, es decir la regla de decisión será

$$\begin{cases} rechazar \quad H_0 \quad si \quad \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha \\ aceptar \quad H_0 \quad si \quad \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \ge -z_\alpha \end{cases}$$




Ejemplo:
Se sabe que la duración, en horas, de un foco de 75 watts tiene una distribución aproximadamente normal, con una desviación estándar de $\sigma = 25$ horas. Se toma una muestra aleatoria de 20 focos, la cual resulta tener una duración promedio de $\bar{x} = 1040$ horas
¿Existe evidencia que apoye la afirmación de que la duración promedio del foco es mayor que 1000 horas?. Utilice $\alpha = 0.05$.

Solución:
La v.a. de interés es X: "duración en horas de un foco tomado al azar"
Asumimos $X \sim N(\mu, 25^2)$
Podemos plantear las hipótesis

$$H_0 : \mu = 1000 \quad \text{contra} \quad H_1 : \mu > 1000$$

Tenemos una muestra de tamaño $n = 20$ que dio un promedio muestral $\bar{x} = 1040$
Como $\alpha = 0.05$ entonces $z_\alpha = z_{0.05} = 1.645$

Por lo tanto la regla de decisión es $\begin{cases} \textit{rechazar } H_0 \textit{ si } \dfrac{\bar{X} - 1000}{25/\sqrt{20}} > 1.645 \\ \textit{aceptar } H_0 \textit{ si } \dfrac{\bar{X} - 1000}{25/\sqrt{20}} \leq 1.645 \end{cases}$

El estadístico toma el valor $Z = \dfrac{\bar{X} - 1000}{25/\sqrt{20}}$ toma el valor $z_0 = \dfrac{1040 - 1000}{25/\sqrt{20}} = 7.1554$

Como $z_0 = 7.1554 > 1.645 = z_{0.05}$ se rechaza $H_0$

**P- valor**

Hasta ahora se dieron los resultados de una prueba de hipótesis estableciendo si la hipótesis nula fue o no rechazada con un valor especificado de $\alpha$ o nivel de significancia.
A menudo este planteamiento resulta inadecuado, ya que no proporciona ninguna idea sobre si el valor calculado del estadístico está apenas en la región de rechazo o bien ubicado dentro de ella. Además, esta forma de establecer los resultados impone a otros usuarios el nivel de significancia predeterminado.

Para evitar estas dificultades, se adopta el enfoque del ***p-valor***. El ***valor p*** o ***p-valor*** es la probabilidad de que el estadístico de prueba tome un valor que sea al menos tan extremo como el valor observado del estadístico de prueba cuando la hipótesis nula es verdadera. Es así como el p-valor da mucha información sobre el peso de la evidencia contra $H_0$, de modo que el investigador pueda llegar a una conclusión para cualquier nivel de significancia especificado.
La definición formal del p-valor es la siguiente:

| |
|---|
| El ***valor p*** es el nivel de significancia más pequeño que conduce al rechazo de la hipótesis nula $H_0$ |

Para las pruebas de distribuciones normales presentadas hasta el momento, es sencillo calcular el p-valor.
Si $z_0$ es el valor calculado del estadístico de prueba *Z*, entonces el p-valor es

a) si las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu \neq \mu_0$

$$p-valor = P(|Z| > |z_0|) = 1 - P(|Z| < |z_0|) = 1 - [\Phi(|z_0|) - \Phi(-|z_0|)] = 1 - [2\Phi(|z_0|) - 1] = 2[1 - \Phi(|z_0|)]$$

b) si las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu > \mu_0$

$$p-valor = P(Z > z_0) = 1 - P(Z \le z_0) = 1 - \Phi(z_0)$$

c) si las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu < \mu_0$

$$p-valor = P(Z < z_0) = \Phi(z_0)$$

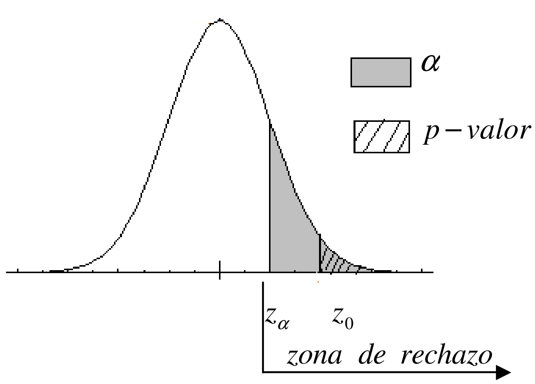
Un p-valor muy chico significa mucha evidencia en contra de $H_0$**;** un p-valor alto significa que no hay evidencia en contra $H_0$

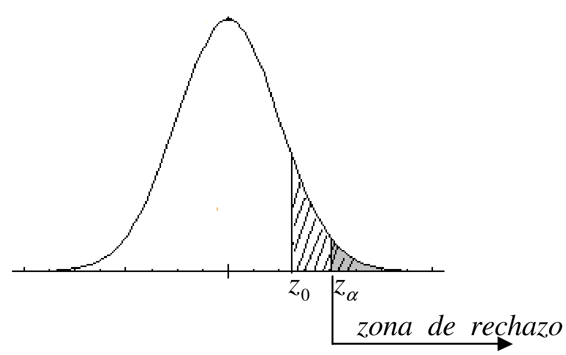
Notar que:

Si $\alpha < p-valor$ entonces se acepta $H_0$ con nivel de significancia $\alpha$

Si $\alpha > p-valor$ entonces se rechaza $H_0$ con nivel de significancia $\alpha$

Esto se ilustra en las siguientes figuras:




<u>Ejemplos</u>:

1- En el ejemplo anteúltimo referido al porcentaje deseado de $SiO_2$ en cierto tipo de cemento aluminoso las hipótesis eran: $H_0: \mu = 5.5$ contra $H_1: \mu \neq 5.5$; y el estadístico de prueba tomó el valor $z_0 = 0.333333 < 2.575 = z_{\frac{0.01}{2}}$; por lo tanto se aceptaba $H_0$.

En esta caso $p-valor = P(|Z| > |z_0|) = 2[1-\Phi(|z_0|)] = 2[1-\Phi(0.33333)] = 2[1-0.62930] = 0.7414$

***Como el p-valor es muy alto no hay evidencia en contra*** $H_0$. Se necesitaría tomar un valor de $\alpha$ mayor a 0.7414 para rechazar $H_0$.

2- En el último ejemplo, sobre la duración, en horas, de un foco de 75 watts, las hipótesis eran $H_0 : \mu = 1000$ contra $H_1 : \mu > 1000$; y el estadístico $Z$ tomó el valor $z_0 = 7.1554 > 1.645 = z_{0.05}$; por lo tanto se rechazaba $H_0$.

En este caso

$p-valor = P(Z > z_0) = 1-\Phi(z_0) = 1-\Phi(7.1554) \approx 0$

Como el p-valor es casi cero hay mucha evidencia en contra de $H_0$. Prácticamente para ningún valor de $\alpha$ se acepta $H_0$

## Error de tipo II y selección del tamaño de la muestra

En la prueba de hipótesis el investigador selecciona directamente la probabilidad del error de tipo I. Sin embargo, la probabilidad $\beta$ de cometer error de tipo II depende del tamaño de la muestra y del valor verdadero del parámetro desconocido.

Supongamos las hipótesis

$$H_0 : \mu = \mu_0 \quad \text{contra} \quad H_1 : \mu \neq \mu_0$$

Entonces si anotamos con $\mu$ al valor verdadero del parámetro

$$\beta = P(aceptar\ \ H_0 / H_0\ \ es\ \ falsa) = P\left( \left| \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \right| \leq z_{\frac{\alpha}{2}} \Big/ \mu \neq \mu_0 \right)$$

Como la hipótesis nula es falsa, entonces $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ ***no tiene distribución*** $N(0,1)$

Por lo tanto hacemos lo siguiente:

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X}-\mu+\mu-\mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} + \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$$ ; y ahora como $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ pues se estandarizó a $\overline{X}$ con el verdadero $\mu$, entonces

$$\beta = P\left( \left| \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \right| \leq z_{\frac{\alpha}{2}} \Big/ \mu \neq \mu_0 \right) = P\left( -z_{\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}} \Big/ \mu \neq \mu_0 \right) =$$

$$= P\left( -z_{\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} + \frac{\mu-\mu_0}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}} \Big/ \mu \neq \mu_0 \right) = P\left( -z_{\frac{\alpha}{2}} - \frac{(\mu-\mu_0)}{\sigma/\sqrt{n}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}} - \frac{(\mu-\mu_0)}{\sigma/\sqrt{n}} \right) =$$

$$=\Phi\left(z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma/\sqrt{n}}\right)-\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma/\sqrt{n}}\right)=\Phi\left(z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)-\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)$$

En consecuencia

> Si las hipótesis son $H_0:\mu=\mu_0$ contra $H_1:\mu\neq\mu_0$, entonces
> $$\beta(\mu)=\Phi\left(z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)-\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)$$

Para un valor específico de $\mu$ y un valor de $\alpha$ dado, podemos preguntarnos qué tamaño de muestra se necesita para que $\beta$ sea menor que un valor dado en particular $\beta_0$.

Por ejemplo si $\mu-\mu_0>0$ entonces podemos aproximar $\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)\approx 0$, y planteamos que

$\beta(\mu)=\Phi\left(z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)<\beta_0$ . Buscamos en la tabla de la $N(0,1)$ para qué $z$ se cumple que $\Phi(z)=\beta_0$ , lo anotamos $-z_{\beta_0}$, y entonces podemos escribir

$$z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}<-z_{\beta_0}\quad\Rightarrow\quad z_{\frac{\alpha}{2}}+z_{\beta_0}<\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\quad\Rightarrow\quad n>\frac{\left(z_{\frac{\alpha}{2}}+z_{\beta_0}\right)^2\sigma^2}{(\mu-\mu_0)^2}$$

En el caso de ser $\mu-\mu_0<0$ entonces podemos aproximar $\Phi\left(z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)\approx 1$, y planteamos que

$\beta(\mu)=1-\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)<\beta_0$ . Es decir $1-\beta_0<\Phi\left(-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\right)$

Buscamos en la tabla de la $N(0,1)$ para qué $z$ se cumple que $\Phi(z)=1-\beta_0$ , lo anotamos $z_{\beta_0}$ , y entonces podemos escribir

$$-z_{\frac{\alpha}{2}}-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}>z_{\beta_0}\quad\Rightarrow\quad z_{\frac{\alpha}{2}}+z_{\beta_0}<-\frac{(\mu-\mu_0)}{\sigma}\sqrt{n}\quad\underbrace{\Rightarrow}_{\mu-\mu_0<0}\quad n>\frac{\left(z_{\frac{\alpha}{2}}+z_{\beta_0}\right)^2\sigma^2}{(\mu-\mu_0)^2}$$

En consecuencia queda la misma fórmula que la anterior
Por lo tanto

> Si las hipótesis son $H_0:\mu=\mu_0$ contra $H_1:\mu\neq\mu_0$, entonces
> $$n>\frac{\left(z_{\alpha/2}+z_{\beta_0}\right)^2\sigma^2}{(\mu-\mu_0)^2}$$

En forma análoga se pude probar que si las hipótesis son

$$H_0 : \mu = \mu_0 \quad \text{contra} \quad H_1 : \mu > \mu_0$$

Entonces

$$\beta = P(aceptar\ \ H_0/H_0\ \ es\ \ falsa) = P\left( \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \le z_\alpha \Big/ \mu \ne \mu_0 \right) =$$

$$= P\left( \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \le z_\alpha \Big/ \mu \ne \mu_0 \right) = P\left( \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z_\alpha - \frac{(\mu - \mu_0)}{\sigma/\sqrt{n}} \right) = \Phi\left( z_\alpha - \frac{(\mu - \mu_0)}{\sigma/\sqrt{n}} \right) = \Phi\left( z_\alpha - \frac{(\mu - \mu_0)}{\sigma}\sqrt{n} \right)$$

Entonces

| |
|---|
| Si las hipótesis son : $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$ entonces<br>$\beta(\mu) = \Phi\left( z_\alpha - \frac{(\mu - \mu_0)}{\sigma}\sqrt{n} \right)$ |

Y si tenemos las hipótesis $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$

$$\beta = P(aceptar\ \ H_0/H_0\ \ es\ \ falsa) = P\left( \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \ge -z_\alpha \Big/ \mu \ne \mu_0 \right) =$$

$$= P\left( \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \ge -z_\alpha \Big/ \mu \ne \mu_0 \right) = P\left( \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ge -z_\alpha - \frac{(\mu - \mu_0)}{\sigma/\sqrt{n}} \right) = 1 - \Phi\left( -z_\alpha - \frac{(\mu - \mu_0)}{\sigma}\sqrt{n} \right)$$

Entonces

| |
|---|
| Si las hipótesis son : $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$ entonces<br>$\beta(\mu) = 1 - \Phi\left( -z_\alpha - \frac{(\mu - \mu_0)}{\sigma}\sqrt{n} \right)$ |

Y además con una deducción análoga al caso de alternativa bilateral:

| |
|---|
| Si las hipótesis son $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$ , (o $H_1 : \mu > \mu_0$ ) entonces<br>$n > \frac{(z_\alpha + z_{\beta_0})^2 \sigma^2}{(\mu - \mu_0)^2}$ |

Ejemplos:

1- En el ejemplo referido al porcentaje deseado de $SiO_2$ en cierto tipo de cemento aluminoso las hipótesis eran: $H_0: \mu = 5.5$ contra $H_1: \mu \neq 5.5$; y el estadístico de prueba tomó el valor $z_0 = 0.333333 < 2.575 = z_{\frac{0.01}{2}}$; por lo tanto se aceptaba $H_0$. Teníamos $n = 16$ y $\sigma = 3$

Si el verdadero promedio de porcentaje es $\mu = 5.6$ y se realiza una prueba de nivel $\alpha = 0.01$ con base en $n$ = 16, ¿cuál es la probabilidad de detectar esta desviación?

¿Qué valor de $n$ se requiere para satisfacer $\alpha = 0.01$ y $\beta(5.6) = 0.01$?

Solución:

La probabilidad de detectar la desviación es la potencia del test cuando $\mu = 5.6$, es decir

$$\pi(5.6) = P\left( rechazar \quad H_0 / H_0 \quad es \quad falsa \right) = 1 - \beta(5.6)$$

Como estamos con hipótesis alternativa bilateral, calculamos

$$\beta(5.6) = \Phi\left( z_{\frac{\alpha}{2}} - \frac{(5.6 - \mu_0)}{\sigma}\sqrt{n} \right) - \Phi\left( -z_{\frac{\alpha}{2}} - \frac{(5.6 - \mu_0)}{\sigma}\sqrt{n} \right) =$$

$$= \Phi\left( 2.575 - \frac{(5.6 - 5.5)}{3}\sqrt{16} \right) - \Phi\left( -2.575 - \frac{(5.6 - 5.5)}{3}\sqrt{16} \right) = \Phi(2.441) - \Phi(-2.708) =$$

$$= 0.99266 - (1 - 0.99664) = 0.9893 \qquad \Rightarrow \qquad \pi(5.6) = 0.0107$$

Ahora se quiere hallar $n$ tal que $\beta(5.6) = 0.01$, como el test es bilateral podemos usar directamente la fórmula con $z_{\beta_0} = z_{0.01} = 2.33$

$$n > \frac{\left( z_{\frac{\alpha}{2}} + z_{\beta_0} \right)^2 \sigma^2}{(\mu - \mu_0)^2} = \frac{(2.575 + 2.33)^2 3^2}{(5.6 - 5.5)^2} = 21653.1225 \qquad \Rightarrow \qquad n \geq 21654$$

2- En el último ejemplo, sobre la duración, en horas, de un foco de 75 watts, las hipótesis eran $H_0: \mu = 1000$ contra $H_1: \mu > 1000$; y el estadístico $Z$ tomó el valor $z_0 = 7.1554 > 1.645 = z_{0.05}$; por lo tanto se rechazaba $H_0$.

En este caso $\sigma = 25$ y $n = 20$

Si la verdadera duración promedio del foco es 1050 horas, ¿cuál es la probabilidad de error de tipo II para la prueba?

¿Qué tamaño de muestra es necesario para asegurar que el error de tipo II no es mayor que 0.10 si la duración promedio verdadera del foco es 1025 hs. ?

Solución:

Como las hipótesis son $H_0: \mu = 1000$ contra $H_1: \mu > 1000$ entonces

$$\beta(\mu) = \Phi\left( z_\alpha - \frac{(\mu - \mu_0)}{\sigma}\sqrt{n} \right) = \Phi\left( 1.645 - \frac{(1050 - 1000)}{25}\sqrt{20} \right) = \Phi(-7.29927) \neq 0$$

Para hallar $n$ tal que $\beta(1025) \leq 0.1$ aplicamos la fórmula con $z_{\beta_0} = z_{0.1} = 1.285$

$$n > \frac{\left(z_\alpha + z_{\beta_0}\right)^2 \sigma^2}{\left(\mu - \mu_0\right)^2} = \frac{(1.645 + 1.285)^2\, 25^2}{(1025 - 1000)^2} = 8.584 \quad \Rightarrow \quad n \geq 9$$

## Relación entre test de hipótesis e intervalos de confianza

Existe una estrecha relación entre la prueba de hipótesis bilateral sobre un parámetro $\mu$ y el intervalo de confianza de nivel $1-\alpha$ para $\mu$.
Específicamente supongamos que tenemos las hipótesis

$$H_0 : \mu = \mu_0 \quad \text{contra} \quad H_1 : \mu \neq \mu_0$$

La regla de decisión es

$$\begin{cases} rechazar \;\; H_0 \;\; si \;\; \left| \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \right| > z_{\frac{\alpha}{2}} \\ aceptar \;\; H_0 \;\; si \;\; \left| \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \right| \leq z_{\frac{\alpha}{2}} \end{cases}$$

Aceptar $H_0$ si $\left| \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \right| \leq z_{\frac{\alpha}{2}}$ es equivalente a: aceptar $H_0$ si $-z_{\frac{\alpha}{2}} \leq \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \leq z_{\frac{\alpha}{2}}$ ; y esto es a su vez equivalente, despejando $\mu_0$, a:

aceptar $H_0$ si $\overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \leq \mu_0 \leq \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n}$; es decir si $\mu_0 \in \left[ \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} ; \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \right]$

Pero resulta que $\left[ \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} ; \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \right]$ es el intervalo de confianza que se construiría para el verdadero parámetro $\mu$ de nivel $1-\alpha$.

Por lo tanto la regla de decisión queda:

$$\begin{cases} rechazar \;\; H_0 \;\; si \;\; \mu_0 \notin \left[ \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} ; \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \right] \\ aceptar \;\; H_0 \;\; si \;\; \mu_0 \in \left[ \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} ; \overline{X} - z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \right] \end{cases}$$

<u>Ejemplo</u>:
En el ejemplo referido al porcentaje deseado de $SiO_2$ en cierto tipo de cemento aluminoso las hipótesis eran: $H_0 : \mu = 5.5$ contra $H_1 : \mu \neq 5.5$;
y teníamos $n = 16$; $\sigma = 3$; un promedio muestral $\bar{x} = 5.25$

Como $\alpha = 0.01$ entonces $z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$

Construimos un intervalo de confianza de nivel $1-\alpha = 1-0.01 = 0.99$

$$\left[ \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] = \left[ 5.25 - 2.575 \frac{3}{\sqrt{16}} ; 5.25 + 2.575 \frac{3}{\sqrt{16}} \right] = [3.31875;\ 7.18125]$$

Entonces la regla de decisión es:

$$\begin{cases} rechazar \ \ H_0 \ \ si \ \ 5.5 \notin [3.31875;\ 7.18125] \\ aceptar \ \ H_0 \ \ si \ \ 5.5 \in [3.31875;\ 7.18125] \end{cases}$$

Como $5.5 \in [3.31875;\ 7.18125]$, entonces se acepta $H_0$.

### 9.3 – <u>Prueba de hipótesis sobre la media, varianza desconocida para muestras grandes</u>

Hasta ahora se ha desarrollado el procedimiento de test de hipótesis para la hipótesis nula $H_0: \mu = \mu_0$ suponiendo que $\sigma^2$ es conocida, pero en la mayoría de las situaciones prácticas $\sigma^2$ es desconocida. En general si $n \geq 30$, entonces la varianza muestral $S^2$ está próxima a $\sigma^2$ en la mayor parte de las muestras, de modo que es posible sustituir $S^2$ por $\sigma^2$. Es decir el estadístico

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \approx N(0,1) \quad \text{aproximadamente, si } n \geq 30 \text{ si } H_0: \mu = \mu_0$$

Además, si no podemos decir que la muestra aleatoria proviene de una población normal, sea $\sigma^2$ conocida o no, por T.C.L. los estadísticos

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \approx N(0,1) \quad \text{aproximadamente, si } n \geq 30 \text{ si } H_0: \mu = \mu_0$$

Y

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \approx N(0,1) \quad \text{aproximadamente, si } n \geq 30 \text{ si } H_0: \mu = \mu_0$$

Las pruebas de hipótesis tendrán entonces un nivel de significancia ***aproximadamente de*** $\alpha$

<u>Ejemplo</u>:
Un inspector midió el volumen de llenado de una muestra aleatoria de 100 latas de jugo cuya etiqueta afirmaba que contenían 12 oz. La muestra tenía una media de volumen de 11.98 oz y desviación estándar de 0.19 oz. Sea $\mu$ la verdadera media del volumen de llenado para todas las latas de jugo recientemente llenadas con esta máquina. El inspector probará $H_0: \mu = 12$ contra $H_1: \mu \neq 12$

a) Determinar el p-valor
b) ¿Piensa que es factible que la media del volumen de llenado es de 12 oz?

Solución:
La v.a. de interés sería $X$: "volumen de llenado de una lata tomada al azar"
No se especifica ninguna distribución para $X$. Anotamos $E(X)=\mu$ y $V(X)=\sigma^2$, ambas desconocidas.
Se toma una muestra de $n=100$ latas y se obtiene $\bar{x}=11.98$ y $s=0.19$
Las hipótesis son $H_0:\mu=12$ contra $H_1:\mu\neq 12$
El estadístico de prueba es

$$Z=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}=\frac{\overline{X}-12}{S/\sqrt{100}} \quad \text{y} \quad \text{si } H_0:\mu=12 \text{ es verdadera entonces } Z\approx N(0,1)$$

El estadístico $Z$ toma el valor $z_0=\dfrac{11.98-12}{0.19/\sqrt{100}}=-1.0526$

Como la hipótesis alternativa es bilateral entonces
$p-valor=P(|Z|>|z_0|)\approx 2[1-\Phi(1.0526)]=2[1-0.85314]=0.29372$
Como el p-valor es mayor que 0.05 se considera que ***no hay evidencia en contra de*** $H_0:\mu=12$
Por lo tanto ***es factible que la media del volumen de llenado sea de 12 oz***

## 9.4 – Prueba de hipótesis sobre la media de una distribución normal, varianza desconocida

Cuando se prueban hipótesis sobre la media $\mu$ de una población cuando $\sigma^2$ es desconocida es posible utilizar los procedimientos de prueba dados anteriormente siempre y cuando el tamaño de la muestra sea grande ($n\geq 30$). Estos procedimientos son aproximadamente válidos sin importar si la población de interés es normal o no. Pero si la muestra es pequeña y $\sigma^2$ es desconocida debe suponerse que la distribución de la variable de interés es normal.
Específicamente, supongamos que la v.a. de interés tiene distribución $N(\mu,\sigma^2)$ donde $\mu$ y $\sigma^2$ son desconocidas.
Supongamos las hipótesis $H_0:\mu=\mu_0$ contra $H_1:\mu\neq\mu_0$
Sea $X_1;X_2,...,X_n$ una muestra aleatoria de tamaño $n$ de la v.a. $X$ y sean $\overline{X}$ y $S^2$ la media y la varianza muestrales respectivamente.
El procedimiento se basa en el estadístico

$$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$$

El cual, ***si la hipótesis nula es verdadera***, tiene distribución ***Student con n-1 grados de libertad***.
Entonces, para un nivel $\alpha$ prefijado, la regla de decisión es

$$\begin{cases} rechazar\ \ H_0\ \ si\ \ |T|>t_{\frac{\alpha}{2},n-1} \\ aceptar\ \ H_0\ \ si\ \ |T|\leq t_{\frac{\alpha}{2},n-1} \end{cases} \quad \text{es decir} \quad \begin{cases} rechazar\ \ H_0\ \ si\ \ \left|\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|>t_{\frac{\alpha}{2},n-1} \\ aceptar\ \ H_0\ \ si\ \ \left|\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|\leq t_{\frac{\alpha}{2},n-1} \end{cases}$$

La lógica sigue siendo la misma, si el estadístico de prueba toma un valor inusual, entonces se considera que hay evidencia en contra $H_0$ y se rechaza la hipótesis nula. Como ahora la distribución del estadístico es Student, nos fijamos si *T* toma un valor $t_0$ en las colas de la distribución Student con *n-1* grados de libertad.

Si la alternativa es $H_1: \mu > \mu_0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; T > t_{\alpha,n-1} \\ aceptar \;\; H_0 \;\; si \;\; T \le t_{\alpha,n-1} \end{cases}$

Si la alternativa es $H_1: \mu < \mu_0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; T < -t_{\alpha,n-1} \\ aceptar \;\; H_0 \;\; si \;\; T \ge -t_{\alpha,n-1} \end{cases}$

Ejemplo:
Antes de que una sustancia se pueda considerar segura para enterrarse como residuo se deben caracterizar sus propiedades químicas. Se toman 6 muestras de lodo de una planta de tratamiento de agua residual en una región y se les mide el pH obteniéndose una media muestral de 6.68 y una desviación estándar muestral de 0.20. ¿Se puede concluir que la media del pH es menor que 7.0? Utilizar $\alpha = 0.05$ y suponer que la muestra fue tomada de una población normal.

Solución:
La v.a. de interés es *X*: "pH de una muestra de lodo tomada al azar"
Asumimos que *X* tiene distribución $N(\mu, \sigma^2)$
Las hipótesis serían $H_0: \mu = 7.0$ contra $H_1: \mu < 7.0$
El estadístico de prueba es $T = \dfrac{\overline{X} - 7.0}{S/\sqrt{6}}$ y toma el valor $t_0 = \dfrac{6.68 - 7.0}{0.20/\sqrt{6}} = -3.919$
Buscamos en la tabla de la distribución Student $t_{\alpha,n-1} = t_{0.05,5} = 2.015$
Entonces como $t_0 = -3.919 < -t_{\alpha,n-1} = -t_{0.05,5} = -2.015$ se rechaza $H_0$, por lo tanto hay evidencia que $\mu < 7.0$

## P-valor de un test *t*

En este caso el cálculo del p- valor se realiza considerando:
Si $t_0$ es el valor calculado del estadístico de prueba *T*, entonces el p-valor es

a) las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu \ne \mu_0$
$p-valor = P(|T| > |t_0|) = 1 - P(|T| \le |t_0|) = 2(1 - P(T \le t_0))$

b) las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu > \mu_0$
$p-valor = P(T > t_0) = 1 - P(T \le t_0)$

c) las hipótesis son $H_0: \mu = \mu_0$ contra $H_1: \mu < \mu_0$
$p-valor = P(T \le t_0)$

Para calcular el p-valor en una prueba t nos encontramos con la dificultad que las tablas de la Student no son completas, por lo tanto en algunas ocasiones se deberá ***acotar*** el p-valor
En el ejemplo anterior para calcular el p-valor de la prueba como es un test con alternativa unilateral
$p-valor = P(T \le t_0) = P(T \le -3.919)$

Buscamos en la tabla de la distribución Student la fila donde figuran $\nu = 5$ grados de libertad y vemos que el valor 3.919 no está tabulado.
Pero $3.365 < 3.919 < 4.032$, y $P(T_5 > 3.365) = 0.01$ y $P(T_5 > 4.032) = 0.005$
Por lo tanto $0.005 < P(T_5 > 3.919) < 0.01$, es decir
$0.005 < p - valor = P(T_5 < -3.919) < 0.01$
Podemos deducir que existe evidencia de que la media del pH es menor que 0.7

## 9.5 – Prueba de hipótesis sobre la diferencia de dos medias, varianzas conocidas

Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:
$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$ y suponemos que las varianzas $\sigma_1^2$ y $\sigma_2^2$ son conocidas.

Sean además
$(X_{11}, X_{12}, ..., X_{1n_1})$ una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, ..., X_{2n_2})$ una muestra aleatoria de tamaño $n_2$ de $X_2$.

El interés recae en probar que $\mu_1 - \mu_2 = \Delta_0$ donde $\Delta_0$ es un valor fijado, por ejemplo si $\Delta_0 = 0$ entonces se querrá probar que $\mu_1 - \mu_2 = 0$ es decir que las medias son iguales.
Ya sabemos que bajo las suposiciones anteriores

$$\begin{cases} \overline{X}_1 = \dfrac{1}{n_1} \sum_{i=1}^{n_1} X_{1i} \sim N\left( \mu_1, \dfrac{\sigma_1^2}{n_1} \right) \\ \overline{X}_2 = \dfrac{1}{n_2} \sum_{i=1}^{n_2} X_{2i} \sim N\left( \mu_2, \dfrac{\sigma_2^2}{n_2} \right) \end{cases}$$

Y además

$$\overline{X}_1 - \overline{X}_2 \sim \mathrm{N}\left( \mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right).$$

Por lo tanto

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1),$$ es decir, tiene distribución normal estandarizada.

Si consideramos las hipótesis

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{contra} \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$$

Entonces usamos como estadístico de prueba a $Z=\dfrac{\overline{X}_1-\overline{X}_2-\Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$

Y $\quad Z=\dfrac{\overline{X}_1-\overline{X}_2-\Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$ si $\quad H_0:\mu_1-\mu_2=\Delta_0$ es verdadera

Por lo tanto la regla de decisión será

$$\begin{cases} rechazar\ \ H_0\ \ si\ \ |Z|>z_{\frac{\alpha}{2}} \\ aceptar\ \ H_0\ \ si\ \ |Z|\le z_{\frac{\alpha}{2}} \end{cases} \quad \text{donde} \quad Z=\frac{\overline{X}_1-\overline{X}_2-\Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$$

Si $H_1:\mu_1-\mu_2>\Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar\ \ H_0\ \ si\ \ Z>z_{\alpha} \\ aceptar\ \ H_0\ \ si\ \ Z\le z_{\alpha} \end{cases}$

Si $H_1:\mu_1-\mu_2<\Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar\ \ H_0\ \ si\ \ Z<-z_{\alpha} \\ aceptar\ \ H_0\ \ si\ \ Z\ge -z_{\alpha} \end{cases}$

Ejemplos:

1- Un diseñador de productos está interesado en reducir el tiempo de secado de una pintura tapaporos. Se prueban dos fórmulas de pintura. La fórmula 1 tiene el contenido químico estándar, y la fórmula 2 tiene un nuevo ingrediente secante que debe reducir el tiempo de secado. De la experiencia se sabe que la desviación estándar del tiempo de secado es 8 minutos, y esta variabilidad no debe verse afectada por la adición del nuevo ingrediente. Se pintan 10 especímenes con la fórmula 1 y otros 10 con la fórmula 2. los tiempos promedio de secado muestrales fueron $\overline{x}_1=121$ minutos y $\overline{x}_2=112$ minutos respectivamente.

¿A qué conclusiones debe llegar el diseñador del producto sobre la eficacia del nuevo ingrediente utilizando $\alpha=0.05$?

Solución:

Aquí las hipótesis son $\quad H_0:\mu_1-\mu_2=0 \quad$ contra $\quad H_1:\mu_1-\mu_2>0$

El estadístico de prueba es $Z=\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{\dfrac{8^2}{10}+\dfrac{8^2}{10}}}$ y toma el valor $\quad z_0=\dfrac{121-112}{\sqrt{\dfrac{8^2}{10}+\dfrac{8^2}{10}}}=2.52$

Buscamos en la tabla de la normal estándar $z_{\alpha}=z_{0.05}=1.645$

Como $z_0 = 2.52 > z_\alpha = z_{0.05} = 1.645$ se rechaza $H_0$ al nivel 0.05 y se concluye que el nuevo ingrediente disminuye el tiempo de secado.

***El cálculo del p-valor y la deducción de $\beta$ la probabilidad de cometer error de tipo II se obtienen de manera análoga a los casos anteriores.*** Por ejemplo para la alternativa bilateral la expresión para $\beta$ es la siguiente donde anotamos $\mu_1 - \mu_2 - \Delta = \Delta_0 - \Delta = \delta$

$$\beta = P\left(aceptar\ H_0 / H_0\ es\ falsa\right) = \Phi\left( z_{\frac{\alpha}{2}} - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) - \Phi\left( -z_{\frac{\alpha}{2}} - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

En el ejemplo anterior el $p-valor = P(Z > z_0) = P(Z > 2.52) = 1 - \Phi(2.52) = 0 - 0059$

También es posible obtener fórmulas para el tamaño de la muestra necesario para obtener una $\beta$ específica para una diferencia dada en las medias $\mu_1 - \mu_2 - \Delta = \Delta_0 - \Delta = \delta$ y $\alpha$. Si asumimos que $n_1 = n_2 = n$ entonces

| | | |
|---|---|---|
| Para $H_1 : \mu_1 - \mu_2 \neq \Delta_0$ | es | $n > \dfrac{\left(z_{\alpha/2} + z_{\beta_0}\right)^2 \left(\sigma_1{}^2 + \sigma_2{}^2\right)}{\delta^2}$ |
| Para $H_1 : \mu_1 - \mu_2 > \Delta_0$ o $H_1 : \mu_1 - \mu_2 < \Delta_0$ | es | $n > \dfrac{\left(z_\alpha + z_{\beta_0}\right)^2 \left(\sigma_1{}^2 + \sigma_2{}^2\right)}{\delta^2}$ |

## 9.6 – Prueba de hipótesis sobre la diferencia de dos medias, varianzas desconocidas

**Caso 1:** $\sigma_1^2 \neq \sigma_2^2$

Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$ y las varianzas $\sigma_1^2$ y $\sigma_2^2$ son ***desconocidas*** .

y además

$(X_{11}, X_{12}, ..., X_{1n_1})$ es una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, ..., X_{2n_2})$ es una muestra aleatoria de tamaño $n_2$ de $X_2$ .

Si las muestras aleatorias se toma de una distribución normal, donde $\sigma_1$ y $\sigma_2$ son ***desconocidos***, $n_1 \geq 30$ y $n_2 \geq 30$, entonces se puede probar que al reemplazar $\sigma_1$ por $S_1$ y $\sigma_2$ por $S_2$, el estadístico

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_1}}} \approx N(0,1). \quad \textbf{\textit{aproximadamente}}$$

Por lo tanto si anotamos $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_1}}}$ valen las reglas de decisión vistas en la sección anterior, con la diferencia que el nivel de significancia del test será ***aproximadamente*** $1-\alpha$

Si ahora $n_1$ ***o*** $n_2$ ***no son mayores que 30***, entonces

$$T^* = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_1}}}$$

tiene distribución aproximadamente Student con $\nu$ grados de libertad bajo la hipótesis $H_0: \mu_1 - \mu_2 = \Delta_0$ donde

$$\nu = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^1/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$ si $\nu$ no es entero, ***se toma el entero más próximo a*** $\nu$

Por lo tanto, si las hipótesis son

$H_0: \mu_1 - \mu_2 = \Delta_0$ contra $H_1: \mu_1 - \mu_2 \neq \Delta_0$ entonces la regla de decisión es

$$\begin{cases} rechazar \quad H_0 \quad si \quad \left|T^*\right| > t_{\frac{\alpha}{2},\nu} \\ aceptar \quad H_0 \quad si \quad \left|T^*\right| \leq t_{\frac{\alpha}{2},\nu} \end{cases}$$

Si $H_1: \mu_1 - \mu_2 > \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad T^* > t_{\alpha,\nu} \\ aceptar \quad H_0 \quad si \quad T^* \leq t_{\alpha,\nu} \end{cases}$

Si $H_1: \mu_1 - \mu_2 < \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad T^* < -t_{\alpha,\nu} \\ aceptar \quad H_0 \quad si \quad T^* \geq -t_{\alpha,\nu} \end{cases}$

<u>Ejemplo</u>:
Un fabricante de monitores prueba dos diseños de microcircuitos para determinar si producen un flujo de corriente equivalente. El departamento de ingeniería ha obtenido los datos siguientes:

| | | | |
|---|---|---|---|
| *Diseño 1* | $n_1 = 15$ | $\bar{x}_1 = 24.2$ | $s_1^2 = 10$ |
| *Diseño 2* | $n_2 = 10$ | $\bar{x}_2 = 23.9$ | $s_2^2 = 20$ |

Con $\alpha = 0.10$ se desea determinar si existe alguna diferencia significativa en el flujo de corriente medio entre los dos diseños, donde se supone que las poblaciones son normales.

<u>Solución</u>:
Las variables aleatorias de interés son
$X_1$ : "flujo de corriente en diseño 1"
$X_2$ : "flujo de corriente en diseño 2"
Asumimos que $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ donde los parámetros son desconocidos
Las hipótesis serían $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_1 - \mu_2 \neq 0$
El estadístico de prueba es

$$|T^*| = \left| \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{15} + \frac{S_1^2}{10}}} \right| \text{ que en este caso toma el valor } t_0^* = \left| \frac{24.2 - 23.9}{\sqrt{\frac{10}{15} + \frac{20}{10}}} \right| = 0.18$$

Debemos buscar en la tabla de la distribución Student $t_{\frac{\alpha}{2},\nu} = t_{\frac{0.10}{2},\nu}$ entonces calculamos

$$\nu = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^1/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}} = \frac{(10/15 + 20/10)^2}{\frac{(10/15)^2}{15-1} + \frac{(20/10)^2}{10-1}} = 14.9333 \quad \Rightarrow \quad \nu = 15$$

Por lo tanto $t_{\frac{\alpha}{2},\nu} = t_{0.05,15} = 1.753$

Como $t_0^* = 0.18 < t_{0.05,15} = 1.753$ entonces se acepta $H_0 : \mu_1 - \mu_2 = 0$
No hay evidencia fuerte que las medias de los dos flujos de corriente sean diferentes.
Si calculamos el p-valor
$p - valor = P\left(|T^*| > t_0^*\right) = P\left(|T^*| > 0.18\right) > 0.40$

**<u>Caso 2:</u>** $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$$ y las varianzas $\sigma_1^2$ y $\sigma_2^2$ son ***desconocidas*** pero ***iguales***.

y además
$(X_{11}, X_{12}, ..., X_{1n_1})$ es una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, ..., X_{2n_2})$ es una muestra aleatoria de tamaño $n_2$ de $X_2$.

Sean $\overline{X}_1$ y $\overline{X}_2$ las medias muestrales y $S_1^2$ y $S_2^2$ las varianzas muestrales. Como $S_1^2$ y $S_2^2$ son los estimadores de la varianza común $\sigma^2$, entonces construimos un ***estimador combinado*** de $\sigma^2$. Este estimador es

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

Se puede comprobar que es un estimador insesgado de $\sigma^2$.
Ya vimos que se puede probar que el estadístico

$$T = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$ tiene distribución Student con $n_1+n_2-2$ grados de libertad

Por lo tanto, si las hipótesis son

$H_0: \mu_1 - \mu_2 = \Delta_0$ contra $H_1: \mu_1 - \mu_2 \neq \Delta_0$ entonces la regla de decisión es

$$\begin{cases} rechazar \quad H_0 \quad si \quad |T| > t_{\frac{\alpha}{2}, n_1+n_2-2} \\ aceptar \quad H_0 \quad si \quad |T| \leq t_{\frac{\alpha}{2}, n_1+n_2-2} \end{cases}$$

Si $H_1: \mu_1 - \mu_2 > \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad T > t_{\alpha, n_1+n_2-2} \\ aceptar \quad H_0 \quad si \quad T \leq t_{\alpha, n_1+n_2-2} \end{cases}$

Si $H_1: \mu_1 - \mu_2 < \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad T < -t_{\alpha, n_1+n_2-2} \\ aceptar \quad H_0 \quad si \quad T \geq -t_{\alpha, n_1+n_2-2} \end{cases}$

Ejemplo:
Se tienen las mediciones del nivel de hierro en la sangre de dos muestras de niños: un grupo de niños sanos y el otro padece fibrosis quística. Los datos obtenidos se dan en la siguiente tabla:

| | | | |
|---|---|---|---|
| *sanos* | $n_1 = 9$ | $\overline{x}_1 = 18.9$ | $s_1^2 = 5.9^2$ |
| *enfermos* | $n_2 = 13$ | $\overline{x}_2 = 11.9$ | $s_2^2 = 6.3^2$ |

Podemos asumir que las muestras provienen de poblaciones normales independientes con iguales varianzas.
Es de interés saber si las dos medias del nivel de hierro en sangre son iguales o distintas. Utilizar $\alpha = 0.05$

Solución:
Las variables de interés son
$X_1$: "nivel de hierro en sangre de un niño sano tomado al azar"
$X_2$: "nivel de hierro en sangre de un niño con fibrosis quística tomado al azar"

Asumimos que $X_1 \sim \mathrm{N}(\mu_1, \sigma^2)$ y $X_2 \sim \mathrm{N}(\mu_2, \sigma^2)$
Consideramos las hipótesis

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{contra} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

Para calcular el valor del estadístico de prueba, primero calculamos

$$S_p = \sqrt{S_p^2} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(9-1)5.9^2 + (13-1)6.3^2}{9 + 13 - 2}} = 6.14$$

El estadístico de prueba es $|T| = \left| \dfrac{\vec{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{9} + \frac{1}{13}}} \right|$ y toma el valor $t_0 = \left| \dfrac{18.9 - 11.9}{6.14\sqrt{\frac{1}{9} + \frac{1}{13}}} \right| = 2.63$

Buscamos en la tabla de la distribución Student $t_{\frac{\alpha}{2}, n_1 + n_2 - 2} = t_{0.025,20} = 2.086$

Como $t_0 = 2.63 > t_{0.025,20} = 2.086$ entonces se rechaza $H_0 : \mu_1 - \mu_2 = 0$

Si calculamos el p-valor de la prueba

$$p - valor = 2(1 - P(T < t_0)) = 2(1 - P(T < 2.63)) = 2P(T > 2.63)$$

Vemos de la tabla de la Student que $t_{0.01,20} = 2.528$ y $t_{0.005,20} = 2.845$ por lo tanto

$$2 \times 0.005 < p - valor = 2P(T > 2.63) < 2 \times 0.01 \quad \text{es decir} \quad 0.01 < p - valor < 0.02$$

## 9.7 – <u>Prueba de hipótesis sobre la diferencia de dos medias para datos de a pares</u>

Ya se vio el caso, cuando se habló de intervalos de confianza para una diferencia de medias, de datos dados de a pares, es decir $(X_{11}, X_{21}); (X_{12}, X_{22}); \ldots; (X_{1n_1}, X_{2n})$.
Las variables aleatorias $X_1$ y $X_2$ tienen medias $\mu_1$ y $\mu_2$ respectivamente.
Consideramos $D_j = X_{1j} - X_{2j}$ con $j = 1, 2, \ldots, n$.
Entonces

$$E(D_j) = E(X_{1j} - X_{2j}) = E(X_{1j}) - E(X_{2j}) = \mu_1 - \mu_2$$

y

$$V(D_j) = V(X_{1j} - X_{2j}) = V(X_{1j}) + V(X_{2j}) - 2Cov(X_{1j}, X_{2j}) = \sigma_1^2 + \sigma_2^2 - 2Cov(X_1, X_2)$$

Estimamos $E(D_j) = \mu_1 - \mu_2$ con $\overline{D} = \dfrac{1}{n}\sum_{j=1}^{n} D_j = \dfrac{1}{n}\sum_{j=1}^{n}(X_{1j} - X_{2j}) = \overline{X}_1 - \overline{X}_2$

En lugar de tratar de estimar la covarianza, estimamos la $V(D_j)$ con $S_D = \dfrac{1}{n-1}\sum_{j=1}^{n}(D_j - \overline{D})^2$

Anotamos $\mu_D = \mu_1 - \mu_2$ y $\sigma^2{}_D = V(D_j)$

Asumimos que $D_j \sim \mathrm{N}(\mu_D, \sigma_D{}^2)$ con $j = 1,2,...,n$

***Las variables aleatorias en pares diferentes son independientes, no lo son dentro de un mismo par***.
Para construir una regla de decisión nuevamente, consideramos el estadístico

$$T = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \quad \text{con distribución } t_{n-1}$$

Si tenemos las hipótesis

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{contra} \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$$

Entonces el estadístico de prueba es

$T = \dfrac{\overline{D} - \Delta_0}{S_D / \sqrt{n}}$ y tiene distribución $t_{n-1}$ si $H_0 : \mu_1 - \mu_2 = \Delta_0$ es verdadera

Por lo tanto, la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; |T| > t_{\frac{\alpha}{2}, n-1} \\ aceptar \;\; H_0 \;\; si \;\; |T| \leq t_{\frac{\alpha}{2}, n-1} \end{cases}$ donde $T = \dfrac{\overline{D} - \Delta_0}{S_D / \sqrt{n}}$

Si $H_1 : \mu_1 - \mu_2 > \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; T > t_{\alpha, n-1} \\ aceptar \;\; H_0 \;\; si \;\; T \leq t_{\alpha, n-1} \end{cases}$

Si $H_1 : \mu_1 - \mu_2 < \Delta_0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; T < -t_{\alpha, n-1} \\ aceptar \;\; H_0 \;\; si \;\; T \geq -t_{\alpha, n-1} \end{cases}$

Ejemplo:
Se comparan dos microprocesadores en una muestra de 6 códigos de puntos de referencia para determinar si hay una diferencia en la rapidez. Los tiempos en segundos utilizados para cada procesador en cada código están dados en la siguiente tabla:

| | **Código** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| **Procesador A** | 27.2 | 18.1 | 27.2 | 19.7 | 24.5 | 22.1 |
| **Procesador B** | 24.1 | 19.3 | 26.8 | 20.1 | 27.6 | 29.8 |

¿Puede concluir que las medias de la rapidez de ambos procesadores son diferentes con nivel de significancia 0.05?

Solución:
Las variables aleatorias de interés son
$X_1$ : "rapidez del procesador A en un código tomado al azar"
$X_2$ : "rapidez del procesador B en un código tomado al azar"
Como ambas variables se miden sobre un mismo código no podemos asumir que son independientes.

Las hipótesis son $H_0: \mu_1 - \mu_2 = 0$ contra $H_1: \mu_1 - \mu_2 \neq 0$
Necesitamos la muestra de las diferencias $D_j$:
3.1, -1.2; 0.4; -0.4; -3.1; -7.7
De esta muestra obtenemos $\overline{d} = -1.483333$ y $s_D = 3.66246$

Además $\alpha = 0.05 \rightarrow t_{\frac{\alpha}{2},n-1} = t_{0.025,5} = 2.571$

El estadístico de prueba es $|T| = \left| \dfrac{\overline{D}}{S_D / \sqrt{6}} \right|$ y toma el valor $t_0 = \left| \dfrac{-1.483333}{3.66246/\sqrt{6}} \right| = 0.99206$

Como $t_0 = 0.99206 < t_{\frac{\alpha}{2},n-1} = t_{0.025,5} = 2.571$ entonces se acepta la hipótesis nula. No hay evidencia de que las medias de la rapidez de ambos procesadores sean diferentes.

## 9.8 – Tests de hipótesis sobre la varianza

Supongamos que se desea probar la hipótesis de que la varianza de una población normal es igual a un valor específico, por ejemplo $\sigma_0^{\ 2}$.
Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria de tamaño $n$ de una v.a. $X$, donde $X \sim N(\mu, \sigma^2)$.
Tomamos como estimador puntual de $\sigma^2$ a $S^2 = \dfrac{1}{n-1} \sum_{1=1}^{n} \left( X_i - \overline{X} \right)^2$

Luego a partir de este estimador puntual construimos el estadístico $X = \dfrac{(n-1)S^2}{\sigma^2}$

Este estadístico contiene al parámetro desconocido a estimar $\sigma^2$ y ya sabemos que tiene una distribución llamada ***ji-cuadrado con n-1 grados de libertad***
Supongamos las hipótesis

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{contra} \quad H_1: \sigma^2 \neq \sigma_0^2$$

Tomamos como estadístico de prueba a

$$X = \frac{(n-1)S^2}{\sigma_0^{\ 2}} \quad \text{y si} \quad H_0: \sigma^2 = \sigma_0^2 \ \text{es verdadera , entonces} \quad X = \frac{(n-1)S^2}{\sigma_0^{\ 2}} \sim \chi^2_{n-1}$$

Nuevamente, el razonamiento es: si el estadístico $X$ que bajo $H_0: \sigma^2 = \sigma_0^2$ tiene distribución $\chi^2_{n-1}$ toma un valor "inusual", se considera que hay evidencia en contra $H_0$
Recordar que la distribución $\chi^2_{n-1}$ es asimétrica. Entonces la regla de decisión es

$$\begin{cases} recahzar \ \ H_0 \ \ si \ \ X > \chi^2_{\frac{\alpha}{2},n-1} \ \ ó \ \ X < \chi^2_{1-\frac{\alpha}{2},n-1} \\ aceptar \ \ H_0 \ \ si \ \ \chi^2_{1-\frac{\alpha}{2},n-1} \leq X \leq \chi^2_{\frac{\alpha}{2},n-1} \end{cases} \quad \text{donde} \quad X = \frac{(n-1)S^2}{\sigma_0^{\ 2}}$$

Si $H_1 : \sigma^2 > \sigma_0^2$ entonces la regla de decisión es $\begin{cases} recahzar \ \ H_0 \ \ si \ \ X > \chi^2_{\alpha,n-1} \\ aceptar \ \ H_0 \ \ si \ \ X \leq \chi^2_{\alpha,n-1} \end{cases}$

Si $H_1 : \sigma^2 < \sigma_0^2$ entonces la regla de decisión es $\begin{cases} recahzar \ \ H_0 \ \ si \ \ X < \chi^2_{1-\alpha,n-1} \\ aceptar \ \ H_0 \ \ si \ \ X \geq \chi^2_{1-\alpha,n-1} \end{cases}$

Para calcular el p-valor, si el estadístico *X* tomó el valor $x_0$, y teniendo en cuenta que no hay simetría en la distribución ji-cuadrado, hacemos:

Si $H_1 : \sigma^2 > \sigma_0^2$ entonces $p - valor = P(X > x_0)$
Si $H_1 : \sigma^2 < \sigma_0^2$ entonces $p - valor = P(X < x_0)$
Si $H_1 : \sigma^2 \neq \sigma_0^2$ entonces $p - valor = 2\min\Big( P(X < x_0) \ \ , P(X > x_0) \Big)$

<u>Ejemplo</u>:
Consideremos nuevamente el ejemplo visto en la sección de intervalos de confianza para la varianza sobre la máquina de llenado de botellas. Al tomar una muestra aleatoria de 20 botellas se obtiene una varianza muestral para el volumen de llenado de $s^2 = 0.0153\,\text{oz}^2$.
Si la varianza del volumen de llenado es mayor que 0.01 $\text{oz}^2$, entonces existe una proporción inaceptable de botellas que serán llenadas con una cantidad menor de líquido. ¿Existe evidencia en los datos muestrales que sugiera que el fabricante tiene un problema con el llenado de las botellas? Utilice $\alpha = 0.05$

<u>Solución</u>:
La variable de interés es *X*: "volumen de llenado de una botella tomada al azar"
Asumimos $X \sim N(\mu, \sigma^2)$
Los datos son $s^2 = 0.0153$ de una muestra de tamaño $n = 20$

Las hipótesis son $H_0 : \sigma^2 = 0.01$ contra $H_1 : \sigma^2 > 0.01$

$\alpha = 0.05 \quad \rightarrow \quad \chi^2_{\alpha,n-1} = \chi^2_{0.05,19} = 30.14$

El estadístico de prueba es $X = \dfrac{(n-1)S^2}{\sigma_0{}^2} = \dfrac{19 \times S^2}{0.01}$ y toma el valor

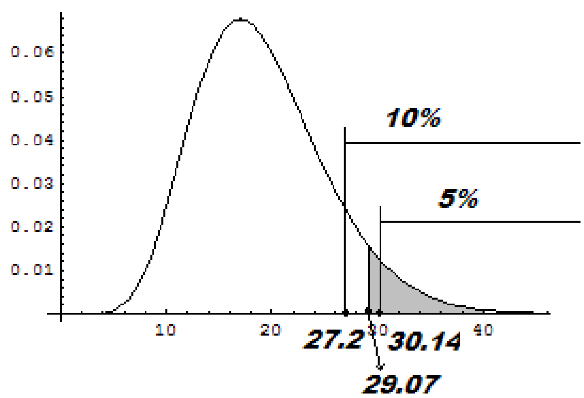
$$x_0 = \frac{19 \times S^2}{0.01} = \frac{19 \times 0.0153}{0.01} = 29.07$$

Como $x_0 = 29.07 < \ \chi^2_{0.05,19} = 30.14$ entonces no hay evidencia fuerte de que la varianza del volumen de llenado sea menor que 0.01
Para calcular el p-valor
$p - valor = P(X > x_0) = P(X > 29.07)$

Buscamos en la tabla de la distribución ji-cuadrado y vemos que en la fila con $\nu = 19$ no figura 29.07, pero 27.20 < 29.07 < 30.14, y además

$$\begin{cases} P(X > 27.20) = 0.10 \\ P(X > 30.14) = 0.05 \end{cases} \quad \Rightarrow \quad 0.05 < p - valor < 0.10$$

En la figura siguiente se ilustra la situación




**9.9 – <u>Tests de hipótesis sobre la igualdad de dos varianzas</u>**

Supongamos que tenemos interés en dos poblaciones normales independientes, donde las medias y las varianzas de la población son desconocidas. Se desea probar la hipótesis sobre la igualdad de las dos varianzas, específicamente:
Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases} \quad \text{y } \mu_1;\ \mu_2;\ \sigma_1^2 \text{ y } \sigma_2^2 \text{ son desconocidos}$$

y además

$(X_{11}, X_{12}, ..., X_{1n_1})$ es una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, ..., X_{2n_2})$ es una muestra aleatoria de tamaño $n_2$ de $X_2$.
Sean $S_1^2$ y $S_2^2$ las varianzas muestrales, $S_1^2$ y $S_2^2$ son los estimadores de $\sigma_1^2$ y $\sigma_2^2$ respectivamente.
Consideramos el estadístico

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$

Notar que $F$ contiene al parámetro de interés $\frac{\sigma_{21}^2}{\sigma_1^2}$ , pues $F = \frac{S_1^2 \times \sigma_2^2}{S_2^2 \times \sigma_1^2}$

Sabemos que $F$ tiene una distribución llamada Fisher con $n_1 - 1$ y $n_2 - 1$ grados de libertad.

Sean las hipótesis $H_0 : {\sigma_1}^2 = \sigma_2^2$ contra $H_1 : {\sigma_1}^2 \neq \sigma_2^2$

Tomamos como estadístico de prueba a $F = \frac{S_1^2}{S_2^2}$

Vemos que $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$ si $H_0 : {\sigma_1}^2 = \sigma_2^2$ es verdadera

Recordando que la distribución Fisher es asimétrica, la regla de decisión es

$$\begin{cases} recahzar \quad H_0 \quad si \quad F > f^2_{\frac{\alpha}{2},n_1-1,n_2-1} \quad ó \quad F < f^2_{1-\frac{\alpha}{2},n_1-1,n_2-1} \\ \quad aceptar \quad H_0 \quad si \quad f^2_{1-\frac{\alpha}{2},n_1-1,n_2-1} \leq F \leq f^2_{\frac{\alpha}{2},n_1-1,n_2-1} \end{cases}$$

Si $H_1 : {\sigma_1}^2 > \sigma_2^2$ entonces la regla de decisión es $\begin{cases} recahzar \quad H_0 \quad si \quad F > f^2_{\alpha,n_1-1,n_2-1} \\ aceptar \quad H_0 \quad si \quad F \leq f^2_{\alpha,n_1-1,n_2-1} \end{cases}$

Si $H_1 : {\sigma_1}^2 < \sigma_2^2$ entonces la regla de decisión es $\begin{cases} recahzar \quad H_0 \quad si \quad F < f^2_{1-\alpha,n_1-1,n_2-1} \\ aceptar \quad H_0 \quad si \quad F \geq f^2_{1-\alpha,n_1-1,n_2-1} \end{cases}$

Para calcular el p-valor, si el estadístico $F$ tomó el valor $f_0$, y teniendo en cuenta que no hay simetría en la distribución Fisher, hacemos:

Si $H_1 : {\sigma_1}^2 > \sigma_2^2$ entonces $p - valor = P(F > f_0)$

Si $H_1 : {\sigma_1}^2 < \sigma_2^2$ entonces $p - valor = P(F < f_0)$

Si $H_1 : {\sigma_1}^2 \neq \sigma_2^2$ entonces $p - valor = 2\min\left(P(F < f_0) \ , P(F > f_0)\right)$

<u>Ejemplo</u>:

En una serie de experimentos para determinar la tasa de absorción de ciertos pesticidas en la piel se aplicaron cantidades medidas de dos pesticidas a algunos especímenes de piel. Después de un tiempo se midieron las cantidades absorbidas (en $\mu g$ ). Para el pesticida A la varianza de las cantidades absorbidas en 6 muestras fue de 2.3; mientras que para el B la varianza de las cantidades absorbidas en 10 especímenes fue de 0.6. Suponga que para cada pesticida las cantidades absorbidas constituyen una muestra aleatoria de una población normal. ¿Se puede concluir que la varianza en la cantidad absorbida es mayor para el pesticida A que para el B? Utilizar $\alpha = 0.05$

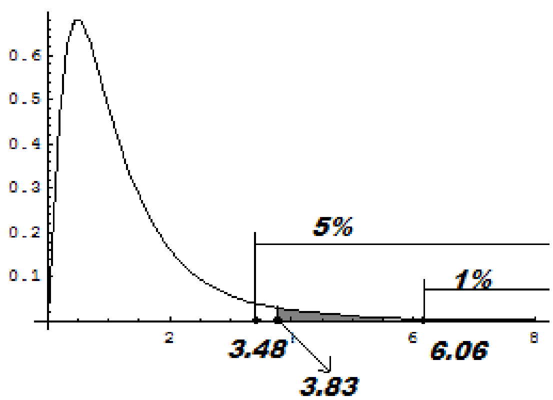
<u>Solución</u>:

Las variables aleatorias de interés son

$X_1$ : “cantidad absorbida de pesticida A en un espécimen de piel tomado al azar”
$X_2$ : “cantidad absorbida de pesticida B en un espécimen de piel tomado al azar”
Asumimos que $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$

Las hipótesis son $H_0 : \sigma_1^2 = \sigma_2^2$ contra $H_1 : \sigma_1^2 < \sigma_2^2$
Los datos son $s_1^2 = 2.3$ y $s_2^2 = 0.6$
$n_1 = 6$ ; $n_2 = 10$
El estadístico de prueba es $F = \dfrac{S_1^2}{S_2^2}$ y toma el valor $f_0 = \dfrac{2.3}{0.6} = 3.83$
Buscamos en la tabla de la distribución Fisher $f_{0.05,5,9} = 3.48$
Como $f_0 = \dfrac{2.3}{0.6} = 3.83 > 3.48 = f_{0.05,5,9}$ se rechaza $H_0 : \sigma_1^2 = \sigma_2^2$
Para saber cuánta evidencia hay contra la hipótesis nula, calculamos el p-valor
De la tabla de la Fisher vemos que $f_{0.05,5,9} = 3.48 < 3.83 < f_{0.01,5,9} = 6.06$
Por lo tanto $0.01 < p - valor < 0.05$

En la figura siguiente se ilustra la situación




**9.10 – <u>Tests de hipótesis sobre una proporción</u>**

En muchos problemas se tiene interés en una variable aleatoria que sigue una distribución binomial. Por ejemplo, un proceso de producción que fabrica artículos que son clasificados como aceptables o defectuosos. Lo más usual es modelar la ocurrencia de artículos defectuosos con la distribución binomial, donde el parámetro binomial $p$ representa la proporción de artículos defectuosos producidos.
En consecuencia, muchos problemas de decisión incluyen una prueba de hipótesis con respecto a $p$.
Consideremos las hipótesis

$$H_0 : p = p_0 \quad \text{contra} \quad H_1 : p \neq p_0$$

Supongamos que consideramos una muestra aleatoria $(X_1, X_2 ..., X_n)$ de tamaño $n$, donde $X_i$ tiene una distribución binomial con parámetros 1 y $p$: $X_i \sim B(1,p)$.

Ya sabemos que $X = X_1 + X_2 + ... + X_n$ , es una v.a. cuya distribución es binomial con parámetros $n$ y $p$: $X \sim B(n,p)$. De acuerdo con esto, la variable aleatoria $\hat{P}$ definida: $\hat{P} = \dfrac{X}{n}$ representa la proporción de individuos de la muestra que verifican la propiedad de interés.

Además

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p \text{ , y } V(\hat{P}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

Consideramos el estadístico de prueba

$$Z = \frac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

Si $H_0 : p = p_0$ es verdadera entonces $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \approx N(0,1)$ aproximadamente por T.C.L.

Por lo tanto la regla de decisión es

$$\begin{cases} rechazar \quad H_0 \quad si \quad |Z| > z_{\frac{\alpha}{2}} \\ aceptar \quad H_0 \quad si \quad |Z| \leq z_{\frac{\alpha}{2}} \end{cases} \qquad \text{donde} \qquad Z = \frac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

Si $H_1 : p > p_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad Z > z_{\alpha} \\ aceptar \quad H_0 \quad si \quad Z \leq z_{\alpha} \end{cases}$

Si $H_1 : p < p_0$ entonces la regla de decisión es $\begin{cases} rechazar \quad H_0 \quad si \quad Z < -z_{\alpha} \\ aceptar \quad H_0 \quad si \quad Z \geq -z_{\alpha} \end{cases}$

Observaciones:

1- La prueba descrita anteriormente requiere que la proporción muestral esté normalmente distribuida. Esta suposición estará justificada siempre que $np_0 > 10$ y $n(1-p_0) > 10$, donde $p_0$ es la proporción poblacional que se especificó en la hipótesis nula.

2- También se podía haber tomado como estadístico de prueba a $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$ donde $X \sim B(n,p)$

Ejemplo:

Un fabricante de semiconductores produce controladores que se emplean en aplicaciones de motores automovilísticos. El cliente requiere que la fracción de controladores defectuosos en uno de los pasos de manufactura críticos no sea mayor que 0.05, y que el fabricante demuestre esta característica del proceso de fabricación con este nivel de calidad, utilizando $\alpha = 0.05$. E fabricante de semiconductores toma una muestra aleatoria de 200 dispositivos y encuentra que 4 de ellos son defectuosos. ¿El fabricante puede demostrar al cliente la calidad del proceso?

Solución:
Sea la v.a. $X$: "número de controladores defectuosos en la muestra"
Entonces $X \sim B(200, p)$ donde $p$ es la proporción de controladores defectuosos en el proceso
Las hipótesis son $H_0: p = 0.05$ contra $H_1: p < 0.05$
Como $\alpha = 0.05$ entonces $-z_\alpha = -z_{0.05} = -1.645$

El estadístico de prueba es $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{\hat{P} - 0.05}{\sqrt{\dfrac{0.05(1-0.05)}{200}}}$ y toma el valor $z_0 = -1.95$

Como $z_0 = -1.95 < -z_\alpha = -z_{0.05} = -1.645$ entonces se rechaza $H_0$, y se concluye que la fracción de controladores defectuosos es menor que 0.05.
Calculamos el p-valor
$p - valor = P(Z < z_0) = P(Z < -1.95) = \Phi(-1.95) = 0.0256$

## Valor de $\beta$ y selección del tamaño de la muestra

Podemos obtener expresiones aproximadas para la probabilidad de cometer error de tipo II de manera análoga a las obtenidas para los test para la media

Si $H_1: p \neq p_0$ entonces

$$\beta(p) = P\left(aceptar \;\; H_0 \big/ H_0 \;\; es \;\; falsa\right) \approx$$

$$\approx \Phi\left(\frac{p_0 - p + z_{\frac{\alpha}{2}}\sqrt{\dfrac{p_0(1-p_0)}{n}}}{\sqrt{\dfrac{p(1-p)}{n}}}\right) - \Phi\left(\frac{p_0 - p + z_{\frac{\alpha}{2}}\sqrt{\dfrac{p_0(1-p_0)}{n}}}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

Si $H_1: p < p_0$ entonces

$$\beta(p) = P\left(aceptar \;\; H_0 \big/ H_0 \;\; es \;\; falsa\right) \approx 1 - \Phi\left(\frac{p_0 - p - z_\alpha\sqrt{\dfrac{p_0(1-p_0)}{n}}}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

Si $H_1: p > p_0$ entonces

$$\beta(p) = P\left(aceptar \;\; H_0 \big/ H_0 \;\; es \;\; falsa\right) \approx \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{\dfrac{p_0(1-p_0)}{n}}}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

Estas ecuaciones pueden resolverse para encontrar el tamaño aproximado de la muestra $n$ para que con un nivel de significancia de $\alpha$ la probabilidad de cometer error de tipo II sea menor o igual que un valor específico $\beta_0$. Las ecuaciones se deducen como en casos anteriores y son

Si $H_1 : p \neq p_0$ entonces $n \geq \left( \dfrac{z_{\frac{\alpha}{2}} \sqrt{p_0(1-p_0)} + z_{\beta_0} \sqrt{p(1-p)}}{p - p_0} \right)^2$

Si $H_1 : p < p_0$ ó $H_1 : p > p_0$ entonces $n \geq \left( \dfrac{z_{\alpha} \sqrt{p_0(1-p_0)} + z_{\beta_0} \sqrt{p(1-p)}}{p - p_0} \right)^2$

Ejemplo:
Volviendo al ejemplo anterior, supongamos que la verdadera proporción de componentes defectuosos en el proceso es $p = 0.03$, ¿cuál es el valor de $\beta$ si $n = 200$ y $\alpha = 0.05$?

Solución:
Ya que la alternativa es $H_1 : p < p_0$ aplicamos la fórmula

$$\beta(p) = P\left( aceptar \quad H_0 \big/ H_0 \quad es \quad falsa \right) \approx 1 - \Phi\left( \frac{p_0 - p - z_\alpha \sqrt{\dfrac{p_0(1-p_0)}{n}}}{\sqrt{\dfrac{p(1-p)}{n}}} \right) =$$

$$= 1 - \Phi\left( \frac{0.05 - 0.03 - 1.645\sqrt{\dfrac{0.05(1-0.05)}{n}}}{\sqrt{\dfrac{0.03(1-0.03)}{200}}} \right) = 1 - \Phi(-0.44) = 0.67$$

Como la probabilidad de aceptar que el proceso tiene la calidad deseada cuando en realidad $p = 0.03$ es bastante alta, podemos preguntar qué tamaño de muestra se necesita para que en el test anterior sea $\beta < 0.1$ si la verdadera proporción de defectuosos es $p = 0.03$. En este caso aplicamos la fórmula donde $z_{\beta_0} = z_{0.1} = 1.28$

$$n \geq \left( \frac{z_\alpha \sqrt{p_0(1-p_0)} + z_{\beta_0} \sqrt{p(1-p)}}{p - p_0} \right)^2 = \left( \frac{1.645\sqrt{0.05(1-0.05)} + 1.28\sqrt{0.03(1-0.03)}}{0.03 - 0.05} \right)^2 \approx 832$$

La muestra requerida es muy grande, pero la diferencia a detectar $p - p_0 = 0.03 - 0.05$ es bastante pequeña.

## 9.11 – Tests de hipótesis sobre dos proporciones

Las pruebas de hipótesis sobre diferencia de medias pueden adaptarse al caso donde tenemos dos parámetros binomiales $p_1$ y $p_2$ de interés.
Específicamente, supongamos que se toman dos muestras aleatorias
$(X_{11}, X_{12}, ..., X_{1n_1})$ es una muestra aleatoria de tamaño $n_1$ de $X_1$
$(X_{21}, X_{22}, ..., X_{2n_2})$ es una muestra aleatoria de tamaño $n_2$ de $X_2$
Donde $X_1 \sim B(1, p_1)$; $X_2 \sim B(1, p_2)$ y $X_1$ y $X_2$ independientes.

Ya sabemos que $\hat{P}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$ y $\hat{P}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$ son estimadores ***insesgados*** de $p_1$ y $p_2$ respectivamente, con varianzas $V(\hat{P}_1) = \frac{p_1(1-p_1)}{n_1}$ y $V(\hat{P}_2) = \frac{p_2(1-p_2)}{n_2}$

Supongamos las hipótesis

$$H_0: p_1 - p_2 = 0 \quad \text{contra} \quad H_1: p_1 - p_2 \neq 0$$

Notar que si la hipótesis nula es verdadera entonces $p_1 = p_2 = p$, donde $p$ es ***desconocido***.

El estadístico $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ tiene distribución ***aproximadamente*** $N(0,1)$ por T.C.L. si $H_0: p_1 - p_2 = 0$ es verdadera. Tomamos como ***estimador de p*** a $\hat{P} = \dfrac{\sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i}}{n_1 + n_2}$ y lo reemplazamos en $Z$

Entonces el estadístico de prueba es $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ que bajo $H_0: p_1 - p_2 = 0$ se puede probar que tiene distribución aproximadamente $N(0,1)$

Entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; |Z| > z_{\frac{\alpha}{2}} \\ aceptar \;\; H_0 \;\; si \;\; |Z| \leq z_{\frac{\alpha}{2}} \end{cases}$ donde $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Si $H_1: p_1 - p_2 > 0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; Z > z_{\alpha} \\ aceptar \;\; H_0 \;\; si \;\; Z \leq z_{\alpha} \end{cases}$

Si $H_1: p_1 - p_2 < 0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; Z < -z_{\alpha} \\ aceptar \;\; H_0 \;\; si \;\; Z \geq -z_{\alpha} \end{cases}$

<u>Ejemplo</u>:

En una muestra de 100 lotes de un producto químico comprado al distribuidor A, 70 satisfacen una especificación de pureza. En una muestra de 70 lotes comprada al distribuidor B, 61 satisfacen la especificación. ¿Pude concluir que una proporción mayor de los lotes del distribuidor B satisface la especificación?

Solución:
Los parámetros de interés son $p_1$ y $p_2$ las verdaderas proporciones de lotes que cumplen las especificaciones de pureza.

Tenemos una muestra aleatoria $\left(X_{11}, X_{12}, ..., X_{1n_1}\right)$ de tamaño $n_1 = 100$ donde $\hat{P}_1 = \dfrac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} = \dfrac{70}{100} = 0.7$

Y otra muestra $\left(X_{21}, X_{22}, ..., X_{2n_2}\right)$ de tamaño $n_2 = 70$ donde $\hat{P}_2 = \dfrac{1}{n_2}\sum_{i=1}^{n_2} X_{2i} = \dfrac{61}{70}$

Las hipótesis son $H_0 : p_1 - p_2 = 0$ contra $H_1 : p_1 - p_2 < 0$

El estadístico de prueba es $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ donde $\hat{P} = \dfrac{\sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i}}{n_1 + n_2}$

En este caso $\hat{P} = \dfrac{\sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i}}{n_1 + n_2} = \dfrac{70+61}{100+70} = \dfrac{131}{170}$

El estadístico toma el valor $z_0 = \dfrac{\dfrac{70}{100} - \dfrac{61}{70}}{\sqrt{\dfrac{131}{170}\left(1 - \dfrac{131}{170}\right)\left(\dfrac{1}{100} + \dfrac{1}{70}\right)}} = -2.6163$

Para saber cuánta evidencia hay contra $H_0 : p_1 - p_2 = 0$ calculamos el p-valor
$p - valor = P\left(Z < z_0\right) = \Phi\left(-2.6163\right) = 0.0045$
Como el p-valor es menor que 0.05, se considera que hay mucha evidencia contra $H_0 : p_1 - p_2 = 0$ y se rechaza la hipótesis nula.

**Valor de $\beta$**

Cuando $H_0 : p_1 - p_2 = 0$ es falsa, la varianza de $\hat{P}_1 - \hat{P}_2$ es

$$V\left(\hat{P}_1 - \hat{P}_2\right) = V\left(\hat{P}_1\right) + V\left(\hat{P}_2\right) = \frac{p_1\left(1-p_1\right)}{n_1} + \frac{p_2\left(1-p_2\right)}{n_2}$$

Anotamos $\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{V\left(\hat{P}_1 - \hat{P}_2\right)} = \sqrt{V\left(\hat{P}_1\right) + V\left(\hat{P}_2\right)} = \sqrt{\dfrac{p_1\left(1-p_1\right)}{n_1} + \dfrac{p_2\left(1-p_2\right)}{n_2}}$

Entonces

Si $H_1: p_1 - p_2 \neq 0$

$$\beta \approx \Phi\left(\frac{z_{\alpha/2}\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\sigma_{\hat{P}_1-\hat{P}_2}}\right)-\Phi\left(\frac{-z_{\alpha/2}\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\sigma_{\hat{P}_1-\hat{P}_2}}\right)$$

Donde $\overline{p} = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ y $\overline{q} = 1 - \overline{p}$

Si $H_1: p_1 - p_2 > 0$ entonces $\beta \approx \Phi\left(\dfrac{z_{\varepsilon}\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\sigma_{\hat{P}_1-\hat{P}_2}}\right)$

Si $H_1: p_1 - p_2 < 0$ entonces $\beta \approx 1-\Phi\left(\dfrac{-z_{\alpha}\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\sigma_{\hat{P}_1-\hat{P}_2}}\right)$

Podemos deducir fórmulas para el tamaño de la muestra, nuevamente asumiendo que $n_1 = n_2 = n$

1) $\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$

2) $\left[\overline{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$

3) $\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$

4) $\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \overline{X}_1 - \vec{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$

5) $\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\quad \overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$

6) $\overline{X}_1 - \overline{X}_2 - t_{\alpha/2, n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}};\quad \overline{X}_1 - \overline{X}_2 + t_{\alpha/2, n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

7) $\left[\overline{X}_1 - \overline{X}_2 - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\quad \overline{X}_1 - \overline{X}_2 + t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] \qquad \nu = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$

8) $\left[\overline{D} - t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}};\ \overline{D} + t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}}\right]$

9) $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}};\ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$

10) $\left[\frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2},n_2-1,n_1-1};\ \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2},n_2-1,n_1-1}\right]$

11) $\left[\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}},\quad \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$

12) $\left[\hat{P}_1 - \hat{P}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}};\ \hat{P}_1 - \hat{P}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\right]$

# RESUMEN DE PROCEDIMIENTOS PARA LA PRUEBA DE HIPÓTESIS.

| Caso | Hipótesis nula | Estadístico de prueba | Hipótesis alternativa | Criterio de rechazo |
|---|---|---|---|---|
| **1** | $H_0 : \mu = \mu_0$<br>$\sigma^2$ conocida<br>distribución normal | $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ | $H_1 : \mu \neq \mu_0$<br>$H_1 : \mu > \mu_0$<br>$H_1 : \mu < \mu_0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |
| **2** | $H_0 : \mu = \mu_0$<br>$\sigma^2$ desconocida<br>distribución normal | $T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}$ | $H_1 : \mu \neq \mu_0$<br>$H_1 : \mu > \mu_0$<br>$H_1 : \mu < \mu_0$ | $\lvert T \rvert > t_{\alpha/2, n-1}$<br>$T > t_{\alpha, n-1}$<br>$T < -t_{\alpha, n-1}$ |
| **3** | $H_0 : \mu = \mu_0$<br>$\sigma^2$ desconocida<br>distribución NO NECESARIAMENTE normal<br>muestra grande | $T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}$ | $H_1 : \mu \neq \mu_0$<br>$H_1 : \mu > \mu_0$<br>$H_1 : \mu < \mu_0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |
| **4** | $H_0 : \mu_1 - \mu_2 = \Delta_0$<br>$\sigma_1{}^2, \sigma_2{}^2$<br>conocidas<br>distribuciones normales | $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ | $H_1 : \mu_1 - \mu_2 \neq \Delta_0$<br>$H_1 : \mu_1 - \mu_2 > \Delta_0$<br>$H_1 : \mu_1 - \mu_2 < \Delta_0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |
| **5** | $H_0 : \mu_1 - \mu_2 = \Delta_0$<br>$\sigma_1{}^2, \sigma_2{}^2$<br>desconocidas<br>distribuciones NO NECESARIAMENTE normales<br>muestras grandes | $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ | $H_1 : \mu_1 - \mu_2 \neq \Delta_0$<br>$H_1 : \mu_1 - \mu_2 > \Delta_0$<br>$H_1 : \mu_1 - \mu_2 < \Delta_0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |
| **6** | $H_0 : \mu_1 - \mu_2 = \Delta_0$<br>$\sigma_1{}^2 = \sigma_2{}^2$<br>desconocida<br>distribuciones normales | $T = \dfrac{\vec{X}_1 - \overline{X}_2 - \Delta_0}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ | $H_1 : \mu_1 - \mu_2 \neq \Delta_0$<br>$H_1 : \mu_1 - \mu_2 > \Delta_0$<br>$H_1 : \mu_1 - \mu_2 < \Delta_0$ | $\lvert T \rvert > t_{\alpha/2, n_1 + n_2 - 2}$<br>$T > t_{\alpha, n_1 + n_2 - 2}$<br>$T < -t_{\alpha, n_1 + n_2 - 2}$ |

| | | | | |
|---|---|---|---|---|
| **7** | $H_0 : \mu_1 - \mu_2 = \Delta_0$<br>${\sigma_1}^2 \neq {\sigma_2}^2$<br>desconocidas<br>distribuciones<br>normales | $T^* = \dfrac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$<br>$\nu = \dfrac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\dfrac{\left(S_1^1/n_1\right)^2}{n_1 - 1} + \dfrac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$ | $H_1 : \mu_1 - \mu_2 \neq \Delta_0$<br>$H_1 : \mu_1 - \mu_2 > \Delta_0$<br>$H_1 : \mu_1 - \mu_2 < \Delta_0$ | $\lvert T^* \rvert > t_{\alpha/2, \nu}$<br>$T^* > t_{\alpha, \nu}$<br>$T^* < -t_{\alpha, \nu}$ |
| **8** | **Datos pareados**<br>$H_0 : \mu_1 - \mu_2 = \Delta_0$<br>la diferencia de las muestras tiene distribución normal | $T = \dfrac{\overline{D} - \Delta_0}{S_D / \sqrt{n}}$ | $H_1 : \mu_1 - \mu_2 \neq \Delta_0$<br>$H_1 : \mu_1 - \mu_2 > \Delta_0$<br>$H_1 : \mu_1 - \mu_2 < \Delta_0$ | $\lvert T \rvert > t_{\alpha/2, n-1}$<br>$T > t_{\alpha, n-1}$<br>$T < -t_{\alpha, n-1}$ |
| **9** | $H_0 : \sigma^2 = \sigma_0^2$<br>distribución normal | $X = \dfrac{(n-1)S^2}{{\sigma_0}^2}$ | $H_1 : \sigma^2 \neq \sigma_0^2$<br>$H_1 : \sigma^2 > \sigma_0^2$<br>$H_1 : \sigma^2 < \sigma_0^2$ | $X > \chi^2_{\alpha/2, n-1}$ *ó* $X < \chi^2_{1-\alpha/2, n-1}$<br>$X > \chi^2_{\alpha, n-1}$<br>$X < \chi^2_{1-\alpha, n-1}$ |
| **10** | $H_0 : {\sigma_1}^2 = \sigma_2^2$<br>distribuciones<br>normales | $F = \dfrac{S_1^2}{S_2^2}$ | $H_1 : {\sigma_1}^2 \neq \sigma_2^2$<br>$H_1 : {\sigma_1}^2 > \sigma_2^2$<br>$H_1 : {\sigma_1}^2 < \sigma_2^2$ | $F > f_{\alpha/2, n_1-1, n_2-1}$ *ó*<br>$F < f_{1-\alpha/2, n_1-1, n_2-1}$<br>$F > f_{\alpha, n_1-1, n_2-1}$<br>$F < f_{1-\alpha, n_1-1, n_2-1}$ |
| **11** | $H_0 : p = p_0$<br>muestra grande | $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ | $H_1 : p \neq p_0$<br>$H_1 : p > p_0$<br>$H_1 : p < p_0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |
| **12** | $H_0 : p_1 - p_2 = 0$<br>muestras grandes | $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ | $H_1 : p_1 - p_2 \neq 0$<br>$H_1 : p_1 - p_2 > 0$<br>$H_1 : p_1 - p_2 < 0$ | $\lvert Z \rvert > z_{\alpha/2}$<br>$Z > z_\alpha$<br>$Z < -z_\alpha$ |

# 1. Distribución Binomial

$$X \sim \mathrm{Binom}(n, \theta)$$

$$p = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} \theta^k (1-\theta)^{n-k} = 1 - \alpha$$

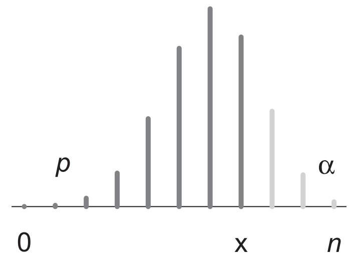



Tabla 1A. Probabilidades acumuladas $p$ de la distribución binomial ($n = 5, 6, 7, 8, 9$).

| | | $\theta$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n=5$ | 0 | 0.951 | 0.774 | 0.590 | 0.328 | 0.237 | 0.168 | 0.078 | 0.031 | 0.010 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.999 | 0.977 | 0.919 | 0.737 | 0.633 | 0.528 | 0.337 | 0.188 | 0.087 | 0.031 | 0.016 | 0.007 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.999 | 0.991 | 0.942 | 0.896 | 0.837 | 0.683 | 0.500 | 0.317 | 0.163 | 0.104 | 0.058 | 0.009 | 0.001 | 0.000 |
| | 3 | 1.000 | 1.000 | 1.000 | 0.993 | 0.984 | 0.969 | 0.913 | 0.813 | 0.663 | 0.472 | 0.367 | 0.263 | 0.081 | 0.023 | 0.001 |
| | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.990 | 0.969 | 0.922 | 0.832 | 0.763 | 0.672 | 0.410 | 0.226 | 0.049 |
| | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n=6$ | 0 | 0.941 | 0.735 | 0.531 | 0.262 | 0.178 | 0.118 | 0.047 | 0.016 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.999 | 0.967 | 0.886 | 0.655 | 0.534 | 0.420 | 0.233 | 0.109 | 0.041 | 0.011 | 0.005 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.998 | 0.984 | 0.901 | 0.831 | 0.744 | 0.544 | 0.344 | 0.179 | 0.070 | 0.038 | 0.017 | 0.001 | 0.000 | 0.000 |
| | 3 | 1.000 | 1.000 | 0.999 | 0.983 | 0.962 | 0.930 | 0.821 | 0.656 | 0.456 | 0.256 | 0.169 | 0.099 | 0.016 | 0.002 | 0.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 0.998 | 0.995 | 0.989 | 0.959 | 0.891 | 0.767 | 0.580 | 0.466 | 0.345 | 0.114 | 0.033 | 0.001 |
| | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.984 | 0.953 | 0.882 | 0.822 | 0.738 | 0.469 | 0.265 | 0.059 |
| | 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n=7$ | 0 | 0.932 | 0.698 | 0.478 | 0.210 | 0.133 | 0.082 | 0.028 | 0.008 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.998 | 0.956 | 0.850 | 0.577 | 0.445 | 0.329 | 0.159 | 0.063 | 0.019 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.996 | 0.974 | 0.852 | 0.756 | 0.647 | 0.420 | 0.227 | 0.096 | 0.029 | 0.013 | 0.005 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 1.000 | 0.997 | 0.967 | 0.929 | 0.874 | 0.710 | 0.500 | 0.290 | 0.126 | 0.071 | 0.033 | 0.003 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 0.995 | 0.987 | 0.971 | 0.904 | 0.773 | 0.580 | 0.353 | 0.244 | 0.148 | 0.026 | 0.004 | 0.000 |
| | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.981 | 0.938 | 0.841 | 0.671 | 0.555 | 0.423 | 0.150 | 0.044 | 0.002 |
| | 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.992 | 0.972 | 0.918 | 0.867 | 0.790 | 0.522 | 0.302 | 0.068 |
| | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n=8$ | 0 | 0.923 | 0.663 | 0.430 | 0.168 | 0.100 | 0.058 | 0.017 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.997 | 0.943 | 0.813 | 0.503 | 0.367 | 0.255 | 0.106 | 0.035 | 0.009 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.994 | 0.962 | 0.797 | 0.679 | 0.552 | 0.315 | 0.145 | 0.050 | 0.011 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 1.000 | 0.995 | 0.944 | 0.886 | 0.806 | 0.594 | 0.363 | 0.174 | 0.058 | 0.027 | 0.010 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 0.990 | 0.973 | 0.942 | 0.826 | 0.637 | 0.406 | 0.194 | 0.114 | 0.056 | 0.005 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.989 | 0.950 | 0.855 | 0.685 | 0.448 | 0.321 | 0.203 | 0.038 | 0.006 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.991 | 0.965 | 0.894 | 0.745 | 0.633 | 0.497 | 0.187 | 0.057 | 0.003 |
| | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.983 | 0.942 | 0.900 | 0.832 | 0.570 | 0.337 | 0.077 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n=9$ | 0 | 0.914 | 0.630 | 0.387 | 0.134 | 0.075 | 0.040 | 0.010 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.997 | 0.929 | 0.775 | 0.436 | 0.300 | 0.196 | 0.071 | 0.020 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.992 | 0.947 | 0.738 | 0.601 | 0.463 | 0.232 | 0.090 | 0.025 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.999 | 0.992 | 0.914 | 0.834 | 0.730 | 0.483 | 0.254 | 0.099 | 0.025 | 0.010 | 0.003 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.999 | 0.980 | 0.951 | 0.901 | 0.733 | 0.500 | 0.267 | 0.099 | 0.049 | 0.020 | 0.001 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 1.000 | 0.997 | 0.990 | 0.975 | 0.901 | 0.746 | 0.517 | 0.270 | 0.166 | 0.086 | 0.008 | 0.001 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.975 | 0.910 | 0.768 | 0.537 | 0.399 | 0.262 | 0.053 | 0.008 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.980 | 0.929 | 0.804 | 0.700 | 0.564 | 0.225 | 0.071 | 0.003 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.990 | 0.960 | 0.925 | 0.866 | 0.613 | 0.370 | 0.086 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 1B. Probabilidades acumuladas $p$ de la distribución binomial ($n = 10, 11, 12, 13, 14$).

| | | | | | | | | | $\theta$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n = 10$ | 0 | 0.904 | 0.599 | 0.349 | 0.107 | 0.056 | 0.028 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.996 | 0.914 | 0.736 | 0.376 | 0.244 | 0.149 | 0.046 | 0.011 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.988 | 0.930 | 0.678 | 0.526 | 0.383 | 0.167 | 0.055 | 0.012 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.999 | 0.987 | 0.879 | 0.776 | 0.650 | 0.382 | 0.172 | 0.055 | 0.011 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.998 | 0.967 | 0.922 | 0.850 | 0.633 | 0.377 | 0.166 | 0.047 | 0.020 | 0.006 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 1.000 | 0.994 | 0.980 | 0.953 | 0.834 | 0.623 | 0.367 | 0.150 | 0.078 | 0.033 | 0.002 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.989 | 0.945 | 0.828 | 0.618 | 0.350 | 0.224 | 0.121 | 0.013 | 0.001 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.988 | 0.945 | 0.833 | 0.617 | 0.474 | 0.322 | 0.070 | 0.012 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.989 | 0.954 | 0.851 | 0.756 | 0.624 | 0.264 | 0.086 | 0.004 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.972 | 0.944 | 0.893 | 0.651 | 0.401 | 0.096 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 11$ | 0 | 0.895 | 0.569 | 0.314 | 0.086 | 0.042 | 0.020 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.995 | 0.898 | 0.697 | 0.322 | 0.197 | 0.113 | 0.030 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.985 | 0.910 | 0.617 | 0.455 | 0.313 | 0.119 | 0.033 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.998 | 0.981 | 0.839 | 0.713 | 0.570 | 0.296 | 0.113 | 0.029 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.997 | 0.950 | 0.885 | 0.790 | 0.533 | 0.274 | 0.099 | 0.022 | 0.008 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 1.000 | 0.988 | 0.966 | 0.922 | 0.753 | 0.500 | 0.247 | 0.078 | 0.034 | 0.012 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.998 | 0.992 | 0.978 | 0.901 | 0.726 | 0.467 | 0.210 | 0.115 | 0.050 | 0.003 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.971 | 0.887 | 0.704 | 0.430 | 0.287 | 0.161 | 0.019 | 0.002 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.967 | 0.881 | 0.687 | 0.545 | 0.383 | 0.090 | 0.015 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.970 | 0.887 | 0.803 | 0.678 | 0.303 | 0.102 | 0.005 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.980 | 0.958 | 0.914 | 0.686 | 0.431 | 0.105 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 12$ | 0 | 0.886 | 0.540 | 0.282 | 0.069 | 0.032 | 0.014 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.994 | 0.882 | 0.659 | 0.275 | 0.158 | 0.085 | 0.020 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.980 | 0.889 | 0.558 | 0.391 | 0.253 | 0.083 | 0.019 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.998 | 0.974 | 0.795 | 0.649 | 0.493 | 0.225 | 0.073 | 0.015 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.996 | 0.927 | 0.842 | 0.724 | 0.438 | 0.194 | 0.057 | 0.009 | 0.003 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.999 | 0.981 | 0.946 | 0.882 | 0.665 | 0.387 | 0.158 | 0.039 | 0.014 | 0.004 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.996 | 0.986 | 0.961 | 0.842 | 0.613 | 0.335 | 0.118 | 0.054 | 0.019 | 0.001 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.999 | 0.997 | 0.991 | 0.943 | 0.806 | 0.562 | 0.276 | 0.158 | 0.073 | 0.004 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.985 | 0.927 | 0.775 | 0.507 | 0.351 | 0.205 | 0.026 | 0.002 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.981 | 0.917 | 0.747 | 0.609 | 0.442 | 0.111 | 0.020 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.980 | 0.915 | 0.842 | 0.725 | 0.341 | 0.118 | 0.006 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.986 | 0.968 | 0.931 | 0.718 | 0.460 | 0.114 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 13$ | 0 | 0.878 | 0.513 | 0.254 | 0.055 | 0.024 | 0.010 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.993 | 0.865 | 0.621 | 0.234 | 0.127 | 0.064 | 0.013 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.975 | 0.866 | 0.502 | 0.333 | 0.202 | 0.058 | 0.011 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.997 | 0.966 | 0.747 | 0.584 | 0.421 | 0.169 | 0.046 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.994 | 0.901 | 0.794 | 0.654 | 0.353 | 0.133 | 0.032 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.999 | 0.970 | 0.920 | 0.835 | 0.574 | 0.291 | 0.098 | 0.018 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.993 | 0.976 | 0.938 | 0.771 | 0.500 | 0.229 | 0.062 | 0.024 | 0.007 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.982 | 0.902 | 0.709 | 0.426 | 0.165 | 0.080 | 0.030 | 0.001 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.968 | 0.867 | 0.647 | 0.346 | 0.206 | 0.099 | 0.006 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.992 | 0.954 | 0.831 | 0.579 | 0.416 | 0.253 | 0.034 | 0.003 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.989 | 0.942 | 0.798 | 0.667 | 0.498 | 0.134 | 0.025 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.987 | 0.936 | 0.873 | 0.766 | 0.379 | 0.135 | 0.007 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.990 | 0.976 | 0.945 | 0.746 | 0.487 | 0.122 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 14$ | 0 | 0.869 | 0.488 | 0.229 | 0.044 | 0.018 | 0.007 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.992 | 0.847 | 0.585 | 0.198 | 0.101 | 0.047 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.970 | 0.842 | 0.448 | 0.281 | 0.161 | 0.040 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.996 | 0.956 | 0.698 | 0.521 | 0.355 | 0.124 | 0.029 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 1.000 | 0.991 | 0.870 | 0.742 | 0.584 | 0.279 | 0.090 | 0.018 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.999 | 0.956 | 0.888 | 0.781 | 0.486 | 0.212 | 0.058 | 0.008 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.988 | 0.962 | 0.907 | 0.692 | 0.395 | 0.150 | 0.031 | 0.010 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.998 | 0.990 | 0.969 | 0.850 | 0.605 | 0.308 | 0.093 | 0.038 | 0.012 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.992 | 0.942 | 0.788 | 0.514 | 0.219 | 0.112 | 0.044 | 0.001 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.982 | 0.910 | 0.721 | 0.416 | 0.258 | 0.130 | 0.009 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.971 | 0.876 | 0.645 | 0.479 | 0.302 | 0.044 | 0.004 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.960 | 0.839 | 0.719 | 0.552 | 0.158 | 0.030 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.992 | 0.953 | 0.899 | 0.802 | 0.415 | 0.153 | 0.008 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.993 | 0.982 | 0.956 | 0.771 | 0.512 | 0.131 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 1C. Probabilidades acumuladas $p$ de la distribución binomial ($n = 15, 16, 17, 18$).

| | | | | | | | | $\theta$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n = 15$ | 0 | 0.860 | 0.463 | 0.206 | 0.035 | 0.013 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.990 | 0.829 | 0.549 | 0.167 | 0.080 | 0.035 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 1.000 | 0.964 | 0.816 | 0.398 | 0.236 | 0.127 | 0.027 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.995 | 0.944 | 0.648 | 0.461 | 0.297 | 0.091 | 0.018 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.999 | 0.987 | 0.836 | 0.686 | 0.515 | 0.217 | 0.059 | 0.009 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.998 | 0.939 | 0.852 | 0.722 | 0.403 | 0.151 | 0.034 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 1.000 | 0.982 | 0.943 | 0.869 | 0.610 | 0.304 | 0.095 | 0.015 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.996 | 0.983 | 0.950 | 0.787 | 0.500 | 0.213 | 0.050 | 0.017 | 0.004 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.985 | 0.905 | 0.696 | 0.390 | 0.131 | 0.057 | 0.018 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.966 | 0.849 | 0.597 | 0.278 | 0.148 | 0.061 | 0.002 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.991 | 0.941 | 0.783 | 0.485 | 0.314 | 0.164 | 0.013 | 0.001 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.982 | 0.909 | 0.703 | 0.539 | 0.352 | 0.056 | 0.005 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.973 | 0.873 | 0.764 | 0.602 | 0.184 | 0.036 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.965 | 0.920 | 0.833 | 0.451 | 0.171 | 0.010 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.987 | 0.965 | 0.794 | 0.537 | 0.140 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 16$ | 0 | 0.851 | 0.440 | 0.185 | 0.028 | 0.010 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.989 | 0.811 | 0.515 | 0.141 | 0.063 | 0.026 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.957 | 0.789 | 0.352 | 0.197 | 0.099 | 0.018 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.993 | 0.932 | 0.598 | 0.405 | 0.246 | 0.065 | 0.011 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.999 | 0.983 | 0.798 | 0.630 | 0.450 | 0.167 | 0.038 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.997 | 0.918 | 0.810 | 0.660 | 0.329 | 0.105 | 0.019 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.999 | 0.973 | 0.920 | 0.825 | 0.527 | 0.227 | 0.058 | 0.007 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.993 | 0.973 | 0.926 | 0.716 | 0.402 | 0.142 | 0.026 | 0.007 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.999 | 0.993 | 0.974 | 0.858 | 0.598 | 0.284 | 0.074 | 0.027 | 0.007 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.993 | 0.942 | 0.773 | 0.473 | 0.175 | 0.080 | 0.027 | 0.001 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.981 | 0.895 | 0.671 | 0.340 | 0.190 | 0.082 | 0.003 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.962 | 0.833 | 0.550 | 0.370 | 0.202 | 0.017 | 0.001 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.989 | 0.935 | 0.754 | 0.595 | 0.402 | 0.068 | 0.007 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.982 | 0.901 | 0.803 | 0.648 | 0.211 | 0.043 | 0.001 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.974 | 0.937 | 0.859 | 0.485 | 0.189 | 0.011 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.990 | 0.972 | 0.815 | 0.560 | 0.149 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 17$ | 0 | 0.843 | 0.418 | 0.167 | 0.023 | 0.008 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.988 | 0.792 | 0.482 | 0.118 | 0.050 | 0.019 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.950 | 0.762 | 0.310 | 0.164 | 0.077 | 0.012 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.991 | 0.917 | 0.549 | 0.353 | 0.202 | 0.046 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.999 | 0.978 | 0.758 | 0.574 | 0.389 | 0.126 | 0.025 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.995 | 0.894 | 0.765 | 0.597 | 0.264 | 0.072 | 0.011 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.999 | 0.962 | 0.893 | 0.775 | 0.448 | 0.166 | 0.035 | 0.003 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.989 | 0.960 | 0.895 | 0.641 | 0.315 | 0.092 | 0.013 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.997 | 0.988 | 0.960 | 0.801 | 0.500 | 0.199 | 0.040 | 0.012 | 0.003 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.987 | 0.908 | 0.685 | 0.359 | 0.105 | 0.040 | 0.011 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.997 | 0.965 | 0.834 | 0.552 | 0.225 | 0.107 | 0.038 | 0.001 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.989 | 0.928 | 0.736 | 0.403 | 0.235 | 0.106 | 0.005 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.975 | 0.874 | 0.611 | 0.426 | 0.242 | 0.022 | 0.001 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.954 | 0.798 | 0.647 | 0.451 | 0.083 | 0.009 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.988 | 0.923 | 0.836 | 0.690 | 0.238 | 0.050 | 0.001 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.981 | 0.950 | 0.882 | 0.518 | 0.208 | 0.012 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.992 | 0.977 | 0.833 | 0.582 | 0.157 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 18$ | 0 | 0.835 | 0.397 | 0.150 | 0.018 | 0.006 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.986 | 0.774 | 0.450 | 0.099 | 0.039 | 0.014 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.942 | 0.734 | 0.271 | 0.135 | 0.060 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.989 | 0.902 | 0.501 | 0.306 | 0.165 | 0.033 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.998 | 0.972 | 0.716 | 0.519 | 0.333 | 0.094 | 0.015 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.994 | 0.867 | 0.717 | 0.534 | 0.209 | 0.048 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.999 | 0.949 | 0.861 | 0.722 | 0.374 | 0.119 | 0.020 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.984 | 0.943 | 0.859 | 0.563 | 0.240 | 0.058 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.996 | 0.981 | 0.940 | 0.737 | 0.407 | 0.135 | 0.021 | 0.005 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 | 0.979 | 0.865 | 0.593 | 0.263 | 0.060 | 0.019 | 0.004 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.942 | 0.760 | 0.437 | 0.141 | 0.057 | 0.016 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.980 | 0.881 | 0.626 | 0.278 | 0.139 | 0.051 | 0.001 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.952 | 0.791 | 0.466 | 0.283 | 0.133 | 0.006 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.985 | 0.906 | 0.667 | 0.481 | 0.284 | 0.028 | 0.002 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.967 | 0.835 | 0.694 | 0.499 | 0.098 | 0.011 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.992 | 0.940 | 0.865 | 0.729 | 0.266 | 0.058 | 0.001 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.986 | 0.961 | 0.901 | 0.550 | 0.226 | 0.014 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.994 | 0.982 | 0.850 | 0.603 | 0.165 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 1D. Probabilidades acumuladas $p$ de la distribución binomial ($n = 19, 20, 21$).

| | | $\theta$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n = 19$ | 0 | 0.826 | 0.377 | 0.135 | 0.014 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.985 | 0.755 | 0.420 | 0.083 | 0.031 | 0.010 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.933 | 0.705 | 0.237 | 0.111 | 0.046 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.987 | 0.885 | 0.455 | 0.263 | 0.133 | 0.023 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.998 | 0.965 | 0.673 | 0.465 | 0.282 | 0.070 | 0.010 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.991 | 0.837 | 0.668 | 0.474 | 0.163 | 0.032 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.998 | 0.932 | 0.825 | 0.666 | 0.308 | 0.084 | 0.012 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.977 | 0.923 | 0.818 | 0.488 | 0.180 | 0.035 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.993 | 0.971 | 0.916 | 0.667 | 0.324 | 0.088 | 0.011 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.998 | 0.991 | 0.967 | 0.814 | 0.500 | 0.186 | 0.033 | 0.009 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.989 | 0.912 | 0.676 | 0.333 | 0.084 | 0.029 | 0.007 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.965 | 0.820 | 0.512 | 0.182 | 0.077 | 0.023 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.988 | 0.916 | 0.692 | 0.334 | 0.175 | 0.068 | 0.002 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.968 | 0.837 | 0.526 | 0.332 | 0.163 | 0.009 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.990 | 0.930 | 0.718 | 0.535 | 0.327 | 0.035 | 0.002 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.977 | 0.867 | 0.737 | 0.545 | 0.115 | 0.013 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.954 | 0.889 | 0.763 | 0.295 | 0.067 | 0.001 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.990 | 0.969 | 0.917 | 0.580 | 0.245 | 0.015 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.986 | 0.865 | 0.623 | 0.174 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 20$ | 0 | 0.818 | 0.358 | 0.122 | 0.012 | 0.003 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.983 | 0.736 | 0.392 | 0.069 | 0.024 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.925 | 0.677 | 0.206 | 0.091 | 0.035 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.984 | 0.867 | 0.411 | 0.225 | 0.107 | 0.016 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.997 | 0.957 | 0.630 | 0.415 | 0.238 | 0.051 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.989 | 0.804 | 0.617 | 0.416 | 0.126 | 0.021 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.998 | 0.913 | 0.786 | 0.608 | 0.250 | 0.058 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 1.000 | 0.968 | 0.898 | 0.772 | 0.416 | 0.132 | 0.021 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.990 | 0.959 | 0.887 | 0.596 | 0.252 | 0.057 | 0.005 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.997 | 0.986 | 0.952 | 0.755 | 0.412 | 0.128 | 0.017 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.983 | 0.872 | 0.588 | 0.245 | 0.048 | 0.014 | 0.003 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 | 0.943 | 0.748 | 0.404 | 0.113 | 0.041 | 0.010 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.979 | 0.868 | 0.584 | 0.228 | 0.102 | 0.032 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.942 | 0.750 | 0.392 | 0.214 | 0.087 | 0.002 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.979 | 0.874 | 0.584 | 0.383 | 0.196 | 0.011 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.949 | 0.762 | 0.585 | 0.370 | 0.043 | 0.003 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.984 | 0.893 | 0.775 | 0.589 | 0.133 | 0.016 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.965 | 0.909 | 0.794 | 0.323 | 0.075 | 0.001 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.992 | 0.976 | 0.931 | 0.608 | 0.264 | 0.017 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.997 | 0.988 | 0.878 | 0.642 | 0.182 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 21$ | 0 | 0.810 | 0.341 | 0.109 | 0.009 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.981 | 0.717 | 0.365 | 0.058 | 0.019 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.915 | 0.648 | 0.179 | 0.075 | 0.027 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.981 | 0.848 | 0.370 | 0.192 | 0.086 | 0.011 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.997 | 0.948 | 0.586 | 0.367 | 0.198 | 0.037 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 1.000 | 0.986 | 0.769 | 0.567 | 0.363 | 0.096 | 0.013 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.997 | 0.891 | 0.744 | 0.551 | 0.200 | 0.039 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 0.999 | 0.957 | 0.870 | 0.723 | 0.350 | 0.095 | 0.012 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.986 | 0.944 | 0.852 | 0.524 | 0.192 | 0.035 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.996 | 0.979 | 0.932 | 0.691 | 0.332 | 0.085 | 0.009 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.974 | 0.826 | 0.500 | 0.174 | 0.026 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.991 | 0.915 | 0.668 | 0.309 | 0.068 | 0.021 | 0.004 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.965 | 0.808 | 0.476 | 0.148 | 0.056 | 0.014 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.988 | 0.905 | 0.650 | 0.277 | 0.130 | 0.043 | 0.001 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.961 | 0.800 | 0.449 | 0.256 | 0.109 | 0.003 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.987 | 0.904 | 0.637 | 0.433 | 0.231 | 0.014 | 0.000 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.963 | 0.802 | 0.633 | 0.414 | 0.052 | 0.003 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.989 | 0.914 | 0.808 | 0.630 | 0.152 | 0.019 | 0.000 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.973 | 0.925 | 0.821 | 0.352 | 0.085 | 0.001 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.981 | 0.942 | 0.635 | 0.283 | 0.019 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.991 | 0.891 | 0.659 | 0.190 |
| | 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 1E. Probabilidades acumuladas de la distribución binomial ($n = 22, 23$).

| | | $\theta$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n = 22$ | 0 | 0.802 | 0.324 | 0.098 | 0.007 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.980 | 0.698 | 0.339 | 0.048 | 0.015 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.999 | 0.905 | 0.620 | 0.154 | 0.061 | 0.021 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.978 | 0.828 | 0.332 | 0.162 | 0.068 | 0.008 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.996 | 0.938 | 0.543 | 0.323 | 0.165 | 0.027 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 0.999 | 0.982 | 0.733 | 0.517 | 0.313 | 0.072 | 0.008 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.996 | 0.867 | 0.699 | 0.494 | 0.158 | 0.026 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 0.999 | 0.944 | 0.838 | 0.671 | 0.290 | 0.067 | 0.007 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.980 | 0.925 | 0.814 | 0.454 | 0.143 | 0.021 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.994 | 0.970 | 0.908 | 0.624 | 0.262 | 0.055 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.998 | 0.990 | 0.961 | 0.772 | 0.416 | 0.121 | 0.014 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.986 | 0.879 | 0.584 | 0.228 | 0.039 | 0.010 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.945 | 0.738 | 0.376 | 0.092 | 0.030 | 0.006 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.979 | 0.857 | 0.546 | 0.186 | 0.075 | 0.020 | 0.000 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 0.933 | 0.710 | 0.329 | 0.162 | 0.056 | 0.001 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.974 | 0.842 | 0.506 | 0.301 | 0.133 | 0.004 | 0.000 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.992 | 0.928 | 0.687 | 0.483 | 0.267 | 0.018 | 0.001 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.973 | 0.835 | 0.677 | 0.457 | 0.062 | 0.004 | 0.000 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.992 | 0.932 | 0.838 | 0.668 | 0.172 | 0.022 | 0.000 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.979 | 0.939 | 0.846 | 0.380 | 0.095 | 0.001 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.985 | 0.952 | 0.661 | 0.302 | 0.020 |
| | 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.993 | 0.902 | 0.676 | 0.198 |
| | 22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 23$ | 0 | 0.794 | 0.307 | 0.089 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.978 | 0.679 | 0.315 | 0.040 | 0.012 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.998 | 0.895 | 0.592 | 0.133 | 0.049 | 0.016 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.974 | 0.807 | 0.297 | 0.137 | 0.054 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.995 | 0.927 | 0.501 | 0.283 | 0.136 | 0.019 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 0.999 | 0.977 | 0.695 | 0.468 | 0.269 | 0.054 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.994 | 0.840 | 0.654 | 0.440 | 0.124 | 0.017 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 0.999 | 0.928 | 0.804 | 0.618 | 0.237 | 0.047 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.973 | 0.904 | 0.771 | 0.388 | 0.105 | 0.013 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.991 | 0.959 | 0.880 | 0.556 | 0.202 | 0.035 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.997 | 0.985 | 0.945 | 0.713 | 0.339 | 0.081 | 0.007 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 | 0.979 | 0.836 | 0.500 | 0.164 | 0.021 | 0.005 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.993 | 0.919 | 0.661 | 0.287 | 0.055 | 0.015 | 0.003 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.965 | 0.798 | 0.444 | 0.120 | 0.041 | 0.009 | 0.000 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.987 | 0.895 | 0.612 | 0.229 | 0.096 | 0.027 | 0.000 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.953 | 0.763 | 0.382 | 0.196 | 0.072 | 0.001 | 0.000 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.983 | 0.876 | 0.560 | 0.346 | 0.160 | 0.006 | 0.000 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.946 | 0.731 | 0.532 | 0.305 | 0.023 | 0.001 | 0.000 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.981 | 0.864 | 0.717 | 0.499 | 0.073 | 0.005 | 0.000 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 | 0.946 | 0.863 | 0.703 | 0.193 | 0.026 | 0.000 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.984 | 0.951 | 0.867 | 0.408 | 0.105 | 0.002 |
| | 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.988 | 0.960 | 0.685 | 0.321 | 0.022 |
| | 22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.911 | 0.693 | 0.206 |
| | 23 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 1F. Probabilidades acumuladas de la distribución binomial ($n = 24, 25$).

| | | | | | | | | | $\theta$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x$ | 0.01 | 0.05 | 0.1 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 | 0.99 |
| $n = 24$ | 0 | 0.786 | 0.292 | 0.080 | 0.005 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.976 | 0.661 | 0.292 | 0.033 | 0.009 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.998 | 0.884 | 0.564 | 0.115 | 0.040 | 0.012 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.970 | 0.786 | 0.264 | 0.115 | 0.042 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.994 | 0.915 | 0.460 | 0.247 | 0.111 | 0.013 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 0.999 | 0.972 | 0.656 | 0.422 | 0.229 | 0.040 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.993 | 0.811 | 0.607 | 0.389 | 0.096 | 0.011 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 0.998 | 0.911 | 0.766 | 0.565 | 0.192 | 0.032 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.964 | 0.879 | 0.725 | 0.328 | 0.076 | 0.008 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.987 | 0.945 | 0.847 | 0.489 | 0.154 | 0.022 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.996 | 0.979 | 0.926 | 0.650 | 0.271 | 0.053 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 0.999 | 0.993 | 0.969 | 0.787 | 0.419 | 0.114 | 0.012 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.988 | 0.886 | 0.581 | 0.213 | 0.031 | 0.007 | 0.001 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.947 | 0.729 | 0.350 | 0.074 | 0.021 | 0.004 | 0.000 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.978 | 0.846 | 0.511 | 0.153 | 0.055 | 0.013 | 0.000 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.992 | 0.924 | 0.672 | 0.275 | 0.121 | 0.036 | 0.000 | 0.000 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.968 | 0.808 | 0.435 | 0.234 | 0.089 | 0.002 | 0.000 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.989 | 0.904 | 0.611 | 0.393 | 0.189 | 0.007 | 0.000 | 0.000 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.960 | 0.771 | 0.578 | 0.344 | 0.028 | 0.001 | 0.000 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.987 | 0.889 | 0.753 | 0.540 | 0.085 | 0.006 | 0.000 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.958 | 0.885 | 0.736 | 0.214 | 0.030 | 0.000 |
| | 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.988 | 0.960 | 0.885 | 0.436 | 0.116 | 0.002 |
| | 22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.991 | 0.967 | 0.708 | 0.339 | 0.024 |
| | 23 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 | 0.920 | 0.708 | 0.214 |
| | 24 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $n = 25$ | 0 | 0.778 | 0.277 | 0.072 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.974 | 0.642 | 0.271 | 0.027 | 0.007 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 2 | 0.998 | 0.873 | 0.537 | 0.098 | 0.032 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 3 | 1.000 | 0.966 | 0.764 | 0.234 | 0.096 | 0.033 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 4 | 1.000 | 0.993 | 0.902 | 0.421 | 0.214 | 0.090 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 5 | 1.000 | 0.999 | 0.967 | 0.617 | 0.378 | 0.193 | 0.029 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 6 | 1.000 | 1.000 | 0.991 | 0.780 | 0.561 | 0.341 | 0.074 | 0.007 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 7 | 1.000 | 1.000 | 0.998 | 0.891 | 0.727 | 0.512 | 0.154 | 0.022 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 8 | 1.000 | 1.000 | 1.000 | 0.953 | 0.851 | 0.677 | 0.274 | 0.054 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 9 | 1.000 | 1.000 | 1.000 | 0.983 | 0.929 | 0.811 | 0.425 | 0.115 | 0.013 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 10 | 1.000 | 1.000 | 1.000 | 0.994 | 0.970 | 0.902 | 0.586 | 0.212 | 0.034 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 11 | 1.000 | 1.000 | 1.000 | 0.998 | 0.989 | 0.956 | 0.732 | 0.345 | 0.078 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 12 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.983 | 0.846 | 0.500 | 0.154 | 0.017 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.922 | 0.655 | 0.268 | 0.044 | 0.011 | 0.002 | 0.000 | 0.000 | 0.000 |
| | 14 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.966 | 0.788 | 0.414 | 0.098 | 0.030 | 0.006 | 0.000 | 0.000 | 0.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.987 | 0.885 | 0.575 | 0.189 | 0.071 | 0.017 | 0.000 | 0.000 | 0.000 |
| | 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.946 | 0.726 | 0.323 | 0.149 | 0.047 | 0.000 | 0.000 | 0.000 |
| | 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.978 | 0.846 | 0.488 | 0.273 | 0.109 | 0.002 | 0.000 | 0.000 |
| | 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 0.926 | 0.659 | 0.439 | 0.220 | 0.009 | 0.000 | 0.000 |
| | 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.971 | 0.807 | 0.622 | 0.383 | 0.033 | 0.001 | 0.000 |
| | 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.991 | 0.910 | 0.786 | 0.579 | 0.098 | 0.007 | 0.000 |
| | 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.967 | 0.904 | 0.766 | 0.236 | 0.034 | 0.000 |
| | 22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.991 | 0.968 | 0.902 | 0.463 | 0.127 | 0.002 |
| | 23 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.993 | 0.973 | 0.729 | 0.358 | 0.026 |
| | 24 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.928 | 0.723 | 0.222 |
| | 24 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

# 2. Distribución Poisson

$$X \sim \text{Poisson}(\lambda)$$

$$p = P(X \leq x) = \sum_{k=0}^{x} \frac{\lambda^k e^{-\lambda}}{k!} = 1 - \alpha$$

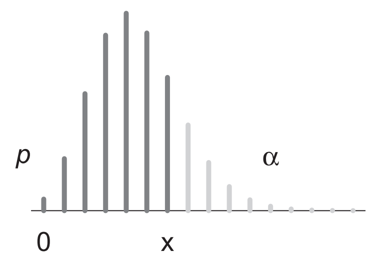



Tabla 2A. Probabilidades acumuladas $p$ de la distribución Poisson.

| | $\lambda$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| 0 | 0.905 | 0.819 | 0.741 | 0.670 | 0.607 | 0.549 | 0.497 | 0.449 | 0.407 | 0.368 |
| 1 | 0.995 | 0.982 | 0.963 | 0.938 | 0.910 | 0.878 | 0.844 | 0.809 | 0.772 | 0.736 |
| 2 | 1.000 | 0.999 | 0.996 | 0.992 | 0.986 | 0.977 | 0.966 | 0.953 | 0.937 | 0.920 |
| 3 | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.997 | 0.994 | 0.991 | 0.987 | 0.981 |
| 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.999 | 0.998 | 0.996 |
| 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |
| 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 2B. Probabilidades acumuladas $p$ de la distribución Poisson.

| | $\lambda$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 |
| 0 | 0.135 | 0.050 | 0.018 | 0.007 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.406 | 0.199 | 0.092 | 0.040 | 0.017 | 0.007 | 0.003 | 0.001 | 0.000 | 0.000 | 0.000 |
| 2 | 0.677 | 0.423 | 0.238 | 0.125 | 0.062 | 0.030 | 0.014 | 0.006 | 0.003 | 0.000 | 0.000 |
| 3 | 0.857 | 0.647 | 0.433 | 0.265 | 0.151 | 0.082 | 0.042 | 0.021 | 0.010 | 0.000 | 0.000 |
| 4 | 0.947 | 0.815 | 0.629 | 0.440 | 0.285 | 0.173 | 0.100 | 0.055 | 0.029 | 0.001 | 0.000 |
| 5 | 0.983 | 0.916 | 0.785 | 0.616 | 0.446 | 0.301 | 0.191 | 0.116 | 0.067 | 0.003 | 0.000 |
| 6 | 0.995 | 0.966 | 0.889 | 0.762 | 0.606 | 0.450 | 0.313 | 0.207 | 0.130 | 0.008 | 0.000 |
| 7 | 0.999 | 0.988 | 0.949 | 0.867 | 0.744 | 0.599 | 0.453 | 0.324 | 0.220 | 0.018 | 0.001 |
| 8 | 1.000 | 0.996 | 0.979 | 0.932 | 0.847 | 0.729 | 0.593 | 0.456 | 0.333 | 0.037 | 0.002 |
| 9 | 1.000 | 0.999 | 0.992 | 0.968 | 0.916 | 0.830 | 0.717 | 0.587 | 0.458 | 0.070 | 0.005 |
| 10 | 1.000 | 1.000 | 0.997 | 0.986 | 0.957 | 0.901 | 0.816 | 0.706 | 0.583 | 0.118 | 0.011 |
| 11 | 1.000 | 1.000 | 0.999 | 0.995 | 0.980 | 0.947 | 0.888 | 0.803 | 0.697 | 0.185 | 0.021 |
| 12 | 1.000 | 1.000 | 1.000 | 0.998 | 0.991 | 0.973 | 0.936 | 0.876 | 0.792 | 0.268 | 0.039 |
| 13 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.987 | 0.966 | 0.926 | 0.864 | 0.363 | 0.066 |
| 14 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.983 | 0.959 | 0.917 | 0.466 | 0.105 |
| 15 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.992 | 0.978 | 0.951 | 0.568 | 0.157 |
| 16 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.989 | 0.973 | 0.664 | 0.221 |
| 17 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.995 | 0.986 | 0.749 | 0.297 |
| 18 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.993 | 0.819 | 0.381 |
| 19 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.997 | 0.875 | 0.470 |
| 20 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.917 | 0.559 |
| 21 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.947 | 0.644 |
| 22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.967 | 0.721 |
| 23 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.981 | 0.787 |
| 24 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 0.843 |
| 25 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.888 |
| 26 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.922 |
| 27 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 0.948 |
| 28 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.966 |
| 29 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.978 |
| 30 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.987 |
| 31 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.992 |
| 32 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.995 |
| 33 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 |
| 34 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |
| 35 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |
| 36 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

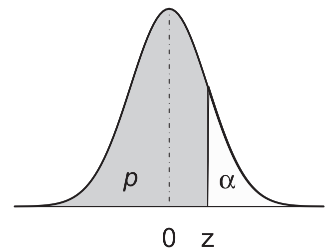
# 3. Distribución Normal Estándar

$$Z \sim \mathrm{N}(0,1)$$

$$p = P(Z \leq z) = \int_{-\infty}^{z} \phi_Z(u)du = 1 - \alpha$$

donde

$$\phi_Z(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}$$




*Nota*: Si $X \sim \mathrm{N}(\mu, \sigma^2)$, entonces $Z = (X - \mu)/\sigma \sim \mathrm{N}(0,1)$. Luego,

$$P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

Tabla 3A. Probabilidades acumuladas $p$ de la distribución normal estándar.

| $z$ | 0.09 | 0.08 | 0.07 | 0.06 | 0.05 | 0.04 | 0.03 | 0.02 | 0.01 | 0.00 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.4 | 0.0002 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 |
| −3.3 | 0.0003 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0005 | 0.0005 | 0.0005 |
| −3.2 | 0.0005 | 0.0005 | 0.0005 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0007 | 0.0007 |
| −3.1 | 0.0007 | 0.0007 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0009 | 0.0009 | 0.0009 | 0.0010 |
| −3.0 | 0.0010 | 0.0010 | 0.0011 | 0.0011 | 0.0011 | 0.0012 | 0.0012 | 0.0013 | 0.0013 | 0.0013 |
| −2.9 | 0.0014 | 0.0014 | 0.0015 | 0.0015 | 0.0016 | 0.0016 | 0.0017 | 0.0018 | 0.0018 | 0.0019 |
| −2.8 | 0.0019 | 0.0020 | 0.0021 | 0.0021 | 0.0022 | 0.0023 | 0.0023 | 0.0024 | 0.0025 | 0.0026 |
| −2.7 | 0.0026 | 0.0027 | 0.0028 | 0.0029 | 0.0030 | 0.0031 | 0.0032 | 0.0033 | 0.0034 | 0.0035 |
| −2.6 | 0.0036 | 0.0037 | 0.0038 | 0.0039 | 0.0040 | 0.0041 | 0.0043 | 0.0044 | 0.0045 | 0.0047 |
| −2.5 | 0.0048 | 0.0049 | 0.0051 | 0.0052 | 0.0054 | 0.0055 | 0.0057 | 0.0059 | 0.0060 | 0.0062 |
| −2.4 | 0.0064 | 0.0066 | 0.0068 | 0.0069 | 0.0071 | 0.0073 | 0.0075 | 0.0078 | 0.0080 | 0.0082 |
| −2.3 | 0.0084 | 0.0087 | 0.0089 | 0.0091 | 0.0094 | 0.0096 | 0.0099 | 0.0102 | 0.0104 | 0.0107 |
| −2.2 | 0.0110 | 0.0113 | 0.0116 | 0.0119 | 0.0122 | 0.0125 | 0.0129 | 0.0132 | 0.0136 | 0.0139 |
| −2.1 | 0.0143 | 0.0146 | 0.0150 | 0.0154 | 0.0158 | 0.0162 | 0.0166 | 0.0170 | 0.0174 | 0.0179 |
| −2.0 | 0.0183 | 0.0188 | 0.0192 | 0.0197 | 0.0202 | 0.0207 | 0.0212 | 0.0217 | 0.0222 | 0.0228 |
| −1.9 | 0.0233 | 0.0239 | 0.0244 | 0.0250 | 0.0256 | 0.0262 | 0.0268 | 0.0274 | 0.0281 | 0.0287 |
| −1.8 | 0.0294 | 0.0301 | 0.0307 | 0.0314 | 0.0322 | 0.0329 | 0.0336 | 0.0344 | 0.0351 | 0.0359 |
| −1.7 | 0.0367 | 0.0375 | 0.0384 | 0.0392 | 0.0401 | 0.0409 | 0.0418 | 0.0427 | 0.0436 | 0.0446 |
| −1.6 | 0.0455 | 0.0465 | 0.0475 | 0.0485 | 0.0495 | 0.0505 | 0.0516 | 0.0526 | 0.0537 | 0.0548 |
| −1.5 | 0.0559 | 0.0571 | 0.0582 | 0.0594 | 0.0606 | 0.0618 | 0.0630 | 0.0643 | 0.0655 | 0.0668 |
| −1.4 | 0.0681 | 0.0694 | 0.0708 | 0.0721 | 0.0735 | 0.0749 | 0.0764 | 0.0778 | 0.0793 | 0.0808 |
| −1.3 | 0.0823 | 0.0838 | 0.0853 | 0.0869 | 0.0885 | 0.0901 | 0.0918 | 0.0934 | 0.0951 | 0.0968 |
| −1.2 | 0.0985 | 0.1003 | 0.1020 | 0.1038 | 0.1056 | 0.1075 | 0.1093 | 0.1112 | 0.1131 | 0.1151 |
| −1.1 | 0.1170 | 0.1190 | 0.1210 | 0.1230 | 0.1251 | 0.1271 | 0.1292 | 0.1314 | 0.1335 | 0.1357 |
| −1.0 | 0.1379 | 0.1401 | 0.1423 | 0.1446 | 0.1469 | 0.1492 | 0.1515 | 0.1539 | 0.1562 | 0.1587 |
| −0.9 | 0.1611 | 0.1635 | 0.1660 | 0.1685 | 0.1711 | 0.1736 | 0.1762 | 0.1788 | 0.1814 | 0.1841 |
| −0.8 | 0.1867 | 0.1894 | 0.1922 | 0.1949 | 0.1977 | 0.2005 | 0.2033 | 0.2061 | 0.2090 | 0.2119 |
| −0.7 | 0.2148 | 0.2177 | 0.2206 | 0.2236 | 0.2266 | 0.2296 | 0.2327 | 0.2358 | 0.2389 | 0.2420 |
| −0.6 | 0.2451 | 0.2483 | 0.2514 | 0.2546 | 0.2578 | 0.2611 | 0.2643 | 0.2676 | 0.2709 | 0.2743 |
| −0.5 | 0.2776 | 0.2810 | 0.2843 | 0.2877 | 0.2912 | 0.2946 | 0.2981 | 0.3015 | 0.3050 | 0.3085 |
| −0.4 | 0.3121 | 0.3156 | 0.3192 | 0.3228 | 0.3264 | 0.3300 | 0.3336 | 0.3372 | 0.3409 | 0.3446 |
| −0.3 | 0.3483 | 0.3520 | 0.3557 | 0.3594 | 0.3632 | 0.3669 | 0.3707 | 0.3745 | 0.3783 | 0.3821 |
| −0.2 | 0.3859 | 0.3897 | 0.3936 | 0.3974 | 0.4013 | 0.4052 | 0.4090 | 0.4129 | 0.4168 | 0.4207 |
| −0.1 | 0.4247 | 0.4286 | 0.4325 | 0.4364 | 0.4404 | 0.4443 | 0.4483 | 0.4522 | 0.4562 | 0.4602 |
| −0.0 | 0.4641 | 0.4681 | 0.4721 | 0.4761 | 0.4801 | 0.4840 | 0.4880 | 0.4920 | 0.4960 | 0.5000 |

Tabla 3B. Probabilidades acumuladas $p$ de la distribución normal estándar.

| $z$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |

**Tabla III.** Valores críticos de la distribución Normal estándar

$$P[Z > Z_\alpha] = \int_{Z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp^{-z^2/2} dz = \alpha$$

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| 0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| 0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| 0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2910 | 0.2776 |
| 0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| 0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| 0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| 0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| 1.0 | 0.1597 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| 1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0995 |
| 1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| 1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| 1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| 1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| 1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0394 | 0.0375 | 0.0367 |
| 1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| 1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| 2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| 2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| 2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| 2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| 2.4 | 0.0082 | 0.00798 | 0.00776 | 0.00755 | 0.00734 | 0.00714 | 0.00695 | 0.00676 | 0.00657 | 0.00639 |
| 2.5 | 0.00621 | 0.00604 | 0.00587 | 0.00570 | 0.00554 | 0.00539 | 0.00523 | 0.00508 | 0.00494 | 0.00480 |
| 2.6 | 0.00466 | 0.00453 | 0.00440 | 0.00427 | 0.00415 | 0.00402 | 0.00391 | 0.00379 | 0.00368 | 0.00357 |
| 2.7 | 0.00347 | 0.00336 | 0.00326 | 0.00317 | 0.00307 | 0.00298 | 0.00289 | 0.00280 | 0.00272 | 0.00264 |
| 2.8 | 0.00256 | 0.00248 | 0.00240 | 0.00233 | 0.00226 | 0.00219 | 0.00212 | 0.00205 | 0.00199 | 0.00193 |
| 2.9 | 0.00187 | 0.00181 | 0.00175 | 0.00169 | 0.00164 | 0.00159 | 0.00154 | 0.00149 | 0.00144 | 0.00139 |
| 3.0 | 0.00135 | 0.00131 | 0.00126 | 0.00122 | 0.00118 | 0.00114 | 0.00111 | 0.00107 | 0.00103 | 0.00100 |
| 3.1 | 0.00097 | 0.00094 | 0.00091 | 0.00087 | 0.00084 | 0.00082 | 0.00079 | 0.00076 | 0.00074 | 0.00071 |
| 3.2 | 0.00069 | 0.00066 | 0.00064 | 0.00062 | 0.0006 | 0.00058 | 0.00056 | 0.00054 | 0.00052 | 0.00050 |
| 3.3 | 0.00048 | 0.00047 | 0.00045 | 0.00043 | 0.00042 | 0.00041 | 0.00032 | 0.00038 | 0.00036 | 0.00035 |
| 3.4 | 0.00034 | 0.00032 | 0.00031 | 0.0003 | 0.00029 | 0.00028 | 0.00027 | 0.00026 | 0.00025 | 0.00024 |
| 3.5 | 0.00023 | 0.00022 | 0.00022 | 0.00021 | 0.00020 | 0.00019 | 0.00019 | 0.00018 | 0.00017 | 0.00017 |
| 3.6 | 0.00016 | 0.00015 | 0.00015 | 0.00014 | 0.00014 | 0.00013 | 0.00013 | 0.00012 | 0.00012 | 0.00011 |
| 3.7 | 0.00011 | 0.00010 | 0.00010 | 0.00010 | 0.00009 | 0.00009 | 0.00009 | 0.00008 | 0.00008 | 0.00008 |
| 3.8 | 0.00007 | 0.00007 | 0.00007 | 0.00006 | 0.00006 | 0.00006 | 0.00006 | 0.00005 | 0.00005 | 0.00005 |
| 3.9 | 0.00005 | 0.00005 | 0.00004 | 0.00004 | 0.00004 | 0.00004 | 0.00004 | 0.00004 | 0.00003 | 0.00003 |

**Tabla IV.** Valores críticos de la distribución $t$ de Student:

Abcisas $t_{\alpha;\nu}$ que dejan a su derecha un área $\alpha$ en una $t$ con $\nu$ grados de libertad.

| $\nu$ | 0.4 | 0.25 | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0025 | 0.001 | 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.325 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 127.32 | 318.31 | 636.62 |
| 2 | 0.289 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 14.089 | 22.326 | 31.598 |
| 3 | 0.277 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 7.453 | 10.213 | 12.924 |
| 4 | 0.271 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | 0.267 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | 0.265 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | 0.263 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | 0.262 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | 0.261 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | 0.260 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | 0.260 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | 0.259 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | 0.259 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | 0.258 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | 0.258 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | 0.258 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | 0.257 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | 0.257 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.197 | 3.610 | 3.922 |
| 19 | 0.257 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | 0.257 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | 0.257 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | 0.256 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23 | 0.256 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.104 | 3.485 | 3.767 |
| 24 | 0.256 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25 | 0.256 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26 | 0.256 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27 | 0.256 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28 | 0.256 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29 | 0.256 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30 | 0.256 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |
| 40 | 0.255 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 2.971 | 3.307 | 3.551 |
| 60 | 0.254 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 2.915 | 3.232 | 3.460 |
| 120 | 0.254 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 2.860 | 3.160 | 3.373 |
| $\infty$ | 0.253 | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 2.807 | 3.090 | 3.291 |

**Tabla V.** Valores críticos de la distribución $\chi^2$ de Pearson:

Abcisas $\chi^2_{\alpha;\nu}$ que dejan a su derecha un área $\alpha$ bajo la $\chi^2$ con $\nu$ grados de libertad.

| $\nu$ | 0.995 | 0.99 | 0.975 | 0.95 | 0.9 | 0.75 | 0.5 | 0.25 | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | 0.016 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 | 10.8 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 | 13.8 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 | 16.3 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 | 18.5 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 | 20.5 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 | 22.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 | 24.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 | 26.1 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 | 21.9 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 | 29.6 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 | 31.3 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 | 32.9 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 | 34.5 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 | 36.1 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 | 37.7 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 | 39.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 | 40.8 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 | 42.3 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 | 43.8 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 | 45.3 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 | 46.8 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 | 48.3 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 | 49.7 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 | 51.2 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 | 52.6 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 | 54.1 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 | 55.5 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 | 56.9 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 | 58.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 | 59.7 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}25$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 5.83 | 7.50 | 8.20 | 8.58 | 8.82 | 8.98 | 9.10 | 9.19 | 9.26 | 9.32 | 9.41 | 9.49 | 9.58 | 9.63 | 9.67 | 9.71 | 9.76 | 9.80 | 9.85 |
| 2 | 2.57 | 3.00 | 3.15 | 3.23 | 3.28 | 3.31 | 3.34 | 3.35 | 3.37 | 3.38 | 3.39 | 3.41 | 3.43 | 3.43 | 3.44 | 3.45 | 3.46 | 3.47 | 3.48 |
| 3 | 2.02 | 2.28 | 2.36 | 2.39 | 2.41 | 2.42 | 2.43 | 2.44 | 2.44 | 2.44 | 2.45 | 2.46 | 2.46 | 2,46 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 |
| 4 | 1.81 | 2.00 | 2.05 | 2.06 | 2.07 | 2.08 | 2.08 | 2.09 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 |
| 5 | 1.69 | 1.85 | 1.88 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.88 | 1.88 | 1.88 | 1.88 | 1.87 | 1.87 | 1.87 |
| 6 | 1.62 | 1.76 | 1.78 | 1.79 | 1.79 | 1.78 | 1.78 | 1.78 | 1.77 | 1.77 | 1.77 | 1.76 | 1.76 | 1.75 | 1.75 | 1.75 | 1.74 | 1.74 | 1.74 |
| 7 | 1.57 | 1,70 | 1.72 | 1.72 | 1.71 | 1.71 | 1.70 | 1.70 | 1.69 | 1.69 | 1.68 | 1.68 | 1.67 | 1.67 | 1.66 | 1.66 | 1.65 | 1.65 | 1.65 |
| 8 | 1.54 | 1.66 | 1.67 | 1.66 | 1.66 | 1.65 | 1.64 | 1.64 | 1.63 | 1.63 | 1.62 | 1.62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.58 |
| 9 | 1.51 | 1.62 | 1.63 | 1.63 | 1,62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.57 | 1.56 | 1.56 | 1.55 | 1.54 | 1.54 | 1.53 | 1.53 |
| l0 | 1.49 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.57 | 1.56 | 1.56 | 1.55 | 1.54 | 1.53 | 1.52 | 1.52 | 1.51 | 1.51 | 1.50 | 1.49 | 1.48 |
| 11 | 1.47 | 1.58 | 1.58 | 1.57 | 1.56 | 1.55 | 1.54 | 1.53 | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 | 1.46 | 1.45 |
| 12 | 1.46 | 1.56 | 1.56 | 1.55 | 1.54 | 1.53 | 1.52 | 1,51 | 1.51 | 1.50 | 1.49 | 1.48 | 1.47 | 1.46 | 1.45 | 1.45 | 1.44 | 1.43 | 1.42 |
| 13 | 1.45 | 1.55 | 1.55 | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 | 1.41 | 1.40 |
| 14 | 1.44 | 1.53 | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.48 | 1.47 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.41 | 1.41 | 1.40 | 1,39 | 1.38 |
| 15 | 1.43 | 1.52 | 1.52 | 1.51 | 1.49 | 1.48 | 1.47 | 1.46 | 1.46 | 1.45 | 1.44 | 1.43 | 1.41 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 |
| 16 | 1.42 | 1.51 | 1.51 | 1.50 | 1.48 | 1.47 | 1.46 | 1.45 | 1.44 | l.44 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 |
| 17 | 1.42 | 1.51 | 1,50 | 1.49 | 1.47 | 1.46 | 1.45 | 1.44 | 1.43 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 |
| 18 | 1.41 | 1.50 | 1.49 | 1.48 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 |
| 19 | 1.41 | 1.49 | 1.49 | 1.47 | 1.46 | l.44 | 1.43 | 1.42 | 1.41 | 1.41 | 1.40 | 1.38 | 1,37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.30 |
| 20 | 1.40 | 1.49 | 1.48 | 1.47 | 1.45 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.29 |
| 21 | 1.40 | 1.48 | 1.48 | 1.46 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.28 |
| 22 | 1.40 | 1.48 | 1.47 | 1.45 | 1.44 | 1.42 | 1.41 | 1.40 | 1.39 | 1.39 | 1.37 | 1.36 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 |
| 23 | 1.39 | 1.47 | 1.47 | 1.45 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.28 | 1.27 |
| 24 | 1,39 | 1.47 | 1.46 | 1.44 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.38 | 1.36 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1.26 |
| 25 | 1,39 | 1.47 | 1.46 | 1.44 | 1.42 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.34 | 1.33 | 1.32 | 1.31 | 1.29 | 1.28 | 1.27 | 1.25 |
| 26 | 1.38 | 1.46 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.37 | 1.35 | 1.34 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1,26 | 1.25 |
| 27 | 1.38 | 1.46 | 1.45 | 1.43 | 1.42 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.28 | 1.27 | 1.26 | 1.24 |
| 28 | 1.38 | 1.46 | 1.45 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.34 | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.25 | 1.24 |
| 29 | 1.38 | 1.45 | 1.45 | 1.43 | 1.41 | 1.40 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.32 | 1.31 | 1.30 | 1.29 | 1.27 | 1.26 | 1.25 | 1.23 |
| 30 | 1.38 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.32 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.24 | 1.23 |
| 40 | 1.36 | 1.44 | 1.42 | 1.40 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.31 | 1.30 | 1.28 | 1.26 | 1.25 | 1.24 | 1.22 | 1.21 | 1.19 |
| 60 | 1.35 | 1.42 | 1.41 | 1.38 | 1.37 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.27 | 1.25 | 1.24 | 1.22 | 1.21 | 1.19 | 1.17 | 1.15 |
| 120 | 1.34 | 1.40 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.26 | 1.24 | 1.22 | 1.21 | 1.19 | 1.18 | 1.16 | 1.13 | 1.10 |
| $\infty$ | 1.32 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.29 | 1.28 | 1.27 | 1.25 | 1.24 | 1.22 | 1.19 | 1.18 | 1.16 | 1.14 | 1.12 | 1.08 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}10$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 39.86 | 49.50 | 53.59 | 55.83 | 57.24 | 58.20 | 58.91 | 59.44 | 59.86 | 60.19 | 60.71 | 61.22 | 61.74 | 62.00 | 62.26 | 62.53 | 62.79 | 63.06 | 63.33 |
| 2 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 | 9.39 | 9.41 | 9.42 | 9.44 | 9.45 | 9.46 | 9.47 | 9.47 | 9.48 | 9.49 |
| 3 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 | 5.23 | 5.22 | 5.20 | 5.18 | 5.18 | 5.17 | 5.16 | 5.15 | 5.14 | 5.13 |
| 4 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 | 3.92 | 3.90 | 3.87 | 3.84 | 3.83 | 3.82 | 3.80 | 3.79 | 3.78 | 3.76 |
| 5 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 | 3.30 | 3.27 | 3.24 | 3.21 | 3.19 | 3.17 | 3.16 | 3.14 | 3.12 | 3.10 |
| 6 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 | 2.94 | 2.90 | 2.87 | 2.84 | 2.82 | 2.80 | 2.78 | 2.76 | 2.74 | 2.72 |
| 7 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 | 2.70 | 2.67 | 2.63 | 2.59 | 2.58 | 2.56 | 2.54 | 2.51 | 2.49 | 2.47 |
| 8 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 | 2.54 | 2.50 | 2.46 | 2.42 | 2.40 | 2.38 | 2.36 | 2.34 | 2.32 | 2.29 |
| 9 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 | 2.42 | 2.38 | 2.34 | 2.30 | 2.28 | 2.25 | 2.23 | 2.21 | 2.18 | 2.16 |
| l0 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 | 2.28 | 2.24 | 2.20 | 2.18 | 2.16 | 2.13 | 2.11 | 2.08 | 2.06 |
| 11 | 3.23 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 | 2.21 | 2.17 | 2.12 | 2.10 | 2.08 | 2.05 | 2.03 | 2.00 | 1.97 |
| 12 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.21 | 2.19 | 2.15 | 2.10 | 2.06 | 2.04 | 2.01 | 1.99 | 1.96 | 1.93 | 1.90 |
| 13 | 3.14 | 2.76 | 2.56 | 2.43 | 2.35 | 2.28 | 2.23 | 2.20 | 2.16 | 2.14 | 2.10 | 2.05 | 2.01 | 1.98 | 1.96 | 1.93 | 1.90 | 1.88 | 1.85 |
| 14 | 3.10 | 2.73 | 2.52 | 2.39 | 2.31 | 2.24 | 2.19 | 2.15 | 2.12 | 2.10 | 2.05 | 2.01 | 1.96 | 1.94 | 1.91 | 1.89 | 1.86 | 1.83 | 1.80 |
| 15 | 3.07 | 2.70 | 2.49 | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.09 | 2.06 | 2.02 | 1.97 | 1.92 | 1,90 | 1.87 | 1.85 | 1.82 | 1.79 | 1.76 |
| 16 | 3.05 | 2.67 | 2.46 | 2.33 | 2.24 | 2.18 | 2.13 | 2.09 | 2.06 | 2.03 | 1.99 | 1.94 | 1.89 | 1.87 | 1.84 | 1.81 | 1.78 | 1.75 | 1.72 |
| 17 | 3.03 | 2.64 | 2.44 | 2.31 | 2.22 | 2.15 | 2.10 | 2.06 | 2.03 | 2.00 | 1.96 | 1.91 | 1.86 | 1.84 | 1.81 | 1.78 | 1.75 | 1.72 | 1.69 |
| 18 | 3.01 | 2.62 | 2.42 | 2.29 | 2.20 | 2.13 | 2.08 | 2.04 | 2.00 | 1.98 | 1.93 | 1.89 | 1.84 | 1.81 | 1.78 | 1.75 | 1.72 | 1.69 | 1.66 |
| 19 | 2.99 | 2.61 | 2.40 | 2.27 | 2.18 | 2.11 | 2.06 | 2.02 | 1.98 | 1.96 | 1.91 | 1.86 | 1.81 | 1.79 | 1.76 | 1.73 | 1.70 | 1,67 | 1.63 |
| 20 | 2.97 | 2.59 | 2.38 | 2.25 | 2.16 | 2.09 | 2.04 | 2.00 | 1.96 | 1.94 | 1.89 | 1.84 | 1.79 | 1.77 | 1.74 | 1.71 | 1.68 | 1.64 | 1.61 |
| 21 | 2.96 | 2.57 | 2.36 | 2.23 | 2.14 | 2.08 | 2.02 | 1.98 | 1.95 | 1.92 | 1.87 | 1.83 | 1.78 | 1.75 | 1.72 | 1.69 | 1.66 | 1.62 | 1.59 |
| 22 | 2.95 | 2.56 | 2.35 | 2.22 | 2.13 | 2.06 | 2.01 | 1.97 | 1.93 | 1.90 | 1.86 | 1.81 | 1.76 | 1.73 | 1.70 | 1.67 | 1.64 | 1.60 | 1.57 |
| 23 | 2.94 | 2.55 | 2.34 | 2.21 | 2.11 | 2.05 | 1.99 | 1.95 | 1.92 | 1.89 | 1.84 | 1.80 | 1.74 | 1.72 | 1.69 | 1.66 | 1.62 | 1.59 | 1.55 |
| 24 | 2.93 | 2.54 | 2.33 | 2.19 | 2.10 | 2.04 | 1.98 | 1.94 | 1.91 | 1.88 | 1.83 | 1.78 | 1.73 | 1.70 | 1.67 | 1.64 | 1.61 | 1.57 | 1.53 |
| 25 | 2.92 | 2.53 | 2.32 | 2.18 | 2.09 | 2.02 | 1.97 | 1.93 | 1.89 | 1.87 | 1.82 | 1.77 | 1.72 | 1.69 | 1.66 | 1.63 | 1.59 | 1.56 | 1.52 |
| 26 | 2.91 | 2.52 | 2.31 | 2.17 | 2.08 | 2.01 | 1.96 | 1.92 | 1.88 | 1.86 | 1.81 | 1.76 | 1.71 | 1.68 | 1.65 | 1.61 | 1.58 | 1.54 | 1.50 |
| 27 | 2.90 | 2.51 | 2.30 | 2.17 | 2.07 | 2.00 | 1.95 | 1.91 | 1.87 | 1.85 | 1.80 | 1.75 | 1.70 | 1.67 | 1.64 | 1.60 | 1.57 | 1.53 | 1.49 |
| 28 | 2.89 | 2.50 | 2.29 | 2.16 | 2.06 | 2.00 | 1.94 | 1.90 | 1.87 | 1.84 | 1.79 | 1.74 | 1.69 | 1.66 | 1.63 | 1.59 | 1.56 | 1.52 | 1.48 |
| 29 | 2.89 | 2.50 | 2.28 | 2.15 | 2.06 | 1.99 | 1.93 | 1.89 | 1.86 | 1.83 | 1.78 | 1.73 | 1.68 | 1.65 | 1.62 | 1.58 | 1.55 | 1.51 | 1.47 |
| 30 | 2.88 | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.93 | 1.88 | 1.85 | 1.82 | 1.77 | 1.72 | 1.67 | 1.64 | 1.61 | 1.57 | 1.54 | 1.50 | 1.46 |
| 40 | 2.84 | 2.44 | 2.23 | 2.09 | 2.00 | 1.93 | 1.87 | 1.83 | 1.79 | 1.76 | 1.71 | 1.66 | 1.61 | 1.57 | 1.54 | 1.51 | 1.47 | 1.42 | 1.38 |
| 60 | 2.79 | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.82 | 1.77 | 1.74 | 1.71 | 1.66 | 1.60 | 1.54 | 1.51 | 1.48 | 1.44 | 1.40 | 1.35 | 1.29 |
| 120 | 2.75 | 2.35 | 2.13 | 1.99 | 1.90 | 1.82 | 1.77 | 1.72 | 1.68 | 1.65 | 1.60 | 1.55 | 1.48 | 1.45 | 1.41 | 1.37 | 1.32 | 1.26 | 1.19 |
| $\infty$ | 2.71 | 2.30 | 2.08 | 1.94 | 1.85 | 1.77 | 1.72 | 1.67 | 1.63 | 1.60 | 1.55 | 1.49 | 1.42 | 1.38 | 1.34 | 1.30 | 1.24 | 1.17 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}05$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 | 241.9 | 243.9 | 245.9 | 248.0 | 249.1 | 250.1 | 251.1 | 252.2 | 253.3 | 254.3 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 | 19.41 | 19.43 | 19.45 | 19.45 | 19.46 | 19.47 | 19.48 | 19.49 | 19.50 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.74 | 8.70 | 8.66 | 8,64 | 8.62 | 8.59 | 8.57 | 8.55 | 8.53 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.69 | 5.66 | 5.63 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.43 | 4.40 | 4.36 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.74 | 3.70 | 3.67 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.30 | 3.27 | 3.23 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.28 | 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 3.01 | 2.97 | 2.93 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2 79 | 2.75 | 2.71 |
| l0 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.62 | 2.58 | 2.54 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.79 | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.49 | 2.45 | 2.40 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.38 | 2.34 | 2.30 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 | 2.60 | 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 | 2.21 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 | 2.53 | 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 1.22 | 2.18 | 2.13 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.16 | 2.11 | 2.07 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2,49 | 2.42 | 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.11 | 2.06 | 2.01 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.38 | 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 | 1.96 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.34 | 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.02 | 1.97 | 1.92 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.31 | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 1.98 | 1.93 | 1.88 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.28 | 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.95 | 1.90 | 1.84 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 | 2.25 | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.87 | 1.81 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.23 | 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.78 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 | 2.20 | 2.13 | 2.05 | 2.01 | 1.96 | 1.91 | 1.86 | 1.81 | 1.76 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.79 | 1.73 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.82 | 1.77 | 1.71 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.15 | 2.07 | 1.99 | 1.95 | 1.90 | 1.85 | 1.80 | 1.75 | 1.69 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 | 2.13 | 2.06 | 1.97 | 1.93 | 1.88 | 1.84 | 1.79 | 1.73 | 1.67 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2,12 | 2.04 | 1.96 | 1.91 | 1.87 | 1.82 | 1.77 | 1.71 | 1.65 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 | 2.10 | 2.03 | 1.94 | 1.90 | 1.85 | 1.81 | 1.75 | 1.70 | 1.64 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.74 | 1.68 | 1.62 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.64 | 1.58 | 1.51 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.53 | 1.47 | 1.39 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.09 | 2.02 | 1.96 | 1.91 | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.43 | 1.35 | 1.25 |
| $\infty$ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.75 | 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.32 | 1.22 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}025$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 647.8 | 799.5 | 864.2 | 899.6 | 921.8 | 937.1 | 948.2 | 956.7 | 963.3 | 968.6 | 976.7 | 984.9 | 993.1 | 997.2 | 1001 | 1006 | 1010 | 1014 | 1018 |
| 2 | 38.51 | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 | 39.36 | 39.37 | 39.39 | 39.40 | 39.41 | 39.43 | 39.45 | 39.46 | 39.46 | 39.47 | 39.48 | 39.49 | 39.50 |
| 3 | 17.44 | 16.04 | 15.44 | 15.10 | 14.88 | 14.73 | 14.62 | 14.54 | 14.47 | 14.42 | 14.34 | 14.25 | 14.17 | 14.12 | 14.08 | 14.04 | 13.99 | 13.95 | 13.90 |
| 4 | 12.22 | 10.65 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.90 | 8.84 | 8.75 | 8.66 | 8.56 | 8.51 | 8.46 | 8.41 | 8.36 | 8.31 | 8.26 |
| 5 | 10.01 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.68 | 6.62 | 6.52 | 6.43 | 6.33 | 6.28 | 6.23 | 6.18 | 6.12 | 6.07 | 6.02 |
| 6 | 8.81 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.52 | 5.46 | 5.37 | 5.27 | 5.17 | 5.12 | 5.07 | 5.01 | 4.96 | 4.90 | 4.85 |
| 7 | 8.07 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.82 | 4.76 | 4.67 | 4.57 | 4.47 | 4.42 | 4.36 | 4.31 | 4.25 | 4.20 | 4.14 |
| 8 | 7.57 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.36 | 4.30 | 4.20 | 4.10 | 4.00 | 3.95 | 3.89 | 3.84 | 3.78 | 3.73 | 3.67 |
| 9 | 7.21 | 5.71 | 5.00 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 4.03 | 3.96 | 3.87 | 3.77 | 3.67 | 3.61 | 3.56 | 3.51 | 3.45 | 3.39 | 3.33 |
| l0 | 6.94 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.78 | 3.72 | 3.62 | 3.52 | 3.42 | 3.37 | 3.31 | 3.26 | 3.20 | 3.14 | 3.08 |
| 11 | 6.72 | 5.26 | 4.63 | 4.28 | 4.04 | 3.88 | 3.76 | 3.66 | 3.59 | 3.53 | 3.43 | 3.33 | 3.23 | 3.17 | 3.12 | 3.06 | 3.00 | 2.94 | 2.88 |
| 12 | 6.55 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.44 | 3.37 | 3.28 | 3.18 | 3.07 | 3.02 | 2.96 | 2.91 | 2.85 | 2.79 | 2.72 |
| 13 | 6.41 | 4.97 | 4.35 | 4.00 | 3.77 | 3.60 | 3.48 | 3.39 | 3.31 | 3.25 | 3.15 | 3.05 | 2.95 | 2.89 | 2.84 | 2.78 | 2.72 | 2.66 | 2.60 |
| 14 | 6.30 | 4.86 | 4.24 | 3.89 | 3.66 | 3.50 | 3.38 | 3.29 | 3.21 | 3.15 | 3.05 | 2.95 | 2.84 | 2.79 | 2.73 | 2.67 | 2.61 | 2.55 | 2.49 |
| 15 | 6.20 | 4.77 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 | 3 06 | 2.96 | 2.86 | 2.76 | 2.70 | 2.64 | 2.59 | 2.52 | 2.46 | 2.40 |
| 16 | 6.12 | 4.69 | 4.08 | 3.73 | 3.50 | 3.34 | 3.22 | 3.12 | 3.05 | 2.99 | 2.89 | 2.79 | 2.68 | 2.63 | 2.57 | 2.51 | 2.45 | 2.38 | 2.32 |
| 17 | 6.04 | 4.62 | 4.01 | 3.66 | 3.44 | 3.28 | 3.16 | 3.06 | 2.98 | 2.92 | 2.82 | 2.72 | 2.62 | 2.56 | 2.50 | 2.44 | 2.38 | 2.32 | 2.25 |
| 18 | 5.98 | 4.56 | 3.95 | 3.61 | 3.38 | 3.22 | 3.10 | 3.01 | 2.93 | 2.87 | 2.77 | 2.67 | 2.56 | 2.50 | 2.44 | 2.38 | 2.32 | 2.26 | 2.19 |
| 19 | 5.92 | 4.51 | 3.90 | 3.56 | 3.33 | 3.17 | 3.05 | 2.96 | 2.88 | 2.82 | 2.72 | 2.62 | 2.51 | 2.45 | 2.39 | 2.33 | 2.27 | 2.20 | 2.13 |
| 20 | 5.87 | 4.46 | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 | 2.77 | 2.68 | 2.57 | 2.46 | 2.41 | 2.35 | 2.29 | 2.22 | 2.16 | 2.09 |
| 21 | 5.83 | 4.42 | 3.82 | 3.48 | 3.25 | 3.09 | 2.97 | 2.87 | 2.80 | 2.73 | 2.64 | 2.53 | 2.42 | 2.37 | 2.31 | 2.25 | 2.18 | 2.11 | 2.04 |
| 22 | 5.79 | 4.38 | 3.78 | 3.44 | 3.22 | 3.05 | 2.93 | 2.84 | 2,76 | 2.70 | 2.60 | 2.50 | 2.39 | 2.33 | 2.27 | 2.21 | 2.14 | 2.08 | 2.00 |
| 23 | 5.75 | 4.35 | 3.75 | 3.41 | 3.18 | 3.02 | 2.90 | 2.81 | 2.73 | 2.67 | 2.57 | 2.47 | 2.36 | 2.30 | 2.24 | 2.18 | 2.11 | 2.04 | 1.97 |
| 24 | 5.72 | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 | 2.64 | 2.54 | 2.44 | 2.33 | 2.27 | 2.21 | 2.15 | 2.08 | 2.01 | 1.94 |
| 25 | 5.69 | 4.29 | 3.69 | 3.35 | 3.13 | 2.97 | 2.85 | 2.75 | 2.68 | 2.61 | 2.51 | 2.41 | 2.30 | 2.24 | 2.18 | 2.12 | 2.05 | 1.98 | 1.91 |
| 26 | 5.66 | 4.27 | 3.67 | 3.33 | 3.10 | 2.94 | 2.82 | 2.73 | 2.65 | 2.59 | 2.49 | 2.39 | 2.28 | 2.22 | 2.16 | 2.09 | 2.03 | 1.95 | 1.88 |
| 27 | 5.63 | 4.24 | 3.63 | 3.31 | 3.06 | 2.92 | 2.80 | 2.71 | 2.63 | 2.57 | 2.47 | 2.36 | 2.25 | 2.19 | 2.13 | 2.07 | 2.00 | 1.93 | 1.85 |
| 28 | 5.61 | 4.22 | 3.63 | 3.29 | 3.06 | 2.90 | 2.78 | 2.69 | 2.61 | 2.55 | 2.45 | 2.34 | 2.23 | 2.17 | 2.11 | 2.05 | l.98 | 1.91 | 1.83 |
| 29 | 5.59 | 4.20 | 3.61 | 3.27 | 3.04 | 2.88 | 2.76 | 2.67 | 2.59 | 2.53 | 2.43 | 2.32 | 2.21 | 2.15 | 2.09 | 2.03 | l.96 | 1.89 | 1.81 |
| 30 | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 | 2.41 | 2.31 | 2.20 | 2.14 | 2.07 | 2.01 | 1.94 | 1.87 | 1.79 |
| 40 | 5.42 | 4.05 | 3.46 | 3.13 | 2.90 | 2.74 | 2.62 | 2.53 | 2.45 | 2.39 | 2.29 | 2.18 | 2.07 | 2.01 | 1.94 | 1.88 | 1.80 | 1.72 | 1.64 |
| 60 | 5.29 | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 | 2.27 | 2.17 | 2.06 | 1.94 | 1.88 | 1.82 | 1.74 | 1.67 | 1.58 | 1.48 |
| 120 | 5.15 | 3.80 | 3.23 | 2.89 | 2.67 | 2.52 | 2.39 | 2.30 | 2.22 | 2.16 | 2.05 | 1.94 | 1.82 | 1.76 | 1.69 | 1.61 | 1.53 | 1.43 | 1.31 |
| $\infty$ | 5.02 | 3.69 | 3.12 | 2.79 | 2.57 | 2.41 | 2.29 | 2.19 | 2.11 | 2.05 | 1.94 | 1.83 | 1.71 | 1.64 | 1.57 | 1.48 | 1.39 | 1.27 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}01$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 4052 | 4999.50 | 5403 | 5625 | 5764 | 5859 | 5928 | 5982 | 6022 | 6056 | 6106 | 6157 | 6209 | 6235 | 6261 | 6287 | 6313 | 6339 | 6366 |
| 2 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 | 99.42 | 99.43 | 99.45 | 99.46 | 99.47 | 99.47 | 99.48 | 99.49 | 99.50 |
| 3 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 | 27.05 | 26.87 | 26.69 | 26.60 | 26.50 | 26.41 | 26.32 | 26.22 | 26.13 |
| 4 | 21.20 | 18,00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 | 14.37 | 14.20 | 14.02 | 13.93 | 13.84 | 13.75 | 13.65 | 13.56 | 13.46 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 | 9.89 | 9.72 | 9.55 | 9.47 | 9.38 | 9.29 | 9.20 | 9.11 | 9.02 |
| 6 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.72 | 7.56 | 7.40 | 7.31 | 7.23 | 7.14 | 7.06 | 6.97 | 6.88 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.47 | 6.31 | 6.16 | 6.07 | 5.99 | 5.91 | 5.82 | 5.74 | 5.65 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.67 | 5.52 | 5.36 | 5.28 | 5.20 | 5.12 | 5.03 | 4.95 | 4.86 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.11 | 4.96 | 4.81 | 4.73 | 4.65 | 4.57 | 4.48 | 4.40 | 4.31 |
| l0 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.71 | 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.08 | 4.00 | 3.91 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.40 | 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.78 | 3.69 | 3.60 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.16 | 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.54 | 3.45 | 3.36 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 | 3.96 | 3.82 | 3.66 | 3.59 | 3.51 | 3.43 | 3.34 | 3.25 | 3.17 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.80 | 3.66 | 3.51 | 3.43 | 3.35 | 3.27 | 3.18 | 3.09 | 3.00 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.67 | 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.05 | 2.96 | 2.87 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.55 | 3.41 | 3.26 | 3.18 | 3.10 | 3.02 | 2.93 | 2.84 | 2.75 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.46 | 3.31 | 3.16 | 3.08 | 3.00 | 2.92 | 2.83 | 2.75 | 2.65 |
| 18 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.37 | 3.23 | 3.08 | 3.00 | 2.92 | 2.84 | 2.75 | 2.66 | 2.57 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.30 | 3.15 | 3.00 | 2.92 | 2.84 | 2.76 | 2.67 | 2.58 | 2.49 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3,56 | 3.46 | 3.37 | 3.23 | 3.09 | 2.94 | 2.86 | 2.78 | 2.69 | 2.61 | 2.52 | 2.42 |
| 21 | 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 | 3.17 | 3.03 | 2.88 | 2.80 | 2.72 | 2.64 | 2.55 | 2.46 | 2.36 |
| 22 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.12 | 2.98 | 2.83 | 2.75 | 2.67 | 2.58 | 2.50 | 2.40 | 2.31 |
| 23 | 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3,21 | 3.07 | 2.93 | 2.78 | 2.70 | 2.62 | 2.54 | 2.45 | 2.35 | 2.26 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.03 | 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.40 | 2.31 | 2.21 |
| 25 | 7.77 | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 | 2.99 | 2.85 | 2.70 | 2.62 | 2.54 | 2.45 | 2.36 | 2.27 | 2.17 |
| 26 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 2.96 | 2,81 | 2.66 | 2.58 | 2.50 | 2.42 | 2.33 | 2.23 | 2.13 |
| 27 | 7.68 | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.39 | 3.26 | 3.15 | 3.06 | 2.93 | 2.78 | 2.63 | 2.55 | 2.47 | 2.38 | 2.29 | 2.20 | 2.10 |
| 28 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.90 | 2.75 | 2.60 | 2.52 | 2.44 | 2.35 | 2.26 | 2.17 | 2.06 |
| 29 | 7.60 | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.33 | 3.20 | 3.09 | 3.00 | 2.87 | 2.73 | 2.57 | 2.49 | 2.41 | 2.33 | 2.23 | 2.14 | 2.03 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.84 | 2.70 | 2.55 | 2.47 | 2.39 | 2.30 | 2.21 | 2.11 | 2.01 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.66 | 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.02 | 1.92 | 1.80 |
| 60 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.84 | 1.73 | 1.60 |
| 120 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.34 | 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 1.66 | 1.53 | 1.38 |
| $\infty$ | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.18 | 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.47 | 1.32 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}005$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 16211 | 20000 | 21615 | 22500 | 23056 | 23437 | 23715 | 23925 | 24091 | 24224 | 24426 | 24630 | 24836 | 24940 | 25044 | 25148 | 25253 | 25359 | 25465 |
| 2 | 198.5 | 199.0 | 199.2 | 199.2 | 199.3 | 199.3 | 199.4 | 199.4 | 199.4 | 199.4 | 199.4 | 199.4 | 199.5 | 199.5 | 199.5 | 199.5 | 199.5 | 19.49 | 19.50 |
| 3 | 55.55 | 49.80 | 47.46 | 46.17 | 45.42 | 44.85 | 44.40 | 44.11 | 43.84 | 43.73 | 43.38 | 43.13 | 42.87 | 42.65 | 42.47 | 42.31 | 42.15 | 41.99 | 41.83 |
| 4 | 31.33 | 26.28 | 24.25 | 23.15 | 22.46 | 21.98 | 21.62 | 21.35 | 21.14 | 20.97 | 20.70 | 20.43 | 20.17 | 20.04 | 19.89 | 19.75 | 19.61 | 19.47 | 19.32 |
| 5 | 22.78 | 18.31 | 16.53 | 15.56 | 14.94 | 14.51 | 14.20 | 13.96 | 13.77 | 13.62 | 13.39 | 13.15 | 12.90 | 12.78 | 12.66 | 12.53 | 12.40 | 12.27 | 12.14 |
| 6 | 18.64 | 14.54 | 12.92 | 12.03 | 11.46 | 11.07 | 10.79 | 10.57 | 10.39 | 10.25 | 10.03 | 9.81 | 9.59 | 9.48 | 9.36 | 9.24 | 9.12 | 9.00 | 8.88 |
| 7 | 16.24 | 12.40 | 10.88 | 10.05 | 9.52 | 9.16 | 8.89 | 6.68 | 8.51 | 8.38 | 8.18 | 7.97 | 7.75 | 7.64 | 7.53 | 7.42 | 7.31 | 7.19 | 7.08 |
| 8 | 14.69 | 11.04 | 9.60 | 8.81 | 8.30 | 7.95 | 7.69 | 7.50 | 7.34 | 7.21 | 7.02 | 6.82 | 6.61 | 6.50 | 6.40 | 6.29 | 6.18 | 6.06 | 5.95 |
| 9 | 13.61 | 10.11 | 8.72 | 7.96 | 7.47 | 7.13 | 6.89 | 6.69 | 6.54 | 6.42 | 6.23 | 6.03 | 5.83 | 5.73 | 5.62 | 5.52 | 5.41 | 5.30 | 5.19 |
| 10 | 12.83 | 9.43 | 8.08 | 7.34 | 6.87 | 6.54 | 6.30 | 6.12 | 5.97 | 5.85 | 5.66 | 5.47 | 5.27 | 5.17 | 5.07 | 4.97 | 4.86 | 4.75 | 4.64 |
| 11 | 12.23 | 8.91 | 7.60 | 6.88 | 6.42 | 6.10 | 5.86 | 5.68 | 5.54 | 5.42 | 5.24 | 5.05 | 4.86 | 4.76 | 4.65 | 4.55 | 4.45 | 4.34 | 4.23 |
| 12 | 11.75 | 8.51 | 7.23 | 6.52 | 6.07 | 5.76 | 5.52 | 5.34 | 5.20 | 5.09 | 4.91 | 4.72 | 4.53 | 4.43 | 4.33 | 4.23 | 4.12 | 4.02 | 3.90 |
| 13 | 11.37 | 8.19 | 6.93 | 6.23 | 5.79 | 5.48 | 5.25 | 5.08 | 4.94 | 4.82 | 4.64 | 4.46 | 4.27 | 4.17 | 4.07 | 3.97 | 3.87 | 3.76 | 3.65 |
| 14 | 11.06 | 7.92 | 6.68 | 6.00 | 5.56 | 5.26 | 5.03 | 4.86 | 4.72 | 4.60 | 4.43 | 4.25 | 4.06 | 3.96 | 3.86 | 3.76 | 3.86 | 3.55 | 3.44 |
| 15 | 10.80 | 7.70 | 6.48 | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 | 4.42 | 4.25 | 4.07 | 3.88 | 3.79 | 3.69 | 3.58 | 3.48 | 3.37 | 3.26 |
| 16 | 10.58 | 7.51 | 6.30 | 5.64 | 5.21 | 4.91 | 4.69 | 4.52 | 4.38 | 4.27 | 4.10 | 3.92 | 3.73 | 3.64 | 3.54 | 3.44 | 3.33 | 3.22 | 3.11 |
| 17 | 10.38 | 7.35 | 6.16 | 5.50 | 5.07 | 4.78 | 4.56 | 4.39 | 4.25 | 4.14 | 3.97 | 3.79 | 3.61 | 3.51 | 3.41 | 3.31 | 3.21 | 3.10 | 2.98 |
| 18 | 10.22 | 7.21 | 6.06 | 5.37 | 4.96 | 4.66 | 4.44 | 4.28 | 4.14 | 4.03 | 3.86 | 3.68 | 3.50 | 3.40 | 3.30 | 3.20 | 3.10 | 2.99 | 2.87 |
| 19 | 10.07 | 7.09 | 5.92 | 5.27 | 4.85 | 4.56 | 4.34 | 4.18 | 4.04 | 3.93 | 3.76 | 3.59 | 3.40 | 3.31 | 3.21 | 3.11 | 3.00 | 2.89 | 2.78 |
| 20 | 9.94 | 6.99 | 5.82 | 5.17 | 4.76 | 4.47 | 4.26 | 4.09 | 3.96 | 3.85 | 3.68 | 3.50 | 3.32 | 3.22 | 3.12 | 3.02 | 2.92 | 2.81 | 2.69 |
| 21 | 9.83 | 6.89 | 5.73 | 5.09 | 4.68 | 4.39 | 4.18 | 4.01 | 3.88 | 3.77 | 3.60 | 3.43 | 3.24 | 3.15 | 3.05 | 2.95 | 2.84 | 2.73 | 2.61 |
| 22 | 9.73 | 6.81 | 5.65 | 5.02 | 4.61 | 4.32 | 4.11 | 3.94 | 3.81 | 3.70 | 3.53 | 3.36 | 3.18 | 3.08 | 2.98 | 2.88 | 2.77 | 2.66 | 2.55 |
| 23 | 9.63 | 6.73 | 5.58 | 4.95 | 4.54 | 4.26 | 4.05 | 3.88 | 3.75 | 3.64 | 3.47 | 3.30 | 3.12 | 3.02 | 2.92 | 2.82 | 2.71 | 2.60 | 2.48 |
| 24 | 9.55 | 6.66 | 5.52 | 4.89 | 4.49 | 4.20 | 3.99 | 3.83 | 3.69 | 3.59 | 3.42 | 3.25 | 3.06 | 2.97 | 2.87 | 2.77 | 2.66 | 2.55 | 2.43 |
| 25 | 9.48 | 6.60 | 5.46 | 4.84 | 4.43 | 4.15 | 3.94 | 3.78 | 3.64 | 3.54 | 3.37 | 3.20 | 3.01 | 2.92 | 2.82 | 2.72 | 2.61 | 2.50 | 2.38 |
| 26 | 9.41 | 6.54 | 5.41 | 4.79 | 4.38 | 4.10 | 3.89 | 3.73 | 3.60 | 3.49 | 3.33 | 3.15 | 2.97 | 2.87 | 2.77 | 2.67 | 2.56 | 2.45 | 2.33 |
| 27 | 9.34 | 6.49 | 5.36 | 4.74 | 4.34 | 4.06 | 3.85 | 3.69 | 3.56 | 3.45 | 3.28 | 3.11 | 2.93 | 2.83 | 2.73 | 2.63 | 2.52 | 2.41 | 2.29 |
| 28 | 9.28 | 6.44 | 5.32 | 4.70 | 4.30 | 4.02 | 3.81 | 3.65 | 3.52 | 3.41 | 3.25 | 3.07 | 2.89 | 2.79 | 2.69 | 2.59 | 2.48 | 2.37 | 2.25 |
| 29 | 9.23 | 6.40 | 5.28 | 4.66 | 4.26 | 3.98 | 3.77 | 3.61 | 3.48 | 3.38 | 3.21 | 3.04 | 2.86 | 2.76 | 2.66 | 2.56 | 2.45 | 2.33 | 2.21 |
| 30 | 9.18 | 6.35 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 | 3.34 | 3.18 | 3.01 | 2.82 | 2.73 | 2.63 | 2.52 | 2.42 | 2.30 | 2.18 |
| 40 | 8.83 | 6.07 | 4.94 | 4.37 | 3.99 | 3.71 | 3.51 | 3.35 | 3.22 | 3.12 | 2.95 | 2.78 | 2.60 | 2.50 | 2.40 | 2.30 | 2.18 | 2.06 | 1.93 |
| 60 | 8.49 | 5.79 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 | 2.90 | 2.74 | 2.57 | 2.39 | 2.29 | 2.19 | 2.08 | 1.96 | 1.83 | 1.69 |
| 120 | 8.18 | 5.54 | 4.50 | 3.92 | 3.55 | 3.28 | 3.09 | 2.93 | 2.81 | 2.71 | 2.54 | 2.37 | 2.19 | 2.09 | 1.98 | 1.87 | 1.75 | 1.61 | 1.43 |
| $\infty$ | 7.88 | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 | 2.52 | 2.36 | 2.19 | 2.00 | 1.90 | 1.79 | 1.67 | 1.53 | 1.36 | 1.00 |

**Tabla VI.** Valores críticos de la distribución $F$ de Snedecor:

Abcisas $F_{\alpha;\nu_1,\nu_2}$ que dejan a su derecha un área $\alpha$ bajo la $F$ con $\nu_1$ y $\nu_2$ grados de libertad.

$\alpha = 0{,}001$

| | $\nu_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 2 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 4053* | 5000* | 5404* | 5625* | 5764* | 5859* | 5929* | 5981* | 6023* | 6056* | 6107* | 6158* | 6209* | 6235* | 6261* | 6287* | 6313* | 6340* | 6366* |
| 2 | 998.5 | 999.0 | 999.2 | 999.2 | 999.3 | 999.3 | 999.4 | 999.4 | 999.4 | 999.4 | 999.4 | 999.4 | 999.4 | 999.5 | 999.5 | 999.5 | 999.5 | 999.5 | 999.5 |
| 3 | 167.0 | 148.5 | 141.1 | 137.1 | 134.6 | 132.8 | 131.6 | 130.6 | 129.9 | 129.2 | 128.3 | 127.4 | 126.4 | 125.9 | 125.4 | 125.0 | 124.5 | 124.0 | 123.5 |
| 4 | 74.14 | 61.25 | 56.18 | 53.44 | 51.71 | 50.53 | 49.66 | 49.00 | 48.47 | 48.05 | 47.41 | 46.76 | 46.10 | 45.77 | 45.43 | 45.09 | 44.75 | 44.40 | 44.05 |
| 5 | 47.18 | 37.12 | 33.20 | 31.09 | 29.75 | 28.84 | 28.16 | 27.64 | 27.24 | 26.92 | 26.42 | 25.91 | 25.39 | 25.14 | 24.87 | 24.60 | 24.33 | 24.06 | 23.79 |
| 6 | 35.51 | 27.00 | 23.70 | 21.92 | 20.81 | 20.03 | 19.46 | 19.03 | 18.69 | 18.41 | 17.99 | 17.56 | 17.12 | 16.89 | 16.67 | 16.44 | 16.21 | 15.99 | 15.75 |
| 7 | 29.25 | 21.69 | 18.77 | 17.19 | 16.21 | 15.52 | 15.02 | 14.63 | 14.33 | 14.08 | 13.71 | 13.32 | 12.93 | 12.73 | 12.53 | 12.33 | 12.12 | 11.91 | 11.70 |
| 8 | 25.42 | 18.49 | 15.83 | 14.39 | 13.49 | 12.86 | 12.40 | 12.04 | 11.77 | 11.54 | 11.19 | 10.84 | 10.48 | 10.30 | 10.11 | 9.92 | 9.73 | 9.53 | 9.33 |
| 9 | 22.86 | 16.39 | 13.90 | 12.56 | 11.71 | 11.13 | 10.70 | 10.37 | 10.11 | 9.89 | 9.57 | 9.24 | 8.90 | 8.72 | 8.55 | 8.37 | 8.19 | 8.00 | 7.81 |
| 10 | 21.04 | 14.91 | 12.55 | 11.28 | 10.48 | 9.92 | 9.52 | 9.20 | 8.96 | 8.75 | 8.45 | 8.13 | 7.80 | 7.64 | 7.47 | 7.30 | 7.12 | 6.94 | 6.76 |
| 11 | 19.69 | 13.81 | 11.56 | 10.35 | 9.58 | 9.05 | 8.66 | 8.35 | 8.12 | 7.92 | 7.63 | 7.32 | 7.01 | 6.85 | 6.68 | 6.52 | 6.35 | 6.17 | 6.00 |
| 12 | 18.64 | 12.97 | 10.80 | 9.63 | 8.89 | 8.38 | 8.00 | 7.71 | 7.48 | 7.29 | 7.00 | 6.71 | 6.40 | 6.25 | 6.09 | 5.93 | 5.76 | 5.59 | 5.42 |
| 13 | 17.81 | 12.31 | 10.21 | 9.07 | 8.35 | 7.86 | 7.49 | 7.21 | 6.98 | 6.80 | 6.52 | 6.23 | 5.93 | 5.78 | 5.63 | 5.47 | 5.30 | 5.14 | 4.97 |
| 14 | 17.14 | 11.78 | 9.73 | 8.62 | 7.92 | 7.43 | 7.08 | 6.80 | 6.58 | 6.40 | 6.13 | 5.85 | 5.56 | 5.41 | 5.25 | 5.10 | 4.94 | 4.77 | 4.60 |
| 15 | 16.59 | 11.34 | 9.34 | 8.25 | 7.57 | 7.09 | 6.74 | 6.47 | 6.26 | 6.08 | 5.81 | 5.54 | 5.25 | 5.10 | 4.95 | 4.80 | 4.64 | 4.47 | 4.31 |
| 16 | 16.12 | 10.97 | 9.00 | 7.94 | 7.27 | 6.81 | 6.46 | 6.19 | 5.98 | 5.81 | 5.55 | 5.27 | 4.99 | 4.85 | 4.70 | 4.54 | 4.39 | 4.23 | 4.06 |
| 17 | 15.72 | 10.66 | 8.73 | 7.68 | 7.02 | 6.56 | 6.22 | 5.96 | 5.75 | 5.58 | 5.32 | 5.05 | 4.78 | 4.63 | 4.48 | 4.33 | 4.18 | 4.02 | 3.85 |
| 18 | 15.38 | 10.39 | 8.49 | 7.46 | 6.81 | 6.35 | 6.02 | 5.76 | 5.56 | 5.39 | 5.13 | 4.87 | 4.59 | 4.45 | 4.30 | 4.15 | 4.00 | 3.84 | 3.67 |
| 19 | 15.08 | 10.16 | 8.28 | 7.26 | 6.62 | 6.18 | 5.85 | 5.59 | 5.39 | 5.22 | 4.97 | 4.70 | 4.43 | 4.29 | 4.14 | 3.99 | 3.84 | 3.68 | 3.51 |
| 20 | 14.82 | 9.95 | 8.10 | 7.10 | 6.46 | 6.02 | 5.69 | 5.44 | 5.24 | 5.08 | 4.82 | 4.56 | 4.29 | 4.15 | 4.00 | 3.86 | 3.70 | 3.54 | 3.38 |
| 21 | 14.59 | 9.77 | 7,94 | 6.95 | 6.32 | 5.88 | 5.56 | 5.31 | 5.11 | 4.95 | 4.70 | 4.44 | 4.17 | 4.03 | 3.88 | 3.74 | 3.58 | 3.42 | 3.26 |
| 22 | 14.38 | 9.61 | 7.80 | 6.81 | 6.19 | 5.76 | 5.44 | 5.19 | 4.99 | 4.83 | 4.58 | 4.33 | 4.06 | 3.92 | 3.78 | 3.63 | 3.48 | 3.32 | 3.15 |
| 23 | 14.19 | 9.47 | 7.67 | 6.69 | 6.08 | 5.65 | 5.33 | 5.09 | 4.89 | 4.73 | 4.48 | 4.23 | 3.96 | 3.82 | 3.68 | 3.53 | 3.38 | 3.22 | 3.05 |
| 24 | 14.03 | 9.34 | 7.55 | 6.59 | 5.98 | 5.55 | 5.23 | 4.99 | 4.80 | 4.64 | 4.39 | 4.14 | 3.87 | 3.74 | 3.59 | 3.45 | 3.29 | 3.14 | 2.97 |
| 25 | 13.88 | 9.22 | 7.45 | 6.49 | 5.88 | 5.46 | 5.15 | 4.91 | 4.71 | 4.56 | 4.31 | 4.06 | 3.79 | 3.66 | 3.52 | 3.37 | 3.22 | 3.06 | 2.89 |
| 26 | 13.74 | 9.12 | 7.36 | 6.41 | 5.80 | 5.38 | 5.07 | 4.83 | 4.64 | 4.48 | 4.24 | 3.99 | 3.72 | 3.59 | 3.44 | 3.30 | 3.15 | 2.99 | 2.82 |
| 27 | 13.61 | 9.02 | 7.27 | 6.33 | 5.73 | 5.31 | 5.00 | 4.76 | 4.57 | 4.41 | 4.17 | 3.92 | 3.66 | 3.52 | 3.38 | 3.23 | 3.08 | 2.92 | 2.75 |
| 28 | 13.50 | 8.93 | 7.19 | 6.25 | 5.66 | 5.24 | 4.93 | 4.69 | 4.50 | 4.35 | 4.11 | 3.86 | 3.60 | 3.46 | 3.32 | 3.18 | 3.02 | 2.86 | 2.69 |
| 29 | 13.39 | 8.85 | 7.12 | 6.19 | 5.59 | 5.18 | 4.87 | 4.64 | 4.45 | 4.29 | 4.05 | 3.80 | 3.54 | 3.41 | 3.27 | 3.12 | 2.97 | 2,81 | 2.64 |
| 30 | 13.29 | 8.77 | 7,05 | 6.12 | 5.53 | 5.12 | 4.82 | 4.58 | 4.39 | 4.24 | 4.00 | 3.75 | 3.49 | 3.36 | 3.22 | 3.07 | 2.92 | 2.76 | 2.59 |
| 40 | 12.61 | 8.25 | 6.60 | 5.70 | 5.13 | 4.73 | 4.44 | 4.21 | 4.02 | 3.87 | 3.64 | 3.40 | 3.15 | 3.01 | 2.87 | 2.73 | 2.57 | 2.41 | 2.23 |
| 60 | 11.97 | 7.76 | 6.17 | 5.31 | 4.76 | 4.37 | 4.09 | 3.87 | 3.69 | 3.54 | 3.31 | 3,08 | 2.83 | 2.69 | 2.55 | 2.41 | 2.25 | 2.08 | 1.89 |
| 120 | 11.38 | 7.32 | 5.79 | 4.95 | 4.42 | 4.04 | 3.77 | 3.55 | 3.38 | 3.24 | 3.02 | 2.78 | 2.53 | 2.40 | 2.26 | 2.11 | 1.95 | 1.76 | 1.54 |
| $\infty$ | 10.83 | 6.91 | 5.42 | 4.621 | 4.10 | 3.74 | 3.47 | 3.27 | 3.10 | 2.96 | 2.74 | 2.51 | 2.27 | 2.13 | 1.99 | 1.84 | 1.66 | 1.45 | 1.00 |

* Multiplicar estas celdas por 100

# Matemática 3

**<u>Practica Nº 1</u>:** Espacios muestrales y eventos - Asignación de probabilidades.

1) Un experimento implica lanzar un par de dados, uno verde y uno rojo, y registrar los números que salen. Si $x$ es igual al resultado en el dado verde e $y$ es el resultado en el dado rojo, describa el espacio muestral S
   a) por extensión
   b) por comprensión

2) Un experimento consiste en lanzar un dado y después lanzar una moneda una vez, si el número en el dado es par. Si el número en el dado es impar, la moneda se lanza dos veces. Use la notación 4C, por ejemplo, para denotar el resultado de que el dado muestre 4 y después la moneda salga cara, y 3CS para denotar el resultado de que el dado muestre 3 seguido por una cara y después por una ceca. Construya un diagrama de árbol para mostrar los 18 elementos del espacio muestral S.

3) Se seleccionan al azar cuatro estudiantes de una clase de química y se clasifican como femenino o masculino.
   a) Liste los elementos del espacio muestras $S_1$ usando la letra F para femenino y la letra M para masculino.
   b) Defina un segundo espacio muestral $S_2$ donde los elementos representen el número de mujeres seleccionadas.

4) Para el espacio muestral del ejercicio 1) liste los elementos del eventos:
   a) A: “la suma de los números es mayor que 8”
   b) B: “ocurre un dos en cualquiera de los dos dados”
   c) C: “sale un número mayor que cuatro en el dado verde”
   d) A I C
   e) A I B
   f) B I C

5) Para el espacio muestral del ejercicio 2) liste los elementos del eventos:
   a) A: “en el dado sale un número menor que 3”
   b) B: “ocurren dos cecas”
   c) $A^C$
   d) $A^C$ I B
   e) A Y B

6) Suponga que los dos dados del ejercicio 1) son normales. Entonces cada resultado del espacio muestral S tienen la misma probabilidad de ocurrir (S es equiprobable).
   Encuentre las siguientes probabilidades:
   a) P(A) ; b) P(B) ; P(C) ; P(A I C )

7) Si se toman 3 libros al azar de un estante que contiene 5 novelas, 3 libros de poemas y 1 diccionario, ¿cuál es la probabilidad de que
   a) se seleccione el diccionario?
   b) se seleccionen 2 novelas y 1 libro de poemas?

8) Un dado octaedro (de ocho caras) tiene el número 1 pintado en dos de sus caras, el 2 en tres de sus caras, el 3 en dos de sus caras y el 4 en una cara. Se lanza el dado. Suponga que cada cara tiene la misma probabilidad de salir.
   a) Determine el espacio muestral de este experimento.

b) Calcular la probabilidad de que salga número par.
**c**) Si el dado estuviera cargado de tal forma que la cara con el número 4 tuviera el doble de probabilidad de salir que cada una de las otras siete caras
c1) ¿cambiaría esto el espacio muestral? Explique.
c2) ¿cambiaría esto la probabilidad de que salga número par? Explique.

9) Se lanzan un dado normal 5 veces. Encuentre la probabilidad de obtener 4 números iguales.

10) Se selecciona una carta al azar entre 50 cartas numeradas de 1 a 50.
Hallar la probabilidad de que el número de la carta sea:
i) divisible por 5, ii) termine en 2.

11) Tres parejas de casados han comprado boletos para el teatro y se sientan en una fila formada por solo seis asientos. Si toman sus asientos de un modo totalmente aleatorio
a) ¿Cuál es la probabilidad de que Pablo y María (marido y mujer), se sienten en los dos asientos de la extrema izquierda?
b) ¿Cuál es la probabilidad de que Pablo y María terminen sentados uno junto a otro?

12) De acuerdo con un trabajo de investigación, la ubicación probable de las PC en una casa son:

| | |
|---|---|
| Dormitorio de adultos: | 0.03 |
| Dormitorio de niños: | 0.15 |
| Otro dormitorio: | 0.14 |
| Oficina o estudio: | 0.40 |
| Otra habitación: | 0.28 |

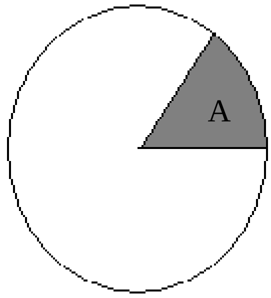
a) ¿Cuál es la probabilidad de que una PC esté en un dormitorio?
b) ¿Cuál es la probabilidad de que una PC no esté en un dormitorio?

13) El interés se enfoca en la vida de un componente electrónico. Suponga que se sabe que la probabilidad de que el componente funcione más de 6000 horas es 0.42. Suponga además que la probabilidad de que el componente no dure más de 4000 horas es 0.04.
a) ¿Cuál es la probabilidad de que la vida del componente sea menor o igual a 6000 horas?
b) ¿Cuál es la probabilidad de que la vida del componente sea mayor que 4000 horas?
c) Sea A el evento de que el componente falle en una prueba específica y B el evento de que el componente se deforma pero no falla. Supongamos que P(A) = 0.20 y P(B) = 0.35
c1) ¿Cuál es la probabilidad de que el componente no falle en la prueba?
c2) ¿Cuál es la probabilidad de que el componente funcione perfectamente (no se deforma ni falla en la prueba)?
c3) ¿Cuál es la probabilidad de que el componente falle o se deforme en la prueba?

14) Sean A y B eventos con $P(A \cup B) = 3/4$, $P(A^C) = 2/3$, y $P(A \cap B) = 1/4$.
Hallar P(A), P(B), $P(A \cap B^C)$.

15) Se escoge al azar un punto interior a un triángulo equilátero de lado 3. hallar la probabilidad de que su distancia a un vértice sea mayor que 1.

(Recordar que:
Si la circunferencia tiene radio r y el sector sombreado A tiene un ángulo de abertura $\alpha$ entonces el área del sector sombreado es
$\frac{\pi r^2 \alpha}{360}$ )




# Matemática 3

**<u>Práctica 2</u>**: Probabilidad condicional – Independencia

1) Se lanza un par de dados normales. Hallar la probabilidad de que la suma de sus números sea 10 o mayor si
   a) aparece un 5 en el primer dado
   b) aparece un 5 en uno de los dos dados por lo menos.

2) Se lanzan 3 monedas normales. Hallar la probabilidad de que sean todas caras si
   a) la primera de las monedas es cara
   b) una de las monedas es cara

3) Se escogen dos dígitos al azar del 1 al 9. si la suma es par, hallar la probabilidad de que ambos números sean impares.

4) Sean los eventos A y B con P(A) = ½ , P(B) = 1/3 y P(A I B) = ¼. Hallar:
   a) P(A/B) ; b) P(B/A) ; c) P(A Y B) ; d) $P(A^C / B^C)$ ; e) $P(B^C / A^C)$

5) Una clase tiene 12 niños y 4 niñas. Si se escogen 3 estudiantes de la clase al azar, ¿cuál es la probabilidad de que sean todos niños?

6) Una urna contiene 3 bolas rojas y 7 blancas. Se saca una bola de la urna y se reemplaza por una del otro color. Se saca de la urna una segunda bola.
   a) Hallar la probabilidad de que la segunda bola sea roja
   b) Si ambas bolas son del mismo color, ¿cuál es la probabilidad de que las dos sean blancas?

7) Una ciudad tiene dos carros de bomberos que operan en forma independiente. La probabilidad de que un carro específico esté disponible cuando se lo necesite es 0.96.
   a) ¿Cuál es la probabilidad de que ninguno esté disponible cuando se les necesite?
   b) ¿Cuál es la probabilidad de que un carro de bomberos esté disponible cuando se le necesite?

8) Una caja contiene 2 caramelos de coco y 3 de chocolate. Una segunda caja contiene 3 caramelos de coco, 2 caramelos de chocolate y 1 de dulce de leche. Si se saca un caramelo al azar de cada caja, encuentre la probabilidad de que

a) ambos caramelos sean de coco.
b) ningún caramelo sea de coco.
c) los dos caramelos sean diferentes.

9) En una prueba de opción múltiple, un estudiante contesta una pregunta que ofrece cuatro posibles respuestas, de las cuales sólo una es correcta. Suponga que la probabilidad de que el estudiante sepa la respuesta a la pregunta es 0.8 y que conteste al azar es 0.2

a) Cuál es la probabilidad de que conteste correctamente la pregunta?
b) Si contesta correctamente la pregunta. Cuál es la probabilidad de que realmente sepa la respuesta correcta?

10) Se lanza cinco veces un dado normal. Hallar la probabilidad de que:

a) en ninguna tirada salga el 1
b) salga el 1 una sola vez.
c) salga el 1 al menos una vez.

11) a) Si $P(A / B) = 0.4$, $P(B) = 0.8$ y $P(A) = 0.6$, ¿puede decirse que los eventos A y B son independientes?

b) Si, $P(A / B) = 0.3$ , $P(B) = 0.8$ y $P(A) = 0.3$, ¿puede decirse que los eventos $A^C$ y B son independientes?

12) En una cierta estación de servicio, el 40% de los clientes utilizan nafta normal sin plomo, 35% utilizan nafta extra sin plomo, y el 25% utilizan nafta premium sin plomo. De los clientes que consumen nafta normal, solo 30% llenan sus tanques, de los que consumen nafta extra, 60% llenan sus tanques, en tanto que de los que usan premium, 50% llenan sus tanques.

a) ¿Cuál es la probabilidad de que el siguiente cliente pida nafta extra sin plomo y llene su tanque?
b) ¿Cuál es la probabilidad de que el siguiente cliente llene el tanque?
c) Si el siguiente cliente llena el tanque, ¿cuál es la probabilidad de que pida nafta normal?.
d) ¿Qué propiedades utiliza para resolver los incisos a), b) y c)?

13) El 10% de los chips informáticos vendidos en el mercado son producidos por una empresa "pirata".

Para un chip "pirata" la probabilidad de que sea defectuosos es del 50% mientras que si el chip no es "pirata" la probabilidad de que sea defectuoso desciende al 5%.

**a)** Definir los sucesos convenientes, junto con sus probabilidades.
**b**) Determinar el porcentaje total de chips defectuosos que salen al mercado.
**c)** Se compra un chip y resulta ser defectuoso. Calcular la probabilidad de que proceda de la empresa "pirata".

14) Se utilizan dos líneas de producción para empaquetar azúcar en bolsas de 5 kg. La línea 1 produce el doble de bolsas que la línea 2. Uno por ciento de las bolsas de la línea 1 están defectuosas ya que no cumplen con una especificación de calidad, mientras que 3% de las bolsas de la línea 2 están defectuosas. Se elige aleatoriamente una bolsa para inspeccionarla.

a) ¿Cuál es la probabilidad de que provenga de la línea 1?
b) ¿Cuál es la probabilidad de que esté defectuosa?
c) Si la bolsa está defectuosa, ¿cuál es la probabilidad de que venga de la línea 1?

d) Si la bolsa no está defectuosa, ¿cuál es la probabilidad de que venga de la línea 1?

# Matemática 3

**<u>Práctica 3</u>**: Variables aleatorias discretas. Funciones de distribución Binomial, Geométrica, Hipergeométrica, Poisson.

1) Clasifique las siguientes variables aleatorias como discretas o continuas:
   a) X: “el número de accidentes automovilísticos por año en la ciudad de La Plata”
   b) Y: “el tiempo en horas que tarda en quemarse una lamparita”
   c) Z: “la cantidad de leche en litros que una vaca específica produce anualmente”
   d) W: “el número de huevos que una gallina pone mensualmente”
   e) N: “el número de permisos de construcción que emiten cada mes en una ciudad”
   f) Q: “el peso del grano producido por acre”

2) Un embarque de 10 automóviles extranjeros contiene 4 que tienen ligeras manchas de pintura. Si una agencia recibe 6 de estos automóviles al azar, sea X: “nº de automóviles que la agencia compra con manchas de pintura”.
   a) Hallar la f.d.p. de X
   b) Determine $P(X = 0)$ ; $P(X = 2)$ ; $P(X \le 2)$ ; $P(X \ge 2)$

3) El espesor de un entablado de madera (en pulgadas) que algún cliente ordena, es una v.a. X que tiene la siguiente F.d.a. :

$$F(x)=\begin{cases} 0, & x<1/8 \\ 0.2, & 1/8\le x<1/4 \\ 0.9, & 1/4\le x<3/8 \\ 1, & x\ge 3/8 \end{cases}$$

Determine las siguientes probabilidades:
a) $P(X \le 1/8)$
b) $P(X \le 1/4)$
c) $P(X \le 5/16)$
d) Hallar la función de distribución de X.

4) La distribución de probabilidad de X: “nº de imperfecciones por 10 metros de una tela sintética en rollos continuos de ancho uniforme, está dada por:
   a) Hallar la función de distribución acumulada de X
   b) Determine F(2) y F(3.1)

| *x* | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *f(x)* | 0.41 | 0.37 | 0.16 | 0.05 | 0.01 |

5) Para las variables aleatorias de los ejercicios 2) y 4) hallar E(X), $E(X^2)$, V(X).

6) Una compañía de materiales químicos envía cierto disolvente en tambores de 10 galones. Sea X: “número de tambores pedidos por un cliente elegido aleatoriamente”. Suponga que X tiene la f.d.p.
   a) Hallar E(X), V(X) y desviación estándar de X.
   b) Sea Y: “número de galones ordenados”
      b1) hallar la f.d.p. de Y
      b2) hallar E(Y) , V(Y) y la desviación estándar de Y

| *x* | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| *p(x)* | 0.4 | 0.2 | 0.2 | 0.1 | 0.1 |

7) En cierto servicio telefónico, la probabilidad de que una llamada sea contestada en menos de 30 segundos es 0.75. Suponga que las llamadas son independientes.

a) Si una persona llama 10 veces, ¿cuál es la probabilidad de que exactamente 9 de las llamadas sean contestadas en un espacio de 30 segundos?

b) Si una persona llama 20 veces, ¿cuál es la probabilidad de que al menos 16 de las llamadas sean contestadas en un espacio de 30 segundos?

c) Si una persona llama 20 veces, ¿cuál es el número promedio de llamadas que serán contestadas en menos de 30 segundos?

8) Un amigo que trabaja en una gran ciudad tiene dos automóviles, uno pequeño y uno grande. Tres cuartas partes del tiempo utiliza el automóvil pequeño para trabajar, y la cuarta parte restante usa el automóvil grande. Si utiliza el automóvil pequeño, por lo general no tiene problemas para estacionarse y, por lo tanto, llega a su trabajo tiempo con una probabilidad de 0.9. Si utiliza el automóvil grande, llega a tiempo su trabajo con una probabilidad de 0.6.

a) ¿Cuál es la probabilidad de que llegue a tiempo al trabajo?.

b) ¿Cuál es la probabilidad de que llegue a tiempo al trabajo en 6 de 10 mañanas, suponiendo que hay independencia entre un día y otro?

9) De un lote de 25 artículos, 5 de los cuales son defectuosos, se eligen 4 al azar. Sea X: "el número de artículos defectuosos entre los elegidos".

a) Obtener la función de distribución de probabilidad de X si los artículos se eligen con sustitución.

b) ¿Cuál es la E(X) y la V(X)?.

10) Con los datos del ejercicio 7), sea Y: "número de veces que hay que llamar hasta obtener la primer respuesta en menos de 30 segundos"

a) ¿Cuál es la probabilidad de tener que llamar 4 veces para obtener la primera respuesta en menos de 30 segundos?

b) ¿Cuál es el número promedio de llamadas que hay que hacer hasta tener una respuesta en menos de 30 segundos?

11) La probabilidad de que una computadora que corre cierto sistema operativo se descomponga en determinado día es 0.1. Determine la probabilidad de que la máquina se descomponga por primera vez en el duodécimo día, después de la instalación del sistema operativo, asumiendo independencia entre los días. Determine la media y la varianza del número de días hasta que el sistema operativo se descompone.

12) El número de solicitudes de asistencia recibido por un servicio de remolque de vehículos es un proceso de Poisson con tasa c=4 por hora.

a) Calcule la probabilidad de que se reciban 10 solicitudes entre las 16 y las 17 hs.

b) Si los operadores de las grúas se toman un descanso de 30 min. Cuál es la probabilidad de que no se pierda ninguna llamada de asistencia durante ese período?

13) El número de visitas realizadas en un día entre semana en una determinada página web se decide modelizar por una variable de Poisson de media 8.

**a)** ¿Cuál es la probabilidad de que en un día se reciban más de 4 visitas? Y ¿entre 7 y 10 visitas (ambos incluidos)?

**b**) ¿Cuál es la probabilidad de que al escoger al azar una semana laboral (de lunes a viernes), haya 3 días con más de 4 visitas?

(Sugerencia: considerar X: "nº de días en la semana laboral con más de 4 visitas")

14) En un lote de 10 microcircuitos, 3 están defectuosos. Se elige aleatoriamente cuatro microcircuitos para ser probados. Sea X: "número de circuitos probados que son defectuosos".
   a) Determine $P(X = 2)$
   b) Determine $E(X)$ y $V(X)$.

15) En referencia al ejercicio anterior, suponga que el lote tiene 1000 microcircuitos, 300 defectuosos. Se elige aleatoriamente 4 microcircuitos para ser probados. Sea X: "número de circuitos probados que son defectuosos".
   a) Determine $P(X = 2)$
   b) Considere que hay independencia entre las extracciones, vuelva a calcular $P(X = 2)$ usando distribución binomial, ¿qué observa?

# Matemática 3

**Práctica 4**: Variables aleatorias continuas. Funciones de distribución de probabilidad uniforme, exponencial, normal

1) El tiempo total, medido en unidades de 100 horas, que un adolescente utiliza su estéreo en un período de un año es una v.a. continua *X* con f.d.p. dada por

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \\ 0, & \text{caso contrario} \end{cases}$$

Encuentre la probabilidad de que en un período de un año el adolescente utilice su estéreo
a) menos de 120 horas
b) entre 50 y 100 horas

2) Suponga que la distancia X entre un blanco puntual y un disparo dirigido al punto, en un juego de tiro al blanco accionado por monedas, es una v.a. continua con f.d.p. dada por:

$$f(x) = \begin{cases} 0.75(1 - x^2) & -1 \le x \le 1 \\ 0 & \text{caso contrario} \end{cases}$$

a) Calcular $P(X > 0)$
b) Calcular $P(-0.5 \le X \le 0.5)$
c) Calcular $P(X < -0.25 \text{ ó } X > 0.25)$
d) Hallar la F.d.a. de X.

3) Considere la función de densidad

$$f(x) = \begin{cases} k\sqrt{x} & 0 < x < 1 \\ 0 & c.c. \end{cases}$$

a) Evalúe *k*
b) Encuentre *F*(*x*)
c) Evalúe P(0.3 < X <0.6) utilizando *F*(*x*)

4) Para las variables aleatorias de los ejercicios anteriores hallar su esperanza y desviación estándar.

**5)** Para la v.a. del ejercicio 1), sea la v.a. *Y* el número de kilowatts-hora que el adolescente gasta al año; se tiene que $Y = 60X^2 + 39X$ . Calcule la esperanza de *Y*. Explique qué propiedad utiliza.

6) Una barra de 12 pulgadas, que está sujeta por ambos extremos, debe someterse a un creciente cantidad de esfuerzo hasta que se rompa.
Sea X: "distancia desde el extremo izquierdo en el que ocu rre la rotura" y suponga que la f.d.p. de X es

$$f(x) = \begin{cases} \dfrac{1}{24} x \left(1 - \dfrac{x}{12}\right) & 0 < x < 12 \\ 0 & \text{en caso contrario} \end{cases}$$

Calcular:
**a)** la F.d.a de X
b) P(X ≤ 4) ; P(X > 6) ; P(4 < X < 6)
c) E(X)
d) la probabilidad de que el punto de ruptura ocurra a más de 2 pulgadas del punto esperado de ruptura.

7) La cantidad de café diaria, en litros, que sirve una máquina que se localiza en el vestíbulo de un aeropuerto es una v.a. X con distribución uniforme continua en (7, 10).
Encuentre la probabilidad de que en un día dado la cantidad de café que sirve esta máquina sea
a) a lo sumo 8.8 litros.
b) más de 7.4 litros, pero menos de 9.5 litros.
c) al menos 8.5 litros.
d) Hallar E(X) y V(X).

8) La variable Z tiene distribución normal estándar.
a) Calcular las siguientes probabilidades:
a1) $P(Z \leq 2.24)$
a2) $P(Z > 1.36)$
a3) $P(0 < Z < 1.5)$
a4) $P(0.3 < Z < 1.56)$
a5) $P(-0.51 < Z < 1.54)$
b) Hallar los valores de z que verifiquen:
b1) $P(Z > z) = 0.5$
b2) $P(Z < z) = 0.8485$
b3) $P(Z < z) = 0.0054$
b4) $P(-z < Z < z) = 0.90$

9) Si X es una variable aleatoria con distribución normal con parámetros: $\mu=10$ y $\sigma^2=36$
Calcular: a) $P(X > 6.4)$
b) $P(4.2 < X < 16)$
c) $P(X \leq 8.14)$

10) En la elaboración de un determinado medicamento en forma de comprimido interviene 1 producto químico cuya cantidad sigue aproximadamente una distribución normal con media 3 grs y desviación estándar 0.05 grs.
**a)** Calcular la probabilidad de que un comprimido pese más de 3.025 grs.
**b)** Un comprimido se considera defectuoso cuando su peso difiere de la media en más de 0.075 grs.
Calcular la proporción de comprimidos defectuosos que se fabrican.
**c)** Estos comprimidos se envasan en cajas de 10 unidades. Si un envase contiene 2 o más comprimidos defectuosos se elimina del mercado. Determinar el porcentaje de cajas que se retiran del mercado.
(Sugerencia: considere X: "nº de comprimidos defectuosos en una caja")

11) Un estudio de cierto sistema de computadoras revela que el tiempo de respuesta, en segundos, tiene una distribución exponencial con una media de 3 segundos.
a) ¿Cuál es la probabilidad de que el tiempo de respuesta exceda 5 segundos?
b) ¿Cuál es la probabilidad de que el tiempo de respuesta exceda 10 segundos?

12) El número de visitas a un sitio web sigue un proceso de Poisson con una razón de 3 por minuto.
a) ¿Cuál es la probabilidad de que transcurra más de un minuto sin recibir una visita?
b) Si transcurren dos minutos sin una visita, ¿cuál es la probabilidad que se dé una visita en el siguiente minuto?

13) El tiempo en horas empleado diariamente en transporte por los trabajadores de una gran ciudad es una v.a. continua con densidad exponencial con media 0.25.

**a)** Calcular la probabilidad de que un trabajador emplee más de media hora en transporte.
**b)** Si los trabajadores emplean al menos una hora, ¿cuál es la probabilidad de que no superen la hora y media?
**c)** Hallar el tiempo mínimo que emplea el 50% de los trabajadores que más tiempo pierden en transporte.

14) Cierto tipo de componente puede ser comprado nuevo o viejo. El 50% de los componentes nuevos duran más de 5 años, pero solo 30% de los usados duran más de 5 años. ¿Sería posible que las duraciones de los componentes se distribuyan exponencialmente?. Explique.

# Matemática 3

**Práctica 5**: Distribución conjunta, suma y promedios de variables aleatorias. Ley de los grandes números. Teorema Central del Límite.

1) Se analizaron las longitudes y los anchos de la bandeja de plástico rectangular para un CD que está instalada en una computadora personal. Las mediciones se redondearon al milímetro mas cercano.
Sean X: "la longitud medida" e Y: " el ancho medido".
La f.d.p. conjunta de (X , Y) está dada por

| Y \ X | ***129*** | ***130*** | ***131*** |
|---|---|---|---|
| ***15*** | 0.12 | 0.42 | 0.06 |
| ***16*** | 0.08 | 0.28 | 0.04 |

a) Determine la probabilidad de que la cubierta del CD tenga una longitud de 129 mm.
b) Determine la probabilidad de que una cubierta de CD tenga ancho de 16 mm.
c) Hallar las distribuciones marginales de X e Y.
d) Hallar E(X), E(Y), V(X), V(Y).

2) En el ejercicio anterior, calcule la f.d.p. condicional $p_{Y/X}$ ($y/x$ = 130)
¿Son X e Y independientes?. Explique.

3) Un software puede hacer llamadas a dos subrutinas A y B. En una ejecución elegida al azar, sean
X: "número de llamadas hechas a la subrutina A"
Y: "número de llamadas hechas a la subrutina B"
La f.d.p. conjunta de (X , Y) esta dada por

| | **Y** | | |
|---|---|---|---|
| **X** | 1 | 2 | 3 |
| 1 | 0.15 | 0.10 | 0.10 |
| 2 | 0.10 | 0.20 | 0.15 |
| 3 | 0.05 | 0.05 | 0.10 |

a) Determine las f.d.p. marginales de X e Y
b) Determine E(X), E(Y), V(X), V(Y)
c) Determine cov(X , Y)
d) ¿Son X e Y independientes?. Explique.

4) Con referencia al ejercicio anterior
a) Determine E(X+Y)
b) Determine V(X+Y) y $\sigma_{X+Y}$
c) Determine P(X+Y = 4)
d) Suponga que cada ejecución de la subrutina A tarda 100 ms y que cada ejecución de la

subrutina B tarda 200 ms.

d1) Determine el número medio de milisegundos de todas las llamadas realizadas a las dos subrutinas.

d2) Encuentre la desviación estándar del número medio de milisegundos de todas las llamadas realizadas a las dos subrutinas.

5) Dos computadoras trabajan en forma independiente.
Sean X: “número de fallas semanales de la computadora 1” e
Y: “número de fallas semanales de la computadora 2”.
Las distribuciones están dadas por

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p(x)$ | 0.25 | 0.25 | 0.30 | 0.20 |

| $y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p(y)$ | 0.15 | 0.20 | 0.40 | 0.25 |

a) Determine $P(X = Y)$ , es decir ambas computadoras tienen el mismo número de fallas

b) Determine $P(X > Y)$, es decir el número de fallas de la computadora 1 es mayor que el de la computadora 2.

6) El tiempo de vida de cierto componente, en años, tiene una función de densidad $f(x) = e^{-x}$ si $x > 0$ , 0 si $x \leq 0$. Están disponibles dos de dichos componentes, cuyos tiempos de vida son independientes. Tan pronto como falle el primer componente, éste se reemplaza por el segundo. Sean las variables aleatorias:
X: “tiempo de vida del primer componente” e Y: “tiempo de vida del segundo componente”

a) Determine $P(X \leq 1 , Y \leq 1)$
b) Determine $E(X)$ , $E(Y)$
c) Determine $E(X+Y)$

7) Una instalación de luz tiene dos focos A y B. La duración del foco A se puede considerar una v.a. X con distribución normal con media 800 hs. y desviación estándar de 100 hs. La duración del foco B se puede considerar una v.a. Y con distribución normal con media 900 hs. y desviación estándar de 150 hs. Suponga que las duraciones de los focos son independientes.

a) ¿Cuál es la probabilidad de que el foco B dure más que el foco A?.
(Sugerencia: piense cómo interpreta el evento $\{Y – X > 0\}$ y qué distribución tiene $Y - X$)

b) Otra instalación de luz tiene solo un foco. Se pone uno del tipo A y cuando se funde se instala otro de tipo B. ¿Cuál es la probabilidad de que el la duración total de ambos sea mayor que 2000 hs?.
(Sugerencia: piense cómo interpreta el evento $\{Y+X > 2000\}$ y qué distribución tiene $Y + X$)

8) El peso de un caramelo pequeño tiene una distribución normal con media 2.835 gramos y desviación estándar de 0.2835 gramos. Suponga que se colocan 16 caramelos en un paquete y que los pesos de éstos son independientes.

a) ¿Cuáles son la media y la varianza del peso neto del paquete?
b) ¿Cuál es la probabilidad de que el peso neto del paquete sea menor que 45.5 gramos?

9) El tiempo para que un sistema automatizado localice una pieza en un almacén, tiene una distribución normal con media de 45 segundos y desviación estándar de 30 segundos. Suponga que se hacen pedidos independientes por 10 piezas.

a) ¿Cuál es la probabilidad de que el tiempo promedio necesario para localizar las 10 piezas sea mayor que 60 segundos? ¿Cuál es la probabilidad de que el tiempo total necesario para localizar las 10 piezas sea mayor que 600 segundos?

b) Enuncie la propiedad teórica que utiliza para resolver el inciso anterior.

10) El centro de cálculo de una universidad dispone de un servidor para gestionar las páginas web personales de profesores y alumnos. Supongamos que la cantidad de memoria ocupada por una de estas páginas puede considerarse como una variable aleatoria con una media de 1.3 MB y una desviación estándar de 0.3. Si el servidor va a gestionar un total de 500 páginas, calcular aproximadamente la probabilidad de que la cantidad total de memoria necesaria supere los 660 MB.
(Sugerencia: considerar las v.a. $X_i$: "cantidad de memoria ocupada por la página i", i=1,2,…,500)

11) El tiempo de vida (en horas) de un componente electrónico viene determinado por la siguiente función de densidad:

$$f(x) = \begin{cases} 2ke^{-\frac{x}{5}} & x > 0 \\ 0 & c.c. \end{cases}$$

**a**) Calcular k y la función de distribución acumulada asociada.
**b)** ¿Qué porcentaje de componentes de este tipo duran entre 2 y 10 horas?. ¿Y más de un día?. Determinar la esperanza y la varianza.
**c)** Si se consideran 40 componentes del tipo anterior, obtener aproximadamente la probabilidad de que la vida media de los 40 componentes esté comprendida entre 2 y 10 horas.
(Sugerencia: considere las v.a. $X_i$ :" duración en horas del componente electrónico i" , i = 1, 2, ..,40)

12) La resistencia a la ruptura de un remache tiene un valor medio de 10000 lb/pulg$^2$ y una desviación estándar de 500 lb/pulg$^2$.
a) ¿Cuál es la probabilidad de que la resistencia promedio a la ruptura de la muestra, para una muestra de 40 remaches, sea entre 9900 y 10200?
b) Si el tamaño de la muestra hubiera sido 15 en lugar de 40, ¿podría calcularse la probabilidad pedida en la parte a) a partir de la información dada?

13) Si el 3% de las válvulas manufacturadas por una compañía son defectuosas, hallar la probabilidad de que en una muestra de 100 válvulas:
i) 0, ii) más de 5, iii) entre 1 y 3 , sean defectuosas.
a) Use la aproximación normal a la binomial.
b) Use la distribución binomial y haga el cálculo con la ayuda de un software de matemática (o una calculadora).

14) Una máquina fabrica piezas cuyas longitudes se distribuyen según una normal de media 32 y desviación estándar 0.3 milímetros, considerándose aceptables aquellas cuya medida se encuentra dentro del intervalo (31.1, 32.6).
**a)** ¿Cuál es la probabilidad de que una pieza fabricada por esta máquina sea defectuosa?
**b)** Calcular la probabilidad de que un lote de 500 piezas contenga más de 15 defectuosas.
**(**Sugerencia: considere la v.a. Y: "nº de piezas defectuosas en el lote", piense qué distribución tiene X.)

# Matemática 3

**<u>Práctica 6</u>:** Estimación puntual

**1)** Suponga que se tiene una muestra aleatoria de tamaño *n* tomada de una población $X$ , que $E(X)=\mu$ y $V(X)=\sigma^2$ . Sean

$$\overline{X}_1 = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i \qquad \text{y} \qquad \overline{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Dos estimadores de $\mu$ . ¿Cuál es el mejor estimador de $\mu$ ?. Explique su elección.

**2)** Sea $X_1, X_2, \ldots\ldots, X_7$ una muestra aleatoria de una población que tiene media $\mu$ y varianza $\sigma^2$ . Considere los siguientes estimadores de $\mu$ :

$$\hat{\Theta}_1 = \frac{X_1 + X_2 + \ldots\ldots + X_7}{7} \qquad \hat{\Theta}_2 = \frac{2X_1 - X_6 + X_4}{2} \qquad \hat{\Theta}_3 = \frac{2X_1 - X_7 + X_3}{3}$$

a) ¿Alguno de estos estimadores es insesgado?
b) Hallar el error cuadrático medio de los estimadores.
c) ¿Cuál estimador es el “mejor”?. ¿En qué sentido es mejor?

**3)** Sea $X_1, X_2, \ldots\ldots, X_n$ una muestra aleatoria de tamaño *n*.

a) Demuestre que $\overline{X}^2$ es un estimador sesgado de $\mu^2$ .
b) Determine la magnitud del sesgo de este estimador.
c) ¿Qué sucede con el sesgo a medida que aumenta el tamaño *n* de la muestra?.

**4)** El número diario de desconexiones accidentales de un servidor sigue una distribución de Poisson. En cinco días se observan: 2 , 5 , 3 , 3 , 7  desconexiones accidentales.

a) Obtenga el estimador de máxima verosimilitud de $\lambda$ . ¿El estimador es insesgado?, ¿es consistente?
b) Obtenga la estimación de $\lambda$ a partir de la muestra dada.
c) Encuentre el estimador de máxima verosimilitud de la probabilidad de que ocurrirán 3 o más desconexiones accidentales y encuentre la estimación de dicha probabilidad a partir de los datos.

**5)** a) Sea $X_1, X_2, \ldots\ldots, X_n$ una muestra aleatoria de una v.a. $B(1, p)$ . Hallar un estimador de máxima verosimilitud (E.M.V.) de *p*.

b) Se selecciona una muestra aleatoria de *n* chips fabricados por cierta compañía. Sea *X* = el número entre los *n* que tienen defectos y *p* = P(el chip tiene defecto). Supongamos que solo se observa *X* ( el número de chips con defectos).

$b_1$) Si *n* = 100 y *x* = 5, ¿cuál es la estimación de *p*?

$b_2$) Si *n* = 100 y *x* = 5, ¿cuál es el E.M.V. de la probabilidad $(1-p)^6$, de que ninguno de los siguientes 6 chips que se examinen tenga defectos?

**6)** Denotemos por *X* la proporción de tiempo asignado que un estudiante seleccionado al azar emplea trabajando en cierta prueba de actitud, y supongamos que la f.d.p. de *X* es:

$$f(x) = \begin{cases} (2\theta+1)x^{2\theta}, & 0 \le x \le 1 \\ 0, & c.c \end{cases} \qquad \text{donde } \theta > -\frac{1}{2}$$

Una muestra aleatoria de diez estudiantes produce la siguiente información:
0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.73, 0.97, 0.94, 0.77.

a) Utilice el método de los momentos para obtener un estimador de $\theta$ y luego calcule la estimación para esta información.

b) Obtenga el E.M.V. de $\theta$ y luego calcule la estimación para la información dada.

**7)** Sea $X_1, X_2, \ldots\ldots, X_n$ una muestra aleatoria de una v.a. $N(\mu, \sigma^2)$.

a) Hallar los estimadores de $\mu$ y $\sigma^2$ por el método de momentos. ¿Los estimadores son insesgados?

b) Hallar los estimadores de $\mu$ y $\sigma^2$ por el método de máxima verosimilitud. ¿Los estimadores son insesgados?

c) Se determina la resistencia al corte de cada una de diez soldaduras eléctricas por puntos de prueba, dando los siguientes datos (lb/plg$^2$):
392, 376, 401, 367, 389, 362, 409, 415, 358, 375.
Si se supone que la resistencia al corte esta normalmente distribuida, estime la verdadera media de resistencia al corte y desviación estándar de resistencia al corte usando el método de máxima verosimilitud y el método de momentos.

d) Estime la probabilidad de que la resistencia al corte de una soldadura al azar sea menor que 420.

# Matemática 3

**Práctica 7**: Intervalos de Confianza

1) Un proceso novedoso para elaborar gasolina ecológica toma biomasa en la forma de sacarosa y la convierte en gasolina usando reacciones catalíticas. En un paso en un proceso de la planta piloto, un ingeniero químico mide la salida de cadenas de carbono de longitud tres. Nueve corridas con el mismo catalizador dieron los rendimientos (en galones)
0.63 2.64 1.85 1.68 1.09 1.67 0.73 1.04 0.68
Suponga que el rendimiento tiene una distribución normal.
a) ¿Qué puede afirmar el ingeniero químico con 95% de confianza acerca del error máximo, si usa la media muestral para estimar el verdadero rendimiento medio?
b) Obtenga un intervalo de confianza del 95% para el verdadero rendimiento medio del proceso de la planta piloto.

2) Se calculan tres intervalos de confianza para la media de la fuerza de corte (en ksi) de pernos de anclaje de un tipo dado, todos de la misma muestra.
Los intervalos son: ( 4.01, 6.02 ) ; ( 4.20 , 5.83 ) y (3.57 , 6.46 ).
Los niveles de los intervalos son 90%, 95% y 99%. ¿Qué intervalo tiene cada nivel?. Justifique.

3) En una muestra aleatoria de 100 baterías producidas por cierto método, el promedio del tiempo de vida fue de 150 horas y la desviación estándar de 25 horas.
**a)** Determine un intervalo de confianza de 95% para la media del tiempo de vida de las baterías

producidas por este método.

**b)** Un ingeniero afirma que la media del tiempo de vida está entre 147 y 153 horas. ¿Con qué nivel de confianza se puede hacer esta afirmación?.

4) Las siguientes mediciones se registraron para el tiempo de secado, en horas, de cierta marca de pintura látex:

3.4 2.5 4.8 2.9 3.6 2.8 3.3 5.6 3.7 2.8 4.4 4.0 5.2 3.0 4.8

Suponiendo que las mediciones representan una muestra aleatoria de una población normal, encuentre un intervalo de confianza de nivel 99% para la media de los tiempos de secado.

5) Se prueban dos fórmulas diferentes de un combustible oxigenado para motor en cuanto al octanaje. La varianza del octanaje para la fórmula 1 es $\sigma_1^2 = 1.5$, mientras que para la fórmula 2 es $\sigma_2^2 = 1.2$. Se prueban dos muestras aleatorias de tamaño $n_1 = 15$ y $n_2 = 20$. Los octanajes promedio observados son $\bar{x}_1 = 89.6$ y $\bar{x}_2 = 92.5$.

a) Construya un intervalo de confianza del 95% para la diferencia en el octanaje promedio.

b) Si tomamos $n_1 = n_2$, ¿qué tamaño de muestra se necesitaría para que la longitud del intervalo se reduzca a la mitad del encontrado en a)?

6) Se comparan las resistencias de dos clases de hilo. Cincuenta piezas de cada clase de hilo se prueban bajo condiciones similares. La marca A tiene una resistencia a la tensión promedio de 78.3 kg con una desviación estándar de 5.6 kg; en tanto que la marca B tiene una resistencia a la tensión promedio de 87.2 kg con una desviación estándar de 6.3 kg. Construya un intervalo de confianza de 95% para la diferencia de las medias poblacionales.

7) Una determinada empresa de material fungible puede adquirir los cartuchos de tóner de impresora de dos proveedores distintos. Con el fin de determinar a qué proveedor comprar se toma una muestra de tamaño 12 de cada uno de los proveedores obteniendo los siguientes resultados (número de hojas impresas):

Proveedor A: $\bar{x}_A = 5459$ $s_A^2 = 33703$ Proveedor B: $\bar{x}_B = 5162$ $s_B^2 = 199928$

Si suponemos que las poblaciones son normales con varianzas iguales construir un intervalo de confianza de nivel 95% para la diferencia entre el número medio de hojas que imprime el cartucho de cada proveedor.

8) Dos empresas competidoras (A y B) en un mismo sector han puesto en marcha, casi simultaneamente, páginas de internet para la venta electrónica. Se han elegido al azar ocho clientes que han visitado la página A y, de manera independiente, otros ocho que han visitado la B y se han medido el tiempo (en minutos) de la duración de la visita de cada cliente. Los resultados fueron los siguientes:

Página A: 2.3 3.5 4.2 3.2 4.4 2.1 1.6 5.3

Página B: 1.3 2.3 4.4 3.7 2.8 6.5 3.6 4.5

Suponer que los datos provienen de poblaciones normales.

Construir un intervalo de confianza de nivel 99% para la diferencia entre los tiempos medios.

9) Una muestra de 10 camiones diesel fue operada tanto caliente como fría para calcular la diferencia en el ahorro de combustible. Los resultados, en millas/galón, se presentan en la tabla siguiente

| **Camión** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **caliente** | 4.56 | 4.46 | 6.49 | 5.37 | 6.25 | 5.90 | 4.12 | 3.85 | 4.15 | 4.69 |
| **frío** | 4.26 | 4.08 | 5.83 | 4.96 | 5.87 | 5.32 | 3.92 | 3.69 | 3.74 | 4.19 |

Determine un intervalo de confianza de 98% para la diferencia en la media del millaje entre motores calientes y fríos. Asuma que la muestra de las diferencias entre motores calientes y fríos es aproximadamente normal

10) Para los datos del ejercicio 4)
   a) construya un intervalo de confianza de 99% para la varianza del tiempo de vida real de estas baterías.
   b) construya un intervalo de confianza de 99% para la desviación estándar del tiempo de vida real de estas baterías.

11) Para los datos del ejercicio 8) hallar un intervalo de confianza de nivel de 95% para el cociente de las varianzas de los tiempos de visita.

12) Un fabricante de calculadoras electrónicas está interesado en estimar la fracción de unidades defectuosas que se producen. Una muestra aleatoria de 800 calculadoras incluye 18 defectuosas. Calcule un intervalo de confianza de nivel 99% para la verdadera fracción de unidades defectuosas.

13) a) Suponga que se quiere estimar qué porcentaje de todos los conductores excede el límite de velocidad de 80 km/h en cierto tramo del camino. ¿Qué tan grande debe ser la muestra para tener al menos 99% de confianza de que el error de su estimación es a lo sumo de 3.5%?.
   b) ¿Cómo se vería afectado el tamaño de la muestra requerida, si se sabe que el porcentaje a estimar es a lo sumo de 40%?.

14) En una prueba del efecto de la humedad en conexiones eléctricas, se probaron 100 conexiones eléctricas bajo condiciones húmedas y 150 en condiciones secas. Veinte de las primeras fallaron y solo diez de las segundas no pasaron la prueba. Determine un intervalo de confianza de 90% para la diferencia entre las proporciones de las conexiones que fallaron, húmedas y secas.

## Matemática 3

**<u>Práctica 8</u>**: Test de Hipótesis

Para cada uno de los ejercicios, modelice la situación y responda las siguientes preguntas:
a) ¿cuál es la hipótesis nula y cuál es la alternativa?
b) ¿cuál es el estadístico que utiliza y qué distribución tiene bajo H0?
c) ¿cuál es la zona de rechazo? Dibújela.
d) ¿cuál es su conclusión para los datos observados? Recuerde responder en relación al enunciado.
e) ¿Puede dar una idea del p-valor? ¿es exacto o aproximado?

1) Para cada una de las siguientes aseveraciones, exprese si es una hipótesis estadística legítima y por qué:
   a) H: $\sigma > 0$ b) H: $s \leq 0.20$ c) H: $\overline{X} - \overline{Y} = 5$ d) H: $\sigma_1/\sigma_2 < 1$ f) $\mu \leq 0.1$

2) Sea el estadístico de prueba Z con una distribución normal estándar cuando $H_0$ es verdadera. Dé el nivel de significancia en cada una de las siguientes situaciones:

a) $H_1$: $\mu > \mu_0$ , región de rechazo $z \geq 1.88$

b) $H_1$: $\mu < \mu_0$ , región de rechazo $z \leq -2.75$

c) $H_1$: $\mu \neq \mu_0$ , región de rechazo $z \geq 2.88$ o $z \leq -2.88$

3) Se supone que una máquina que llena cajas de cereal está calibrada, por lo que la media del peso de llenado es de 340 gr.

Sea $\mu$ la media verdadera del peso de llenado. Suponga que en una prueba de hipótesis $H_0$: $\mu = 340$ contra $H_1$: $\mu \neq 340$, el p-valor es 0.30.

a) ¿Se debe rechazar $H_0$ con base en esta prueba?. Explique

b) ¿Puede concluir que la máquina está calibrada y decir que la media del peso de llenado es de 340 gr?. Explique.

4) Un proceso de fabricación produce cojinetes de bola con diámetros que tienen una distribución normal y una desviación estándar de $\sigma = 0.04$ cm. Los cojinetes de bola que tienen diámetros que son muy pequeños o muy grandes son indeseables. Para poner a prueba la hipótesis nula de que $\mu = 0.5$ cm se selecciona al azar una muestra de 25 y se encuentra que la media muestral es 0.51

**a)** Establezca las hipótesis nula y alternativa tales que el rechazo de la hipótesis nula implicará que los cojinetes de bola son indeseables.

b) Con $\alpha = 0.02$ , ¿cuál es el valor crítico para el estadístico de prueba?. Realice el test.

5) Cuando está operando adecuadamente, una planta química tiene una media de producción diaria de por lo menos 740 toneladas. La producción se mide en una muestra aleatoria simple de 60 días. La muestra tenía una media de 715 toneladas por día y desviación estándar de 24 toneladas por día. Sea $\mu$ la media de la producción diaria de la planta. Un ingeniero prueba que $H_0$: $\mu \geq 740$ contra $H_1$: $\mu < 740$.

a) Determine el p-valor

b) ¿Piensa que es factible que la planta esté operando adecuadamente o está convencido de que la planta no funciona en forma adecuada?. Explique su razonamiento.

6) Pruebe la hipótesis de que el contenido medio de los envases de un lubricante específico es de 10 litros, si los contenidos de una muestra aleatoria de 10 envases son:

10.2 9.7 10.1 10.3 10.1 9.8 9.9 10.4 10.3 9.8

Utilice un nivel de significancia de 0.01 y suponga que la distribución del contenido es normal.

7) Para determinar el efecto del grado de combustible en la eficiencia del combustible, 80 nuevos automóviles de la misma marca, con motores idénticos, fueron conducidos cada uno durante 1000 millas. Cuarenta de los automóviles funcionaron con combustible regular y otros 40 con combustible de grado Premium; los primeros tenían una media de 27.2 milla/galón, con desviación estándar de 1.2 milla/galón. Los segundos tenían una media de 28.1 milla/galón y una desviación estándar de 2.0 milla/galón. ¿Puede concluir que este tipo de automóvil tiene mejor millaje con combustible Premium? Utilice el p-valor.

8) Se probó la velocidad en cierta aplicación de 50 chips nuevos de computadora, con otra cantidad igual de diseño viejo. La velocidad promedio, en MHz, de los nuevos chips fue de

495.6, y la desviación estándar de 19.4. La velocidad promedio de los chips viejos fue de 481.2, y la desviación estándar fue de 14.3.

a) ¿Se puede concluir que la media de la velocidad de los nuevos es mayor que la de los chips viejos?. Establezca las hipótesis nula y alternativa adecuadas y después encuentre el p-valor.

b) Una muestra de 60 chips aún más viejos tenía velocidad promedio de 391.2 MHz, con desviación estándar de 17.2 MHz. Alguien afirma que los nuevos chips tienen tienen una velocidad promedio mayor a 100 MHz que los más viejos. ¿Los datos proporcionan evidencias convincentes para esta afirmación? . Establezca las hipótesis nula y alternativa y después determine el p-valor.

9) Se considera usar dos marcas diferentes de pintura látex. El tiempo de secado en horas se mide en especímenes de muestras del uso de las dos pinturas. Se seleccionan 15 especímenes de cada una y los tiempos de secado son lo siguientes:

**Pintura A:**

3.5, 2.7, 3.9, 4.2, 3.6, 2.7, 3.3, 5.2, 4.2, 2.9, 4.4, 5.2, 4.0, 4.1, 3.4

**Pintura B:**

4.7, 3.9, 4.5, 5.5, 4.0, 5.3, 4.3, 6.0, 5.2, 3.7, 5.5, 6.2, 5.1, 5.4, 4.8

Suponga que el tiempo de secado se distribuye normalmente con $\sigma_A = \sigma_B$, y que ambos tiempos de secado son independientes.

a) Encuentre un intervalo de confianza para la diferencia de las medias $\mu_A - \mu_B$ de nivel 95%.

b) Utilice el intervalo usado en a) para hacer un test para decidir si las medias difieren.

10) Se estudia el flujo de transito en dos intersecciones transitadas entre las 4 P.M. y las 6 P.M. para determinar la posible necesidad de señales de vuelta. Se descubrió que en 21 días laborales hubo en promedio 247.3 automóviles que se aproximaron a la primera intersección desde el sur y dieron vuelta a la izquierda, mientras que en 11días laborales hubo en promedio 254.1 automóviles que se aproximaron a la segunda intersección desde el sur y dieron vuelta a la izquierda. Las desviaciones estándar muestrales correspondientes son $s_1 = 15.2$ y $s_2 = 18.7$

Suponga que las distribuciones son normales y que hay independencia entre ambas muestras. Pruebe la hipótesis nula $\mu_1 - \mu_2 = 0$ contra la alternativa $\mu_1 - \mu_2 \neq 0$ con nivel de significan cia $\alpha = 0.01$

11) La directiva de una compañía de taxis está tratando de decidir si debe cambiar de neumáticos normales a neumáticos radiales para mejorar el ahorro de combustible. Se equiparon cada uno de los diez taxis con uno de los dos tipos de neumáticos y se condujeron en una trayectoria de prueba. Sin cambiar de conductores, se seleccionó el tipo de neumáticos y se repitió la trayectoria de prueba. El ahorro de combustible (en milla/galón) para los diez automóviles es:

| | **Automóvil** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **radial** | 32.1 | 36.1 | 32.3 | 29.5 | 34.3 | 31.9 | 33.4 | 34.6 | 35.2 | 32.7 |
| **normal** | 27.1 | 31.5 | 30.4 | 26.9 | 29.9 | 28.7 | 30.2 | 31.8 | 33.6 | 29.9 |

Asuma que la diferencia en ahorro de combustible entre ambos neumáticos es aproximadamente normal.

a) Debido a que el cambio de neumáticos en la flota de taxis es caro, la directiva no quiere cambiar a menos que una prueba de hipótesis proporcione evidencias de que mejorará el millaje. Establezca la hipótesis nula y alternativa adecuadas, y encuentre el p-valor.
b) Un análisis costo-beneficio muestra que será provechoso cambiar a neumáticos radiales si la media de la mejora del millaje es mayor a dos millas /galón. Establezca la hipótesis nula y alternativa adecuadas, y encuentre el p-valor, para una prueba de hipótesis diseñada como base de la decisión de cambiar.

12) El departamento de seguridad de un gran edificio de oficinas quiere probar la hipótesis nula de que $\sigma = 2.0$ minutos para el tiempo que tarda un guardia en realizar su rondín contra la hipótesis alternativa de que $\sigma \neq 2.0$ minutos. ¿Qué se puede concluir con un nivel de significancia de 0.01, si una muestra aleatoria de tamaño $n = 31$ da como resultado $s = 1.8$ minutos?. Asuma que la muestra proviene de una distribución normal.

13) Con referencia al ejercicio 10) use el nivel de significancia de 0.05 para probar la afirmación de que existe una mayor variabilidad en el número de automóviles que dan vuelta a la izquierda aproximándose desde el sur entre 4 P.M. y 6 P.M. en la segunda intersección.

14) Un taller acaba de recibir una maquina nueva y busca ajustarla correctamente. Según el técnico vendedor de la máquina, la maquina está ajustada para que no produzca más de 4% de piezas defectuosas. Al tomar una muestra de 350 piezas producidas, encuentra 10 defectuosas. La empresa no puede permitirse un nivel de defectuosos mayor de 5%. Razonar que tipo de test se debe realizar con el fin de determinar si la maquina se encuentra mal ajustada, y realizar dicho contraste. (tomar $\alpha = 0.05$).

15) En una muestra de 100 lotes de un producto químico comprado al distribuidor A, 70 satisfacen una especificación de pureza. En una muestra de 70 lotes comprada al distribuidor B, 61 satisfacen la especificación. ¿Puede concluir que una proporción mayor de los lotes del distribuidor B satisface la especificación?. Utilice el p-valor.