# Estadística Descriptiva

## Estadística

El término estadística se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad situaciones.

En un sentido amplio, la estadística se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos.

Esto ayuda a las personas que deben tomar decisiones una mejor comprensión del entorno, permitiéndoles así tomar mejores decisiones con base en mejor información.

## Estadística Descriptiva

La mayor parte de la información estadística en periódicos, revistas, informes de empresas y otras publicaciones consta de datos que se resumen y presentan en una forma fácil de leer y de entender. A estos resúmenes de datos, que pueden ser tabulares, gráficos o numéricos se les conoce como estadística descriptiva.

Mediante estadísticos se pueden describir y comprender un entorno.

## Estadística Inferencial

Una de las principales contribuciones de la estadística es emplear datos de una muestra para hacer estimaciones y probar hipótesis acerca de las características de una población mediante un proceso al que se le conoce como inferencia estadística.

## Población y muestra

Cuando se examina un grupo entero o universo completo de observaciones, se lo denomina población. Cuando se examina una pequeña parte del grupo, lo llamamos muestra.

El concepto de población en Estadística va más allá de la clásica definición que se da en la Demografía.

En la actividad estadística una población puede estar constituida por elementos de cualquier tipo, no solamente por seres humanos.

Por ejemplo, se puede hablar de la población de viviendas de un barrio; de la población de comprobantes contables de una empresa; de la población de alumnos en Henry, etc.

## Distribución de frecuencias

La distribución de frecuencias es una forma de presentación de los datos que facilita su tratamiento conjunto y permite una comprensión diferente de ellos. Es una tabla de datos en base a observaciones (frecuencias). La frecuencia es el número de casos

que pertenecen a un valor determinado.

Edades de los compradores de automóviles

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Edades | | Nº de autos  *J;* | Verdadero (.imite  VI. | Punto Medio  x, | Frecuencia Acumulada Menor que  ,..,,., | Frecuencia Acumulada Mayor que  *F /+J* | Frecuencia Relativa  *'11* | Frecuencia Relativa Acumulada Menor que  *H/"'* | Frecuencia Relativa Acumulada Mayor que  *H/+'* |
| 25 - 29 | | 6 | 24,5 | 27 | *6* | *80* | , 7 50% | 7,50% | 10, 0 00% |
| 30 | 34 | 9 | 29,5 | 32 | 15 | 74 | 11,25% | 18,75% | 92,50% |
| 35 | 39 | 15 | 34,5 | 37 | 30 | 65 | 1, 8 75% | 37,50% | 81,25% |
| 40 | 44 | 18 | 39,5 | 42 | 48 | 50 | 2, 2 50% | 60,00% | 62,50% |
| 45-49 | | 15 | 44,5 | 47 | 63 | 32 | 18,75% | 78 ,75% | 40,00% |
| 50-54 | | 10 | 49,5 | 52 | 73 | 17 | 12,50% | 91 ,25% | 21,25% |
| *55* - 59 | | *1* | 54,5 | *51* | 80 | *1* | 8,75% | 100,00% | 8,75% |
|  | | 80 |  | | | | 100,00 % |  | |

## Histograma

El Histograma es un gráfico de la distribución de frecuencias, que se construye con rectángulos de superficie proporcional al producto de la amplitud por la frecuencia absoluta (o relativa) de cada uno de los intervalos de clase.

Edad de lo s compradores

62

S7

-

47 S2

22 27 32 37 42

►

. b

.

"

.

2

o

.,

,.

cS

!;

►

<

"

Edad de los compradores de autos

20

18

16

14

12

10

8

6

4

g

s

-g

e

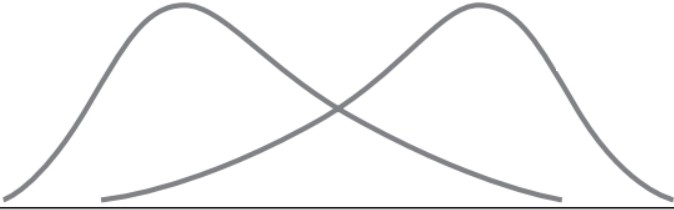
r

## Medidas de tendencia central

La tendencia central se refiere al punto medio de una distribución.

El sesgo se produce cuando al trazar una línea vertical que pase por el punto más alto de la curva dividirá su área en dos partes que no son iguales. Cuando se da el caso de que cada parte es una imagen de espejo de la otra, esta curva se denomina simétrica. Si la curva esta sesgada hacia la derecha, se considera positivamente sesgada y si el sesgo se pronuncia hacia la izquierda, se denomina negativamente sesgada.

## Sesgos



Curva A:

sesgada a la derecha

*\*

*/*

CurvaB: sesgada a

**Media aritmética (Promedio, mean)**

Es la suma de los valores de todas las observaciones, divido la cantidad de elementos de la muestra.

|  |  |  |
| --- | --- | --- |
| **Media aritmética de la poblaci6n**  Suma **de** los valores **de** todas las obseivaciones  *µ. = -kX*  *N*  Número de elementos de la población | |  |
|  |
|  | **Media aritmética de la muestra**  **Suma** de los valores de todaslas obseivaciones  Número deelementos de la muestra | |
| -*x - -LX*  *n* |

## Ventajas – Desventajas

- Un solo número que representa a un conjunto de datos completo.

- Concepto familiar.

- Es única debido a que cada conjunto de datos posee una y sólo una media.

- Es útil para la comparación de medias de varios conjuntos de datos.

- Puede verse afectada por valores extremos.

- Para grandes conjuntos de datos que no están agrupados en frecuencias relativas resulta un cálculo tedioso.

- Cuando existen valores de clase extremos abiertos ("60 años o más", "18 años o menos", etc.) no se puede calcular.

## La mediana

Es un solo valor del conjunto de datos que mide la observación central del conjunto. Es decir que esta sola observación es el elemento que está más al centro del conjunto de números, la mitad de los elementos están por arriba de este punto y la otra mitad está por debajo. Entendiendo al conjunto de datos como la cantidad total de observaciones.

**Mediana**

/.

Número de elementos del arreglo

Mediana = (*n* ; 1 ) -ésimo término del arreglo de datos

## Ventajas – Desventajas

- No se ve afectada por valores extremos.

- Es fácil de entender y se puede calcular a partir de cualquier tipo de datos.

- La podemos encontrar incluso cuando nuestros datos son descripciones cualitativas, en lugar de números.

- Ciertos procedimientos estadísticos que utilizan la mediana son más complejos que aquellos que utilizan la media.

- Debemos ordenar los datos antes de llevar a cabo cualquier cálculo.

## La moda

La moda es el valor que más se repite en el conjunto de datos. Como en todos los demás aspectos de la vida, el azar puede desempeñar un papel importante en la organización de datos. En ocasiones, el azar hace que un solo elemento no representativo se repita lo suficiente para ser el valor más frecuente del conjunto de datos. Es por esto que rara vez utilizamos la moda de un conjunto de datos no agrupados como una medida de tendencia central.

## Ventajas – Desventajas

- Se puede utilizar como una posición central para datos tanto cualitativos como cuantitativos.

- También, al igual que la mediana, los valores extremos no afectan indebidamente a la moda.

-La podemos utilizar aun cuando una o más clases sean de extremo abierto.

- No se utiliza tan a menudo como medida de tendencia central.

- A veces, no existe un valor modal debido a que el conjunto de datos no contiene valores que se presenten más de una vez.

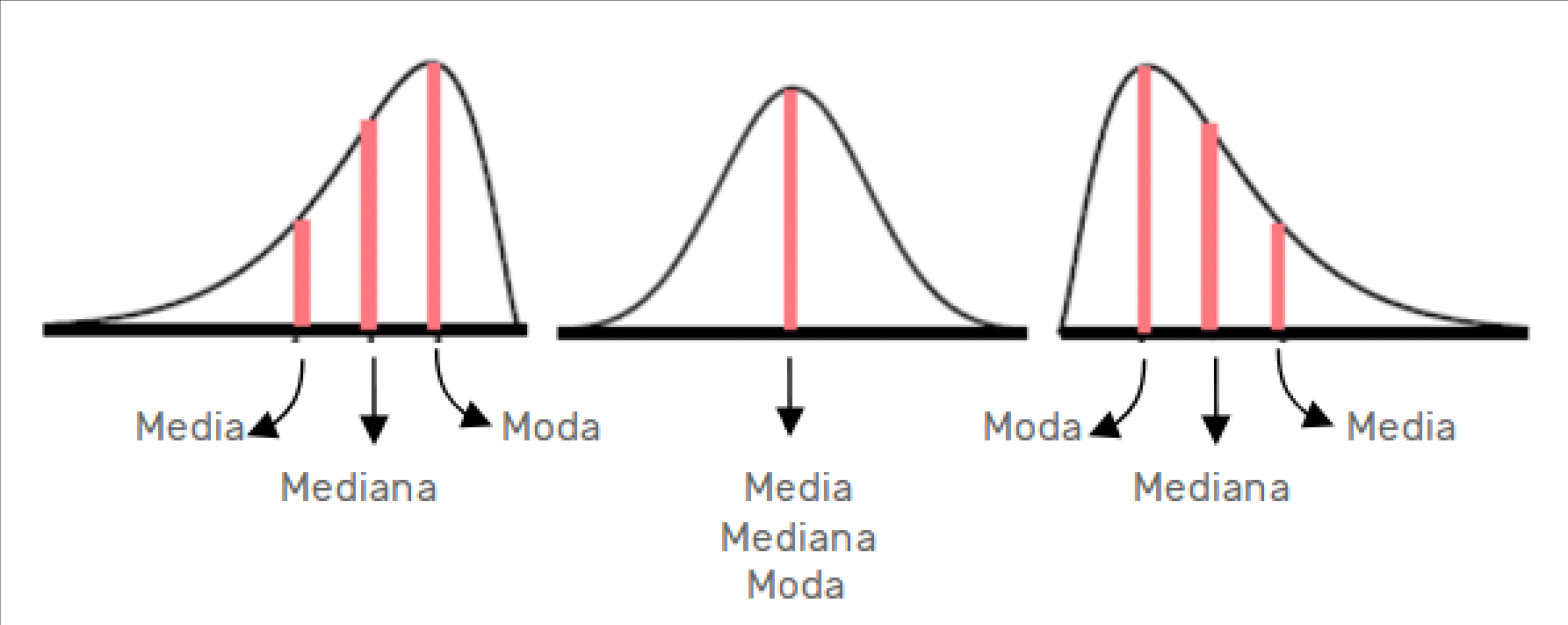
- Cuando los conjuntos de datos contienen dos, tres o más modas, es difícil interpretarlos y compararlos.

## Resumen

Las distribuciones simétricas que sólo contienen una moda siempre tienen el mismo valor para la media, la mediana y la moda.

*Sesgo positivo* (es decir, sesgada a la derecha): la moda se encuentra en el punto más alto de la distribución, la mediana está a la derecha de la moda y la media se encuentra todavía más a la derecha de la moda y la mediana.

*Sesgo negativo* (es decir, sesgada a la izquierda), la moda se encuentra en el punto más alto de la distribución, la mediana está a la izquierda y la media se encuentra todavía más a la izquierda de la moda y la mediana.



Cuando la población está sesgada negativa o positivamente, la mediana suele ser la mejor medida de posición, debido a que siempre está entre la moda y la media. La frecuencia de ocurrencia de un solo valor no incluye mucho en la mediana como es el caso de la moda, ni la distorsionan los valores extremos como la media. En cualquier otro caso, no existen guías universales para la aplicación de la media, la mediana o la moda como medidas de tendencia central. Toca a los analistas decidir cual describe mejor el conjunto de datos.

## La dispersión

Al igual que sucede con cualquier conjunto de datos, la media, la mediana y la moda sólo nos revelan una parte de la información que debemos conocer acerca de las características de los datos.

Para aumentar nuestro entendimiento del patrón de los datos, debemos medir también su dispersión, separación o variabilidad.

## El rango

El rango es la diferencia entre el más alto y el más pequeño de los valores observados. Sólo toma en cuenta los valores más alto y más bajo de una distribución y ninguna otra observación del conjunto de datos. Como resultado, ignora la naturaleza de la variación entre todas las demás observaciones, y tiene una gran influencia de los valores extremos.

## La varianza

Es la suma de los cuadrados de las distancias entre la media y cada elemento de la población, dividido entre el número total de observaciones. Al elevar al cuadrado cada distancia, logramos que todos los números sean positivos y, al mismo tiempo, asignamos más peso a las desviaciones más grandes (desviación es la distancia entre

la media y un valor).

La fórmula para calcular la varianza es:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Varianza de población** | | | | | | |  |
| *el*= | *1(x* -  *N* | *µ,)2*= | 1 | x2  *N* | \_ | µ, 2 | |

donde:

* *el* = varianza de la población
* *x* = elemento u observación
* *µ,* = media de la población
* *N* = número total de elementosde la población
* 1 = suma de todos los valores *(x* - *µ,)2,* o to dos los valores *x2*

## La Desviación Estándar

Para poder realizar una interpretación intuitiva de la varianza, debemos hacer un cambio significativo. Esta medida se conoce como la desviación estándar y es la raíz cuadrada de la varianza.

La desviación estándar, entonces, queda en las mismas unidades que los datos originales.

## Coeficientes de variación

El coeficiente de variación es una medida relativa de dispersión. Relaciona la desviación estándar y la media, expresando la desviación estándar como porcentaje de la media. La unidad de medida, entonces, es “porcentaje”, en lugar de las unidades de

los datos originales.

# Introducción a la teoría de probabilidad

## Probabilidad

Es la rama de las matemáticas que se ocupa de medir o determinar cuantitativamente la posibilidad de que ocurra un determinado suceso.

Un modelo matemático es una representación simbólica de un fenómeno cualquiera, realizada con el fin de estudiarlo mejor, dichas representaciones puede ser fenómenos físicos, económicos, sociales, etcétera.

Los modelos matemáticos pueden clasificarse en determinísticos y probabilísticos.

## Modelos determinísticos

Cuando se realiza el modelo matemático de un fenómeno y en este se pueden manejar los factores que intervienen en su estudio con el propósito de predecir sus resultados, se llamará modelo determinístico.

#Función de interes compuesto def intere(scapitallnicial,i,n):

valorFinal = capitallnicial\*(l+i)\*\*n

return valorFinal

#Valore de cálculo capital = 120000

i= 0.08

n = 5

#Implementación print(intere(scapital,i,n))

## Modelos probabilísticos

En los modelos probabilísticos, no podemos controlar los factores que intervienen en dichos modelos. A partir de lo cual surge la definición de modelo probabilístico o estocástico. Además de que dichos factores ocurren de tal manera que no es posible predecir sus resultados.

- Lanzamiento de una moneda.

- Estimación de productos defectuosos de una línea de producción.

## Experimentos

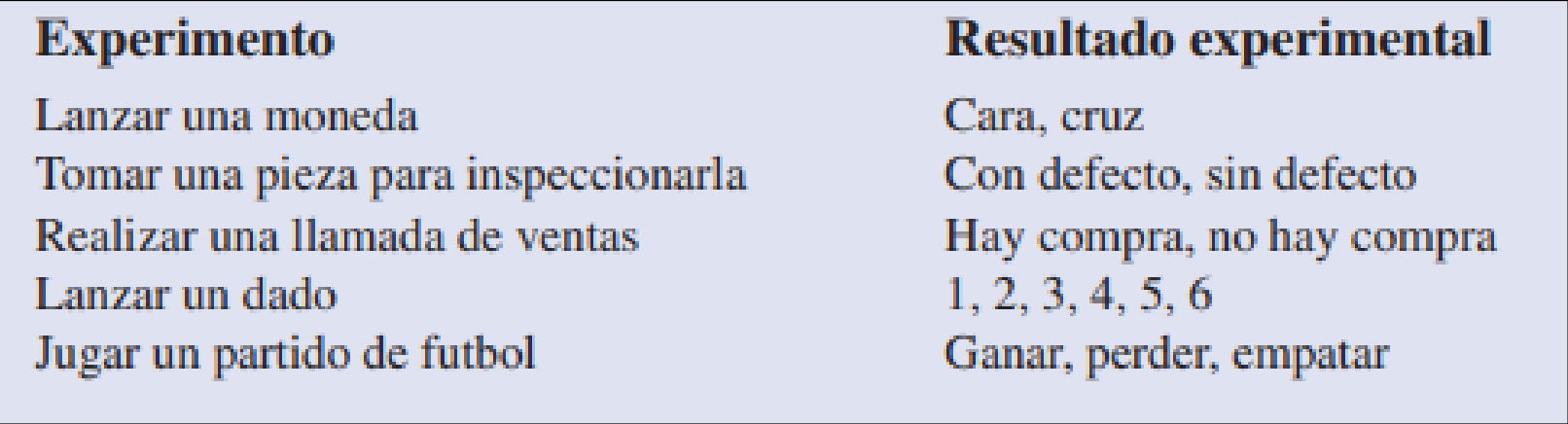
Un experimento aleatorio es el proceso de obtención de una observación en que se cumple alguna de las siguientes condiciones:

a) Todos los resultados posibles son conocidos.

b) Antes de realizar el experimento el resultado es desconocido.

c) Es posible repetir el experimento en condiciones ideales.

Existen también experimentos del tipo determinísticos que se basan en la aplicación de modelos como el desarrollado en el ejemplo.



## Espacio muestral

El espacio muestral es el conjunto de todos los resultados posibles de un experimento, la colección de todos los posibles eventos. La forma en que se subdivide el espacio muestral depende del tipo de probabilidades que se va a determinar.

Hay varias formas alternativas de observar un espacio muestral:

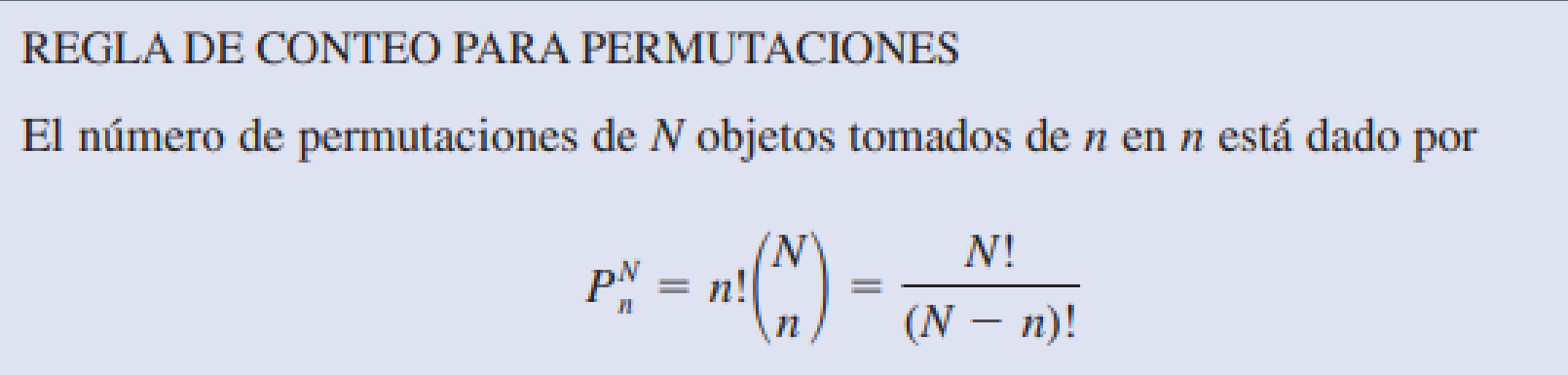
* Clasificación cruzada de los eventos en una tabla llamada tabla de contingencias o tabla de probabilidad.
* Representación gráfica de los diversos eventos como uniones o intersecciones de círculos en un diagrama de Venn.

## Reglas de conteo

La regla de conteo para experimentos de pasos múltiples permite determinar el número de resultados experimentales sin tener que enumerarlos.

Otra regla de conteo útil le permite contar el número de resultados experimentales cuando el experimento consiste en seleccionar n objetos de un conjunto (usualmente mayor) de N objetos.

La tercera regla de conteo que suele ser útil, es para permutaciones. Dicha regla permite calcular el número de resultados experimentales cuando se seleccionan n objeto un conjunto de N objetos y el orden de selección es relevante. Los mismos n objetos seleccionados en orden diferente se consideran un resultado experimental diferente.



## Interpretaciones de la probabilidad.

### Corriente clásica:

Se consideran espacios muestrales uniformes, es decir, se asignan probabilidades a eventos con base en resultados equiprobables (igualmente verosímiles). Si un experimento admite una cierta cantidad de resultados posibles, entonces las probabilidades de un suceso es el cociente entre el número de casos del suceso y el número total de casos del experimento.

"La probabilidad de un evento que se está llevando a cabo se calcula dividiendo el número de resultados favorables por el número de resultados posibles".

### Corriente frecuentista:

Se asigna un valor de probabilidad a un evento, a partir del cual se considera que ocurrirá. La definición o interpretación de la probabilidad está basada, como su nombre lo indica, en la frecuencia relativa con la cual se obtendría el evento, para esto el experimento se repite una gran cantidad de veces. El éxito del lanzamiento de un cohete, no se puede obtener a partir de una gran cantidad de lanzamientos de cohetes; por tanto, la probabilidad se obtiene en forma frecuentista del éxito de un lanzamiento.

### Corriente subjetivista:

En la corriente subjetivista (interpretación de la probabilidad que es muy empleada en el estudio del análisis de decisiones) se asignan probabilidades a eventos basándose en el conocimiento o experiencia que cada persona tiene sobre el experimento; por tanto, la probabilidad asignada está sujeta al conocimiento que el científico tenga con respecto al fenómeno estudiado. De este modo, para un mismo experimento las probabilidades asignadas por diferentes personas pueden ser distintas.

### Corriente bayesiana:

Se asignan probabilidades a los eventos después del experimento. Es decir, la asignación de probabilidades está basada en el conocimiento de la ocurrencia de eventos que estén en dependencia con el evento de estudio.

En tal situación decimos que la información obtenida influyó en la asignación de probabilidades.

## Diferencia entre Estadística y Probabilidad

La estadística se basa en el estudio de los datos para analizarlos e intentar obtener conclusiones sobre fenómenos que ocurren de forma aleatoria.

La probabilidad se encarga del estudio de variables aleatorias para medir la frecuencia con la que se consigue un resultado determinado en un fenómeno aleatorio que en la mayoría de ocasiones depende del azar.

# Espacios muestrales y sucesos

## Clasificación de los sucesos

Los elementos básicos de la teoría de la probabilidad son los resultados de un experimento aleatorio. Un experimento es un ensayo o juego que puede constar de uno o más intentos y cuyo resultado es la ocurrencia de uno, y sólo uno de los varios resultados posibles y no se sabe cuál ocurrirá.

Ou.-ación (meses) Número de proyectos

|  |  |  |  |
| --- | --- | --- | --- |
| **Etapa 1 Dis eño** | **Etapa** 2  **Construcción** | **Punto muestra!** | **que tuvieron esta duración** |
| 2 | 6 | (2, 6) | 6 |
| 2 | 7 | (2, 7) | 6 |
| 2 | 8 | *(2,* 8) | 2 |
| 3 | 6 | (3, 6) | 4 |
| 3 | 7 | (3, 7) | 8 |
| 3 | 8 | (3, 8) | 2 |
| 4 | 6 | (4, 6) | 2 |
| 4 | 7 | (4, 7) | 4 |
| 4 | 8 | (4, 8) | 6 |
|  |  |  | Total 40 |

La probabilidad de cualquier evento es igual a la suma de las probabilidades de los puntos muestrales que forman el evento.

|  |  |  |
| --- | --- | --- |
| **Punto muestral** | **Tiempo de terminación del proyecto** | **Probabilidad del punto muestral** |
| (2, 6) | 8 meses | P(2, 6) = 6/40 = 0.1S |
| (2, 7) | 9 meses | P(2, 7) = 6/40 = 0.1S |
| (2, 8) | JO meses | P(2, 8) = 2/40 = O.OS |
| (3 , 6) | 9 meses | *P(3,* 6) = 4/40 = O. JO |
| (3 , 7) | JO meses | *P(3,* 7) = 8/40 = 0.20 |
| (3, 8) | **11** m eses | *P(3,* 8) = 2/40 = O.OS |
| (4 , 6) | JO meses | *P(4,* 6) = 2/40 = O.OS |
| (4 , 7) | **11** m eses | *P(4,* 7) = 4/40 = O. JO |
| (4 , 8 ) | 12 meses | *P(4,* 8) = 6/40 = 0.1S |
|  |  | Total 1.00 |

### Simple:

Es un evento que puede describirse con una característica única.

Ejemplos: el auto sea Nacional, el comprador tenga más de 40 años, el cliente pague con tarjeta de débito, el precio de la acción suba.

### Sucesos Excluyentes.

En una sola realización de un experimento aleatorio dos sucesos A y B son mutuamente excluyentes cuando no se pueden presentar simultáneamente, es decir, cuando la ocurrencia de uno cualquiera de ellos imposibilita la ocurrencia de los otros.

Ejemplo: auto nacional o importado.

Todos los sucesos opuestos son excluyentes, pero no todos los sucesos excluyentes son opuestos. Un conjunto de eventos es colectivamente exhaustivo si uno de los eventos debe ocurrir. Que el auto sea nacional o importado, son sucesos colectivamente exhaustivos. Uno de ellos debe ocurrir. Si no ocurre nacional, debe ocurrir importado y viceversa.

Dos sucesos son compatibles cuando pueden ocurrir al mismo tiempo. Ejemplo: nacional o más de 40 años. Esta definición no indica que estos sucesos deban necesariamente ocurrir en forma conjunta.

Dos sucesos son compatibles cuando es posible que ocurran al mismo tiempo. Obsérvese que esta definición no indica que esos eventos deban ocurrir necesariamente en forma conjunta.

## Regla de la adición

Se utiliza cuando se desea determinar la probabilidad de que ocurra un evento u otro o ambos en una sola observación. Nos permite encontrar la probabilidad del evento “A ó B”: considera la ocurrencia de cualquiera de los eventos, evento A o evento B o ambos A y B.

La ley de la adición proporciona una manera de calcular la probabilidad de que ocurra el evento A o el evento B o ambos. En otras palabras, la ley de la adición se emplea para calcular la probabilidad de la unión de los dos eventos.

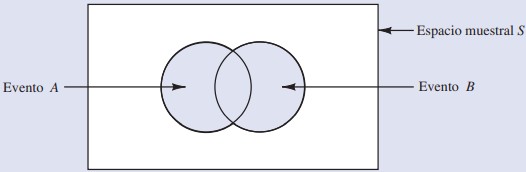
LEY DE LA ADICIÓN

P(A U B) = P(A) U P(B) - P(A ∩ B)

# Regla de la adición

P(A U B) = P(A) U P(B) - P(A ∩ B)

LA UNIÓN DE LOS EVENTOS *A* Y *B* APARECE SOMBREADA



## Regla de la adición

P(A U B) = P(A) + P(B)

## Regla de la multiplicación

### Sucesos Independientes

Dos eventos son independientes cuando la ocurrencia o no ocurrencia de un suceso o evento no tiene ningún efecto en la probabilidad de ocurrencia de otro suceso o evento.

Cuando para dos sucesos o eventos cualquiera A y B la P(A/B)=P(A), ambos sucesos son independientes. En ese caso también ocurre que P(B/A)=P(B), y a partir de estas dos últimas igualdades, se veri􀁽ca que, si A y B son independientes P(A y B) = P(A) \* P(B).

#Probabilidad de que dos clientes sucesivos, el primero pague en efectivo y el segundo con tarjeta de crédito.

efectivo= 6/15 tarjetaCredito = 7/15

probalidad = efectvio \* tarjetaCreidto

print(probabilidad)

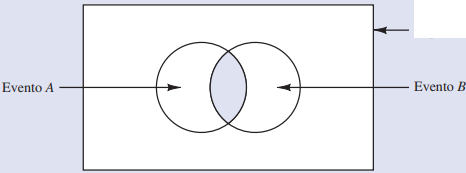
#Probabilidad de que dos clientes sucesivos, los dos paguen en efectivo: efectivo= 6/15

probalidad = efectvio \* efectivo

print(probabilidad)

Mientras que la ley de las sumas de probabilidades sirve para calcular la probabilidad de la unión de dos eventos, la ley de la multiplicación es útil para calcular la probabilidad de la intersección de dos eventos. La ley de la multiplicación se basa en la definición de probabilidad condicional.

**FIGURA 4.6** LA INTERSECCIÓN DE LOS EVENTOS *A* Y *8* APARECE SOMBREADA

 Espacio muestral *S*

LEY DE LA MULTIPLICACIÓN

*P(A* n *B)* = *P(B)P(A* I *B)*

o

*P(A* n *B)* = *P(A)P(B* I *A)*

La notación | indica que se está considerando la probabilidad del evento A dada la condición de que el evento B ha ocurrido. Por tanto, la notación P(A | B) se lee “la probabilidad de A dado B”. Dos sucesos o eventos son condicionales cuando la ocurrencia o no ocurrencia de un suceso o evento afecta la probabilidad de ocurrencia del otro.

PROBABILIDAD CONDICIONAL

*P(A* n *B)*

*P(A* I *B)* = *P(B)*

o

*P(B* I *A)* = *P(A*n *B)*

*P(A)*

Cuando para dos sucesos o eventos cualquiera A y B la P(A/B)=P(A), ambos sucesos son independientes. En ese caso también ocurre que P(B/A)=P(B), y a partir de estas dos últimas igualdades, se veri􀁽ca que, si A y B son independientes:

P(AyB) = P(A)\*P(B).

# Distribuciones de probabilidad

Cuando a todos los posibles valores numéricos de una variable aleatoria se le asignan valores de probabilidad, ya sea mediante un listado o una función matemática el resultado es una distribución de probabilidad. La suma de las probabilidades de todos los resultados numéricos posibles debe ser igual a 1.

Una función de probabilidad puede asignar valores de probabilidad a cada estado del espacio muestral.

## Características

Las características generales de las distribuciones de probabilidad difieren según el tipo de variable aleatoria, discreta o continua, que se encuentre bajo estudio.

Si la variable aleatoria es continua, no pueden listarse todos los posibles valores de la variable, motivo por el cual las probabilidades que se determinan por medio de una función matemática son gráficamente representadas por una función de densidad de probabilidad, o curva de probabilidad.

### Variables discretas

- Puede tomar solamente algunos valores dentro de un intervalo definido.

- Las probabilidades se representan con los símbolos pi o p(xi).

- El gráfico de la distribución de probabilidad se denomina gráfico de bastones.

- Las probabilidades se calculan mediante la aplicación de las reglas provenientes de la teoría clásica de probabilidad como de fórmulas específicas.

- La condición de cierre se verifica realizando la sumatoria de las probabilidades p = 1.

- La distribución de probabilidad en el caso de una variable aleatoria discreta se denomina genéricamente función de probabilidad.

## Distribución Binomial

Se utilizan para variables del tipo binario (lanzar una moneda) en donde los eventos son igualmente probables.

p= probabilidad de éxito.

q= Probabilidad de fracaso.

n= espacio muestral.

k= número de éxitos.

### Ejemplo

Una novela ha tenido un gran éxito, y se estima que el 80% de un grupo de lectores ya la han leído.

En un grupo de 4 amigos aficionados a la lectura:

¿Cuál es la probabilidad de que en el grupo hayan leído la novela 2 personas? Caso binario: Haber leído la novela - No haber leído la novela.

#Probabilidad de que 2 integrantes del grupo hayan leído la novela con una probabilidad de éxito del 0.8.

print(funcion\_binomial(2,4,0.8))

## Distribución Poisson

Se utiliza para describir sucesos raros en donde se considera que la probabilidad del suceso es muy pequeña. La variable aleatoria es el número de veces que ocurre un evento en un intervalo de tiempo, distancia, área, volumen u otra similar.

lamda= espacio muestral.

k= número de éxitos.

### Ejemplo

La probabilidad de que en el lapso de una semana en el taller de la concesionaria uno de los autos vendidos tenga problemas cubiertos por la garantía es 0,02. Suponiendo que en el taller se atienden 450 autos semanalmente. ¿Cuál es la probabilidad de que se presenten 5 autos con problemas por semana?

## Distribución Hipergeométrica

En un experimento de características hipergeométricas el resultado de una observación es afectado por los resultados de las observaciones previas, por tanto, las probabilidades son condicionales.

### Ejemplo

Una empresa que importa los autos que vende una concesionaria, desea hacer una encuesta de satisfacción a los compradores de estos autos. De una muestra de 80 autos, 30 son importados. Si se seleccionan 9 clientes. ¿Cuál es la probabilidad de que haya 2 que compraron autos importados?

## Variables Continuas

- Puede tomar cualquier valor en un determinado campo de variación.

- La probabilidad se representa con los símbolos 􀁽 o f(x).

- En un punto la probabilidad no tiene sentido. Sólo tiene sentido en un intervalo particular de la variable aleatoria xi, por más pequeño que éste sea.

- En el gráfico, se ve como una función continua f(x), y la probabilidad en sí misma, denominada A, se representa como un área entre los puntos x1 y x2.

## Distribución normal

Esta dada por una función de densidad y la probabilidad se obtiene en base a una variable aleatoria xi que se encuentra entre dos valores arbitrarios de x1 y x2, la cual está dada por el área A bajo la curva cuyo valor se encuentra integrando la función f(x) entre ambos valores, es decir que en tanto la probabilidad en un punto cualquiera no tiene sentido. La solución práctica para obtener esas probabilidades consiste en utilizar la Tabla de Probabilidades apropiada para calcular cualquier probabilidad en el caso normal, sin que importe cuáles son los valores particulares de la variable aleatoria ni los parámetros de la distribución.

## Estandarización

La distribución normal requiere la estandarización de las variables mediante la siguiente fórmula:

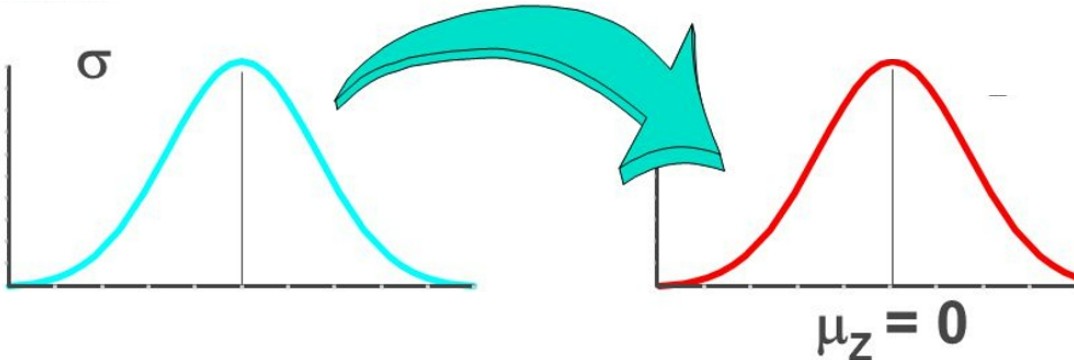
X = Variable aleatoria.

mu = Media.

sigma = Desvío estándar.

Distribución Normal

*z x\_-=\_µ*



O'z = 1

µ

*X*

Distribución Normal Estandarizada

z

### Ejemplo

Luego de estandarizar las variables, se debe buscar el valor de Z en la tabla de distribución normal y determinar la probabilidad en base al área delimitada por el experimento.

## Resumen

***Camerísticas de las Distribuciones de Probabilidad***

|  |  |  |
| --- | --- | --- |
| ***Tipo ,le vllriable*** | ***Discretll*** | ***Co11ti1111a*** |
| ***Dentro de un i11tervalo*** | Pu ede tomar sólo algu- nos valores | Puede tomar cualquier valor |
| ***Simbología ,Le la probabilidad*** | p¡ o p(x) | f¡ o ftx) |
| ***Co11ce pto de***  ***probabilidad*** | En un punto | En un interva lo  (en un punto no tiene sentido) |
| ***Cltlculo de la*** |  | *X* |
| ***Probabilitlad*** | *P( x.* = *x )* = *p*  1 11 *11* | *P(x 5 x 5 x* ) = *{ f (x)dx* = *A*  1 , *2* ***X*** |
|  |  | 1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Tipo de variable*** | ***Discreta*** | | ***Co11ti11ua*** | |
| ***Gráfico*** | De ba s tone s | | De á reas | |
|  | |  | |
| **p**  **pl** [  **pi** pr  1 1 | **x,** |
|  | ( |
| **l l** "N  **X X** |
|  | | **x,** X, **x1** | |
| ***Valor de la probabilidad*** | Vale p¡ (bastón) o cero | | Vale A (área) | |
| ***Co1u/ició11 tfe cierre*** | *N*  *LP* = 1  i• I *I* | | ...,  ***ff ( x)dx*** = 1  -«> | |
| ***Denomi11ació11 ge nérica*** | Función de probabi lidad | | Función de Densidad | |

# Test Estadísticos

## Tests

Por lo general, en la práctica, se tienen que tomar decisiones sobre poblaciones, partiendo de la información muestral de las mismas. Tales decisiones se llaman, decisiones estadísticas. Para tomar decisiones conviene hacer determinados supuestos y tales supuestos son formulados respecto del valor de algún parámetro, que pueden ser o no ciertos. A estos los denominamos Hipótesis Estadísticas y, en general, lo son sobre las distribuciones de probabilidad de las poblaciones.

## H0 y Ha

En el procedimiento de test estadísticos, se utilizan las denominadas pruebas de hipótesis y en ellas se usan datos de una muestra para probar dos afirmaciones contrarias indicadas por H0 (hipótesis nula) y Ha (la hipótesis alternativa).

## Hipótesis en Investigaciones

La puntuación promedio de un Henry Challenge es de 78, Henry determina que a través de métodos innovadores puede aumentar esa media.

En este caso, se establece un grupo de investigación que busca evidencias para concluir que el nuevo sistema aumenta la media del rendimiento.

La hipótesis de investigación es, entonces, que el nuevo sistema proporciona un rendimiento medio mayor. Es decir, μ > 78. Como lineamiento general, una hipótesis de investigación se debe plantear como hipótesis alternativa.

- H0: μ <= 78.

- Ha: μ > 78.

## Hipótesis en Afirmación

Cuando lo que realizamos es una afirmación, en este caso sería que quienes rinden un HC obtienen por lo menos 78 puntos en promedio, tratamos de corroborar que esa afirmación es correcta.

- H0: μ >= 78.

- Ha: μ < 78.

Es decir que en toda situación en la que se desee probar la validez de una afirmación, la hipótesis nula se suele basar en la suposición de que la afirmación sea verdadera. Entonces, la hipótesis alternativa se formula de manera que rechazar H0 proporcione la evidencia estadística de que la suposición establecida es incorrecta.

## Hipótesis en Decisiones / Alternativa

Existen además otras formas de realizar el planteo de H0 y Ha, Como cuando se debe tomar una decisión. Por ejemplo, controlar la calidad de un determinado repuesto en donde debe medir obligatoriamente 10 cm.

- H0: μ = 10.

- Ha: μ ! = 10.

Lo que determina solo dos alternativas.

## Resumen

Alternativas al plantear la hipótesis nula y alternativa:

Investigación, afirmación y decisiones/alternativa.

## Metodología

1. Formular la hipótesis nula.

2. Formular la hipótesis alternativa.

3. Especificar el nivel de significación.

4. Determinar el tamaño de la muestra.

5. Determinar el estadístico de prueba.

6. Establecer los valores críticos que dividen las zonas de rechazo y de no rechazo.

7. Obtener los datos y calcular los estadísticos.

8. Determinar el estadístico de prueba ha caído en la región de rechazo o en la de no rechazo.

9. Determinar la decisión estadística.

10. Expresar la decisión estadística en términos del problema.

## Especificar el nivel de significación

La distribución muestral del estadístico analizado, suele seguir una distribución estadística conocida, como la distribución normal estandarizada, la distribución t o la distribución chi cuadrado, éstas se utilizan como ayuda para determinar si la hipótesis nula es cierta.

Existen dos tipos de errores:

Error de tipo I: Es la probabilidad de que se rechace la hipótesis nula cuando es verdadera.

Error de tipo II: Es la probabilidad de aceptar la hipótesis nula cuando es falsa.

## Tamaño de la muestra

El tamaño de la muestra se determina al tomar en cuenta la importancia de Error de Tipo I y Error de Tipo II y al considerar las restricciones presupuestarias al efectuar el estudio.

Generalmente las muestras grandes, permiten detectar incluso diferencias pequeñas entre los valores hipotéticos los parámetros poblacionales. Para un nivel de Error I dado, aumentar el tamaño de la muestra reducirá Error II y así se incrementará el poder de la prueba para detectar que la hipótesis nula es falsa.

## Estadístico de prueba

Una vez definidas las hipótesis nula y alternativa, y el tamaño de la muestra se puede establecer la distribución a utilizar: normal, t- student ó chi cuadrado.

## Zonas de rechazo y de no rechazo.

Se representa como un área (como toda probabilidad en una función de densidad), que se ubica a la derecha, a la izquierda o a ambos lados (en este caso, con la mitad del Error de Tipo I en cada lado) según como se haya de􀁽nido la Hipótesis alternativa.

Esta probabilidad permite encontrar, en la tabla correspondiente (normal, t de Student o chi cuadrado), un valor de la variable (z, t ó chi cuadrado) denominado “valor crítico” simbolizado con zc (o eventualmente tc ó chi c), que divide al eje de las abscisas en dos zonas: la “zona de rechazo”, que se extiende por debajo del error de Tipo I , y la “zona de no rechazo”, que se extiende a lo largo del resto del eje.

## Obtener los datos y calcular los estadísticos

Este paso está reservado a la efectiva realización de la investigación muestral. Es decir que en este momento es cuando se realiza el estudio tendiente a obtener los valores muestrales y calcular los estadísticos.

## Estadístico

Se debe determinar la técnica a utilizar para determinar si el estadístico muestral ha caído en la región de rechazo o en la de no rechazo, es decir, el modo en que el estadístico de la muestra se va a comparar con el parámetro hipotético. El estadístico de prueba puede ser el estadístico muestral (el estimador insesgado del parámetro que se prueba) o una versión transformada de ese estadístico muestral.

Se debe construir una variable estandarizada a partir de los resultados obtenidos.

## Determinar la decisión estadística

Se determina la decisión de la prueba de hipótesis - si z1 > zc entonces z1 cae en la “zona de rechazo” y se considera que las diferencias entre z1 y zc son significativas entonces Rechazo la Hipótesis nula. - si z1 ≤ zc entonces z1 cae en la “zona de no rechazo” y se considera que las diferencias entre z1 y zc no son significativas entonces No Rechazo la Hipótesis nula.

## Pruebas de una cola

Cuando nos encontramos frente a una prueba de hipótesis del tipo: "H0: μ<= x" y "H1: μ> x" o "H0: μ>= x" y "H1: μ< x" , la denominamos prueba de una cola.

## Pruebas de dos colas

Cuando nos encontramos frente a una prueba de hipótesis del tipo: "H0: μ= x" y "H1: μ!= 1x", la denominamos prueba de dos colas.