**TEMA 2:**

1. Indica cómo es cada una de estas variables: cuantitativa discreta (CD), cuantitativa continua (CC), cualitativa nominal (CLN), cualitativa ordinal (CLO).

| ***Variables:*** | ***Solución:*** |
|---|---|
| Velocidad de red | |
| Tiempo dedicado a jugar con la consola a la semana | |
| Número de discos duros defectuosos en un pedido de 950 | |
| Número de conexiones a Internet en un periodo de tiempo | |
| Navegador más utilizado | |
| Marca personal en 100 metros lisos | |
| Altura de los jóvenes entre 18 y 25 años | |
| Estado de conservación de diferentes ordenadores (malo, regular, bueno, muy bueno) | |
| Velocidad en Hz. de un microprocesador | |

2. A partir de la siguiente tabla de números aleatorios extrae una muestra aleatoria simple de 7 números entre el 1 y el 850. Supón que aleatoriamente has obtenido como punto de inicio la fila 5 y columna 4 y se va a hacer el muestreo aleatorio simple avanzando de arriba hacia abajo.

0781 0454 6226 5735 4274 9857 5450 2593 0333 3375 8797

0219 1607 1024 8066 0735 0572 2998 4823 7002 4545 7372

0747 5816 0128 9876 2757 5086 6283 9888 2936 6431 6606

8997 5094 9229 7900 2415 3189 8646 8855 0265 8987 0091

4636 1631 9032 **4857** 5423 7069 8031 5103 4885 7811 6102

0096 7565 9378 9321 1712 4353 9441 8244 6290 2138 9030

9654 6586 3314 6857 8626 1217 9186 8338 9790 8295 8564

3509 1672 7787 8389 5892 4113 4896 0087 4198 4443 4566

5215 0929 2585 4884 3827 2306 3618 7247 8194 8639 4355

9895 0659 9411 3338 6897 4115 1717 4111 9004 8498 5323





La muestra obtenida es:

3. Obtén una muestra sistemática de 5 elementos entre los números 1 y 248.

   N= n= k=

   Nota que si N/n da decimal para obtener k hay que ---------------- el valor que nos ha dado.
   h=número aleatorio entre 1 y ---------- (pon el valor numérico). Vamos a suponer que nos ha dado h=3. Entonces la muestra sistemática se obtendría de la siguiente forma:
   h, h+k, h+2k, h+3k, h+4k.

   Sustituyendo los valores de h y k, la muestra obtenida es:

4. Una gran compañía ha hecho una compra de 31500 ordenadores portátiles, para sus trabajadores, a tres empresas (A, B y C). De la empresa A compró 12900 portátiles, de la empresa B compró 7234 y de la empresa C compró el resto. El director de la compañía desea realizar un control de calidad sobre 200 de esos portátiles obtenidos mediante un muestreo aleatorio estratificado. Índica (haciendo todos los pasos como se ha hecho para la empresa A) cuántos ordenadores tiene que muestrear de cada empresa.
   nA= tamaño de la muestra de la empresa A.

$$\frac{12900}{31500} = \frac{nA}{200} \rightarrow nA = \frac{200 \cdot 12900}{31500} = 81.9 \sim 82$$

   nB= tamaño de la muestra de la empresa B.

   nC= tamaño de la muestra de la empresa C.

Observa que para obtener los tamaños muestrales, si estos dan con decimales, habrá que ---------------------, teniendo en cuenta que la suma de todos los tamaños muestrales debe dar el tamaño de la muestra total que en este caso es: -------------. Comprobación:

5. Una empresa acaba de recibir un pedido de portátiles para ponerlos a la venta entre sus clientes. Dichos portátiles vienen numerados con códigos desde el 354 al 597. El gerente de dicho centro está preocupado por la calidad de dichos portátiles y decide obtener una muestra sistemática de 6 portátiles y someterlos a varias pruebas. Ayúdale a obtener la muestra.

   N= n= k=

   h=número aleatorio entre 1 y ------ (pon el valor numérico). Vamos a suponer que nos ha dado h=2. Entonces la muestra sistemática entre 1 y N= --------, se obtendría de la siguiente forma: h, h+k, h+2k, h+3k, h+4k, h+5k, y sustituyendo los valores correspondientes daría:

   Pero como la muestra sistemática que nos piden se identifica con códigos entre 354 y 597, tendremos que sumar --------------- a cada uno de los elementos obtenidos anteriormente, consiguiendo la siguiente muestra sistemática de códigos:

**TEMA 4:**

1. Consideramos el experimento aleatorio de lanzar dos dados. Realiza las siguientes cuestiones:

Obtén el espacio muestral del experimento.





Indica si las siguientes frases son verdaderas o falsas:

- Los sucesos A={obtener un 1 en el primer lanzamiento}={11, 12, 13, 14, 15, 16} es un suceso aleatorio. ¿V o F? ----.
- El suceso A={Obtener un 1 en el primer lanzamiento} es un suceso elemental. ¿V o F? ----.
- El suceso A={Obtener un 1 en el primer lanzamiento y un 3 en el segundo} es un suceso elemental. ¿V o F? ----.
- El suceso A={La suma de las puntuaciones de los dados es mayor que 12} es un suceso seguro. ¿V o F? -------- ¿Si has indicado que es Falso cómo llamarías a este suceso?-----------------------------.

Vamos a trabajar los conceptos de **sucesos compatibles, incompatibles y complementarios**. Completa las siguientes frases:

Los sucesos A={Sacar un número par menor que 4 en el primer lanzamiento} y B={Sacar un número impar en el segundo lanzamiento} son sucesos -------------------------------- ya que la intersección nos da:

$A \cap B = \{ \qquad \qquad \qquad \}$,

que es distinto del conjunto ------------------------.

Los sucesos A={Sacar un número par menor que 4 en el primer lanzamiento} y B={La suma de los dados es 9} son sucesos -------------. Ya que A∩B=-------------, es decir el conjunto ---------------.

Sea A={La suma de los dados es menor que 9} y B={La suma de los dados es mayor que 9}. Estos sucesos ¿Son o no complementarios? ¿Por qué? ----
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
-----------------------------------------------------------------.

Los sucesos A={La suma de los dados es menor que 5} y B={La suma de los dados es mayor o igual que 5} son sucesos ----------------------- porque la unión de ellos nos da todo el --------------------------------- y la intersección es --------------------------------.

2. Consideramos el experimento aleatorio de lanzar dos dados. Teniendo en cuenta su espacio muestral, obtenido en el anterior ejercicio, calcula las siguientes probabilidades:

Probabilidad de que salga un 1 en el primer lanzamiento:

$$P=\frac{Número\ de\ casos\ favorables}{Número\ de\ casos\ posibles}=$$

Probabilidad de que la suma de los dados sea superior a 8:

$$P=\frac{Número\ de\ casos\ favorables}{Número\ de\ casos\ posibles}=$$

3. Repasamos **las propiedades de la probabilidad.**

- Las probabilidades siempre toman valores entre ----------- y ------------.
- La probabilidad de un suceso es igual a 1 menos la probabilidad de su ----------------------------. Por tanto:
  P(A)=1-P(   )   y   P($\bar{A}$)=1-P(     ).
- P(A∪B)= . Notemos que si A y B son incompatibles, obtendremos que
  P(A∪B)= , ya que la intersección de dos sucesos incompatibles es el ------------------------ y P(∅) =
- P(A∩$\overline{B}$)=P(A-B)= .Notemos que si B está contenido en A, podremos calcular esta probabilidad como P(A-B)= , ya que en este caso A∩B=
- Sean A y B dos sucesos tales que P(A)>0. Denotamos por P(B|A) a la probabilidad de B dado (o condicionada a) que ha ocurrido A. Entonces:





P(B|A)= y por tanto despejando obtenemos:
P(A∩B)=P(        )P(     ).

- Sean A y B dos sucesos tales que P(B)>0. Denotamos por P(A|B) a la probabilidad de A dado que B ha ocurrido. Entonces:
  P(A|B)= y por tanto despejando obtenemos
  P(A∩B)=P(        )P(     ).
- Por definición A y B son independientes si P(B|A)= ------ y P(A|B)=------.
- A y B son independientes si y sólo si P(A∩B)=P(   )P(   ).
- $P(\bar{A}\cap\bar{B})$ =P(            )=1-P(            ).
- $P(\bar{A}\cup\bar{B})$ =P(            )= 1-P(            ).

4. Consideremos los sucesos A, B tal que P(A) =1/3, P(B) = 2/4 y P(A∩B)= 1/8. Calcula las siguientes probabilidades.
   Probabilidad de que se verifique alguno de los sucesos.

   Probabilidad de que no suceda B.

   Probabilidad de que no ocurra ni A ni B.

   Probabilidad de que bien no ocurra A o bien no ocurra B.

5. Si A y B son dos sucesos tal que $P(A)=0.15$, $P(A\cup B)=0.75$ y $P(A\cap B)=0.03$. Calcula la probabilidad de A-B y de B-A.

P(A-B)=

P(B-A)=

6. Consideremos dos sucesos A, B independientes tal que $P(A) =0.7$ y $P(A\cap B)= 0.4$. Calcula $P(A\cup B)$.

$P(A\cup B)=$

**Recordemos el teorema de la probabilidad total y el teorema de Bayes.**

Sean $A_1, A_2,..., A_n$, sucesos incompatibles dos a dos y cuya unión es el espacio muestral. Sea B otro suceso. Entonces:

**Teorema de la probabilidad total:**

$$P(B) = P(B|A_1)\, P(A_1) + P(B|A_2)P(A_2) + ... + P(B|A_n)P(A_n)$$

**Teorema de Bayes:**

$P(A_k|B) = P(B|A_k)P(A_k)/P(B)=$

$=P(B|A_k)P(A_k)/ (P(B|A_1)\, P(A_1) + P(B|A_2)P(A_2) + ... + P(B|A_n)P(A_n))$

7. Se sabe que el 65% de los accidentes de tráfico que se producen durante la noche de los sábados se deben a la ingesta excesiva de alcohol, el 25% se deben a la imprudencia del conductor y el resto a otras causas. En estos accidentes, el resultado es nefasto el 30% de las veces en el primer caso, el 20% en el segundo y el 5% en el tercero.

- Calcula la probabilidad de que un sábado por la noche se produzca un accidente con resultado nefasto.





Definimos $A_1$={accidente por ingesta excesiva de alcohol}, $A_2$={accidente por imprudencia del conductor}, $A_3$={accidente por otras causas}. Sea B={accidente con resultado nefasto}. Entonces:

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

$P(B|A_1)=$ $P(B|A_2)=$ $P(B|A_3)=$

P(B)=

- Si un sábado por la noche se produce un accidente con resultado nefasto, calcula la probabilidad de que la causa de dicho accidente sea la ingesta excesiva de alcohol.

$P(\quad|B)=$

- Si un sábado por la noche se produce un accidente con resultado nefasto, calcula la probabilidad de que la causa de dicho accidente no sea la ingesta excesiva de alcohol.

$P(\quad|B)=$

- Si un sábado por la noche se produce un accidente con resultado no nefasto, calcula la probabilidad de que la causa de dicho accidente sea la ingesta excesiva de alcohol.

$P(\quad|\overline{B})=$

8. En una caja hay 10 bolas rojas y 20 bolas negras; 5 de las bolas rojas y 4 de las bolas negras tienen marcado un punto blanco. Si se escoge una bola al azar, completa la siguiente tabla y calcula la probabilidad de los siguientes sucesos. Obtén los resultados en forma de fracción irreducible.

| | Roja | Negra | Total fila |
|---|---|---|---|
| Con punto blanco | | | |
| Sin punto blanco | | | |
| Total columna | | | Total= |

Sea roja y tenga marcado un punto blanco.

No tenga marcado un punto blanco o sea negra.

Sea negra, sabiendo que tiene marcado un punto blanco.

9. **Repasemos el análisis combinatorio:**
   **Factorial de n:** $n! =$

   $0! =$

   Calcula:

   $4! =$

   $\frac{5!2!}{4!3!} =$

   $\frac{4!2!5!}{7!3!} =$





**Variaciones sin repetición (o simplemente variaciones):**

Importa el orden y los elementos no se pueden repetir.

Número de variaciones de m elementos tomados de n en n:

$V_{m,n} =$

Ejemplo: $V_{7,2} =$

Obtén todas las variaciones posibles de las letras A, B, C tomadas de 2 en 2.

AB BA AC CA BC CB ¿Cuántas hay? $V_{3,2} = 3 \cdot 2 = 6$.

**Variaciones con repetición:**

Importa el orden y los elementos pueden repetirse.

Número de variaciones con repetición de m elementos tomados de n en n: $VR_{m,n} =$

Ejemplo: $VR_{7,2} =$

Obtén todas las variaciones con repetición de las letras A, B, C tomadas de 2 en 2.

AB BA AC CA BC CB AA BB CC ¿Cuántas hay? $VR_{3,2} = 3^2 = 9$.

**Permutaciones sin repetición (o simplemente permutaciones):**

Importa el orden, los elementos no se pueden repetir y se utilizan todos.

Numero de permutaciones de m elementos: $P_m = V_{m,m} =$

Ejemplo: $P_4 =$

Obtén todas las permutaciones posibles de las letras A, B, C

ABC ACB BAC BCA CAB CBA

¿Cuántas hay? $P_3 = V_{3,3} = 3! = 3 \cdot 2 \cdot 1 = 6$.

**Permutaciones con repetición:**

Tenemos m elementos y cada uno se repite una serie de veces a1, a2,…,ar tal que a1+a2+…+ar=m. Importa el orden y se utilizan todos los elementos (incluidos todos los repetidos) para hacer la permutación de los elementos.

Número de permutaciones con repetición: $PRm^{a1,\ a2,\ \dots,\ ar} =$

Ejemplo: $PR_5^{2,\ 2,\ 1} =$

Obtén todas las permutaciones posibles de las letras A, A, B, C.

AABC AACB ABAC ABCA ACAB ACBA BAAC BACA BCAA CAAB CABA CBAA

¿Cuántas hay? $PR4^{2,\ 1,\ 1} = \frac{4!}{2!1!1!} = 12.$

**Combinaciones sin repetición (o simplemente combinaciones):**

No importa el orden y los elementos no se pueden repetir.

Número de combinaciones de m elementos tomados de n en n:

$Cm,n = \binom{m}{n} =$

Ejemplo: $C7,2 =$

Obtén las combinaciones posibles de las letras A, B, C tomadas de 2 en 2.

{A B}, {A C}, {B C}

¿Cuántas hay? $C3,2 = \binom{3}{2} = 3.$

**Combinaciones con repetición:**

No importa el orden y los elementos se pueden repetir

Número de combinaciones con repetición de m elementos tomados de n en n:

$CRm,n =$

$CR7,2 =$





Obtén las combinaciones con repetición posibles de las letras A, B, C tomadas de 2 en 2.

{A B}, {A C}, {B C}, {A A}, {B B}, {C C} → $CR3,2 = \binom{3+2-1}{2} = 6.$

10. Cuántos números de cuatro cifras se pueden formar con los dígitos 1, 2, 3, 4 y 5, si no se repite ninguna cifra.

Cuántos de estos números son múltiplos de 2.

11. Cuántos números de tres cifras se pueden formar con los dígitos 1, 2, 3, 4 y 5.

Cuántos de estos números empiezan por 3 o 5.

12. En el diseño de un videojuego de carreras, un circuito de velocidad debe incluir 11 curvas, de las que cinco deben ser hacia la derecha y seis hacia la izquierda. Calcula de cuántas formas distintas pueden distribuirse las curvas en el circuito.

13. Una urna contiene 15 bolas, de las cuales 7 son azules, 5 son rojas y 3 verdes. Se extraen tres bolas al azar (sin devolución). Calcula la probabilidad de que al menos una sea roja.

**TEMA 5:**

1. **Vamos a profundizar sobre las variables aleatorias discretas:**

Sea X una variable aleatoria discreta. Entonces se define su f**unción de cuantía como:**

$f(x)=$ , para todo x que pertenece al rango de la variable aleatoria $X$ (es decir, para todo valor x que puede tomar la variable X). El rango de una variable X lo denotaremos por ---------------.

La función de cuantía toma valores entre -------- y ----------. Y la suma de la función de cuantía para todos los valores que puede tomar la variable debe dar -----, matemáticamente se puede expresar como:

**Función de distribución de una v.a. discreta:**

$F(x)=$ , para todo x perteneciente al conjunto de los números reales.

**Esperanza y propiedades de una v.a. discreta:**

$E(X)=$

$E(h(X))=$ h(X) es una función de X (por ejemplo $X^2$, $X^3+1$, $X^5-2$, …).

$E(X^2)=$

$E(aX+b)=$

$E(X_1+X_2+\ldots+X_n)=$

$E(a_1X_1+a_2X_2+\ldots+a_nX_n+b)=$

Sean n variables independientes $X_1, X_2,\ldots, X_n$, entonces:

$E(X_1X_2\ldots X_n)=$





**Varianza, desviación típica de una v.a. discreta y propiedades:**

**Varianza:** $\sigma^2 = Var(X)=$

**Desviación típica:** $\sigma =$

$Var(aX+b)=$

Sean n variables independientes $X_1, X_2,\ldots, X_n$:

$Var(X_1+X_2+\ldots+X_n)=$

$Var(a_1X_1+a_2X_2+\ldots+a_nX_n+b)=$

Sean $X_1$ y $X_2$ variables independientes:

$Var(X_1+X_2)=$ $Var(X_1-X_2)=$

- Un técnico informático recibe un salario mensual de 1200 euros más una comisión de 50 euros por cada ordenador que instala. El número de ordenadores que instala al mes es una variable aleatoria X cuya función de cuantía viene definida en la tabla.

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X=x_i)$ | 0.1 | 0.2 | k | 0.2 | 0.1 |

Calcula k:

Calcula la función de distribución de la variable aleatoria discreta X=número de ordenadores instalados al mes. Y completa las siguientes frases:

La probabilidad de que en un mes se instalen menos de 3 ordenadores es igual a --------, que corresponde al valor de la función de distribución para x=------------, es decir F( ). La probabilidad de que en un mes se instalen 3 o menos ordenadores es igual a --------, que corresponde al valor de la función de distribución para x= ----------, es decir F( ).

La probabilidad de que en un mes se instalen entre 2 y 4 ordenadores es igual a --------, que usando la función de distribución se puede expresar como: $P(2\leq X\leq 4)=P(X\leq 4)-P(X \quad 2)= P(X\leq 4)-P(X \quad 1)=F(\ )-F(\ )= \quad - \quad = \quad .$

Calcula el número esperado de ordenadores instalados al mes.

Calcula el salario mensual esperado de dicho técnico.

Y=salario mensual→ Y=---X+------- → E(Y)=E( )=

=---E( )+--------------=-------------------------= euros.

Calcula la desviación típica del salario mensual de dicho técnico.

Y=salario mensual→$\sigma_Y=\sqrt{Var(Y)} =$





Calcula $E(X^3+1)$, para la variable aleatoria X definida en el ejercicio.

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $x_i^3+1$ | | | | | |
| $P(X=x_i)$ | 0.1 | 0.2 | **0.4** | 0.2 | 0.1 |

E(Y)=

2. **Vamos a profundizar sobre la distribución Binomial:**
   Se considera un experimento con sólo dos posibles resultados (éxito y fracaso). Se realiza de forma independiente el experimento n veces.
   Consideramos la variable X= número de éxitos en n experiencias, con p=P(éxito) constante. Entonces X se distribuye binomial con parámetros n y p. Y se denota por: $X\sim B(n,p)$
   Además sabemos que $E(X)=$ y $Var(X)=$

   Las variables que se distribuyen binomial son ---------------- (¿discretas o continuas?).

   **Función de cuantía de $X\sim B(n,p)$:**

   $f(k)=P(X=k)=$ para k=0,…, n.

- Sea $X\sim B(6,0.2)$, Calcula mediante la fórmula:

  $P(X=4)=$

- Sea $X \sim B(6,0.2)$ ¿Con qué funciones y parámetros se podrían calcular con el SPSS las siguientes probabilidades? Posteriormente veremos en prácticas cómo obtener el resultado final:

$P(X=4)=$

$P(X\leq 2)=$

$P(X<2)=$

$P(2\leq X\leq 4)=$

- Un examen consta de 10 preguntas a las que hay que contestar sí o no. Suponiendo que las personas que realizan dicho examen no saben contestar a ninguna de las preguntas y en consecuencia contestan al azar, calcula la probabilidad de obtener algún acierto:





3. **Vamos a profundizar sobre la distribución Poisson:**

Modela el número de veces que se verifica algún fenómeno por unidad de tiempo, espacio, superficie o volumen. Por ejemplo: sea X= número de acontecimientos en una unidad de tiempo. Sea $\lambda$= número medio de veces que ocurre ese acontecimiento en dicha unidad de tiempo. Entonces X se distribuye Poisson con parámetro $\lambda$, y se denota por: $X \sim Po(\lambda)$. Además sabemos que $E(X)=$ y $Var(X)=$ .Las variables que se distribuyen Poisson son ------------------- (¿discretas o continuas?).

**Propiedad:** Si $X_1,\ldots, X_k$ son variables aleatorias independientes $Po(\lambda_i)$, $i=1,\ldots,k \rightarrow X_1+\cdots+X_k$ es $Po(\qquad)$.

**Función de cuantía de** $X \sim Po(\lambda)$

$f(k)=P(X=k)=$ para k=0, 1, …

- Sea $X \sim Po(6)$, calcula mediante la fórmula:

$P(X=4)=$

- Sea $X \sim Po(6)$, ¿Con qué funciones y parámetros se podrían calcular con el SPSS las siguientes probabilidades? Posteriormente veremos en prácticas cómo obtener el resultado final:

$P(X=3)=$

$P(X\leq 4)=$

$P(X>5)=$

$P(3\leq X\leq 5)=$

- Los mensajes que llegan a una computadora utilizada como servidor lo hacen de acuerdo con una distribución Poisson con una tasa promedio de 0.1 mensajes por minuto. Definiendo las variables necesarias e indicando en cada caso cómo se distribuyen, calcula la probabilidad de que lleguen:

Entre 3 y 5 mensajes en media hora.

Más de 2 mensajes y menos de 4 mensajes en media hora.

Más de 2 mensajes en media hora y menos de 4 en la media hora siguiente.





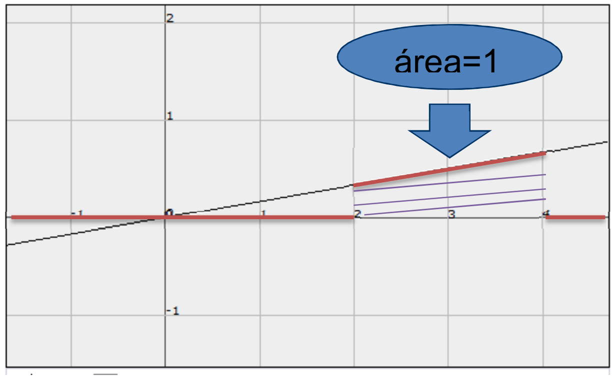
4. **Vamos a profundizar sobre las variables aleatorias continuas:** Una variable aleatoria continua viene definida por su función de densidad. La función de densidad cumple las siguientes condiciones:

- 
- 

- Sea X una variable aleatoria continua. Halla k para que la siguiente función sea la función de densidad de X.

$$\begin{rcases} f(x)=kx, \ 2<x<4 \\ f(x)=0, \text{ en el resto} \end{rcases}$$

Para la variable anterior, cuya función de densidad está representada gráficamente aquí, calcula las siguientes probabilidades y represéntalas gráficamente:




P(X<4)=

P(1<X<3)=

P(X>3)=

P(X=2)=

Nota que si X es una variable aleatoria continua, **P(X=a)=0** para todo a perteneciente a R.





Se define **la función de distribución** de una variable aleatoria continua como: F(x)=P(X≤x)= , para todo x perteneciente a R.

- Calcula la función de distribución de la anterior variable aleatoria X cuya función de densidad era:

f(x)=x/6, 2<x<4

f(x)=0, en el resto

Calcula las siguientes probabilidades con ayuda de la función de distribución obtenida:

P(X<4)=

P(1<X<3)=

P(X>3)=

$P(X>3|X>2.5)=$

$P(X<3|X>2.5)=$

**Esperanza de una v.a. continua y sus propiedades**

$E(X)=$

$E(h(X))=$ donde h(X) es una función de la variable X.

$E(X^2)=$





**Varianza, desviación típica y propiedades de una v.a. continua:**

**Varianza:**

$\sigma^2 = Var(X)=$

**Desviación típica:**

$\sigma =$

**Las propiedades de la esperanza y la varianza son las mismas en el caso discreto y continuo.**

- Calcula para la variable continua de los ejemplos anteriores:

$E(X)=$

$Var(X)=$

5. **Vamos a profundizar sobre la distribución Normal.**
   Sea X una v.a. que se distribuye N(μ,σ)→ Dibuja su función de densidad:

Las variables que se distribuyen normal son ---------------------.

E(X)= y Var(X)=

$Z=\frac{X-\mu}{\sigma} \sim$

- Sea $X \sim N(55,12)$, expresa con las funciones del SPSS las siguientes probabilidades:

$P(X \leq 58)=$

$P(X > 50.8)=$

$P(49 < X < 61)=$

Para la misma variable, calcula k tal que $P(X<k)=0.75$ y represéntalo gráficamente.

k=-----------------------------------------------------------=





Denominamos $Z\alpha$ = percentil $1-\alpha$ de una N(0,1). Represéntalo gráficamente.

¿Cómo se calcularían en este caso con el SPSS estos valores críticos?

$Z_{0.025}=$

$Z_{0.85}=$

- La duración de cierto tipo de baterías sigue una distribución normal con media de 3 años y desviación típica de 0.5 años. Una batería se considera fuera de los límites de control de calidad si su duración es inferior a 2 años. Calcula la probabilidad de que una batería esté fuera de los límites de control.

- La longitud X en milímetros de ciertos componentes electrónicos, sigue una distribución N(15,σ). Si el 80% de la producción tiene una longitud inferior a 18 milímetros, calcula el valor de σ. Usa dos decimales en los cálculos intermedios.

- Sean $X_1$, $X_2$,..., $X_n$, variables aleatorias independientes con distribución $N(\mu_i, \sigma_i)$, i=1, 2,..., n, respectivamente, entonces la variable $a+b_1X_1+b_2X_2+\ldots+b_nX_n$, con a, $b_1$, $b_2$, …, $b_n$ constantes, se distribuye:

- La resistencia eléctrica equivalente de una asociación en serie de resistencias es la suma aritmética de éstas. En un circuito de un aparato electrónico hay tres resistencias soldadas en serie, extraídas de poblaciones normales con medias 250, 300 y 200 ohmios, respectivamente y desviaciones típicas 2, 3 y 2 ohmios, respectivamente. Para que el aparato funcione bien la resistencia equivalente de esta parte del circuito debe estar comprendida entre 740 y 760 ohmios. Calcula el porcentaje de aparatos que no funcionarán bien, debido a un valor inadmisible de la citada resistencia equivalente.





- Sean $X_1$, $X_2$,..., $X_n$, variables aleatorias independientes con distribución $N(\mu, \sigma)$, i=1, 2,..., n, respectivamente, entonces:

  $X_1+X_2+\ldots+X_n$ se distribuye

  $\bar{X}=$ se distribuye

- Las alturas en centímetros de los alumnos de Ingeniería Multimedia siguen una distribución N(170,7). Se eligen 2 alumnos al azar. Calcula la probabilidad de que la suma de sus alturas sea inferior a 350 centímetros.

- Las alturas en centímetros de los estudiantes de Ingeniería Multimedia siguen una distribución N(170,7). Se eligen 16 alumnos al azar. Calcula la probabilidad de que la media de sus alturas sea superior a 176 centímetros.

**TEMA 6:**

**La estimación puntual es un valor que se mide a partir de una muestra y que se usa como estimación de un valor poblacional o parámetro.** Sea $x_1$, $x_2$,..., $x_n$ una muestra extraída de una población. Completa la siguiente tabla:

| **Parámetro poblacional** | **Estimación muestral** |
|---|---|
| **μ=**media poblacional | $\bar{X}$=media muestral<br>Fórmula: |
| **$\sigma^2$=** | $S^2$=<br>Fórmula: |
| **σ=** | S=desviación típica muestral<br>Fórmula: |

El problema de una estimación puntual es que no transmite la sensación de exactitud, ya que no refleja el tamaño de la muestra ni el grado de variabilidad de la población. Esta estimación puede ser muy exacta o estar muy equivocada dependiendo de esos dos factores. Por tanto, en la mayoría de los casos, es preferible estimar los parámetros poblacionales mediante lo que se denomina un intervalo de confianza. La estimación por intervalo establece un intervalo dentro del cual es muy probable que se encuentre el parámetro poblacional. El **coeficiente de confianza** (1-α) se usa para indicar la probabilidad de que una estimación por intervalo contenga el parámetro poblacional. El **nivel de confianza** es el coeficiente de confianza expresado como un porcentaje ((1- α)·100%)

**Intervalos de confianza para la media poblacional:**

Sea $x_1$, $x_2$, . . . , $x_n$, es una muestra aleatoria simple de tamaño n de **una población normal con media μ y desviación típica σ**.

**Si σ es conocida**, entonces:

(1) $\mathbf{I_\mu^{(1-\alpha)\cdot 100\%}=}$





**Si σ es desconocida**, entonces:

(2) $\mathbf{I_\mu^{(1-\alpha)\cdot 100\%}=}$

Si la muestra no proviene de una normal pero es grande (n mayor o igual que 30) se obtienen los siguientes intervalos de confianza:

**Si σ es conocida**, entonces:

(3) $\mathbf{I_\mu^{(1-\alpha)\cdot 100\%}=}$

**Si σ es desconocida**, entonces:

(4) $\mathbf{I_\mu^{(1-\alpha)\cdot 100\%}=}$

**Plantea y resuelve los siguientes ejercicios en folios aparte y completa lo indicado aquí:**

1. Se ha medido la velocidad (Km/h) sobre una muestra de 5 coches que circulan por una carretera, obteniéndose 98, 85, 77, 92, 112. Calcula razonadamente un intervalo de confianza al 95 por ciento para la velocidad media de los vehículos en dicha carretera suponiendo que la muestra ha sido obtenida por muestreo aleatorio simple sobre una población normal.

Define el parámetro poblacional:

¿Usamos fórmulas de Normalidad o de muestra grande?

¿La desviación típica poblacional σ es conocida o desconocida?

**Usamos entonces la fórmula:**

**Datos obtenidos:**

**Solución:**

2. Se ha medido la velocidad (Km/h) sobre una muestra de 5 coches que circulan por una carretera, obteniéndose 98, 85, 77, 92, 112. **Sabiendo que la desviación típica de la velocidad de los coches se asume que es igual a 14 Km/h,** calcula razonadamente un intervalo de confianza al 95 por ciento para la velocidad media de los vehículos en dicha carretera suponiendo que la muestra ha sido obtenida por muestreo aleatorio simple sobre una población normal.

Define el parámetro poblacional:

¿Usamos fórmulas de Normalidad o de muestra grande?

¿La desviación típica poblacional σ es conocida o desconocida?

**Usamos entonces la fórmula:**

**Datos obtenidos:**

**Solución:**

3. La velocidad media de una muestra de 200 conexiones a Internet es 3.6 Mb/sg. Sabiendo que la desviación típica de la velocidad de las conexiones es 0.7 Mb/sg. Halla el intervalo de confianza para la media poblacional para un nivel de confianza del 99 por ciento.

Define el parámetro poblacional:

¿Usamos fórmulas de Normalidad o de muestra grande?

¿La desviación típica poblacional σ es conocida o desconocida?





**Usamos entonces la fórmula:**

**Datos obtenidos:**

**Solución:**

4. Se ha calculado la velocidad media y la desviación típica de una muestra de 200 conexiones a Internet obteniendo de media 3.6 Mb/sg y desviación típica 0.6 Mb/sg. Halla el intervalo de confianza para la media poblacional para un nivel de confianza del 99 por ciento.

Parámetro poblacional:

¿Usamos fórmulas de Normalidad o de muestra grande?

¿La desviación típica poblacional σ es conocida o desconocida?

**Usamos entonces la fórmula:**

**Datos obtenidos:**

**Solución:**

**Intervalos de confianza para la desviación típica poblacional y la varianza poblacional:**

Al igual que para la media poblacional, se pueden obtener intervalos de confianza para la varianza y la desviación típica poblacional. En clase se han dado las fórmulas (también podéis encontrarlas en los capítulos del libro on-line). Ayudándote de dichas fórmulas resuelve el siguiente ejercicio **en folios aparte y completa lo indicado aquí.**

5. El precio de un determinado móvil en los comercios de una ciudad sigue una distribución normal. Se toma una muestra aleatoria simple de 5 comercios y se observa el precio de dicho artículo, obteniendo las siguientes observaciones: 132, 125, 130, 139, 126. Obtén al nivel de confianza del 90% un intervalo de confianza para la varianza poblacional y otro para la desviación típica poblacional.

¿Usamos fórmulas de Normalidad o de muestra grande?

¿La media poblacional μ es conocida o desconocida?

**Para varianza poblacional**

**Parámetro poblacional: $\sigma^2$=**

**Datos obtenidos:**

$$\mathbf{I}_{\sigma^2}^{(1-\alpha)\cdot 100\%} =$$

**Para la desviación típica poblacional**

**Parámetro poblacional: $\sigma$=**

**Datos obtenidos:**

$$\mathbf{I}_{\sigma}^{(1-\alpha)\cdot 100\%} =$$





**Intervalos de confianza para proporciones:**

Se tiene muestra aleatoria simple de tamaño n de una distribución Bernoulli de parámetro p ($0 \le p \le 1$). Es decir se tiene una distribución Binomial del tipo B(1,p). Entonces se puede calcular el siguiente intervalo de confianza para la proporción poblacional, con nivel de confianza $(1-\alpha)\cdot 100\%$.

Si la muestra es pequeña:

$$\mathbf{I}_{p}^{(1-\alpha)\cdot 100\%} =$$

Si la muestra es grande:

$$\mathbf{I}_{p}^{(1-\alpha)\cdot 100\%} =$$

En ambas fórmulas, $\bar{p}$ es la estimación muestral de p.

¿En caso de que nos pidieran calcular intervalos de confianza para porcentajes poblacionales qué haríamos para obtener los resultados a partir de los intervalos de confianza para proporciones?

6. Se está probando un algoritmo que identifica las matrículas de los coches. Se obtiene una muestra aleatoria de 20 matrículas y el algoritmo consigue la identificación correcta de 18 de ellas. Deseamos calcular un intervalo de confianza al 98 por ciento para p, donde p es la proporción poblacional de matrículas leídas correctamente por el algoritmo.

¿Usamos fórmulas de muestras pequeñas o de muestra grande?

**Resolución:**

El intervalo obtenido en este caso hay que corregirlo ya que las proporciones sólo toman valores entre --------------y-----------------. Por lo que la **solución** sería:

$$\mathbf{I_p^{(1-\alpha)\cdot 100\%} =}$$

Hemos visto que el anterior intervalo de confianza es muy pobre ya que da poca información. Lo lógico sería utilizar muestras grandes siempre que sea posible.

7. Supongamos que para el ejercicio anterior se obtiene una muestra aleatoria de 200 matrículas y el algoritmo consigue la identificación correcta de 180 de ellas. Calcula un nuevo intervalo de confianza al 98 por ciento para p, donde p es la proporción poblacional de matrículas leídas correctamente por el algoritmo.

$$\mathbf{I_p^{(1-\alpha)\cdot 100\%} =}$$

¿Cuál sería el intervalo de confianza en este caso para el porcentaje poblacional de matrículas identificadas correctamente por el algoritmo?





**Intervalos de confianza diferencia de medias poblacionales**

Sean $x_1, x_2, \ldots, x_m$, y $y_1, y_2, \ldots, y_n$, dos muestras aleatorias simples, independientes y que provienen de distribuciones normales con parámetros $\mu_1$, $\sigma_1$ y $\mu_2$, $\sigma_2$, respectivamente. Los denotaremos de la siguiente forma: $I\mu_1\text{-}\mu_2^{(1-\alpha)\cdot 100\%}$, donde $(1-\alpha)\cdot 100\%$ es el nivel de confianza.

En clase se han dado las fórmulas (también podéis encontrarlas en los capítulos del libro on-line). Ayudándote de dichas fórmulas resuelve el siguiente ejercicio.

8. Las notas de un examen de estadística (puntuado sobre 100) obtenidas por una muestra de 15 estudiantes de Ingeniería Multimedia y una muestra de 12 estudiantes de Ingeniería Informática ofrecen los siguientes resultados.

Ingeniería Multimedia: media 44, desviación típica 9.

Ingeniería Informática: media 42, desviación típica 12.

Se sabe que estas muestras son aleatorias y provienen de una distribución normal. Por la experiencia adquirida se sabe de antemano que las desviaciones típicas poblacionales son iguales. Calcula el intervalo al 95 por ciento de confianza para la diferencia de medias μ1-μ2 donde μ1 es la nota media del examen de Estadística de los alumnos de Ingeniería Multimedia y μ2 la de los alumnos de Ingeniería Informática.

¿Usamos fórmulas de Normalidad o de muestra grande?

¿Las desviaciones típicas poblacionales $\sigma_1$ y $\sigma_2$ son conocidas o desconocidas?

¿Las desviaciones típicas poblacionales $\sigma_1$ y $\sigma_2$ pueden considerarse iguales?

**Usamos entonces la fórmula:**

**Resolución:**

**Tamaño muestral para estimación de medias**

Si la muestra proviene de una distribución normal, el tamaño apropiado de la muestra para estimar una media con una confianza del (1-α)100% se puede obtener:

no=

con σ=desviación típica de la población o si no se conoce se estima, ε=error tolerable máximo y $Z_{\frac{\alpha}{2}}$=

que con el SPSS se obtendría como:

Cuando trabajamos con poblaciones finitas (supongamos de tamaño N), se debe obtener el tamaño de la muestra corregido: n=





**Tamaño muestral para estimación de proporciones**

El procedimiento para determinar el tamaño de la muestra apropiado para estimar una proporción es similar al que se usa para las medias muestrales. El tamaño apropiado de la muestra para estimar una proporción con una confianza del (1-α)100% se puede obtener:

no=

con p=proporción de la población o si no se conoce se estima, ε=error tolerable máximo y $Z_{\frac{\alpha}{2}}$=

que con el SPSS se obtendría como:

Queremos hacer notar que si no se tiene ninguna información sobre p, se toma p=1/2 obteniendo:

no=

Cuando trabajamos con poblaciones finitas (supongamos de tamaño N), se debe obtener el tamaño de la muestra corregido n=

9. Se desea hacer un estudio sobre el porcentaje de empresas de un cierto sector que ofrecen comercio electrónico. Si se desea cometer un error máximo del 2 por ciento con una confianza del 97 por ciento, índica qué tamaño de muestra deberíamos elegir si se sabe que hay 3450 empresas en dicho sector y en un estudio piloto preliminar se ha obtenido que el 10 por ciento de las empresas del sector ofrecen comercio electrónico.

**TEMA 7:**

En los análisis estadísticos se utilizan con frecuencia los contrastes paramétricos, en los que se requiere decidir si una declaración acerca de un parámetro de una distribución de probabilidad es verdadera o falsa. La declaración se denomina hipótesis, y el procedimiento de tomar la decisión acerca de la falsedad o veracidad de la hipótesis se llama contraste de hipótesis. Los pasos a realizar son los siguientes:

Se establece la hipótesis nula y la hipótesis alternativa. La hipótesis denominada hipótesis nula $H_0$ especifica el valor que se desea contrastar para ver si se ha de rechazar o no rechazar esta hipótesis nula. En caso de rechazar la hipótesis nula se acepta la hipótesis alternativa $H_1$.

Se supone que la hipótesis nula es cierta y se determina la distribución muestral para esta suposición. A partir de esta distribución se obtendrá el estadístico para la prueba.

Se obtiene una muestra aleatoria a partir de la población que se está estudiando, y con los datos obtenidos, se calcula el valor concreto del estadístico para la prueba. Por último, se evalúa la posibilidad de seleccionar ese valor del estadístico para la prueba a partir de la distribución muestral propuesta. Si parece poco probable que el valor obtenido del estadístico provenga de dicha distribución muestral se rechaza la hipótesis nula. Si no es así, no habrá evidencias para sugerir que la hipótesis nula sea falsa.

**TABLA 1: Contrastes de hipótesis bajo normalidad**

| Hipótesis nula | Estadístico de prueba | Hipótesis alternativa | Criterios de rechazo | P-valor |
|---|---|---|---|---|
| $H_0: \mu = \mu_0$ | | $H_1: \mu \neq \mu_0$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| $\sigma^2$ conocida | $Z_0 = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ | $H_1: \mu > \mu_0$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| | | $H_1: \mu < \mu_0$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |
| $H_0: \mu = \mu_0$ | | $H_1: \mu \neq \mu_0$ | $\lvert t_0\rvert > t_{\frac{\alpha}{2},n-1}$ | $2P(t_{n-1} \geq \lvert t_0\rvert)$ |
| $\sigma^2$ desconocida | $t_0 = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ | $H_1: \mu > \mu_0$ | $t_0 > t_{\alpha,n-1}$ | $P(t_{n-1} \geq t_0)$ |
| | | $H_1: \mu < \mu_0$ | $t_0 < -t_{\alpha,n-1}$ | $P(t_{n-1} \leq t_0)$ |
| $H_0: \sigma^2 = \sigma_0^2$ | | $H_1: \sigma^2 \neq \sigma_0^2$ | $\chi_0^2 > \chi^2_{\frac{\alpha}{2},n-1}$ ó $\chi_0^2 < \chi^2_{1-\frac{\alpha}{2},n-1}$ | $2P(\text{cola menor})$ |
| | $\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$ | $H_1: \sigma^2 > \sigma_0^2$ | $\chi_0^2 > \chi^2_{\alpha,n-1}$ | $P(\chi^2_{n-1} \geq \chi_0^2)$ |
| | | $H_1: \sigma^2 < \sigma_0^2$ | $\chi_0^2 < \chi^2_{1-\alpha,n-1}$ | $P(\chi^2_{n-1} \leq \chi_0^2)$ |





**TABLA 2: Contrastes de hipótesis bajo normalidad**

| Hipótesis nula | Estadístico de prueba | Hipótesis alternativa | Criterios de rechazo | P-valor |
|---|---|---|---|---|
| $H_0: \mu_1 = \mu_2$ | | $H_1: \mu_1 \neq \mu_2$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| $\sigma_1^2,\ \sigma_2^2$ | $Z_0 = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ | $H_1: \mu_1 > \mu_2$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| conocidas | | $H_1: \mu_1 < \mu_2$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |
| $H_0: \mu_1 = \mu_2$ | (7.1) | $H_1: \mu_1 \neq \mu_2$ | $\lvert t_0\rvert > t_{\frac{\alpha}{2},n_1+n_2-2}$ | $2P(t_{n_1+n_2-2} \geq \lvert t_0\rvert)$ |
| $\sigma_1^2 = \sigma_2^2$ | $t_0 = \frac{\overline{X}_1-\overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ | $H_1: \mu_1 > \mu_2$ | $t_0 > t_{\alpha,n_1+n_2-2}$ | $P(t_{n_1+n_2-2} \geq t_0)$ |
| desconocidas | | $H_1: \mu_1 < \mu_2$ | $t_0 < -t_{\alpha,n_1+n_2-2}$ | $P(t_{n_1+n_2-2} \leq t_0)$ |
| $H_0: \mu_1 = \mu_2$ | (7.2) | $H_1: \mu_1 \neq \mu_2$ | $\lvert t_0\rvert > t_{\frac{\alpha}{2},\nu}$ | $2P(t_\nu \geq \lvert t_0\rvert)$ |
| $\sigma_1^2 \neq \sigma_2^2$ | $t_0 = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ | $H_1: \mu_1 > \mu_2$ | $t_0 > t_{\alpha,\nu}$ | $P(t_\nu \geq t_0)$ |
| desconocidas | | $H_1: \mu_1 < \mu_2$ | $t_0 < -t_{\alpha,\nu}$ | $P(t_\nu \leq t_0)$ |
| $H_0: \sigma_1^2 = \sigma_2^2$ | | $H_1: \sigma_1^2 \neq \sigma_2^2$ | $F_0 > F_{\frac{\alpha}{2},n_1-1,n_2-1}$ ó $F_0 < F_{1-\frac{\alpha}{2},n_1-1,n_2-1}$ | $2P(\text{cola menor})$ |
| | $F_0 = \frac{S_1^2}{S_2^2}$ | | | |
| | | $H_1: \sigma_1^2 > \sigma_2^2$ | $F_0 > F_{\alpha,n_1-1,n_2-1}$ | $P(F_{n_1-1,n_2-1} \geq F_0)$ |

(7.1)

$$S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}},$$

(7.2)

$$\nu = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

**TABLA 3: Contrastes de hipótesis para muestras grandes**

| Hipótesis nula | Estadístico de prueba | Hipótesis alternativa | Criterios de rechazo | P-valor |
|---|---|---|---|---|
| $H_0: \mu = \mu_0$ | $Z_0 = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ | $H_1: \mu \neq \mu_0$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| | | $H_1: \mu > \mu_0$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| | | $H_1: \mu < \mu_0$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |
| $H_0: \mu_1 = \mu_2$ | $Z_0 = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ | $H_1: \mu_1 \neq \mu_2$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| | | $H_1: \mu_1 > \mu_2$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| | | $H_1: \mu_1 < \mu_2$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |
| $H_0: p = p_0$ | (7.3) $Z_0 = \frac{p^*-p_0}{\sqrt{\frac{(1-p_0)p_0}{n}}}$ | $H_1: p \neq p_0$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| | | $H_1: p > p_0$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| | | $H_1: p < p_0$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |
| $H_0: p_1 = p_2$ | (7.4) $Z_0 = \frac{p_1^*-p_2^*}{\sqrt{p^*(1-p^*)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ | $H_1: p_1 \neq p_2$ | $\lvert Z_0\rvert > Z_{\frac{\alpha}{2}}$ | $2P(Z \geq \lvert Z_0\rvert)$ |
| | | $H_1: p_1 > p_2$ | $Z_0 > Z_\alpha$ | $P(Z \geq Z_0)$ |
| | | $H_1: p_1 < p_2$ | $Z_0 < -Z_\alpha$ | $P(Z \leq Z_0)$ |

(7.3)

$p^* = \frac{X}{n}$, es la aproximación de la proporción poblacional $p$,

donde $X$ es el número de elementos que cumplen la condición que se está estudiando y $n$ es el tamaño de la muestra, y

(7.4)

$$p_1^* = \frac{X_1}{n_1}, \quad p_2^* = \frac{X_2}{n_2}, \quad p^* = \frac{X_1+X_2}{n_1+n_2},$$

donde $n_1$ y $n_2$ son los tamaños muestrales y $p_1^*$ y $p_2^*$ son las aproximaciones de $p_1$ y $p_2$, respectivamente.





Con la ayuda de las tablas anteriores realiza los siguientes ejercicios en hojas aparte.

1. El control de calidad de una fábrica de pilas y baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Hasta ahora se asumía que el tiempo medio de duración en conversación de dichas baterías era 300. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que en una muestra de 100 baterías el tiempo medio de duración en conversación fue de 290 minutos con desviación típica de 30 minutos. Contrasta a un nivel de significación del 1% si las sospechas del control de calidad son ciertas. Representa gráficamente la zona de aceptación y de rechazo de este contraste de hipótesis, indicando además cómo se distribuye el estadístico de prueba bajo la hipótesis nula y donde se sitúa el valor de dicho estadístico para los datos del problema.
2. En municipios de 2000 a 5000 habitantes de todo el territorio nacional se calcula que la media de horas vistas de televisión por hogar al día es de 4.31 horas. Sea una zona de la que se tomó una muestra de 120 municipios de aquellas dimensiones, con un promedio de 3.86 horas de televisión al día y con desviación típica de 1.02. Bajo normalidad, contrasta con un nivel de significación del 2 por ciento si en dicha zona se puede considerar que el promedio es igual al de todo el territorio.
3. Una empresa de desarrollo de videojuegos está analizando el funcionamiento de un nuevo videojuego que va a lanzar al mercado. En uno de los niveles de dicho videojuego hay que lanzar dos dados y atendiendo a la suma de las puntuaciones de los dados aparece un escenario diferente en el que el jugador tiene que sortear una serie de obstáculos. La dificultad del escenario está asociada a la puntuación obtenida. El director de la empresa cree que la implementación de dicha parte del videojuego no se ha realizado correctamente ya que parece que el escenario asociado a la suma de la puntuación de los dados igual a 10 no sigue las reglas de aleatoriedad correspondientes. Para estudiarlo se prueba dicha parte de la implementación simulando el lanzamiento de los dados 1300 veces y se obtiene 90 veces dicha suma. Realizando de forma razonada el contraste de hipótesis adecuado explica si el director de la empresa está en lo cierto.
4. En un intento de innovar y seguir mejorando, se diseñaron dos nuevas infraestructura de red, denominadas de tipo A y de tipo B. Para estudiar las prestaciones de los diseños iniciales, se realizaron pruebas de transferencia de ficheros, transfiriendo un fichero cada segundo durante dos minutos, obteniendo una media de 3.21 Mbps con una desviación típica de 0.7 Mbps para el tipo A y una media de 3.45 Mbps y una desviación típica de 0.9 Mbps para el tipo B. Mediante el contraste de hipótesis adecuado, explica si se podría dar por cierta la afirmación de

que la velocidad de transferencias de ficheros de ambas redes es similar. En caso contrario calculando los P-valores correspondientes explica cuál de las dos redes obtiene mayor velocidad media de transferencia de ficheros.

5. Una empresa de marketing e Internet ha prestado sus servicios a una gran multinacional con el fin de conseguir un mejor posicionamiento de sus sitios web y aumentar el número de visitas. Entre los aspectos tratados se analizó la proporción de visitas nuevas de dos de sus sitios web, uno de ellos dedicado a la venta on-line de tecnología y otro dedicado a la venta on-line de ropa de deporte. Se tomó una muestra de 25000 visitas del sitio web sobre tecnología, obteniendo que el 57.5 por ciento eran visitas nuevas mientras que de la muestra de 45000 visitas del sitio web de venta de ropa de deporte se obtuvo que 25432 eran visitas nuevas. Mediante el contraste de hipótesis adecuado, explica si se podría dar por cierta la afirmación de que la proporción de visitas nuevas de dichos sitios web son iguales. En caso contrario calculando los P-valores correspondientes explica cuál de los dos sitios web ha tenido más visitas nuevas.





**TEMA 8:**

**Contraste de homogeneidad:** Se plantea con una variable X que define estratos (o subpoblaciones) en la población, habiendo fijado el número de casos por estrato. Para cada estrato se observa otra variable Y, y realizamos el siguiente contraste:

$H_0$: La variable Y se distribuye igual (homogéneamente) en todos los estratos (o subpoblaciones) generados por X,

$H_1$: no $H_0$.

**Contraste de independencia (o contingencia):** Se plantea cuando los individuos son clasificados respecto a dos variables X e Y. El objetivo es estudiar si ambas variables son o no independientes, por lo que planteamos el siguiente contraste de hipótesis:

$H_0$: La variables X e Y son independientes,

$H_1$: no $H_0$ (es decir, existe relación entre X e Y).

**Estos contrastes de hipótesis se basan en la siguiente regla:** Rechazar $H_0$ si

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(\theta_{ij} - E_{ij})^2}{E_{ij}} \geq \chi^2_{\alpha,\nu},$$

Otra forma de estudiar si se rechaza la hipótesis nula es usando el P-valor. Así, se rechaza la hipótesis nula para todo α tal que

$$\alpha > P(\chi_\nu^2 \geq \chi_0^2)$$

donde $\nu=(r-1)\cdot(c-1)$. El valor r es el número de filas de la tabla y c el número de columnas. El valor $\theta_{ij}$ es la frecuencia absoluta observada en la celda (i, j) y $E_{ij}$ es la frecuencia esperada bajo la hipótesis nula en la celda (i, j) y se obtiene con la siguiente fórmula: $E_{ij}=R_i\cdot C_j/T$, donde $R_i$ es el ------------- de la fila ---------, $C_j$ es el total de la --------------------- j y T el total de la muestra. Por último, α es el nivel de significación y $\chi^2_{\alpha,\upsilon}$ es el percentil 1-α de una Chi-cuadrado con υ grados de libertad.

Recordemos que, en caso de no darnos el nivel de significación, si el P-valor es pequeño rechazaremos la hipótesis nula.

**Ejemplo:** Una gran compañía con dos cadenas de televisión ha realizado un estudio para analizar si existe alguna relación entre la cadena de televisión de la compañía preferida y la edad del telespectador. La siguiente tabla muestra la clasificación de datos obtenidos para la muestra utilizada en la investigación. Plantea el contraste de hipótesis y ve completando la tabla paso a paso:

$H_0$: la edad del espectador y la cadena preferida son -----------------------.

$H_1$: no $H_0$ (es decir, ----------------------------------------------------------)

| | | | Cadena | | |
|---|---|---|---|---|---|
| | | | Cadena 1 | Cadena 2 | Total filas |
| Grupos de edad | Niño | Frecuencia observada | 300 | 180 | |
| | | Frecuencia esperada | | | |
| | | Residuos tipificados | | | |
| | | Símbolo | | | |
| | Joven | Recuento | 210 | 240 | |
| | | Frecuencia esperada | | | |
| | | Residuos tipificados | | | |
| | | Símbolo | | | |
| | Adulto | Recuento | 114 | 200 | |
| | | Frecuencia esperada | | | |
| | | Residuos tipificados | | | |
| | | Símbolo | | | |
| Total columnas | | Recuento | | | |

Un ejemplo del significado de frecuencias observadas:

$\theta_{32}$=**----------**, esto significa que en nuestra muestra hay **-----------------** personas que son --------------------- y les gusta la cadena ----------.

Y ahora de frecuencias esperadas:

---------=$R_2 \cdot C_1/T$= . Notemos que $E_{21}$ es el valor que se esperaba obtener, si la edad y la cadena preferida fueran ------------, en la casilla correspondiente a ---------------- y cadena ------ preferida.





Una vez obtenidas las frecuencias observadas y las esperadas bajo la hipótesis nula, calcula:

$\chi_0^2 =$

Observa que coincide con el valor obtenido con el SPSS:

| | Valor | gl | Sig. asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | **55,542** | **2** | **,000** |
| Razón de verosimilitudes | 56,167 | 2 | ,000 |
| Asociación lineal por lineal | 54,652 | 1 | ,000 |
| N de casos válidos | 1244 | | |

A la vista de los datos de la tabla se tiene por tanto:

$\chi^2_0=$

y P-valor=P(X≥ )= , donde X es una -------------------------- con ----------------------grados de ----------------------.

Por tanto como el P-valor es ------------------------------- (en este caso concretamente vale ---) rechazamos $H_0$ lo que implica que ------------------ relación entre el grupo de edad y la cadena de la compañía preferida.

A continuación vamos a estudiar las casillas de interés e interpretar los resultados. Recordemos: En caso de rechazar la hipótesis nula hay que estudiar qué casillas han sido las que han contribuido en mayor medida a esto.

Para ello calculamos los residuos tipificados $Z_{ij}$:

$$Z_{ij} = \frac{(\theta_{ij} - E_{ij})}{\sqrt{E_{ij}}},$$

y asignamos símbolos a las casillas de la siguiente forma:

- Si $|Z_{ij}| \leq 1.645$ asignamos el símbolo $\cdot$.
- Si $1.645 < |Z_{ij}| \leq 1.96$ asignamos el símbolo o.
- Si $1.96 < |Z_{ij}| \leq 2.576$ asignamos el símbolo O.
- Si $|Z_{ij}| > 2.576$ asignamos el símbolo @.

Completa la tabla anterior con los residuos tipificados y los símbolos. El SPSS obtiene la tabla que hemos ido completando pero hay que añadir manualmente los símbolos. Para analizar las casillas, tenemos que tener en cuenta que las diferencias interesantes vienen marcadas por @ seguidas de las marcadas por O. En caso de no existir ningún símbolo del tipo @, se analizarán las del tipo O. Además habrá que analizar el signo de los $Z_{ij}$.

A la vista de la tabla obtenida analiza las casillas de interés:

Hay --------------------personas que son niños y les gusta la cadena 1 de lo que cabría esperar si las variables edad y cadena preferida fueran independientes.

Hay ----------------- personas que son niños y les gusta la cadena 2 de lo que cabría esperar si las variables edad y cadena preferida fueran independientes.

Hay --------------------personas que son adultos y les gusta la cadena 1 de lo que cabría esperar si las variables edad y cadena preferida fueran independientes.

Hay ----------------- personas que son adultos y les gusta la cadena 2 de lo que cabría esperar si las variables edad y cadena preferida fueran independientes.

Resuelve:

Tamaño de la muestra=----------. Número de personas a las que les gusta la cadena 1=---------. Número de personas a las que les gusta la cadena 2=-----.

P(guste la cadena 1)=





P(guste la cadena 2)=

Ahora bien, si nos pidieran la probabilidad de que guste la cadena 1 siendo joven, tendríamos que calcular: P( | )=

Pero si lo que nos piden es la probabilidad de ser joven sabiendo que le gusta la cadena 1, tendríamos que calcular: P( | )=

**El test de bondad de ajuste Ji-cuadrado:** El test de bondad de ajuste determina si una muestra procede de una distribución específica. Es decir, se plantea el siguiente contraste de hipótesis:

$H_0$: La muestra procede de la población especificada

$H_1$: no $H_0$ (La muestra no procede de la población especificada)

El test de bondad de ajuste se basa en la regla: Rechazar $H_0$ si

$\chi^2_0 \geq \chi^2_{\alpha, k-1}$ donde

$$\chi^2_0 = \sum_{i=1}^{k} \frac{(\theta_i - E_i)^2}{E_i}$$

donde $\theta_i$ es la frecuencia observada en cada categoría, $E_i$ la frecuencia esperada de cada categoría bajo la hipótesis nula y k el número de categorías. Otra forma de estudiar si se rechaza el test es con el P-valor. Así se rechaza la hipótesis nula para todo $\alpha$ tal que:

$$\alpha > P(\chi^2_{k-1} \geq \chi^2_0).$$

Al igual que se ha hecho en los contrastes de hipótesis de homogeneidad y contingencia, cuando se rechaza la hipótesis nula en un test de bondad de ajuste Ji-cuadrado hay que estudiar qué casillas han contribuido en mayor medida a esto. La forma de hacerlo es análoga a la vista anteriormente. Obtenemos los residuos tipificados:

$$Z_i = \frac{(\theta_i - E_i)}{\sqrt{E_i}},$$

y asignamos símbolos a las casillas de la siguiente forma:

- Si $|Z_i| \leq 1.645$ asignamos el símbolo $\cdot$.
- Si $1.645 < |Z_i| \leq 1.96$ asignamos el símbolo o.
- Si $1.96 < |Z_i| \leq 2.576$ asignamos el símbolo O.
- Si $|Z_i| > 2.576$ asignamos el símbolo @.

**Ejemplo:** Una empresa de ordenadores está a punto de introducir en el mercado, un nuevo modelo de ordenador dirigido al público infantil. Para ello ha preparado 4 anuncios publicitarios que se emitirán por televisión en distintas franjas horarias. Para decidir la franja horaria en la que se emitirá cada anuncio piden a un grupo de niños, que vean los anuncios e indiquen su favorito. Los resultados son:

| Anuncio | A | B | C | D |
|---|---|---|---|---|
| Frecuencia | 85 | 71 | 45 | 33 |

Contrasta si se prefiere un anuncio u otro, y en su caso estudia las casillas de interés.

Definimos:

$p_A$=proporción de niños que les gusta el anuncio A

$p_B$=proporción de niños que les gusta el anuncio B

$p_C$=proporción de niños que les gusta el anuncio C

$p_D$=proporción de niños que les gusta el anuncio D

Entonces el contraste de hipótesis que se debe plantear es:
$H_0$: $p_A=p_B=p_C=p_D$=------

$H_1$: no $H_0$

Introducimos los datos en el SPSS como se ha explicado en clase y obtenemos la siguiente tabla:





**Estadísticos de contraste**

| | VAR00006 |
|---|---|
| Chi-cuadrado | 28,906[a] |
| gl | 3 |
| Sig. asintót. | ,000 |

¿Qué calculos se han hecho para obtener $\chi_0^2$?

$\chi^2_0=$

P-valor=P( ≥28,906)=------, donde X es una una Chi-cuadrado con -------------grados de ----------------------------.
Por tanto como el ---------------------------- es -------------------- (en este caso igual a 0) rechazamos Ho, lo que implica que los anuncios ----------- han -------------------- por igual.

Vamos a estudiar las casillas de interés, para ello hay que obtener los residuos -----------------, podemos ayudarnos también de una hoja de cálculo ya que el SPSS ---------------- obtiene sus valores.

| | $\theta_i$ | $E_i$ | $\theta_i$-$E_i$ | $Z_i$ | Símbolo |
|---|---|---|---|---|---|
| **A** | 85 | | | | |
| **B** | 71 | | | | |
| **C** | 45 | | | | |
| **D** | 33 | | | | |
| Total | | | | | |

$E_i=$ , para todo i=A, B, C, D.

$Z_{A=}$ $Z_{B=}$ $Z_{C=}$ $Z_{D=}$

A partir de los datos obtenidos, concluimos que hay ------------ niños que les gusta el anuncio A y -------- niños que les gusta el anuncio ------------- de lo que cabría esperar si todos los anuncios gustaran por ------------.
Esto nos lleva a pensar que el anuncio A debería emitirse en el horario con mayor público infantil y dejar el anuncio D para la franja horaria con menos público infantil.