Nombre y Apellidos:

CUESTIONARIO

1. La media aritmética de una característica de una población ...
   a) nunca puede ser negativa.
   b) siempre es mayor que la mediana.
   c) es muy sensible a valores atípicos y extremos.
   d) siempre es mejor medida de centralización que la mediana.

2. Cuando sumamos la media y la varianza, obtenemos:
   a) El tercer cuartil.
   b) Un sinsentido, porque estamos sumando cantidades con unidades distintas.
   c) Una cota superior para el 75 % de los datos.
   d) Un número que se compara con el rango para determinar la asimetría de los datos.

3. La recta de regresión entre dos variables es $\hat{y} = 0{,}5 - 1{,}97x$. Esto quiere decir que:
   a) La covarianza entre las variables es negativa.
   b) Hemos cometido un error en los cálculos.
   c) Existe una relación de tipo lineal y directa entre las variables.
   d) Ninguna de las respuestas anteriores es cierta.

4. Si queremos estudiar el grado de asociación entre dos variables cualitativas, debemos utilizar:
   a) El coeficiente de contingencia.
   b) El coeficiente de correlación de Spearman.
   c) El coeficiente de correlación de Pearson.
   d) La covarianza entre las variables.

5. Dados los sucesos A y B pertenecientes al mismo espacio de sucesos con probabilidad no nula e incompatibles. En este caso podemos afirmar que $P(B/A)$ es igual a:
   a) La $P(A)$
   b) La $P(B)$
   c) 0
   d) 1

6. Dados los sucesos A y B pertenecientes al mismo espacio de sucesos tales que $P(A) = 0{,}4$, $P(B) = 0{,}5$ y $P(\overline{A} \cup \overline{B}) = 0{,}7$. La $P(A \cup B)$ es:
   a) 0,4
   b) $\boxed{0{,}6}$
   c) 0,9
   d) 0,3
7. Sea X una variable aleatoria discreta que toma los valores $\{1, 2, 3, 4, 5\}$ con probabilidades respectivas $\{1/10, 3/10, 3/10, 2/10, 1/10\}$. El valor esperado de X es igual a:
   a) 1,0
   b) 3,3
   c) 3,0
   d) $\boxed{2{,}9}$
8. Sea X variable aleatoria continua con función de densidad $f(x)$ definida en el intervalo $[0, 5]$ ¿Cuál de las siguientes afirmaciones no es necesariamente verdadera?
   a) $f(6) = 0$
   b) La integral de $f(x)$ en el intervalo $[0, 5]$ vale 1.
   c) $\boxed{f(5) = 1}$
   d) $f(x)$ es mayor o igual que cero en el intervalo $[0, 5]$.
9. Sea X una variable aleatoria que sigue una distribución Chi-cuadrado con 8 grados de libertad. El valor a que verifica que $P[X < a] = 0{,}99$ vale:
   a) 26,12
   b) $\boxed{20{,}09}$
   c) 21,95
   d) 18,48
10. La probabilidad de error α de un contraste de hipótesis, es la probabilidad de:
   a) $\boxed{\text{Rechazar } H_0 \text{ cuando es verdadera } H_0.}$
   b) Aceptar H_0 cuando es verdadera H_1.
   c) Equivocarse cuando se rechaza H_1.
   d) Equivocarse cuando se acepta H_0.

CUADRO DE RESPUESTAS

PREGUNTA	1	2	3	4	5	6	7	8	9	10
(a)										
(b)										
(c)										
(d)										

Nota: Cada respuesta correcta suma 0,20 y cada respuesta errónea resta 0,067.

Nombre y Apellidos:

PROBLEMAS

1. El departamento de control de calidad de una empresa que fabrica pantallas LED, analiza el número de píxeles que presentan imperfecciones en cada pantalla. El número de píxeles erróneos encontrados en una muestra de 50 pantallas se muestran en la siguiente tabla:

Nº Píxeles Erróneos	n_i	N_i
0		20
1	10	
2		42
3	6	
4		

a) **(0.5 ptos.) Completa la tabla de frecuencias.**

b) **(0.5 ptos.) Calcula el número medio de píxeles erróneos. ¿Qué número de errores es el más frecuente?**

c) **(0.5 ptos.) Calcula la mediana y el percentil 70 de la distribución.**

SOLUCIÓN:

a) En este apartado completamos la tabla de frecuencias absolutas sabiendo que la suma total de estas frecuencias debe ser 50:

Nº Píxeles Erróneos	n_i	N_i	$x_i \cdot n_i$
0	20	20	0
1	10	30	10
2	12	42	24
3	6	48	18
4	2	50	8

b) Nos piden en este apartado calcular la media y la moda de la distribución de frecuencias. Para determinar la media, añadimos a la tabla del apartado anterior una columna donde calculamos $x_i \cdot n_i$. Para determinar la moda basta elegir el dato que tiene mayor frecuencia:

$$\overline{X} = \frac{\sum x_i \cdot n_i}{n} = \frac{60}{50} = \boxed{1.2 \ \text{ errores}} \qquad\qquad Mo = \boxed{0 \ \text{ errores}}$$

c) Debemos calcular la mediana o percentil 50 y el percentil 70. En ambos casos el procedimiento es el mismo: 1º) determinamos la posición que ocupa el percentil k mediante la expresión $pos = \frac{k \cdot n}{100}$; 2º) elegimos el primer dato cuya frecuencia acumulada es mayor o igual a la posición que se ha obtenido en el paso anterior.

$$P_{50} \qquad \Rightarrow pos = \frac{50 \cdot 60}{100} = 30 \qquad \Rightarrow P_{50} = \boxed{1 \ \text{ errores}}$$

$$P_{70} \qquad \Rightarrow pos = \frac{70 \cdot 60}{100} = 42 \qquad \Rightarrow P_{70} = \boxed{2 \ \text{ errores}}$$

2. **Una empresa electrónica observa que el número de componentes que fallan antes de cumplir 100 horas de funcionamiento es una variable de Poisson de media 8. Calcular**

   a) **(0.5 ptos.) ¿Cuál es la probabilidad de que falle una componente en 25 horas?**

   b) **(0.5 ptos.) ¿Cuál es la probabilidad de como máximo fallen 2 componentes en 50 horas?**

   c) **(0.5 ptos.) ¿cuál es la probabilidad de que fallen por lo menos 20 componentes en 200 horas?**

   SOLUCIÓN: Para este problema vamos a definir la variable aleatoria que denotamos por X_{100}, y que contabiliza el número de componentes que fallan en las primeras 100 horas de funcionamiento. Esta variable se ajusta a una distribución Poisson de parámetro 8. Tomando esta variable como referencia pasamos a resolver los distintos apartados:

   a) En este apartado utilizamos la variable X_{25}, que contabiliza el número de componentes que fallan en las primeras 25 horas de funcionamiento. Esta variable se ajusta a una distribución Poisson de media 2:

   $$P[X_{25} = 1] = \frac{e^{-2} \cdot (2)^1}{1!} = \boxed{0.2707}$$

   b) En este apartado utilizamos la variable X_{50}, que contabiliza el número de componentes que fallan en las primeras 50 horas de funcionamiento. Esta variable se ajusta a una distribución Poisson de media 4:

   $$P[X_{50} \leq 2] = \frac{e^{-4} \cdot (4)^0}{0!} + \frac{e^{-4} \cdot (4)^1}{1!} + \frac{e^{-4} \cdot (4)^2}{2!} = \boxed{0.2381}$$

   c) En este apartado utilizamos la variable X_{200}, que contabiliza el número de componentes que fallan en las primeras 200 horas de funcionamiento. Esta variable se ajusta a una distribución Poisson de media 16 y podemos aproximarla por una $N(16; 4)$:

   $$P[X_{200} \geq 20] \overset{(c.c.)}{=} P[X_{200} \geq 19{,}5] = P[Z \geq 0{,}875] = 1 - F_Z(0{,}875) = \boxed{0.1894}$$

3. **Dos empresas A y B que prestan servicios de internet aseguran mediante publicidad que tienen una velocidad de conexión similar. Seleccionamos aleatoriamente una muestra de usuarios de cada una de estas empresas y obtenemos los siguientes datos relativos a la velocidad de conexión (mbps):**

Empresa A	**21**	**25**	**26**	**25**	**19**	**22**	**22**	**28**	**29**	**18**	**16**	**21**	**30**
Empresa B	**26**	**15**	**18**	**12**	**24**	**16**	**18**	**13**	**15**	**21**			

   Suponiendo normalidad en los datos, ¿podemos afirmar que las empresas A y B ofrecen en promedio la misma velocidad de conexión?

   SOLUCIÓN: En primer lugar y haciendo uso de la calculadora, obtenemos los estadísticos muestrales que necesitamos para realizar este problema:

   $$\begin{array}{lll} n_1 = 13 & \overline{x}_1 = 23{,}231 & S_{c_1} = 4{,}343 \\ n_2 = 10 & \overline{x}_2 = 17{,}8 & S_{c_2} = 4{,}614 \end{array}$$

   Para estudiar si existen diferencias significativas entre la velocidad de conexión de ambas empresas, suponiendo normalidad en los datos, debemos contrastar primero la igualdad de varianzas.

Esto es debido a que las varianzas muestrales son desconocidas y los tamaños muestrales son pequeños ($n_1 + n_2 \leq 30$):

$$\left.\begin{array}{l} H_0 \equiv \sigma_1^2 = \sigma_2^2 \\ H_1 \equiv \sigma_1^2 \neq \sigma_2^2 \end{array}\right\} \qquad f_{exp} = \frac{S_{c_1}^2}{S_{c_2}^2} = \frac{4{,}343^2}{4{,}614^2} = 0{,}886$$

La región crítica de este contraste para un valor $\alpha = 0{,}05$ tiene la expresión:

$$\text{R.C.} = \{f_{exp} < F_{0{,}025}(12,9)\} \cup \{f_{exp} > F_{0{,}975}(12,9)\}$$

siendo el percentil $F_{0{,}975}(12,9) = 3{,}868$, según la tabla, y utilizando que

$$F_{\frac{\alpha}{2}}(n_1-1, n_2-1) = \frac{1}{F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)} \quad \Rightarrow \quad F_{0{,}025}(12,9) = \frac{1}{F_{0{,}975}(9,12)} = \frac{1}{3{,}436} = 0{,}291$$

se tiene que

$$\text{R.C.} = \{f_{exp} < 0{,}291\} \cup \{f_{exp} > 3{,}868\} \Rightarrow H_0$$

luego no podemos rechazar la hipótesis de igualdad de las varianzas poblacionales.

Con este resultado, utilizaremos ahora el contraste de comparación de medias suponiendo varianzas iguales:

$$\left.\begin{array}{l} H_0 \equiv \mu_1 = \mu_2 \\ H_1 \equiv \mu_1 \neq \mu_2 \end{array}\right\} \qquad t_{exp} = \frac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{23{,}231 - 17{,}8}{\sqrt{19{,}902}\sqrt{\frac{1}{13} + \frac{1}{10}}} = 2{,}894$$

donde hemos calculado

$$S_p^2 = \frac{(n_1-1)S_{c_1}^2 + (n_2-1)S_{c_2}^2}{n_1+n_2-2} = 19{,}902$$

En este caso el estadístico t_{exp} se ajusta a una distribución $t(n_1+n_2-2)$. La región crítica del contraste al 5 % de significación es

$$\text{R.C.} = \{|t_{exp}| > t_{0{,}975}(21)\} = \{|t_{exp}| > 2{,}080\} \quad \Rightarrow \quad H_1,$$

Por tanto podemos rechazar H_0 para $\alpha = 0{,}05$, es decir, podemos afirmar que la diferencia de velocidad de conexión entre ambas empresas es significativa para $\alpha = 0{,}05$.

4. **Es conocido que la temperatura de la placa base de un ordenador está condicionada por múltiples factores, como son las aplicaciones que utilizamos, las condiciones ambientales, las componentes, etc. Con objeto de estudiar esta variable en un ordenador se ha tomado una muestra de 60 mediciones en diferentes condiciones de trabajo, obteniéndose una media de 42^oC y una cuasi-desviación típica de 8^oC. Suponiendo normalidad en los datos, ¿Con qué nivel de confianza puede admitirse que la temperatura media en la placa de este ordenador se encuentra en el intervalo $(42 \pm 2{,}5)$?**

SOLUCIÓN: En este problemas nos dan un intervalo de confianza para la media de una población normal y tenemos que determinar su nivel de confianza $(1-\alpha)$. Puesto que el tamaño muestral $n > 30$, aproximamos la distribución t-student por la Normal estándar. La expresión para el intervalo de confianza sería:

$$IdC_{1-\alpha}(\mu) = \left[\overline{x} \pm Z_{1-\alpha/2}\frac{S_c}{\sqrt{n}}\right] = \left[42 \pm Z_{1-\alpha/2}\frac{8}{\sqrt{60}}\right] = [42 \pm 2{,}5]$$

Identificando términos nos queda que:

$$Z_{1-\alpha/2}\frac{8}{\sqrt{60}} = 2{,}5 \qquad \Rightarrow \qquad Z_{1-\alpha/2} = 2{,}42$$

Buscamos en la tabla de la Z el valor 2,42 y comprobamos que $(1-\alpha/2) = 0{,}9922$. Por tanto el nivel de confianza debe ser $1-\alpha = 0{,}9844$ o del $98{,}44\,\%$.