**NOMBRE Y APELLIDOS:**

1. Marca la respuesta correcta: (1/2 cada una; las respuestas incorrectas restan 1/6)

a) La función de distribución de una variable aleatoria es siempre …

- [ ] Menor que 1
- [ ] Mayor ó igual que 1
- [ ] 1
- [x] Menor ó igual que 1

b) Para una muestra $\{x_1, x_2, \cdots, x_n\}$, la expresión $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ se llama …

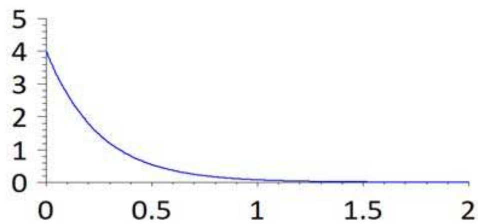
- [ ] Media
- [x] Varianza muestral
- [ ] Mediana
- [ ] Desviación típica muestral

c) La distribución F de Fisher se utiliza en la construcción de intervalos de confianza para …

- [ ] La comparación de medias de dos poblaciones normales independientes
- [x] La comparación de varianzas de dos poblaciones normales independientes
- [ ] La media de una población normal
- [ ] La varianza de una población normal

d) El gráfico adjunto corresponde a la función de densidad …

- [ ] N(4,1)
- [ ] t de Student con 4 grados de libertad
- [x] Exponencial de parámetro 4
- [ ] $\chi^2$ con 4 grados de libertad




2. Un programa detecta mensajes SPAM con probabilidad 0,98. La probabilidad de que marque como SPAM un mensaje que no lo es es 0,04. Si el 1% de los mensajes recibidos son SPAM: (1)
   (a) Calcula la probabilidad de que un mensaje sea marcado como SPAM.
   (b) Un mensaje ha sido marcado como SPAM. Calcula la probabilidad de que lo sea.

$S = \text{"SPAM"}$ $\quad \mathrm{P}(S) = 0.01$
$\bar{S} = \text{"No SPAM"}$ $\quad \mathrm{P}(\bar{S}) = 0.99$

$M = \text{"Marcar como SPAM"}$ $\quad \mathrm{P}(M \mid S) = 0.98$
$\bar{M} = \text{"No marcar como SPAM"}$ $\quad \mathrm{P}(M \mid \bar{S}) = 0.04$

(a) $\mathrm{P}(M) = \mathrm{P}(M \cap S) + \mathrm{P}(M \cap \bar{S}) = \mathrm{P}(M \mid S)\cdot \mathrm{P}(S) + \mathrm{P}(M \mid \bar{S})\cdot \mathrm{P}(\bar{S}) =$

$= 0.98 \times 0.01 + 0.04 \times 0.99 = 0.0098 + 0.0396 = 0.0494$

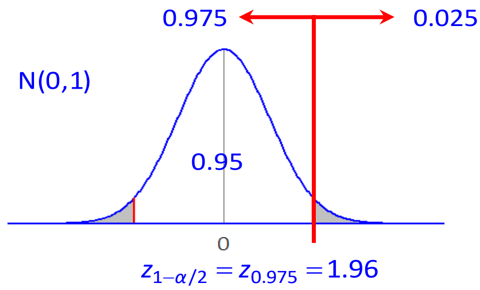
(b) $\mathrm{P}(S \mid M) = \frac{\mathrm{P}(S \cap M)}{\mathrm{P}(M)} = \frac{\mathrm{P}(M \cap S)}{\mathrm{P}(M)} = \frac{0.0098}{0.0494} = 0.198$

3. En una muestra de 400 personas de una población se halló que 80 leen habitualmente las noticias por internet. Construye e interpreta un intervalo de confianza al 95% para la proporción de personas que leen habitualmente las noticias por internet. (1.5)

Intervalo de confianza para $p$: $\hat{p} \pm \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \cdot z_{1-\alpha/2}$

Proporción muestral: $\hat{p} = \dfrac{80}{400} = 0.2$

$$1-\alpha = 0.95 \quad \Rightarrow \quad \alpha = 0.05 \quad \Rightarrow$$
$$\Rightarrow \quad \alpha/2 = 0.025 \quad \Rightarrow \quad 1-\alpha/2 = 0.975$$




$$0.2 \pm \sqrt{\frac{0.2 \times 0.8}{400}} \times 1.96$$
$$0.2 \pm 0.02 \times 1.96$$
$$0.2 \pm 0.04$$

$\boxed{(\ 0.16\ ,\ 0.24\ )}$

4. Se ha medido el tiempo que 10 trabajadores emplearon para el ensamblado de un grupo de piezas, obteniéndose una media muestral de 8,4 minutos y una desviación de 1,2 minutos. ¿Se puede afirmar que el tiempo medio de ensamblado supera los 8 minutos?
Responde mediante el contraste de hipótesis adecuado al nivel 0,01. (1.5)

$$\left.\begin{array}{l} H_0 : \mu = 8 \\ H_1 : \mu > 8 \end{array}\right\} \quad \alpha = 0.01$$

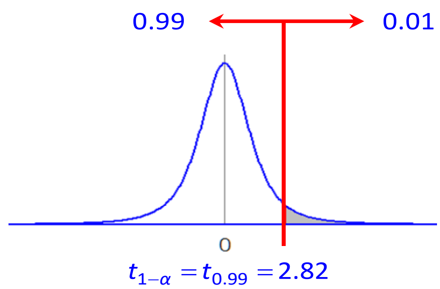
Estadístico de contraste: $T = \dfrac{\overline{X} - \mu_0}{\hat{s}/\sqrt{n}} = \dfrac{8.4-8}{1.2/\sqrt{10}} = 1.05$

Distribución bajo la hipótesis nula: t con 9 grados de libertad

Región crítica a la derecha de nivel 0.01:

$C = [2.82\ ,\ \infty\ )$

$T_{obs} = 1.05 \notin C$ , por tanto, no se rechaza la hipótesis nula.




$\boxed{\text{A la vista de los datos observados, no se puede afirmar que el tiempo medio de ensamblado supere los 8 minutos.}}$

**NOMBRE Y APELLIDOS:**

5. Se han medido las presiones sistólica y diastólica de 12 personas:
   (a) Plantea y construye el modelo de regresión lineal para PRESIÓN DIASTOLICA sobre PRESIÓN SISTÓLICA.
   (b) Construye la tabla de análisis de la varianza y razona si hay relación lineal significativa al 5%. Calcula el coeficiente de correlación lineal.

| | SIST. | DIAST. |
|---|---|---|
| | 120 | 84 |
| | 134 | 97 |
| | 131 | 87 |
| | 120 | 78 |
| | 120 | 77 |
| | 129 | 92 |
| | 145 | 101 |
| | 135 | 92 |
| | 136 | 98 |
| | 132 | 95 |
| | 124 | 84 |
| | 124 | 86 |
| Media | 129.17 | 89.25 |
| Varianza | 55.97 | 55.85 |
| Covarianza | 51.54 | |
| Varianza residual: | | 10.07 |

Pendiente: $\frac{S_{XY}}{S_X^2}$

Ordenada en el origen: $\overline{Y} - \frac{S_{XY}}{S_X^2}\overline{X}$

(2)

(a) Pendiente: $\frac{S_{XY}}{S_X^2} = \frac{51.54}{55.97} =$ **0.9209**

Ordenada: $\overline{Y} - \frac{S_{XY}}{S_X^2}\overline{X} = 89.25 - \frac{51.54}{55.97}129.2 =$ **-29.7**

(b)

| **Fuente de variación** | **SC** | **g.l.** | **SCM** | **F** | **Percentil** | |
|---|---|---|---|---|---|---|
| Modelo | 569.5 | 1 | 569.5 | 56.554 | 4.96 | **R H0** |
| Residual | 100.7 | 10 | 10.07 | | | |
| Total | 670.2 | 11 | | | | |

Coeficiente de correlación lineal: **0.9218**
Se puede calcular como el cociente entre la covarianza y el producto de las desviaciones ó como la raíz cuadradaa del coeficiente de determinación, que es el cociente entre la suma de cuadrados debida al modelo y la total.

Construcción de la tabla:

1. La suma de cuadrados media residual es la varianza residual.
   Sus grados de libertad son n - 2 = 12 - 2 = 10.
   De ahí, la suma de cuadrados residual será 10 * 10.07.
2. La suma de cuadrados total es la varianza de Y multiplicada por el tamaño de la muestra, es decir, 55.85 * 12.
3. La suma de cuadrados debida al modelo es la total menos la residual.

Y ya es fácil completar el resto de la tabla.

(Sigue a la vuelta)

6. Se ha observado el número de clientes que entran en cinco sucursales de un mismo banco situadas en diferentes barrios a lo largo de tres franjas horarias (entre las 8 y las 10, entre las 10 y las 12 y entre las 12 y las 14). Con los datos adjuntos:

(a) Construye la tabla de análisis de la varianza.

(b) ¿Se observa diferencia significativa de la situación de la sucursal sobre el número de clientes? ¿Y de la franja horaria? (Razona las respuestas a partir de la tabla de análisis de la varianza).

(c) Calcula e interpreta los coeficientes de determinación.

| **Fuente de variación** | **Suma de cuadrados media** |
|---|---|
| Entre sucursales | 64.433 |
| Entre franjas horarias | 19.135 |
| Residual | 7.866 |

(2)

(a)

| **Fuente de variación** | **SC** | **g.l.** | **SCM** | **F** | **Percentil (0.95)** | |
|---|---|---|---|---|---|---|
| Sucursales | 257.72 | 4 | 64.43 | 8.1909 | 3.83 | **R H0** |
| Franjas | 38.28 | 2 | 19.14 | 2.4333 | 4.46 | **No R H0** |
| Residual | 62.928 | 8 | 7.866 | | | |
| Total | 358.93 | 14 | | | | |

(b) Efecto de la sucursal sobre el número de clientes:

H0: $\mu_i = \mu$ para todo $i = 1, \ldots, 5$.

H1: Algún $\mu_i$ es distinto de los demás.

El contraste se resuelve mediante la F de sucursales. Su distribución bajo la hipótesis nula es F con 4 y 8 grados de libertad. La región crítica de nivel 0.05 empieza en 3.83. El valor observado de F, 8.191, está dentro de la región crítica y por tanto rechazamos la hipótesis nula. Concluimos que hay efecto significativo al nivel 0.05 de la sucursal sobre el número de clientes.

Efecto de la franja horaria sobre el número de clientes:

H0: $\mu_i = \mu$ para todo $i = 1, \ldots, 3$.

H1: Algún $\mu_i$ es distinto de los demás.

El contraste se resuelve mediante la F de franjas. Su distribución bajo la hipótesis nula es F con 2 y 8 grados de libertad. La región crítica de nivel 0.05 empieza en 4.46. El valor observado de F, 2.433, no está dentro de la región crítica y por tanto no rechazamos la hipótesis nula. Concluimos que no hay efecto significativo al nivel 0.05 de la franja horaria sobre el número de clientes.

(c) Sucursales: **0.718**

Franjas: **0.1067**

El 71.8% de la variabilidad en el número de clientes se debe al efecto de la sucursal.

El 10.7% de la variabilidad en el número de clientes se debe al efecto de la franja horaria.

**NOMBRE Y APELLIDOS:**

1. Marca la respuesta correcta: (1/2 cada una; las respuestas incorrectas restan 1/6)

a) La función de distribución de una variable aleatoria es siempre …

☐ Menor que 1 ☒ Menor ó igual que 1
☐ 1 ☐ Mayor ó igual que 1

b) Para una muestra $\{x_1, x_2, \cdots, x_n\}$, la expresión $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ se llama …

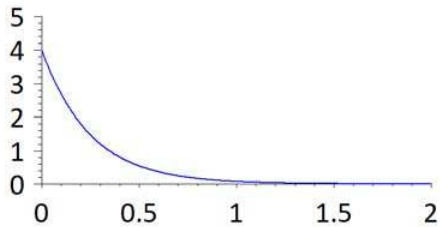
☐ Media ☐ Mediana
☒ Varianza muestral ☐ Desviación típica muestral

c) La distribución F de Fisher se utiliza en la construcción de intervalos de confianza para …

☒ La comparación de varianzas de dos poblaciones normales independientes
☐ La comparación de medias de dos poblaciones normales independientes
☐ La media de una población normal
☐ La varianza de una población normal

d) El gráfico adjunto corresponde a la función de densidad …

☐ N(4,1)
☒ Exponencial de parámetro 4
☐ t de Student con 4 grados de libertad
☐ $\chi^2$ con 4 grados de libertad




2. Un programa detecta mensajes SPAM con probabilidad 0,98. La probabilidad de que marque como SPAM un mensaje que no lo es es 0,04. Si el 1% de los mensajes recibidos son SPAM: (1)

(a) Calcula la probabilidad de que un mensaje sea marcado como SPAM.
(b) Un mensaje ha sido marcado como SPAM. Calcula la probabilidad de que no lo sea.

$S = \text{"SPAM"}$ $\quad \mathrm{P}(S) = 0.01$
$\bar{S} = \text{"No SPAM"}$ $\quad \mathrm{P}(\bar{S}) = 0.99$

$M = \text{"Marcar como SPAM"}$ $\quad \mathrm{P}(M \mid S) = 0.98$
$\bar{M} = \text{"No marcar como SPAM"}$ $\quad \mathrm{P}(M \mid \bar{S}) = 0.04$

(a) $\mathrm{P}(M) = \mathrm{P}(M \cap S) + \mathrm{P}(M \cap \bar{S}) = \mathrm{P}(M \mid S)\cdot \mathrm{P}(S) + \mathrm{P}(M \mid \bar{S}) \cdot \mathrm{P}(\bar{S}) =$

$= 0.98 \times 0.01 + 0.04 \times 0.99 = 0.0098 + 0.0396 = 0.0494$

(b) $\mathrm{P}(\bar{S} \mid M) = \frac{\mathrm{P}(\bar{S} \cap M)}{\mathrm{P}(M)} = \frac{\mathrm{P}(M \cap \bar{S})}{\mathrm{P}(M)} = \frac{0.0396}{0.0494} = 0.802$

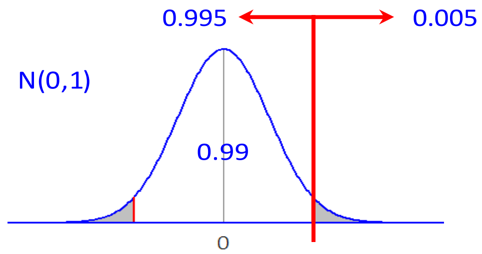
(Sigue a la vuelta)

3. En una muestra de 300 personas de una población se halló que 120 leen habitualmente las noticias por internet. Construye e interpreta un intervalo de confianza al 99% para la proporción de personas que leen habitualmente las noticias por internet. (1.5)

Intervalo de confianza para $p$ : $\hat{p} \pm \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \cdot z_{1-\alpha/2}$

Proporción muestral: $\hat{p} = \dfrac{120}{300} = 0.4$

$$1-\alpha = 0.99 \;\Rightarrow\; \alpha = 0.01 \;\Rightarrow$$
$$\Rightarrow\; \alpha/2 = 0.005 \;\Rightarrow\; 1-\alpha/2 = 0.995$$




$z_{1-\alpha/2} = z_{0.995} = 2.57$

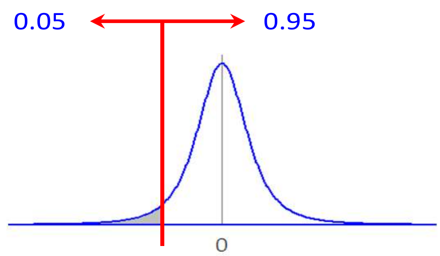
$$0.4 \pm \sqrt{\frac{0.4 \times 0.6}{300}} \times 2.57$$
$$0.4 \pm 0.03 \times 2.57$$
$$0.4 \pm 0.08$$

$$\boxed{(\,0.32\,,\,0.48\,)}$$

4. Se ha medido el tiempo que 16 trabajadores emplearon para el ensamblado de un grupo de piezas, obteniéndose una media muestral de 7,6 minutos y una desviación de 0,8 minutos. ¿Se puede afirmar que el tiempo medio de ensamblado no llega a 8 minutos?
Responde mediante el contraste de hipótesis adecuado al nivel 0,05. (1.5)

$$\left.\begin{array}{l} H_0 : \mu = 8 \\ H_1 : \mu < 8 \end{array}\right\} \quad \alpha = 0.05$$

Estadístico de contraste: $T = \dfrac{\overline{X} - \mu_0}{\hat{s}/\sqrt{n}} = \dfrac{7.6-8}{0.8/\sqrt{16}} = -2$




$t_{\alpha} = t_{0.05} = -1.75$

Distribución bajo la hipótesis nula: t con 15 grados de libertad

Región crítica a la izquierda de nivel 0.05:

$C = (-\infty, -1.75\,]$

$T_{obs} = -2 \in C$ , por tanto, se rechaza la hipótesis nula.

| |
|---|
| A la vista de los datos observados, se puede afirmar que el tiempo medio de ensamblado no llega a 8 minutos. |

**NOMBRE Y APELLIDOS:**

5. Se han medido las presiones sistólica y diastólica de 12 personas:
   (a) Plantea y construye el modelo de regresión lineal para PRESIÓN DIASTOLICA sobre PRESIÓN SISTÓLICA.
   (b) Construye la tabla de análisis de la varianza y razona si hay relación lineal significativa al 1%. Calcula el coeficiente de correlación lineal.

| | SIST. | DIAST. |
|---|---|---|
| | 120 | 84 |
| | 134 | 97 |
| | 131 | 87 |
| | 120 | 78 |
| | 120 | 77 |
| | 129 | 92 |
| | 145 | 101 |
| | 135 | 92 |
| | 136 | 98 |
| | 132 | 95 |
| | 124 | 84 |
| | 124 | 86 |
| Media | 129.17 | 89.25 |
| Varianza | 55.97 | 55.85 |
| Covarianza | 51.54 | |
| Varianza residual: | 10.07 | |

Pendiente: $\frac{S_{XY}}{S_X^2}$

Ordenada en el origen: $\overline{Y} - \frac{S_{XY}}{S_X^2}\overline{X}$

(2)

(a) Pendiente: $\frac{S_{XY}}{S_X^2} = \frac{51.54}{55.97} =$ **0.9209**

Ordenada: $\overline{Y} - \frac{S_{XY}}{S_X^2}\overline{X} = 89.25 - \frac{51.54}{55.97}129.2 =$ **-29.7**

(b)

| **Fuente de variación** | **SC** | **g.l.** | **SCM** | **F** | **Percentil** | |
|---|---|---|---|---|---|---|
| Modelo | 569.5 | 1 | 569.5 | 56.554 | 4.96 | **R H0** |
| Residual | 100.7 | 10 | 10.07 | | | |
| Total | 670.2 | 11 | | | | |

Coeficiente de correlación lineal: **0.9218**

Se puede calcular como el cociente entre la covarianza y el producto de las desviaciones ó como la raíz cuadradaa del coeficiente de determinación, que es el cociente entre la suma de cuadrados debida al modelo y la total.

Construcción de la tabla:

1. La suma de cuadrados media residual es la varianza residual.
   Sus grados de libertad son n - 2 = 12 - 2 = 10.
   De ahí, la suma de cuadrados residual será 10 * 10.07.
2. La suma de cuadrados total es la varianza de Y multiplicada por el tamaño de la muestra, es decir, 55.85 * 12.
3. La suma de cuadrados debida al modelo es la total menos la residual.

Y ya es fácil completar el resto de la tabla.

6. Se ha observado el número de clientes que entran en cinco sucursales de un mismo banco situadas en diferentes barrios a lo largo de tres franjas horarias (entre las 8 y las 10, entre las 10 y las 12 y entre las 12 y las 14). Con los datos adjuntos:
   (a) Construye la tabla de análisis de la varianza.
   (b) ¿Se observa diferencia significativa de la situación de la sucursal sobre el número de clientes? ¿Y de la franja horaria? (Razona las respuestas a partir de la tabla de análisis de la varianza).
   (c) Calcula e interpreta los coeficientes de determinación.

| **Fuente de variación** | **Suma de cuadrados media** |
|---|---|
| Entre sucursales | 64.433 |
| Entre franjas horarias | 19.135 |
| Residual | 7.866 |

(2)

(a)

| **Fuente de variación** | **SC** | **g.l.** | **SCM** | **F** | **Percentil (0.95)** | |
|---|---|---|---|---|---|---|
| Sucursales | 257.72 | 4 | 64.43 | 8.1909 | 3.83 | **R H0** |
| Franjas | 38.28 | 2 | 19.14 | 2.4333 | 4.46 | **No R H0** |
| Residual | 62.928 | 8 | 7.866 | | | |
| Total | 358.93 | 14 | | | | |

(b) Efecto de la sucursal sobre el número de clientes:

H0: $\mu_i = \mu$ para todo $i = 1, \ldots, 5$.

H1: Algún $\mu_i$ es distinto de los demás.

El contraste se resuelve mediante la F de sucursales. Su distribución bajo la hipótesis nula es F con 4 y 8 grados de libertad. La región crítica de nivel 0.05 empieza en 3.83. El valor observado de F, 8.191, está dentro de la región crítica y por tanto rechazamos la hipótesis nula. Concluimos que hay efecto significativo al nivel 0.05 de la sucursal sobre el número de clientes.

Efecto de la franja horaria sobre el número de clientes:

H0: $\mu_i = \mu$ para todo $i = 1, \ldots, 3$.

H1: Algún $\mu_i$ es distinto de los demás.

El contraste se resuelve mediante la F de franjas. Su distribución bajo la hipótesis nula es F con 2 y 8 grados de libertad. La región crítica de nivel 0.05 empieza en 4.46. El valor observado de F, 2.433, no está dentro de la región crítica y por tanto no rechazamos la hipótesis nula. Concluimos que no hay efecto significativo al nivel 0.05 de la franja horaria sobre el número de clientes.

(c) Sucursales: **0.718**

Franjas: **0.1067**

El 71.8% de la variabilidad en el número de clientes se debe al efecto de la sucursal.

El 10.7% de la variabilidad en el número de clientes se debe al efecto de la franja horaria.