

SEGUNDO PARCIAL DE ESTADÍSTICA

APELLIDO y NOMBRE ..... L.U .....



**IMPORTANTE: complete dentro de los espacios destinados a cada pregunta**

1. Un laboratorio de análisis clínicos realiza un estudio sobre “calidad de atención al cliente” y encuentra que el tiempo de espera desde que el paciente ingresa hasta que es atendido por el bioquímico sigue una distribución normal con media 10 minutos y desvío 2.5 minutos. Calcule con ayuda de la App:

a) El porcentaje de pacientes que deberá esperar más de 15 minutos para ser atendidos por el bioquímico.	$P(X > 15) = 0.0227$ ; el 2.27% deberá esperar más de 15 min en ser atendido por el bioquímico.
b) El tiempo máximo de espera del 10% de los pacientes que son atendidos más rápido.	$P_{10} = 6.79$ minutos
c) El cuartil 1 para el tiempo de espera.	$Q_1 = P_{25} = 8.31$ minutos

2. Se quiere estimar mediante un intervalo de confianza del 95%, qué proporción de enfermos de coronavirus tiene alteración del sentido del gusto. Para ello se tomó una muestra de 70 enfermos y 42 dijeron tener ese síntoma.

a) Calcule la estimación puntual de la proporción de enfermos que tiene alteración del sentido del gusto.	$P$ estimado = $42/70 = 0.6$				
b) Observe la salida del programa Epidat e interprete el intervalo en términos de la variable. <i>Intervalo de confianza (95,0%)</i> <table border="1" style="margin-left: 20px;"> <tr> <td><i>Límite inferior (%)</i></td> <td><i>Límite superior (%)</i></td> </tr> <tr> <td style="text-align: center;"><b>48,5</b></td> <td style="text-align: center;"><b>71,5</b></td> </tr> </table>	<i>Límite inferior (%)</i>	<i>Límite superior (%)</i>	<b>48,5</b>	<b>71,5</b>	Entre un 48.5% y un 71.5% de los enfermos de coronavirus sufren alteración del sentido del gusto, con una confianza del 95%
<i>Límite inferior (%)</i>	<i>Límite superior (%)</i>				
<b>48,5</b>	<b>71,5</b>				
c) A partir del intervalo obtenido calcule el error de estimación con que está estimando el parámetro.	Error de estimación = $(71.5 - 48.5)/2 = 11.5$				
d) Indique dos formas de disminuir ese error de estimación.	Aumentando el tamaño de muestra y disminuyendo el nivel de confianza				
e) Indique la distribución del estadístico usado para obtener el intervalo.	Distribución Normal estandar				

3. Se realizó una encuesta para estimar la **diferencia** entre la proporción de jóvenes y la proporción de adultos que están de acuerdo con la incorporación del lenguaje inclusivo. De los 75 jóvenes encuestados 60 dijeron estar de acuerdo, mientras que de los 90 adultos encuestados 54 lo estaban. A continuación se presenta la salida del programa Epidat:

*Intervalo de confianza (95,0%) de (Jóvenes – Adultos)*

<i>Diferencia de proporciones</i>	<i>Límite inferior</i>	<i>Límite superior</i>
<b>0,20</b>	<b>0,06</b>	<b>0,34</b>

a) Indique justificando su respuesta si la diferencia entre ambas proporciones es estadísticamente significativa.	La diferencia entre ambas proporciones es estadísticamente significativa porque el intervalo de confianza no contiene al cero.
b) Interprete el intervalo de confianza en términos del problema.	Hay entre un 6% y un 34% más jóvenes que adultos que están de acuerdo con el lenguaje inclusivo, con una confianza del 95%
c) Indique dos maneras de aumentar la precisión de la estimación	Aumentar el tamaño de las muestras. Disminuir el nivel de confianza.

4. Se realiza un estudio cuyo objetivo es testear si la duración del parto (en minutos) es mayor en mujeres primerizas que en aquellas que ya han tenido hijos. Para ello se toma una muestra de 10 mujeres de cada grupo y se les mide la duración del parto. Se supone que ambas variables tienen distribución normal. A continuación se presentan la tabla con medias y desvíos y la salida del programa Epidat:

	Duración media del parto	Desvío estándar
Primerizas (X)	65 minutos	20 minutos
No Primerizas (Y)	45 minutos	15 minutos

Prueba de comparación de varianzas (F)

Estadístico F	gl numerador	gl denominador	Valor p
1,778	9	9	0,404

Prueba de comparación de medias

Estadístico t	gl	Valor p
2,53	18	0,010

a) Plantee las hipótesis del estudio en términos de los parámetros	$H_0: \mu_x = \mu_y$ $H_1: \mu_x > \mu_y$
--	--

b) Plantee las hipótesis del estudio en forma narrativa.	Ho: La duración <b>media</b> del parto en las primerizas es igual que en las no primerizas. H1: La duración <b>media</b> del parto en las primerizas es <b>mayor</b> que en las no primerizas
c) Plantee en forma narrativa la hipótesis <b>nula</b> para evaluar el supuesto de homocedasticidad en términos de las variables del problema.	Ho= la varianza de la duración del parto en las primerizas es igual que la varianza de la duración del parto en las no primerizas
d) A partir del valor P saque conclusiones respecto al supuesto de homocedasticidad justificando su respuesta; indique la probabilidad de error o el nivel de significación efectivo según corresponda.	Como P=0.404 es mayor a 0.10, no se rechaza Ho. No hay evidencias de que las varianzas sean distintas, trabajando a un nivel de significación efectivo de 0.404. Se verifica el supuesto de homocedasticidad.
e) ¿Qué tipo de error podría estar cometiendo en la conclusión anterior?	Como no se rechaza Ho, se podría estar cometiendo un error de tipo II. (No rechazar Ho cuando es falsa)
f) A partir del valor P saque conclusiones respecto al objetivo principal del estudio trabajando con un nivel de significación de 0.05; indique la probabilidad de error o el nivel de significación efectivo según corresponda.	Como P=0.010 es menor a 0.05, se rechaza Ho. Con una probabilidad de error de 1% decimos que la duración media del parto en primerizas es mayor que en no primerizas.
g) Verifique con ayuda de la app el valor P reportado por el Epidat para la prueba de comparación de medias.	Valor P= P(t18 > 2.53) = 0.01048

5. En un estudio se comparó el DVI (diámetro del ventrículo izquierdo en mm, evaluado ecocardiográficamente) entre pacientes con Cardiopatía Chagásica Crónica (CCC), pacientes con cardiopatía isquémica (CI) y pacientes con cardiopatía dilatada (CD). Se seleccionaron 6 pacientes de cada patología.

Patología	CCC (1)	CI (2)	CD (3)
DVI medio	67	69	74

- a) Complete con los valores correspondientes

$\bar{X}$	N	K	n
70 mm	18	3	6

b)

Hipótesis nula en términos de los parámetros	Ho: $\mu_1 = \mu_2 = \mu_3$
Hipótesis <b>nula en forma narrativa</b> en términos del problema	El diámetro <b>medio</b> del ventrículo izquierdo es el mismo en pacientes con CCC, pacientes con CI y pacientes con CD

- c) Complete la tabla de ANOVA con los datos suministrados y conociendo que  $\sum \bar{x}_i^2 = 14726$ .

FUENTE	SC	gl	CM	F
ENTRE	156	2	78	3.9(*)
DENTRO	300	15	20	
TOTAL	456	17		

(Explícite sólo el cálculo de la SC Entre)

$$SC_{Entre} = n \sum \bar{x}_i^2 - N \bar{X}^2 = 6 * 14726 - 18 * 70^2 = 88356 - 88200 = 156$$

d) Calcule el Valor P con ayuda de la App.	P(F2;15 > 3.9) = 0.0432
e) A partir del valor P saque conclusiones respecto al objetivo del estudio trabajando con el nivel de significación más conveniente; indique la probabilidad de error o el nivel de significación efectivo según corresponda.	Como Valor P = 0.0432 es menor a 0.05, se rechaza Ho. Existen diferencias significativas en el diámetro medio del ventrículo izquierdo entre pacientes con CCC, pacientes con CI y pacientes con CD, con una probabilidad de error de 4.32%.
f) El valor P de la prueba de Levene dio 0.35. Saque conclusiones justificando su respuesta e indicando la probabilidad de error o el nivel de significación efectivo según corresponda	Como Valor P=0.35 es mayor a 0.10, no se rechaza Ho. Con la evidencia disponible, no se pudo demostrar que la variabilidad del DVI sea diferente entre pacientes de cada tipo de patología cardíaca, con un nivel de significación efectivo de 0.35. Se cumple el supuesto de homocedasticidad.

6. Se quiere estimar el contenido calórico (en calorías) de la cerveza, en función del porcentaje de alcohol que contiene. Para ello se seleccionan 6 marcas conocidas, desde cervezas sin alcohol a cervezas que contienen hasta un 8% de alcohol con los siguientes resultados:

Marca de cerveza	A	B	C	D	E	F
Porcentaje de alcohol (%)	0	3	4	5	7	8
Contenido calórico (cal)	25	65	80	150	165	220

$$\hat{Y} = 8.2 + 24.3 X$$

a) Identifique cuál es la variable dependiente y cuál la variable independiente.	Variable Independiente X: porcentaje de alcohol Variable dependiente Y: contenido calórico
b) Interprete los coeficientes de la recta obtenida	Se espera que una cerveza sin alcohol tenga 8.2 calorías. Por cada aumento de un 1% en el contenido de alcohol

	de la cerveza, se espera que el contenido calórico aumente en 24.3 calorías.
c) Indique las unidades de los coeficientes de la recta.	Ordenada al origen: calorías Pendiente: calorías/porcentaje
d) Justifique porqué la recta $\hat{Y} = 40 + 60 X$ no representa un buen ajuste a este conjunto de datos.	Porque si reemplazamos algunos de los valores de X en esta recta, los valores de Y obtenidos están muy alejados de los obtenidos en la muestra.
e) Plantee las hipótesis para testear la significación de la regresión.	Ho: $\beta_1 = 0$ (no existe una regresión significativa) H1: $\beta_1 \neq 0$ (existe una regresión significativa)
f) Saque conclusiones sabiendo que el Valor P del ANOVA aplicado a la regresión dio 0.0018, e indique el error o el nivel de significación efectivo según corresponda.	Como Valor P = 0.0018 es menor a 0.01, se rechaza Ho. Existe una regresión altamente significativa, con una probabilidad de error de 0.18%. La recta es un buen estimador del contenido calórico en función del porcentaje de alcohol.
g) Interprete el valor del coeficiente $R^2 = 0.93$	El porcentaje de alcohol que contiene una cerveza explica un 93% de su contenido calórico, el resto es explicado por otras variables.
h) ¿Cuántas calorías se espera que tenga una cerveza con un porcentaje de alcohol de 6%?	Se espera que una cerveza con 6% de alcohol tenga 154 calorías. $(8.2 + 24.3 * 6)$
i) ¿Cuántas calorías se espera que tenga una cerveza con un porcentaje de alcohol de 10%?	No puede estimarse porque la recta se calculó con cervezas que tenían hasta un 8% de alcohol como máximo.
j) ¿Qué porcentaje de alcohol se espera que tenga una cerveza en cuya etiqueta se lee "90 calorías"?	No puede estimarse porque la recta sirve para estimar calorías (variable aleatoria) en función de porcentaje de alcohol (variable fija), pero no a la inversa.



**RECUERDE CONVERTIR EL ARCHIVO A PDF ANTES DE SUBIRLO A MOODLE**