Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
20/02/2020 **Curso virtual**	**FEBRERO 2020 – TIPO B**

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2020 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Distribución de frecuencias en una prueba de memoria

X_i	n_i
7	3
8	4
9	6
10	15
11	12
12	10

Tabla 2. Datos de 20 alumnos donde:

Sí = Utilizan los cursos virtuales
No = No utilizan los cursos virtuales

	Sí	**No**
Aprobados	7	3
Suspensos	3	7

Supuesto 1. Un test de laboratorio es efectivo al 95% detectando en una muestra de sangre cierta enfermedad cuando esta se padece realmente. Sin embargo, el test también arroja un resultado "falso positivo" en el 1% de personas sanas sobre las que se realiza el test.

1. La variable estado civil tiene un nivel de medida: A) ordinal; **B) nominal**; C) de intervalo.
2. En una distribución de frecuencias relativas acumuladas, el valor para el último intervalo (el correspondiente al valor máximo) es: **A) 1**; B) 100; C) dependiente del tamaño de la muestra.
3. Con los datos de la Tabla 1. ¿Qué porcentaje de sujetos superan la puntuación 10,5? A) 74%; **B) 44%**; C) 56%.
4. ¿Cuál de las siguientes afirmaciones describe la distribución de frecuencias de la Tabla 1? A) es una distribución simétrica; **B) las mayores frecuencias corresponden a valores altos de X**; C) las menores frecuencias corresponden a valores altos de X.
5. ¿Cuál de las siguientes afirmaciones es FALSA? Los percentiles permiten comparar directamente entre sí puntuaciones obtenidas ...: A) …por sujetos diferentes dentro del mismo grupo y para la misma variable; B) …por el mismo sujeto en el mismo grupo y con diferentes variables; **C) …por el mismo grupo en diferentes variables**.
6. Dadas las siguientes puntuaciones: 5, 6, 3, 7, 4. ¿Cuánto vale su varianza? **A) 2**; B) 1,41; C) 6.
7. Si se quiere representar gráficamente la proporción de hombres y mujeres que votan a cada uno de los grandes partidos políticos que se presentan en tu ciudad. ¿Qué tipo de gráfica se utilizaría? A) El diagrama de sectores; B) El diagrama de tallo y hojas; **C) El diagrama de barras conjunto**.
8. Para los datos de la Tabla 2, el valor de χ^2 es: **A) 3,2**; B) 1,6; C) 0,37.
9. Siguiendo con los datos de la Tabla 2, el valor máximo del coeficiente de contingencia será: **A) 0,707**; B) 0,077; C) 0,770.
10. Dos psicólogos ordenan a cinco pacientes en función de su aptitud para un puesto de trabajo. Sabiendo que la suma de las diferencias de los dos órdenes al cuadrado vale 4, el coeficiente de correlación entre los dos órdenes toma el valor: **A) 0,8**; B) 0,2; C) 0,4.
11. ¿Cuál de los siguientes índices es el más adecuado para medir la relación entre dos variables cualitativas? **A)** χ^2; B) Spearman; C) Biserial puntual.

12. Respecto al modelo de regresión, si la varianza de los errores es 0,3 y la varianza de las puntuaciones observadas es 2,64, entonces la desviación típica de las puntuaciones pronosticadas será: A) 2,94; **B) 1,53**; C) 2,34.
13. ¿Cuál de las siguientes afirmaciones es VERDADERA? **A) Una correlación de -0,78 entre dos variables X e Y tiene la misma intensidad que otra correlación de +0,78 entre otras variables U y V**; B) Una correlación de 0,60 entre dos variables X e Y indica el doble de correlación que otra correlación de 0,30 entre otras dos variables U y V; C) Encontrar una relación entre dos variables cuantitativas significa que existe una relación de causa-efecto.
14. De la expresión $P(A \cap B) = P(A)P(B/A) = P(A)P(B)$ se deduce que: A) el suceso B está condicionado al suceso A; **B) los sucesos A y B son independientes**; C) el suceso A está condicionado al suceso B.
15. Se lanzan simultáneamente un dado y una moneda. ¿Cuál es la probabilidad de obtener un número par en el dado y una cara en la moneda? A) 0,5; **B) 0,25**; C) 0,75.
16. En el supuesto 1, la especificidad de la prueba de laboratorio es: **A) 0,99**; B) 0,95; C) 0,01.
17. Sea X una variable aleatoria discreta, cuya función de probabilidad asigna 1/4 a los valores 1, 2, 3 y 4. La función $F(x) = P(X \leq x)$ será: A) $P(x = 1) = 1/4$, $P(x = 2) = 3/4$, $P(x = 3) = 2/4$, $P(x = 4) = 1$; B) $P(x = 1) = P(x = 2) = P(x = 3) = P(x = 4) = 1/4$; **C) $P(x \leq 1) = 1/4$, $P(x \leq 2) = 2/4$, $P(x \leq 3) = 3/4$, $P(x \leq 4) = 1$**.
18. Sea X una variable aleatoria discreta con valores 0, 1, 2, 3. Siendo sus probabilidades respectivamente 0,2, 0,3, 0,3, 0,2. ¿Cuál será la varianza de la variable X? A) 1,50; B) 0,45; **C) 1,05**.
19. Sabiendo que X se distribuye normalmente, que la media de X es 60 y que la puntuación directa 40,8 es superada por el 89,97% de las observaciones. ¿Cuál es el valor aproximado de la desviación típica? **A) 15**; B) 1,28; C) 17,87.
20. En una distribución normal con varianza igual a 4. ¿Cuál es la proporción de observaciones entre la puntuación diferencial $x = -1,5$ y la media aritmética? A) 0,2266; **B) 0,2734**; C) 0,3520.
21. En una distribución F con 8 y 8 grados de libertad. ¿Cuál es el valor aproximado del percentil 1? **A) 0,166**; B) 6,029; C) 0,222.
22. Sabiendo que la media poblacional de una variable X vale 160 y su varianza poblacional vale 9. La distribución muestral de la media para un tamaño de muestra n = 81, tendrá: A) Una desviación típica de 1; **B) Una desviación típica de 0,33**; C) Los mismos valores de la población para la media y la desviación típica de la distribución muestral.
23. Una población se compone de los posibles resultados que se pueden obtener cada vez que se tira un dado no trucado. ¿Cuál es la media poblacional? **A) 3,5**; B) 4; C) 3.
24. Se ha extraído una muestra de 144 personas de una población. Se conoce la varianza poblacional que es igual a 9 y también que el límite superior en la estimación por intervalo de la media es 6,645, con un nivel de confianza del 99%. ¿Cuánto vale la media de la muestra seleccionada? A) 6,15; B) 6,06; **C) 6**.
25. En la construcción del intervalo de confianza para estimar, con una determinada probabilidad, entre qué valores se encontrará la media poblacional a partir de la información proporcionada por una muestra representativa, ¿cuál de las siguientes afirmaciones es verdadera? A) Al aumentar el nivel de confianza disminuye la amplitud del intervalo; **B) Al aumentar el tamaño de la muestra disminuye la amplitud del intervalo**; C) El tamaño de la muestra no afecta a la amplitud del intervalo de confianza.

SOLUCIONES

1. **B**. Página 13.

2. **A**. Página 26.

3. **B**. Se ha de obtener el percentil correspondiente a la puntuación 10,5.

X_i	X	n_i	n_a
7	6,5 – 7,5	3	3
8	7,5 – 8,5	4	7
9	8,5 – 9,5	6	13
10	9,5 – 10,5	15	28
11	10,5 – 11,5	12	40
12	11,5 – 12,5	10	50

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(10,5 - 10,5)\cdot 12}{1} + 28}{50}\right]\cdot 100 = 56\%$$

56% corresponde con el porcentaje de observaciones por debajo de 10,5. Por tanto,

100% - 56% = 44% de observaciones quedarán por encima de la puntuación 10,5

4. **B**. Página 38.

5. **C**. Página 76.

6. **A**. Primero se calcula la media, para así poder calcular la varianza:

$$\bar{X} = \frac{\sum x_i}{n} = 5;\ S_x^2 = \frac{\sum X_i^2}{n} - \bar{X}^2 = \frac{135}{5} - 25 = 2$$

7. **C**. Página 155.

8. **A**. Se calculan las tablas de frecuencias observadas y esperadas y a continuación se halla χ^2:

Frecuencias observadas

	SI	NO	
Aprobado	7	3	10
Suspenso	3	7	10
	10	10	20

Frecuencias esperadas

	SI	NO
Aprobado	5	5
Suspenso	5	5

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{4}{5} + \frac{4}{5} + \frac{4}{5} + \frac{4}{5} = \frac{16}{5} = 3,2$$

9. **A**. El valor máximo del coeficiente de contingencia depende del número de filas y columnas. Siendo k = número de filas = número de columnas. Se halla mediante la siguiente fórmula:

$$Cmax = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{2-1}{2}} = 0{,}707$$

10. **A**. Tenemos dos variables ordinales, luego hemos de aplicar el coeficiente de correlación de Spearman. Sabemos que la suma de los cuadrados de las diferencias entre los rangos es igual a 4, y aplicando la fórmula de Spearman.

$$r_s = 1 - \frac{6 \cdot \sum d^2}{n(n^2-1)} = 1 - \frac{6 \cdot 4}{5(5^2-1)} = 1 - \frac{24}{120} = 1 - 0{,}20 = 0{,}80$$

11. **A**. Página 217.

12. **B**. Partiendo de que la varianza de las puntuaciones en Y es igual a la varianza de los pronósticos más la varianza de los errores se obtiene que:

$$S_Y^2 = S_{Y'}^2 + S_{Y.X}^2; 2{,}64 = S_{Y'}^2 + 0{,}3;\ S_{Y'} = \sqrt{2{,}64 - 0{,}3} = 1{,}53$$

13. **A**. Página 201.

14. **B**. Página 269.

15. **B**. Se definen los eventos Pr = resultado par en el dado y Cr = cara al tirar la moneda. Sabiendo que:

$$P(Pr) = \frac{3}{6} = 0{,}5; y\ P(Cr) = \frac{1}{2} = 0{,}5$$

Se obtiene que:

$$P(Pr \cap Cr) = P(Pr) \cdot P(Cr) = 0{,}5 \cdot 0{,}5 = 0{,}25$$

16. **A**. La especificidad hace referencia a la probabilidad de la prueba parar detectar a los verdaderos negativos. Esto es 1-“falsos positivos”; es decir 1- 0,01= 0,99.

17. **C**. A partir de los valores de la función de probabilidad que figura en la siguiente tabla:

x	1	2	3	4
f(x)	1/4	1/4	1/4	1/4

Se obtiene que:

$$F(1) = P(X \leq 1) = \frac{1}{4}$$

$$F(2) = P(X \leq 2) = P(X=1) + P(X=2) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}$$

$$F(3) = P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

$$F(4) = P(X \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$$

18. **C**. A partir de los valores de la función de probabilidad que figura en la siguiente tabla:

x	0	1	2	3
f(x)	0,2	0,3	0,3	0,2

En primer lugar se calcula la media:

$$\mu_x = \sum x \cdot f(x) = 0 \cdot 0{,}2 + 1 \cdot 0{,}3 + 2 \cdot 0{,}3 + 3 \cdot 0{,}2 = 0 + 0{,}3 + 0{,}6 + 0{,}6 = 1{,}5$$

A continuación se calcula la varianza: $\sigma^2 = V(x) = \sum (x - \mu)^2 \cdot f(x)$

x	***f*(x)**	**(x-μ)**	**$(x-\mu)^2$**	**$(x-\mu)^2 \cdot f(x)$**
0	0,2	-1,5	2,25	0,45
1	0,3	-0,5	0,25	0,075
2	0,3	0,5	0,25	0,075
3	0,2	1,5	2,25	0,45
			∑	1,05

19. **A**. En primer lugar se busca la probabilidad superada por el 89,97% de las observaciones (que deja por debajo el 10,03% de la distribución), 1-0,8997 = 0,1003. A continuación se halla en la tabla de la curva normal su puntuación típica z = -1,28. A partir de ahí se aplica la formula correspondiente:

$$z = \frac{X - \bar{X}}{S_x}\ ;\ -1{,}28 = \frac{40{,}8 - 60}{S_x};\ S_x = \frac{40{,}8 - 60}{-1{,}28} = 15$$

20. **B**. Por debajo de la media se encuentra una proporción de observaciones igual a 0,5000. Y por debajo de una puntación diferencial igual a -1,5:

$$z = \frac{X - \bar{X}}{S_x} = \frac{x}{S_x} = \frac{-1{,}5}{\sqrt{4}} = -0{,}75;$$

Tras calcular la puntuación típica se busca en la tabla de la curva normal su proporción. z = -0,75 (0,2266). Por último se calcula la proporción de observaciones entre la media y x = -1,5.

Esto es: 0,5000-0,2266 = 0,2734.

21. **A**. Se hace uso de la propiedad recíproca de la distribución F. De este modo, dado que los grados de libertad en el numerador y el denominador son los mismos, 8 y 8, bastará con calcular 1/ $F_{8,8}$ para un nivel de confianza 0,99. Esto es,

$$_{0,01}F_{8,8} = 1/\ _{(1-0,01)}F_{8,8} = 1/\ _{0,99}F_{8,8} = 1/6{,}029 = 0{,}166 \text{ (ver tabla VII)}$$

22. **B**. En este caso, la media de la distribución muestral de la media es igual a la media poblacional, es decir 160. El valor de la desviación típica de la distribución muestral, en cambio, será igual a:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\ ; \sigma_{\bar{X}} = \frac{3}{\sqrt{81}}\ ; \sigma_{\bar{X}} = \frac{3}{9} = 0{,}33$$

23. **A**. Se calcula el valor de la media poblacional:

$$\mu = \frac{\sum X_i}{N}\ ; \mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3{,}5$$

24. **C**. Para un nivel de confianza del 99% se utilizará z = 2,58. Por tanto, E_{max} será:

$$E_{max} = Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 2{,}58\frac{3}{12} = 0{,}645$$

Y la media de la muestra seleccionada será:

$$L_{sup} = \bar{X} + E_{max};\ \bar{X} = L_{sup} - E_{max} = 6{,}645 - 0{,}645 = 6$$

25. **B**. Página 459.