



Universidad: **Francisco de Vitoria · UFV**

Grado: **Biomedicina**

Asignatura: **Bioestadística**

Examen Resuelto Enero 2021

Instrucciones

- El examen tendrá una duración de dos horas.
- Material permitido: una calculadora y el formulario y tablas estadísticas que se entregarán junto con el enunciado del examen.
- Entrega del examen: el alumno debe entregar la hoja de enunciados junto con el formulario y las tablas estadísticas y las hojas con el desarrollo de las respuestas.
- Puntuación de las preguntas:
- Tipo test (3 pts): cada respuesta correcta son 0,2 pts. Las respuestas incorrectas no restan.
 - Problema 1 (2,5 pts).
 - Problema 2 (2 pts)
 - Problema 3 (2,5 pts).

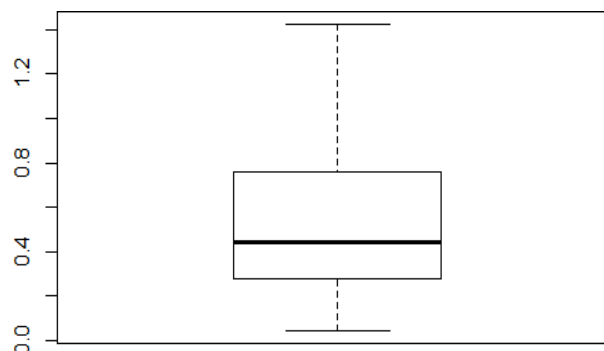
Teoría

En las siguientes cuestiones sólo hay una opción correcta. Marca la opción que corresponda.

1. Se tiene un conjunto de datos cuyo coeficiente de asimetría de Fisher es igual a 1,7. Esto quiere decir que:
a. La mediana es menor que la media aritmética.
b. La mediana es igual a la media aritmética, la distribución es, por tanto, asimétrica.
c. La mediana es igual a la media aritmética, la distribución es, por tanto, simétrica.
d. La mediana es mayor que la media aritmética.
2. La técnica de conteo conocida como variación considera diferentes los conjuntos:
a. Que varían en la naturaleza de los elementos o en el orden de estos.
b. Que varían exclusivamente en la naturaleza de los elementos sin importar el orden de estos.
c. Que varían exclusivamente en el orden de los elementos sin importar la naturaleza de estos.
d. Que varían tanto en la naturaleza de los elementos como en el orden de estos.
3. Una prueba de Bernoulli es:
a. Un experimento que sigue una distribución binomial.
b. Un experimento que consta de una sola prueba y cuyo resultado debe tener dos o más opciones mutuamente excluyentes.
c. Un experimento que consta de una sola prueba y cuyo resultado sólo puede tener dos opciones mutuamente excluyentes.
d. Un experimento con n pruebas independientes con resultado dicotómico.



4. ¿Cuál de las siguientes condiciones no es un requisito para realizar un análisis ANOVA?
 - a. La variable estudiada debe distribuirse normalmente en las K poblaciones o, si esto no se cumple, el tamaño del grupo debe ser mayor que 30.
 - b. Debe cumplirse la condición de homocedasticidad entre los K grupos.
 - c. El muestreo debe ser aleatorio.
 - d. El número total de datos de observaciones debe ser mayor que 30.**
5. Las frecuencias absolutas de las modalidades de una variable cualitativa son:
 - a. La varianza de cada modalidad.
 - b. El porcentaje de casos que presenta cada modalidad.
 - c. La media de cada modalidad.
 - d. El número de casos que presenta cada modalidad.**
6. En un contraste de hipótesis realizado con el estadístico Z de la curva normal, a la hora de comprobar si el resultado es estadísticamente significativo, ¿cuándo se considera el valor $Z\alpha$?
 - a. Cuando el contraste es bilateral.**
 - b. Cuando el contraste es unilateral por la derecha.
 - c. Cuando el contraste es unilateral por la izquierda.
 - d. Siempre.
7. ¿Qué significa $P(A/B)$?
 - a. La probabilidad de que ocurra A sabiendo que ha ocurrido B.
 - b. La probabilidad de que ocurra A sabiendo que no ha ocurrido B.**
 - c. Sabiendo que ocurre A, la probabilidad de que no ocurra B.
 - d. Sabiendo que ocurre A, la probabilidad de que ocurra B.
8. Se tiene la hipótesis de que la proporción de individuos que van a urgencias al menos una vez al año en una determinada ciudad es mayor que 0,2. Se seleccionan de manera aleatoria 10 individuos de dicha ciudad y 5 de ellos afirman haber acudido a urgencias al menos una vez en el último año. Se quiere resolver el contraste con un alfa de 0,05, ¿qué método usarías?
 - a. Utilizaría el estadístico de la t de Student ya que no conocemos la desviación típica de la población.
 - b. Basta con aplicar el estadístico Z por aproximación a la distribución normal.
 - c. No se cumplen los criterios para la aproximación a la normal, por lo que resolvería utilizando la ley de probabilidad de Poisson.
 - d. Resolvería el contraste utilizando la ley binomial.**
9. Se tiene un conjunto de datos que da lugar al siguiente diagrama de cajas:



¿Cuál de las siguientes afirmaciones es cierta?

- a. La distribución podría considerarse normal.
- b. La media de la distribución es aproximadamente 0,4.
- c. La distribución presenta un gran número de valores atípicos.



d. Se aprecia una clara asimetría en la distribución.

10. Cuando la probabilidad de que ocurra un suceso A sabiendo que ha ocurrido un suceso B no cambia respecto a cuando no se tiene información sobre B, $P(A/B) = P(A)$, se dice que:
- Los sucesos A y B son independientes.**
 - Los sucesos A y B son incompatibles.
 - Los sucesos A y B son dependientes.
 - Los sucesos A y B son iguales.
11. La potencia de un contraste de hipótesis es:
- La probabilidad de rechazar la hipótesis nula cuando esta es falsa.**
 - La probabilidad de aceptar la hipótesis nula cuando esta es falsa.
 - La probabilidad de aceptar la hipótesis alternativa cuando esta es falsa.
 - La probabilidad de aceptar la hipótesis alternativa cuando la hipótesis nula es cierta.
- 12.Cuál de las siguientes variables es ordinal:
- Veces a la semana hace deporte una persona.
 - El número de horas de deporte que una persona hace a la semana.
 - Tipo de deporte que practica una persona (categorizado como individual o colectivo).
 - Cuánto deporte hace una persona (categorizado como mucho, regular o poco).**
13. ¿Qué indica el tercer cuartil de una variable?
- El valor por debajo del cual está el 25% de las observaciones de dicha variable.
 - El valor medio de las primeras tres observaciones de dicha variable.
 - El valor por encima del cual está el 25% de las observaciones de dicha variable.**
 - La dispersión del 25% de los valores que están más próximos entre sí.
14. ¿Cuál de las siguientes no es una condición para que un experimento se considere binomial?
- El experimento consiste en n pruebas idénticas.
 - El resultado de cada prueba es dicotómico y puede ser evaluado como "éxito" o "fracaso".
 - Las n pruebas son independientes entre sí.
 - La variable aleatoria binomial, X, es el número de prueba en el que se obtiene "éxito" por primera vez.**
15. ¿En cuál de las siguientes aplicaciones se puede utilizar la prueba de chi cuadrado?
- Para estudiar la asociación entre variables cualitativas.**
 - Para comparar las varianzas de dos muestras independientes.
 - Para contrastar la varianza de una muestra frente a un valor de referencia.
 - Para comparar las medias de k grupos, con $k > 2$.

Problema 1

Un estudio pretende analizar la relación de las pulsaciones por minuto en reposo en función de la altura en cm. Para ello se recogieron ambas variables en un grupo de 8 personas, obteniéndose los siguientes resultados:

Altura (cm)	160	167	173	162	167	175	185	162
Pulsaciones por minuto	68	62	76	64	65	80	80	63

- Determinar la recta de regresión de pulsaciones en función de altura. (0.75 pts)
- Determinar el coeficiente de determinación. (0.75 pts)



- c) En base a los resultados obtenidos en los apartados a) y b), ¿qué conclusiones se pueden sacar sobre la relación entre ambas variables? (1 pts)

Apartado a)

Nos piden la recta de regresión de las pulsaciones por minuto en función de la altura, por tanto, la variable dependiente (y) será las pulsaciones por minuto y la independiente (x) será la altura.

Construimos la tabla con los sumatorios para después calcular los parámetros β_0 y β_1 que definen la recta:

$$y = \beta_0 + \beta_1 \cdot x.$$

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
160	68	10880	25600	4624
167	62	10354	27889	3844
173	76	13148	29929	5776
162	64	10368	26244	4096
167	65	10855	27889	4225
175	80	14000	30625	6400
185	80	14800	34225	6400
162	63	10206	26244	3969
$\sum x_i = 1351$	$\sum y_i = 558$	$\sum x_i y_i = 94611$	$\sum x_i^2 = 228645$	$\sum y_i^2 = 39334$

Ahora procedemos a los cálculos:

$$\beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{8 \cdot 94611 - 1351 \cdot 558}{8 \cdot 228645 - 1351^2} = 0,765$$

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n} = \frac{558 - 0,765 \cdot 1351}{8} = -59,44$$

Apartado b)

Nos piden el coeficiente de determinación, esto es, R^2 .

$$R^2 = \frac{\beta_1^2 \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right)}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}} = \frac{0,765^2 \left(228645 - \frac{1351^2}{8} \right)}{39334 - \frac{558^2}{8}} = 0,7$$

Apartado c)

El coeficiente de determinación nos indica que ambas variables están relacionadas con cierta fuerza, de hecho, la variable altura es capaz de explicar un 70% de la variabilidad que se presenta en las pulsaciones por minuto. La raíz del coeficiente de determinación nos devuelve el coeficiente de correlación de Pearson, r , que tendrá signo positivo ya que la pendiente, β_1 , tiene signo positivo y valor de $r = \sqrt{R^2} = 0,837$. Esto nos indica que además de estar ambas variables relacionadas, lo están de forma lineal y directamente proporcional, por lo que a mayor altura se esperan mayores pulsaciones por minuto.



Problema 2

En una empresa de 750 trabajadores, un 25% es positivo para un marcador genético que indica un riesgo elevado de padecer hipertensión. Si se eligen al azar 10 empleados distintos de la empresa y se analiza la presencia del marcador genético:

- ¿Cuál es la probabilidad de que exactamente 3 lo tengan? (1 pts)
- ¿Cuál es la probabilidad de que haya más de 1 empleado con el marcador genético entre los 10 seleccionados al azar? (1 pts)

El evento presentar el marcador genético en cuestión o no puede considerarse una prueba con resultado dicotómico. Además, podemos considerar que el hecho de que un trabajador presente el marcador no altera la probabilidad de que otro trabajador lo tenga, por tanto, se trata de pruebas idénticas e independientes. Al tomar una muestra al azar de 10 trabajadores podemos definir la variable binomial X como el número de trabajadores que tienen el marcador entre los 10 elegidos, con probabilidad de éxito $p = 0,25$.

Apartado a)

La probabilidad de que exactamente 3 tengan el marcador es la probabilidad de que $X = 3$.

$$P(X = 3) = \binom{10}{3} \cdot 0,25^3 \cdot 0,75^7 = 120 \cdot 0,015625 \cdot 0,13348 = 0,25$$

Apartado b)

La probabilidad de que haya más de un empleado con el marcador es la probabilidad de que $X > 1$.

$$P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - 0,75^{10} - 10 \cdot 0,25 \cdot 0,75^9 = 0,756$$

Problema 3

Un grupo de investigadores estudian la aceleración de redes neuronales celulares obteniendo los siguientes resultados:

2,5	3,3	2,8	3,4	3,5	3,1	3,5	3,2	3,5	3,2	3,5	3,3	3,5
1	5	5	5	1	8	3	5	4	9	4	3	5

Determina el intervalo de confianza con nivel de seguridad del 95% para la aceleración media asumiendo que los datos proceden de una distribución normal.

Nos dicen que los datos proceden de una distribución normal, por lo tanto, aunque $n < 30$ sabemos que la variable aleatoria media de la muestra seguirá una distribución normal.

Por otro lado, no tenemos valores poblacionales. Tan solo una muestra de 13 medidas en redes neuronales celulares. Esto nos obliga a usar el estimador de la desviación estándar de la muestra y, por tanto, trabajaremos con una t de Student. Como los grados de libertad son $n-1 = 12$, no es posible aproximar la t de Student a una normal.

Podemos considerar la fracción muestral como pequeña, ya que la población de estudio, que son las redes neuronales celulares, puede considerarse mucho más grande que la muestra de 13 que se han estudiado.



Comenzamos los cálculos obteniendo la media muestral y el estimador de la desviación típica:

$$\bar{x} = \frac{\sum x_i}{n} = 3,3$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{1,1442}{12}} = 0,31$$

Por lo explicado anteriormente sabemos que el intervalo de confianza con seguridad del 95%, es decir, $\alpha=0,05$, se obtiene mediante la fórmula:

$$\bar{x} \pm t_{12, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

Por lo explicado antes:

$$L_i = 3,3 - 2,179 \cdot \frac{0,31}{\sqrt{13}} = 3,11$$

$$L_s = 3,3 + 2,179 \cdot \frac{0,31}{\sqrt{13}} = 3,49$$

El intervalo de confianza con seguridad del 95% de la media de la aceleración de las redes neuronales celulares es:

$$\mu \in (3,11 - 3,49)$$