P1.- Para cada pregunta responda Verdadero o Falso, justificando su respuesta en cada caso.

1.1 Para hacer inferencia estadística basta tomar cualquier muestra dentro de la población de estudio.
1.2 Un estimador insesgado es siempre mejor que un estimador sesgado.
1.3 La varianza de un estimador insesgado puede valer 0.
1.4 Diremos que un estimador es consistente si su varianza se mantiene siempre por debajo de una cota determinada.
1.5 Si nos restringimos a tomar estimadores insesgados que son combinaciones lineales de los valores de una muestra considerada, entonces el estimador con menor varianza para la media teórica de una variable aleatoria con distribución desconocida es la media muestral.
1.6 En un intervalo de confianza de largo mínimo para la media μ de una distribución Normal, si aumento el nivel de confianza, entonces aumenta el largo del intervalo.
1.7 En una encuesta a nivel nacional sobre una muestra de 1.200 personas, el 65% consideró que la cumbre APEC fue beneficiosa para el país. Con un nivel de confianza del 95% estos resultados permiten concluir que la diferencia absoluta entre la proporción estimada y la proporción poblacional es menor al 3%. NOTA: Suponga normalidad de las observaciones.
1.8 En un test de hipótesis, en el que se utiliza un estadístico "t" para llevar a cabo el contraste, el P-Valor indica la probabilidad de obtener un valor al menos tan extremo como el calculado para el estadístico "t" con la muestra obtenida.
1.9 Si la correlación lineal entre dos variables aleatorias continuas X e Y es cercana a 1 indica una relación funcional fuerte entre X e Y.
1.10 Un test de significación individual, o test t-Student, en un modelo lineal, permite testear la hipótesis $x_i = 0$ versus $x_i \neq 0 \quad \forall i = 1..n$.
1.11 Es correcto indicar que la solución derivada de la técnica de los mínimos cuadrados ordinarios en un modelo lneal simple se podría utilizar para estimar los coeficientes β_0 y β_1 de un modelo de la forma $y = \beta_0 e^{\beta_1 x}$ definido sobre un conjunto de datos (x,y).
1.12 Para hacer un test ANOVA solo basta conocer la varianza total de los datos, la varianza por grupo y el número de observaciones por grupo.

PAUTA:

P1.- Para cada pregunta responda Verdadero o Falso, justificando su respuesta en cada caso.

1.1 Para hacer inferencia estadística basta tomar cualquier muestra dentro de la población de estudio.

FALSO: No basta tomar una muestra cualquiera. Los resultados de las técnicas clásicas de inferencia estadísticas dependen, entre otros supuestos, de la aleatoriedad e independencia de las observaciones, lo que se consigue con muestreo aleatorio.

1.2 Un estimador insesgado es siempre mejor que un estimador sesgado.

FALSO: Ya que un estimador insesgado puede tener una varianza muy grande, lo que lo hace más impreciso. Una manera má apropiada de escoger entre estimadores, es elegir a aquel que posea el menor Error Cuadrático Medio.

1.3 La varianza de un estimador insesgado puede valer 0.

FALSO: Ya que la varianza de todo estimador insesgado se encuentra acotada inferiormente por la Cota de Cramer-Rao.

1.4 Diremos que un estimador es consistente si su varianza se mantiene siempre por debajo de una cota determinada.

FALSO: Un estimador $\hat{\theta}$ de un parámetro θ es consistente, si converge en probabilidad hacia θ cuando el tamaño de la muestra crece. Ahora bien, una prueba más práctica de dicha convergencia es probar que el estimador converge en media cuadrática, esto es que $E(\hat{\theta}) \to \theta$ y $Var(\hat{\theta}) \to 0$ cuando $n \to \infty$.

1.5 Si nos restringimos a tomar estimadores insesgados que son combinaciones lineales de los valores de una muestra considerada, entonces el estimador con menor varianza para la media teórica de una variable aleatoria con <u>distribución</u> desconocida es la media muestral.

VERDADERO: En efecto, si $\tilde{\theta}$ es un estimador para θ, tal que $\tilde{\theta} = \sum_i a_i X_i$, entonces el valor de los a_i que minimizan la varianza es $1/n$, con lo que resulta la media muestral $\overline{X}$. Ya que a_i = 1/n es solución al problema de minimización:

Min $Var(\tilde{\theta}) = \sum_i a_i^2 \sigma^2$ s.a. $E(\tilde{\theta}) = \theta$

1.6 En un intervalo de confianza de largo mínimo para la media μ de una distribución Normal, si aumento el nivel de confianza, entonces aumenta el largo del intervalo.

VERDADERO: Un intervalo de confianza de largo mínimo para la media μ de una Normal es de la forma: $\bar{X} - k\frac{\sigma}{\sqrt{n}}, \bar{X} + k\frac{\sigma}{\sqrt{n}}$ es el caso de varianza conocida y $\bar{X} - t\frac{S_n}{\sqrt{n-1}}, \bar{X} + t\frac{S_n}{\sqrt{n-1}}$ para el caso de varianza desconocida, en que k y t son los valores de las distribuciones Normal y t-Student respectivamente a un nivel de confianza α dado. La forma acampanada de ambas distribuciones nos revela que si aumenta el nivel de confianza 1-α , disminuye el valor de α , y por ende las probabilidades de las colas de la distribución dismiuyen, y el intervalo crece.

1.7 En una encuesta a nivel nacional sobre una muestra de 1.200 personas, el 65% consideró que la cumbre APEC fue beneficiosa para el país. Con un nivel de confianza del 95% estos resultados permiten concluir que la diferencia absoluta entre la proporción estimada y la proporción poblacional es menor al 3%. NOTA: Suponga normalidad de las observaciones.

VERDADERO: Esto equivale a resolver $P(|\hat{p} - p| \leq c) = 0{,}95$, que corresponde al intervalo de confianza para una proporción. Si suponemos normalidad se puede encontrar entonces que $c = k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1{,}96\sqrt{\frac{0{,}65(1-0{,}65)}{1200}} \approx 0{,}03$

1.8 En un test de hipótesis, en el que se utiliza un estadístico "t" para llevar a cabo el contraste, el P-Valor indica la probabilidad de obtener un valor al menos tan extremo como el calculado para el estadístico "t" con la muestra obtenida.

VERDADERO: El P-Valor calculado con un estadístico "t", se obtiene al encontrar P(T>c), donde T es la distribución que sigue el estadístico de prueba "t", y "c" es el valor encontrado para el estadístico "t" con la muestra obtenida.

1.9 Si la correlación lineal entre dos variables aleatorias continuas X e Y es cercana a 1 indica una relación funcional fuerte entre X e Y.

FALSO: Si la correlación lineal entre X e Y es cercana a 1, lo único que se puede afirmar (desde el punto de vista numérico) es que existe una fuerte asociación **lineal**. Es importante recalcar que la existencia de otras tendencias funcionales y/o puntos extremos puede distorsionar esta última interpretación.

1.10 Un test de significación individual, o test t-Student, en un modelo lineal, permite testear la hipótesis $x_i = 0$ versus $x_i \neq 0 \quad \forall i = 1..n$.

FALSO: Un test de significación individual revela si una variable es significativa o no haciendo un test sobre el COEFICIENTE asociado a dicha variable. Lo que corresponde testear es H_0: $\beta_i = 0$ versus H_1: $\beta_i \neq 0$.

1.11 Es correcto indicar que la solución derivada de la técnica de los mínimos cuadrados ordinarios en un modelo lneal simple se podría utilizar para estimar los coeficientes β_0 y β_1 de un modelo de la forma $y = \beta_0 e^{\beta_1 x}$ definido sobre un conjunto de datos (x,y).

VERDADERO: Si se aplican logaritmos a cada lado obtenemo, podemos llevar el modelo a la forma clásica de un modelo lineal simple: $y' = \beta_0' + \beta_1' x$, donde $y' = \ln y$, $\beta_0' = \ln \beta_0$, y $\beta_1' = \beta_1$, el cual se puede resolver con el método propuesto.

1.12 Para hacer un test ANOVA solo basta conocer la varianza total de los datos, la varianza por grupo y el número de observaciones por grupo.

VERDADERO: El test F que se requiere para decidir el test ANOVA es de la forma:

$$F = \frac{B}{q-1} \Big/ \frac{W}{n-q}$$

donde W es la suma de la variación total intra-grupo, la que se deriva a partir de las varianzas individuales S_j^2 de cada grupo, y el tamaño de cada grupo n_j : $W = \sum_{j=1}^{q} n_j S_j^2$, mientras que B, la varianza entre-grupos se puede obtener de la relación T = B + W, donde T es la variación total de los datos, es decir: $T = nS_n^2$