**INFERENCIA ESTADÍSTICA**

**EXAMEN**

13-diciembre-2018

**Profesores**: Jeanette Fuentes; Francisco Javier Leiva

**Ayudantes**: Sofía Muñoz; Francisca Quintanilla; Jaime Sánchez

1. **Preguntas Conceptuales (20pts.)**

Comente la veracidad de las siguientes afirmaciones, y justifique su respuesta.

1. En el método de máxima verosimilitud, lo que se busca es encontrar el máximo valor entre todos los posibles para un determinado indicador.

**RESPUESTA:**

*Falso, en el método de máxima verosimilitud lo que se busca es en encontrar un valor que haga máxima la probabilidad de ocurrencia del valor de un determinado indicador.*

1. Al escoger una muestra con la cual trabajar, siempre será mejor escoger aquella que tenga un tamaño más grande, pues esa muestra será la que tenga un menor nivel de error muestral.

**RESPUESTA:**

*Falso, el nivel de error muestral depende de la desviación estándar y del tamaño de la muestra, por lo que no necesariamente una muestra con tamaño más grande tendrá un nivel de precisión mayor.*

1. En caso de tener dos estimadores para un parámetro, siempre será mejor escoger el estimador con menor varianza.

**RESPUESTA**:

*Falso, en caso de ambos estimadores fueran insesgados, efectivamente se preferirá aquel que tenga menor varianza por ser más eficiente. Sin embargo, si ocurriera que el estimador con menor varianza no fuera insesgado, y el estimador de mayor varianza si lo fuera se preferirá aquel que tenga menor error cuadrático medio.*

1. Según lo que nos plantea el Teorema Central del Límite es posible ver que el promedio muestral de una variable *x* se puede representar con cualquier distribución, pero siempre y cuando la variable *x* se distribuya normal.

**RESPUESTA**:

*Falso, es al revés, es decir: para una variable aleatoria x con cualquier distribución su promedio muestral tendrá una distribución normal, con tamaño muestral suficientemente grande. Por ejemplo:*

*x*

1. **ANOVA (20pts.)**

En un experimento se compararon tres métodos de enseñar un idioma extranjero; para evaluar la instrucción, se administró una prueba de vocabulario de 50 preguntas a los 24 estudiantes del experimento repartidos de a ocho estudiantes por grupo y se midió el número de respuestas correctas. Con esta información sobre el experimento, responda lo siguiente:

* 1. (2pts.) Identifique claramente cuál es la variable dependiente y la variable independiente del análisis.

**RESPUESTA**:

*La variable dependiente es el número de respuestas correctas en la prueba.*

*La variable independiente es cada uno de los tres métodos de enseñanza del idioma extranjero.*

* 1. (3pts.) Plantee la hipótesis nula e hipótesis alternativa, señalando claramente los indicadores involucrados en este experimento.

**RESPUESTA**:

*Hipótesis planteadas sin señalar a qué promedio se refieren, o a qué grupo corresponden, tendrán puntaje incompleto.*

* 1. (10pts.) Complete la siguiente tabla ANOVA (cifras en español, es decir, coma es separador de decimales y punto es separador de miles).

**RESPUESTA**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Suma de Cuadrados (SC) | Grados de Libertad | Media Cuadrática  (CM) | F | Sig.  *p-value* |
| Tratados  (Inter - grupos) | 647,584  (2 puntos) | 2  (1 punto) | 323,792 | 8.360  (2 puntos) | 0,002 |
| Residuos  (Entre-grupos) | 813,374  (2 puntos) | 21  (1 punto) | 38.732  (1 punto) |
| Total | 1.460,958 | 23  (1 punto) |  |

*Procedimiento para completar la tabla:*

*1) Calcular los grados de libertad, en el total son n-1 y n=24, por lo tanto son 23. Los grupos a comparar son 3 por lo tanto los gl Inter son 2, y se verifica que (2+21) son los 23 del total.*

*2) La suma de cuadrados Inter se obtiene multiplicando la media cuadrática por los gl, i.e. 323.792\*2=647.584*

*3) Teniendo la SC Inter, se obtiene la SC Intra restando 1460.958-647.584=813.374 4) Con la SC Intra y los gl se calcula la media cuadrática Intra =813.374/21=38.732 5) Por último con las dos MC se calcula el test F=323.792/38.732=8.360*

* 1. (5pts.) Realice la prueba de hipótesis al 95% y entregue las conclusiones (use el método que prefiera: p-value o valor crítico, pero explique claramente cómo llega a la conclusión).

**RESPUESTA**:

*Se rechaza la hipótesis nula de que todas las medias son iguales. Dar alguna de las dos siguientes explicaciones:*

*Si utiliza método de valor crítico: el valor del test F es mayor al valor crítico correspondiente a 3,467*

*Si utiliza el método de p\_value: la significancia estimada es menor al nivel del p-value*

1. **Intervalos de Confianza (20pts.)**

Suponga que está analizando los resultados del curso de Inferencia Estadística de la Universidad UDP y otra universidad, para ello ha realizado dos muestras (una en cada universidad) de la cual ha obtenido los siguientes resultados sobres las notas finales del curso:

|  |  |  |  |
| --- | --- | --- | --- |
| **Universidad** |  |  |  |
| UDP | 4,8 | 0,7 | 25 |
| Otra universidad | 4,5 | 0,9 | 27 |

* 1. (6pts.) Calcule el intervalo de confianza para la media poblacional de los alumnos de la UDP, a un 99% de confianza.

**RESPUESTA**:

*El intervalo de confianza para la media con desviación estándar desconocida es:*

*El valor t al* 99% *de confianza lo buscamos en la tabla y corresponde a:*

*Por lo tanto:*

* 1. (7pts.) Calcule el intervalo de confianza para la diferencia de medias poblacionales al 90% de confianza. Interprete su resultado.

**RESPUESTA**:

*Los grados de libertad corresponden (redondeado hacia abajo) a:*

*El valor t al* 90% *de confianza lo buscamos en la tabla y corresponde a:*

*El intervalo de confianza para la media con desviación estándar desconocida es:*

*Este resultado quiere decir que con un* 90% *de confianza el promedio de los alumnos de la UDP va desde* 0,0735 *inferior a alumnos de otras universidades a ser mayor en* 0,6735*.*

* 1. (7pts.) Respectivamente, calcule el intervalo de confianza para las varianzas poblacionales al 95% de confianza. Explique a qué se debe la diferencia o similitud de los resultados.

**RESPUESTA**:

*El intervalo de confianza para la media con desviación estándar desconocida es:*

*Por lo tanto, los valores críticos los buscamos en la tabla ji-cuadrado, por lo que para los alumnos UDP corresponderán a:*

*Para los alumnos de otra universidad:*

*Por lo tanto los intervalos para los alumnos UDP y de otra universidad corresponden respectivamente a:*

*En primer lugar se observa que el intervalo para la varianza de los alumnos de otra universidad tiene valores mayores (mínimo y máximo), es decir, necesariamente al mismo nivel de confianza es muy posible que la varianza de los alumnos de otra universidad sea mayor, además, el intervalo es más extenso. La diferencia se debe a todos los factores que implican el cálculo del intervalo (grados de libertad, varianza y valor crítico) pero el más influyente es la varianza de la muestra la cual es mayor en los alumnos de otra universidad.*

1. **Contraste de Hipótesis (20pts.)**

Considerando los datos de la pregunta anterior, conteste las siguientes preguntas:

* 1. (10pts.) Utilizando el método del *p-value*. Testee la hipótesis nula de que la media poblacional de la nota de los alumnos de la UDP menor o igual a 4,51105.

**RESPUESTA**:

*El test que se quiere realizar es:*

*El estadístico corresponde a:*

*Por lo tanto el p-value corresponde a la probabilidad en la tabla t de ser mayor al valor del estadístico, es decir:*

*Este valor se puede buscar en la tabla t con* 24 *grados de libertad y corresponde a:*

*Por lo tanto se rechaza la hipótesis nula a un* 10% *y* 5% *de significancia.*

* 1. (10pts.) Considerando que la desviación estándar poblacional de los alumnos de la otra universidad es de , a un 5% de significancia: ¿es la desviación estándar poblacional de la nota de los alumnos de la UDP menor que la desviación estándar poblacional de la nota de los alumnos de otra universidad?

**RESPUESTA**:

*El test que se quiere realizar es:*

*Para poder resolver, verificaremos el test equivalente:*

*El estadístico corresponde a:*

*El valor crítico lo buscamos en la tabla ji-cuadrado con* 24 *grados de libertad, el que corresponde a:*

*Por lo tanto, como en este caso, es un test de una cola y es de cola inferior, y ocurrió que el estadístico es mayor que el valor crítico, es decir, el estadístico se encuentra en la zona de no rechazo, es decir, no existe evidencia estadística suficiente para, al* 5% *de significancia, rechazar la hipótesis nula de que la varianza poblacional UDP es mayor o igual a* 0,81*, lo que significa que no existe evidencia estadística suficiente para, al* 5% *de significancia, rechazar la hipótesis nula de que la desviación estándar poblacional UDP es mayor o igual a* 0,9*.- Por lo que dada la evidencia estadística contestaríamos negativamente a la pregunta planteada.*

**Formulario**:

; ;























