**Segunda Tarea - Estadísticas para la Gestión II**

**Felipe Herrera, Hernán Romo y David Toro**

**Integrantes: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Instrucciones:**

* Puede ser realizada por un máximo de 3 personas, de distintas secciones.
* Cada respuesta debe estar con el desarrollo correspondiente, puede trabajar en Excel o STATA (pegar comandos en el documento a entregar)
* Si se llega a detectar evidencia de copia entre dos o más trabajos, será evaluado con nota 1.0
* **Plazo de entrega: lunes 17 de diciembre. Debe entregarlo impreso en horario de cátedra.**

**1. Responda V o F. Justifique las falsas.**

a) \_\_\_V\_\_\_ El nivel de significancia representa la proporción de datos que no se encuentran comprendidos en el intervalo de confianza establecido.

b) \_\_\_F\_\_\_ Al confeccionar un intervalo de confianza estamos buscando estimar los valores posibles del estadístico, a partir de un parámetro.

Se estima los valores posibles del parámetro a partir del estadístico.

c) \_\_\_F\_\_\_ A medida que disminuye el nivel de confianza, se incrementa la probabilidad de rechazar la hipótesis nula cuando ésta no debía ser rechazada.

Si se reduce el nivel de confianza, el nivel de significancia sería más alto, por lo que la probabilidad de cometer el error tipo II será mayor.

d) \_\_\_F\_\_\_ Cuando se rechaza la hipótesis nula, se acepta la hipótesis alternativa como irrefutable.

En test de hipótesis no existe lo irrefutable. Siempre son estimaciones, por lo que nunca se “acepta” como verdadera una hipótesis, pero sí puede “rechazarse” o “no rechazarse”.

e) \_\_\_F\_\_\_ En un intervalo de confianza del 95% para diferencia de proporciones, existe una probabilidad de 0,05 de que la diferencia entre ambas muestras sea igual a cero o menos.

La probabilidad de que esto ocurra es del 0,5 (o 50%) pues es el valor que se ubica al centro en un intervalo de confianza para diferencia de proporciones.

**2. Un investigador está interesado en determinar la proporción de población indígena de un territorio determinado. Para ello dispone de un listado con los nombres y direcciones de todas las personas que residen en dicho territorio. Para hacer la estimación el investigador toma una muestra aleatoria de 100 individuos y en ella obtiene que un 25% de la población pertenece a pueblos originarios.**

1. Sobre la base del valor estimado para la muestra ¿Qué se puede afirmar sobre el verdadero valor del parámetro en la población estudiada? (utilice un 95% de Nivel de confianza).



Sobre la base de la evidencia muestral, se puede inferir que, con un nivel de confianza del 95%, el verdadero valor del parámetro en la población estimada se encuentra entre 16,5% y 33,4%

1. Posterior a la realización de la encuesta, el investigador descubre que en dicho territorio se había efectuado 5 años antes un censo en donde se determinó que la población indígena en ese momento ascendía a 30%. Sobre la base de la evidencia muestral, ¿se puede afirmar que la proporción indígena ha aumentado significativamente? Plantee las hipótesis nula y alternativa. Efectúe sus cálculos con 90% de Nivel de confianza.



HO=30%

H1>30%

\*Dado que el valor Pr(Z>z)=0,86 es mayor (en valor absoluto) que el nivel de significación (0,10), entonces la probabilidad de cometer un error tipo I (rechazar la hipótesis nula cuando no debería ser rechazada) es más alta que el 10% y por lo tanto no se rechaza la hipótesis nula, o dicho de otra forma, se debe rechazar la alternativa. Por lo que, dada la evidencia muestral, no se puede afirmar que la proporción indígena ha aumentado significativamente, a un nivel de confianza de 90%.

**3. El jefe del Departamento de Desarrollo Regional de la SUBDERE le pide a usted revisar el ingreso per cápita de Chile. Para ello dispone los siguientes datos para el año 2016.**

|  |  |  |
| --- | --- | --- |
| **Región** | **Ingreso 2016** | **Habitantes** |
| XV De Arica y Parinacota | $ 1.100.000 | 226.068 |
| I De Tarapacá | $ 2.343.000 | 330.558 |
| II De Antofagasta | $ 12.434.000 | 607.534 |
| III De Atacama | $ 3.465.000 | 286.168 |
| IV De Coquimbo | $ 4.100.000 | 757.586 |
| V De Valparaíso | $ 13.455.000 | 1.815.902 |
| RMS Región Metropolitana de Santiago | $ 64.034.000 | 7.112.808 |
| VI Del Libertador General Bernardo O’Higgins | $ 6.556.000 | 914.555 |
| VII Del Maule | $ 4.596.000 | 1.044.950 |
| VIII Del Biobío | $ 9.879.000 | 2.037.414 |
| IX De La Araucanía | $ 4.567.000 | 957.224 |
| XIV De Los Ríos | $ 2.256.000 | 384.837 |
| X De Los Lagos | $ 4.567.000 | 828.708 |
| XI Aysén del General Carlos Ibáñez del Campo | $ 1.116.000 | 103.158 |
| XII De Magallanes y de la Antártica Chilena | $ 1.754.000 | 166.533 |
| (1) Ingreso anual en millones de pesos |  |  |

1. ¿Cuál es el ingreso per cápita de Chile?

**$ 7,75 millones de pesos aprox**

Posteriormente, le solicitan evaluar una política pública implementada en el año 2017 cuyo objetivo era incrementar el ingreso de las 5 regiones con menor nivel de este parámetro. Dispone de los siguientes datos muestrales, los cuales indican el porcentaje de aumento del ingreso per cápita con respecto a la medición poblacional

|  |  |  |
| --- | --- | --- |
| **Región** | **% de aumento del ingreso** | **Encuestados** |
| XV De Arica y Parinacota | 15% | 5 |
| IV De Coquimbo | 20% | 7 |
| VII Del Maule | 16% | 10 |
| VIII Del Biobío | 14% | 20 |
| IX De La Araucanía | 13% | 12 |

1. ¿Cuál sería el ingreso per cápita del conjunto de las regiones antes de la intervención y después de la intervención, suponiendo que la población no ha variado?

**Antes=$4,82 aprox.**

**Después=$5,56 millones aprox.**

1. Genere un intervalo de confianza para la muestra (una vez intervenidas) con un 90% de NC

n=54



1. Si se posee una desviación muestral de 2 millones, esta política habrá aumentado significativamente el ingreso per cápita del conjunto de las regiones intervenidas? (85% NC)

H0: Ingreso per cápita=4,82 millones

H1: Ingreso per cápita>4,82 millones



\*Dado que el valor Pr(T>t)=0,004 es menor (en valor absoluto) que el nivel de significación (0,15), la probabilidad de cometer un error tipo I (rechazar la hipótesis nula cuando no debería ser rechazada) es menor que el 15%, por lo tanto, se rechaza la hipótesis nula, o dicho de otra forma, no se debe rechazar la alternativa. Por lo que, dada la evidencia muestral, se podría afirmar que el ingreso per cápita del conjunto de las regiones intervenidas ha aumentado significativamente.

**4. Considerando la base de datos Casen 2015 sobre las comunas Lumaco y Purén:**

|  |  |
| --- | --- |
| **Variable** | **Codificación** |
| comuna | 9207 = Lumaco  9208 = Purén |
| zona | 0 = Rural  1 = Urbana |
| sexo | 0 = Hombre  1 = Mujer |
| pobreza | 1 = Pobres extremos  2 = Pobres no extremos  3 = No pobres |
| Grupo etario | 1= Niño (1-14 años)  2= Joven (15-29 años)  3= Adulto (30-59 años)  4= Adulto Mayor(60 o más años) |

**a) Genere un intervalo de confianza para la proporción de mujeres de cada comunas (90%NC)**



**b) Genere un intervalo de confianza para la proporción de hombres de cada comuna (95%NC)**

Crear variable sexoh donde hombre sea 1 y mujer 0



**c) Se puede afirmar que la proporción de población urbana en ambas comunas es menor al 50% (80% NC).**

H0: Proporción urbana ambas comunas=50%

H1: Proporción urbana ambas comunas<50%

Dado que el z calculado=-1.8187 es mayor (en términos de valor absoluto) que el z de tabla= 0.8416 (por lo que cae en la zona de rechazo de la hipótesis nula), se debe rechazar la hipótesis nula, o en otras palabras no se debe rechazar la hipótesis alternativa [Método 1]

Dado que el valor Pr(Z<z)=0.0345 es menor que el nivel de significación (0.20), la probabilidad de cometer un error tipo I es menor que el 20% por lo que se debe rechazar la hipótesis nula, o en otras palabras no se debe rechazar la hipótesis alternativa [Método 2]

Por lo anterior, existe evidencia estadísticamente significativa para afirmar, con un 80% de nivel de confianza, que la proporción de población urbana en ambas comunas es menor al 50%

**d) ¿Es posible afirmar que existen diferencias significativas entre la proporción de población mayor a 30 años entre ambas comunas? (99% NC)**

H0: Población mayor de 30 Lumaco= Población mayor de 30 Purén

H1: Población mayor de 30 Lumaco=/=Población mayor de 30 Purén (DOS COLAS)

Generar variable “mayores” donde los adultos y adultos mayores sean “1” y niños y jóvenes “0”

** **

Dado que el z calculado=-1.8369 es menor (en términos de valor absoluto) que el z de tabla= -2.5785 (por lo que cae en la zona de no rechazo de la hipótesis nula), no se debe rechazar la hipótesis nula, o en otras palabras se debe rechazar la hipótesis alternativa [Método 1]

Dado que el valor Pr(Z<z)=0.0662 es mayor que el nivel de significación (0.01), la probabilidad de cometer un error tipo I es mayor que el 1% por lo que no se debe rechazar la hipótesis nula, o en otras palabras, se debe rechazar la hipótesis alternativa [Método 2]

Por lo anterior, no existe evidencia estadísticamente significativa que permita afirmar, con un 99% de nivel de confianza, que la proporción de población mayor a 30 años es distinta entre ambas comunas

**e) Genere un intervalo de confianza para la proporción de pobres extremos para ambas comunas y luego uno para la misma proporción en cada comuna (95% NC)**

**Generar variable “pobrezaextrema”**

Ambas comunas:



Por comuna:



**f) Los datos de la Casen 2012 indican que la proporción de pobres extremos en ambas comunas era de 24% dicho año, ¿Ha disminuido de manera significativa la proporción de pobres extremos en Lumaco y Purén en comparación al año 2012? (93% NC)**

H0: La proporción de pobres extremos se ha mantenido (=24%)

H1: La proporción de pobres extremos ha disminuido (<24%)

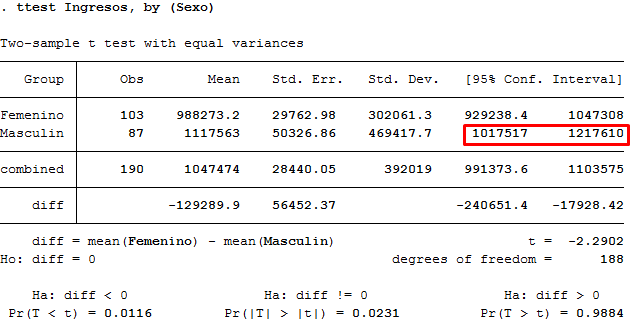
 

Dado que el z calculado=-2.3926 es mayor (en términos de valor absoluto) que el z de tabla= -1.6448 (por lo que cae en la zona de rechazo de la hipótesis nula), se debe rechazar la hipótesis nula, o en otras palabras no se debe rechazar la hipótesis alternativa [Método 1]

Dado que el valor Pr(Z<z)=0.0084 es menor que el nivel de significación (0.05), la probabilidad de cometer un error tipo I es menor que el 5% por lo que se debe rechazar la hipótesis nula, o en otras palabras no se debe rechazar la hipótesis alternativa [Método 2]

Por lo anterior, existe evidencia estadísticamente significativa para afirmar, con un 95% de nivel de confianza, que la proporción de pobres extremos ha disminuido en comparación al año 2012.

**4. Utilizando una muestra de estudiantes egresados de Administración Pública, un investigador quiso verificar estadísticamente si los hombres tienen ingresos mayores a las mujeres. Utilizando una base de datos, realizó los comandos correspondientes en Stata y obtuvo el siguiente output:**



Analice el output y responda:

1. ¿Se puede afirmar, con un 99% de NC que los hombres ganan, en promedio, más que las mujeres? Justifique sus cálculos.



Si los hombres ganan más que las mujeres, la hipótesis alternativa que se elige es diff<0. Dado que el valor Pr(T<t)=0.011 es mayor que el nivel de significación (0.01), la probabilidad de cometer un error tipo I es mayor que el 1% por lo que no se debe rechazar la hipótesis nula, o en otras palabras, se debe rechazar la hipótesis alternativa. Es decir, **no** se puede afirmar con un 99% de nivel de confianza, que los hombres ganan, en promedio, más que las mujeres.

1. ¿Se puede afirmar, con un 95% de NC que las mujeres ganan, en promedio, menos que los hombres? Justifique sus cálculos.



Si las mujeres ganan menos que los hombres, la hipótesis alternativa que se elige es diff<0. Dado que el valor Pr(T<t)=0.011 es menor que el nivel de significación (0.05), la probabilidad de cometer un error tipo I es menor que el 5% por lo que se debe rechazar la hipótesis nula, o en otras palabras, no se debe rechazar la hipótesis alternativa. Es decir, **sí** se puede afirmar, que las mujeres ganan, en promedio, menos que los hombres.

1. ¿Cómo podría interpretarse el intervalo que está en el recuadro rojo?

Dada la evidencia muestral, podemos afirmar con un 95% de NC que el verdadero valor para de la media poblacional del Ingreso de los hombres, debe rondar entre $1017517 y $1217610

1. ¿Cuál es la probabilidad de seleccionar una mujer al azar y que ésta gane menos de $950.000? Justifique sus cálculos.

|  |  |
| --- | --- |
| mu | 988273,2 |
| x | 950000 |
| s | 302061,3 |
| n | 103 |
| t | -1,28593288 |



La probabilidad de que una mujer escogida al azar tenga un ingreso de 950.000 o menos es de 0,1.

CUALQUIERA DE LAS DOS. LA DE ARRIBA ES SÓLO CON MEDIAS, Y LA DE ABAJO CON DIF DE MEDIAS, DAN DISTINTO.

|  |  |
| --- | --- |
| **Muestreo de Diferencia de Medias** |  |
| **Promedio muestral A** | 988273,2 |
| **Promedio muestral B** | 1117563 |
| **Promedio pobl. A** | 0 |
| **Promedio pobl. B** | 0 |
| **Numero Muestra A** | 103 |
| **Numero Muestra B** | 87 |
| **Desv. Esta. Muestral A** | 302061,3 |
| **Desv. Esta. Muestral B** | 469417,7 |
| **Grados de Libertad** | 188 |
| **Valor Sp** | 387689,1507 |
| **Valor T** | -2,290 |

