****

**Universidad de Chile**

Facultad de Ciencias Sociales

Profesora: Catalina Canals

Ayudantes: Daniel Cifuentes, Camila Lizama,

Daniela Olivares, Sofía Villalobos

**Pauta Examen de Segunda Oportunidad**

 **Estadística II, sección 2**

**Ejercicios**

1. **Recomendación de test a utilizar (Total: 1,8 pts. 0,6 pts. c/u)**

**En los siguientes casos: (1) identifique la hipótesis de investigación, (2) indique si tiene toda la información necesaria sobre los datos para decidir qué test utilizar, (3.a) si tiene toda la información necesaria sobre los datos para decidir qué test utilizar: sugiera qué test de hipótesis utilizar y por qué, indique si se cumple cada uno de los supuestos del test sugerido (si no se cumplen todos los supuestos, explique por qué recomienda utilizarlo a pesar de ello), y plantee las hipótesis nula y alternativa del test a realizar, 3.b) si NO tiene toda la información necesaria sobre los datos para decidir qué test utilizar, indique qué información requeriría para saber qué test recomendar, sugiera cómo podría obtenerla, e indique qué test recomendaría –y cuáles serían la hipótesis nula y la alternativa-en cada caso posible.**

1. **A través de una encuesta realizada a muestra probabilística, unos investigadores están interesados en estimar el promedio de horas que las personas dedican al trabajo doméstico en Chile y analizar si el promedio/mediana de horas supera las 8 hrs. semanales. Se disponen de los siguientes datos de la muestra:**

|  |
| --- |
| **Datos Muestra (n=25)** |
| **Valor P en prueba Kolmogorov-Smirnov de normalidad del promedio de horas semanales trabajadas** | **0,02** |

* Hipótesis de investigación: el promedio de horas que las personas dedican al trabajo doméstico en Chile y analizar si el promedio/mediana de horas supera las 8 hrs. **(0,1 pts.)**
* Elección del test: Test T de medias de 1 cola, dado que se trata de una variable cuantitativa (por ende usamos medias), la hipótesis de investigación tiene una direccionalidad clara (por eso es de una cola) y se cumplen los supuestos necesarios (notar que no se cumplen todos necesariamente, pero si los requeridos para aplicar el test en este caso) **(0,2 pts.).**
* Supuestos: muestra probabilística (se cumple por enunciado), variable cuantitativa (se cumple dado que las hrs. Refiere a una variable cuantitativa) y distribución normal. Este último supuesto no sabemos si se cumple (requeriríamos haber hecho el test shapiro-wilks), pero con n superior a 15, el test T es robusto a no normalidad **(0,2 pts.)**
* H0: el promedio de hrs. de trabajo doméstico es Chile =8, H1: el promedio de hrs. De trabajo en Chile >8. **(0,1 pts.)**
1. Unos investigadores quieren comparar si la proporción de desempleados en Chile aumentó entre el 2015 y el 2016, a través de una muestra probabilística de 700 casos del año 2015 y 1000 casos del 2016. Se disponen de los siguientes datos de la muestra:

|  |
| --- |
| Datos Muestra (n=1700) |
| Valor P en prueba Kolmogorov-Smirnov de igualdad de distribuciones | 0,04 |

* Hipótesis de investigación: la proporción de desempleados en Chile aumentó entre el 2015 y el 2016 **(0,1 pts.)**
* Elección del test: Test Z de proporciones para muestras independientes, de 1 cola. Se elige el Test Z porque se cumplen sus supuestos, es para muestras independientes ya que los individuos muestreados no son los mismos en ambos años, de 1 cola porque la hipótesis de investigación tiene direccionalidad clara, y de proporciones porque estar desempleado o no es una variable categórica **(0,2 pts.).**
* Supuestos: muestra probabilística (se cumple por enunciado), variable cualitativa (desempleado) y n grande (700 y 1000 son grandes). **(0,2 pts.)**
* H0: la proporción de desempleados en Chile se mantuvo igual entre el 2015 y el 2016, H1: la proporción de desempleados en Chile aumentó entre el 2015 y el 2016. **(0,1 pts.)**
1. A través de una muestra probabilística de 40 casos unos investigadores quieren analizar si las proporciones de hombres y de mujeres a favor de realizar una Asamblea Constituyente (AC) en Chile son iguales. Se disponen de los siguientes datos de la muestra:

|  |
| --- |
| **Tabla de Frecuencias esperadas en caso de H0 cierta** |
|  | Hombres | Mujeres |
| A Favor de AC | 10 | 10 |
| Contra AC | 10 | 10 |

* Hipótesis de investigación: la proporción de hombres y mujeres a favor de AC es igual **(0,1 pts.)**
* Elección del test: Chi cuadrado de asociación, dado que n es pequeño no podemos hacer test Z para proporciones, pero se cumplen los supuestos para usar éste test. **(0,2 pts.).**
* Supuestos: muestra probabilística (se cumple por enunciado), variable cualitativa (sexo y a favor/contra AC) y no más de un 20% de casillas con FE<5 (se cumple porque no hay ninguna). **(0,2 pts.)**
* H0: no hay asociación entre el sexto y estar a favor/contra la AC; H1: hay asociación entre el sexto y estar a favor/contra la AC. **(0,1 pts.)**
1. **Verdadero (V) o Falso (F) (Total: 2,8 pts., 0,4 pts. c/u)**

**En los siguientes casos, indique si las afirmaciones son verdaderas o falsas. En caso de ser falsas indique porqué lo son, e indique qué sería lo correcto. En caso de ser verdaderas, explique por qué lo son, agregue información que complemente la frase y/o explique qué quiere decir la afirmación. Sugerencia: sea breve, con 4 líneas por respuesta suele ser suficiente.**

1. **Si con una muestra probabilística de 10 estudiantes de educación escolar de Chile, se estima, con 95% de confianza, un intervalo [5%,10%] para la proporción que repite de curso, esto quiere decir que la probabilidad de que el parámetro poblacional (porcentaje de estudiantes de educación escolar en Chile que repite) se encuentre entre 5% y 10% corresponde a 95%.**

Falso. Si el n es tan pequeño, no podemos asegurar que se cumple el TCL ni la LGN tal que la fórmula del intervalo de confianza sea válida.

1. **El Test Z para proporciones sirve para evaluar la igualdad de proporciones en muestras relacionadas.**

Falso. Sirve solo para muestras independientes. El Test de McNemar sirve para test de proporciones con muestras relacionadas.

1. **El Test de Mc Nemar no permite poner a prueba hipótesis de una cola.**

Falso. Permite test de una cola, pero solo cuando n es grande.

1. **A una muestra de 11 estudiantes del Instituto Nacional se le realiza una encuesta en 8° básico, y otra cuando van en I Medio. Luego, al realizar un Test T para medias para comparar el promedio de notas los estudiantes del Instituto Nacional en 8° básico y I Medio, y con la hipótesis alternativa que plantea que ambos promedios son diferentes, si el estadístico T =-3 NO se rechaza la hipótesis nula que plantea que ambos promedios son iguales en la población.**

Falso. Considerando un 95% de confianza, T critico (con 10 gl) corresponde a 1,8. Dado que -3<-1,8 y el test es de dos colas, se rechaza H0.

1. **Al trabajar con una muestra pequeña (ej. 25 casos), si el valor P es inferior a la significancia, podemos estar muy seguros de que nuestra hipótesis nula es falsa.**

Verdadero. Cuando n es pequeño es muy improbable que, aun siendo H0 falsa, podamos rechazarla, por ende, si el Valor P < la significancia, habiendo un n pequeño, H0 podemos estar seguros de su falsedad siempre y cuando el muestreo haya sido adecuado.

1. **La correlación de Pearson sirve para analizar si hay cualquier tipo de relación entre dos variables cuantitativas, mientras que la regresión lineal sirve solo cuando existe relación lineal entre variables.**

Falso. R de Pearson analiza si existe relación lineal entre variables.

1. **Si al realizar un test binomial (en un caso donde se cumplen todos los supuestos) para poner a prueba la hipótesis nula “la proporción de futbolistas chilenos de nivel socioeconómico alto es igual al 10%”, versus la hipótesis alternativa de que “la proporción de futbolistas chilenos de nivel socioeconómico alto es distinta a 10%”, si el Valor P=0,45 rechazamos la hipótesis nula.**

Falso, usando los niveles de confianza utilizados frecuentemente (significancias entre 0,1 y 0,01), NO rechazamos H0 dado que el Valor P> significancia.

**Parte 3: Regresión lineal simple (Total: 1,4 pts.)**

**Un grupo de investigadores realiza un estudio para predecir las horas dedicadas al trabajo doméstico en los chilenos, a partir de la variable edad, obteniendo los siguientes resultados.**

**Observaciones: 1000**

**R cuadrado: 0,08**

**Prob >F: 0,001**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coeficiente** | **P>|t|** |
| **Edad** | **1,1** | **0.01** |
| **Constante** | **0.3** | **0.1** |

1. **Refiérase al ajuste del modelo (interprete el R cuadrado y el test asociado al estadístico F) (0,4 pts.)**

El R cuadrado indica que la edad explica el 8% de la varianza de las hrs. Dedicadas al trabajo doméstico **(0,2 pts),** mientras que el Valor P asociado al estadístico F=0,001, indica que, con 95% de confianza, se rechaza H0 que establece que el efecto de la edad en tal variable es nulo **(0,2 pts.).** Dado lo anterior el ajuste del modelo es bajo (no explica mucho de la varianza), pero tampoco es nulo, en tanto la edad afecta el índice.

1. **Interprete estadísticamente el coeficiente asociado a la variable edad estimado por su modelo. Interprete sociológicamente su resultado (1 pts.).**

Al aumentar en 1 año la edad, las hrs. Dedicadas al trabajo doméstico aumentan en 1,1 (es decir 1 hrs. Y 6 minutos) (0,4 pts.).

Proponer una hipótesis razonable de por qué esto ocurre (ej. En la sociedad chilena los niños suelen participar menos que los adultos en el trabajo doméstico, y los adultos mayores, jubilados pero aún no incapacitados, al no tener actividad laboral tienden a aportar más a estas actividades)

1. **Bonus: ¿Por qué cree que su modelo no explica toda la varianza de las hrs. Dedicadas al trabajo doméstico? (0,3 pts.)**

Proponer que la relación edad/trabajo doméstico no es lineal, y/o proponer otras variables que influyen en el trabajo doméstico, sugiriendo en qué sentido lo hacen (0,3).