# **UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN**

**ESCUELA DE ESTADÍSTICA**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SEGUNDO TRABAJO** | **Pruebas de Hipótesis (10%)** | **SEM. 02 – 2021** | **Cédula** | **1000098328** |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre Completo | Daniel Muñoz Ruiz | D. I. | 1000098328 |
| Nombre Completo | Juan Felipe Murillo Ramírez | D. I. | 1000757193 |

La base de datos con la que le corresponde trabajar, se obtiene como una muestra aleatoria de una gran base de datos. La base original corresponde a la información de 2000 estudiantes de cierta Universidad. Dicha base contiene las variables: **MASA** (en kg), **ESTATURA** (en cm), **GASTO** (semanal en pesos), **GENERO** (Hombre, Mujer) y **CIUDAD** (de procedencia: ARMENIA, BARRANQUILLA, BOGOTA, CALI, CARTAGENA, MEDELLIN).

Cada pareja debe generar una base de datos, que corresponderá a una muestra de tamaño 120 de la base original. Los datos originales están en el archivo: “base.txt”, el cual está disponible en Moodle, pestaña Trabajos – Trabajo 2. Para hacerlo debe seguir los siguientes pasos:

# Para leer el archivo de datos, darle click al archivo “base.txt”, aparecerá una nueva pestaña. Copie el contenido y guárdelo en un block de notas, con el mismo nombre: “base.txt”. En R ejecute el siguiente código:

**uno <- read.table(file.choose(), header=T)**

Se habilitará un explorador y se escoge el archivo “base.txt”

Copiar el siguiente código en R, sin modificar ninguna de sus líneas

**genera <- function(cedula){**

**set.seed(cedula)**

**data <- uno[sample(1:2000,120),]**

**data**

**}**

# Para crear la base de datos con la cual trabajara, debe ejecutar las siguientes líneas:

**datos <- genera(cedula)**

**datos[,4] <- as.factor(datos[,4])**

**datos[,5] <- as.factor(datos[,5])**

# cedula representa el número de cedula de uno de los dos integrantes. En el trabajo debe indicar cuál número de cédula usó.

Ingrese a R Commander y cargue el archivo “datos”.

Con la base ya creada debe resolver las siguientes preguntas.

1. (**20 pts**.) Se tiene la creencia de que la **ESTATURA** media de los estudiantes es inferior a 170 cms. ¿Es esto cierto? Justifique su respuesta.
2. (**30 pts**.) ¿Puede afirmarse que la **MASA** promedio de los Hombres es superior a la de las Mujeres? Justifique su respuesta
3. (**20 pts**.) ¿Se puede afirmar que el porcentaje de Hombres es superior al 52%? Justifique su respuesta.
4. (**30 pts**.) Un investigador afirma que la proporción de estudiantes que proviene de cada una de las 6 ciudades es la misma. ¿Es cierta esta afirmación?

**Consideraciones sobre el trabajo.**

1. Para la solución de cada pregunta se deben plantear todas las hipótesis de interés, definir claramente los parámetros a ser utilizados, probar los supuestos que se requieran, definir claramente sus Estadísticos de prueba y llegar a una conclusión usando **SOLO** el valor P.
2. El trabajo debe contener máximo 7 páginas incluyendo estas dos páginas de instrucciones.
3. La solución a las preguntas debe hacerse **SOLO** en R (con comandos o usando R-Commander). En este caso debe aportar la evidencia de los resultados obtenidos en R.
4. Debe preparar un documento con la solución a estas preguntas, con los respectivos soportes de R, en formato **pdf**, debe ser cargado en la carpeta: “Soporte Trabajo 2”, que aparece en el curso en Moodle, en la pestaña “Trabajos”, pestaña “Trabajo 2”, **antes de las** **6:00 p.m.**, **del 5 de febrero de 2022**.

Solución.

1. Primero debemos verificar si los datos de “Estatura” se distribuyen de forma normal. Para esto usamos la prueba de shapiro-wilks con las siguientes hipótesis:

X: Estatura

H0: XN(µ, σ2). H1: X N(µ, σ2).

Texto

Descripción generada automáticamente

Dado a que tenemos un Vp demasiado grande, se concluye entonces que X se distribuye normalmente.

Ahora, para determinar si la media es mayor a 170cm usamos las siguientes hipótesis:

H0: µ≥170 H1: µ<170

Se usará Tc como

Tenemos los siguientes datos de “estatura”:

Texto

Descripción generada automáticamente con confianza media

Tc =

Tc =

Para determinar nuestra prueba de hipótesis usaremos el valor P, el cual nos arroja como resultado:



Por lo tanto, como el valor P es grande no hay suficiente evidencia para rechazar H0, por lo que se termina concluyendo que la media poblacional de la estatura de la muestra es mayor o igual a 170cm.

1. Para determinar la diferencia de masas entre hombres y mujeres primero debemos determinar si la distribución de masas es normal y luego se determina la variabilidad entre masa de hombres y de mujeres. Entonces se tiene que:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Como podemos observar, la masa de los hombres no se distribuye de forma normal, mientras que la masa de mujeres si. Entonces, dado que no se distribuye normal y que no conocemos sus varianzas, las hipótesis a tratar y el estadístico de prueba son:

X: masa promedio de hombres.

Y: masa promedio de mujeres.

Texto, Carta

Descripción generada automáticamente

|  |  |  |  |
| --- | --- | --- | --- |
| X | Valor | Y | Valor |
|  | 64.48 |  | 62.67 |
|  | 11.96296 |  | 10.8827 |
| n | 54 | m | 66 |

H0: H1:



Teniendo en cuenta que el valor P es grande, no hay suficiente información para rechazar H0, por ende, la masa promedio de hombres es menor o igual a la de mujeres.

1. Nos piden determinar una proporción, por lo que usaremos un estadístico de prueba de la forma:

Siendo entonces:





Como el valor P es muy grande, se concluye que no hay suficiente información para rechazar H0, por ende, es falso afirmar que el porcentaje de hombres es superior al 52%.

1. Tenemos los siguientes datos:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ciudad | Armenia | Barranquilla | Bogotá | Cali | Cartagena | Medellín |
| Frecuencia  Observada | 21 | 22 | 20 | 21 | 18 | 18 |
| Pi | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |
| Frecuencia  esperada | 20 | 20 | 20 | 20 | 20 | 20 |

Probaremos las siguientes hipótesis:

Se calcula entonces un Xc para una muestra de Chi-cuadrado como:

El valor P para este problema es



Por lo tanto, al tener un valor P grande, no se puede rechazar H0, entonces se concluye que la proporción de estudiantes de cada ciudad es la misma.