

Proyecto 2020
ASIGNATURA LIC MI CARRERA
Orientaciones metodológicas:
Este es el tema de mi clase
Estudiante Pérez Pérez *

Ejercicios

Ejercicio 1 Genere una Población normal de tamaño 500, seleccione 8 muestras de tamaños varios (Muy mayor que 30, mayor que 30, 30, 20), 4 muestras con remplazo y 4 sin remplazo.

- a. Calcule para cada una de las muestras los Estadísticos Descriptivos, de la Conferencia 1.
- b. Calcúlelos en la población inicial. Analice las diferencias.
- c. Grafique los resultados.
- d. Para cada muestra calcule los intervalos de confianza para la media y la varianza.
- e. Analice las diferencias en los resultados de las muestras de tamaños similares.

Propuesta de distintos ejercicios de la clase, para desarrollar las habilidades a crear durante la clase.

Ejercicio 2 Aquí va orden del ejercicio 2.

Ejercicio 3 Analizando los datos del archivo "adult.data.csv" (Equipo 10), ¿Hay diferencias significativas entre el promedio de años dedicados a la educación y la cantidad de ingreso de los censados?

*Soy estudiante de X año de mi carrera.

Objetivos

- a. Esta sección va dedicada a los objetivos de la clase, las metas para el encuentro y ciertas especificidades que considere de importancia resaltar durante el transcurso de la clase.
- b. Según la temática se pueden hacer alusión a los medios de enseñanza utilizados convenientemente.

Introducción

(Xmin')

(Como introducir mi clase?)

- Recursos para motivar la clase.
- Recuento por los antecedentes de los resultados o investigadores.
- Esta no tiene que venir acompañada por plecas, solo es un ejemplo.

Teorizando un poco

(Ymin')

- a. **(Un Teorema interesante)** Tras la introducción se podrán construir las secciones que se estimen convenientes para el desarrollo de la clase.
- b. **(Un brillante algoritmo)** Los nombre de cada una de estas secciones quedan a la elección del autor.

Ejercicio 1

(código en documento adjunto)

Se genera una población inicial con 500 valores y una distribución normal con media 0 y varianza 1. Luego de esta se extraen 4 muestras sin reposición, cada una de tamaño 200, 60, 30 y 20 respectivamente. Luego se extraen otras 4 de igual tamaño a las anteriores y con reposición.

La exactitud de los estimadores puntuales fluctúa en cada prueba realizada, la fluctuación era mayor o menor dependiendo del tamaño de las muestras. Las muestras de mayor tamaño presentan sus estimadores son generalmente más exactos, en cambio, las de menor tamaño suelen estar más alejadas del valor real. Esto se debe a que en las muestras más grandes cuando tienen un caso extremo no representativo de la población su impacto queda disminuido por el resto de los datos no extremos, mientras que en las muestras pequeñas la existencia de uno de estos altera considerablemente la información extraída. Esta razón refleja la dependencia directa que existe entre los resultados de una muestra y la calidad de sus datos. Podemos deducir entonces la importancia de que las muestras estén compuestas por datos fiables, e incluso más cuando la muestra es pequeña.

La extracción de muestras con reposición en lo que se refiere a la media no tiene muchos cambios.

Ejercicio 2

Ejercicio 3

El problema trata sobre comparar la medias de años de educación de los grupos de ingresos $\leq 50K$ y $> 50K$, y ver si hay alguna diferencia significativa entre esas medias.

El código referente a este ejercicio está en el archivo exercise_3.R

Para lograr lo anterior se dividió todos los datos del archivo csv en dos grupos: un grupo tiene ingresos $\leq 50K$ y otro grupo tiene ingresos de $> 50K$.

Luego de realizar la división se escogieron dos muestras sin reposición de tamaño $N = 60$ cada una.

Por lo que el problema se reduce a una prueba de hipótesis para la comparación de las medias de dos poblaciones Normales.

Las hipótesis para la prueba que se escogieron fueron:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Donde μ_1 es la media de años de educación de la población de personas que poseen ingresos por debajo de los 50K (inclusivo) y μ_2 es la media de años de educación de la población de personas que poseen ingresos por encima de los 50K.

Se asumió que no se conocían las varianzas de dichos grupos (o que era muy costoso calcularlas). Por lo que para hacer la prueba de hipótesis de la media se hace primero una prueba de hipótesis para la igualdad de las varianzas.

Parte del código para la hipótesis de varianza

```
result <- var.test(sample1,
sample2, alternative = "two.sided")
```

Luego de establecer la igualdad o desigualdad de varianza se procede a realizar la prueba de la media:

Parte del código para la hipótesis de la media

```
result <- t.test(sample1, sample2,
alternative = alt, var.equal = varequal)
```

Finalmente se compara el resultado del p -value de $result$ con un valor α preestablecido y si es menor se rechaza la hipótesis nula.

Analíticamente

Se hace primero la prueba de varianza.

Los datos se tomaron del script referente al ejercicio.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{S_1^2}{S_2^2} = \frac{6.37}{5.13} = 1.24$$

$$F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0.975}(59, 59) = 1.67$$

$$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0.025}(59, 59) = 0.597$$

Como $F > F_{\alpha/2}(n_1 - 1, n_2 - 1)$ y $F < F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$ no se puede descartar H_0 por lo que se asume que $\sigma_1^2 = \sigma_2^2$

$$T_{\bar{X}-\bar{Y}} = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} = 3.7$$

$$t_{1-\alpha/2}(n_1 + n_2 - 2) = t_{0.975}(116) = 1.98$$

Como $|T_{\bar{X}-\bar{Y}}| > t_{1-\alpha/2}(n_1+n_2-2)$ se cumple la región crítica y se descarta H_0 pudiendo afirmarse que se aprecian diferencias significativas.

Conclusiones

(Cierta cantidad de minutos')

Se resumirán los resultados más destacados ejercitados en la actividad.

Se puede hacer mención de aplicaciones del método estudiado, posibles investigaciones o repercusiones en la cotidianidad. Así como los elementos de mayor significación.

Estudio Independiente

(Algun tiempo')

Orientar y comentar los ejercicios siguientes:

Ejercicio 4 *De creerlo conveniente, la asignación de tareas para el estudio independiente, o la asignación de evaluaciones.*

Ejercicio 5 *La cantidad de los mismos es a conveniencia aunque podría ser de ayuda su justificación.*

Ejercicio 3

Para concluir, la solución de los ejercicios propuestos.

Ejercicio 4

El esquema de clase es variable y queda sujeto a la voluntad del participante, lo que si deberá ajustarse a los requisitos de la convocatoria oficial.