

Tarea1-Intervalos_de_Confianza

July 8, 2019

```
In [1]: import pandas as pd
import numpy as np
import scipy.stats as stats
```

0.1 Leer los datos

Primero leemos los datos del archivo de excel y creamos las tablas necesarias con ellos.

```
In [2]: hoja_ejercicio1 = pd.read_excel("./Tarea 1 - Inferencia estadística.xlsx", sheet_name="1")
hoja_ejercicio2 = pd.read_excel("./Tarea 1 - Inferencia estadística.xlsx", sheet_name="2")
hoja_ejercicio3 = pd.read_excel("./Tarea 1 - Inferencia estadística.xlsx", sheet_name="3")

columnas = ["No. Individuo", "cincunferencia abdominal (cm)"]

tabla1 = hoja_ejercicio1.iloc[1:, 0:2]
tabla1.columns = columnas

tabla2 = hoja_ejercicio1.iloc[1:36, 3:5]
tabla2.columns = columnas

tabla3 = hoja_ejercicio2.iloc[1:17, :2]
tabla3.columns = columnas

tabla4 = hoja_ejercicio3.iloc[1:20, :3]
tabla4.columns = columnas + ["Mayor a 75"]
```

0.2 Ejercicio 1

Para el ejercicio 1 usamos la Tabla 1 y la Tabla 2, siendo tabla 1 la población y tabla 2 una muestra. El ejercicio esta dividido en 3 problemas.

0.2.1 Ejercicio1 problema 1

1. La información presentada pertenece a población de estudiantes de Investigación científica seccion C. En la tabla 2 se extrajo una muestra aleatoria de 35 estudiantes. Calcule el intervalo de confianza para la media (confianza 90%).

```
In [3]: media_muestra = tabla2.iloc[:, 1].mean()
media_poblacion = tabla1.iloc[:, 1].mean()
```

```

desviacion_standar_muestra = tabla2.iloc[:,1].std()
desviacion_estandar_poblacion = tabla1.iloc[:,1].std()
n = tabla2.iloc[:,1].count()
confianza = 0.90
significancia = 1.0 - confianza
significancia_medios = significancia/2.0
z = abs(stats.norm.ppf(significancia_medios))

lim_inf = media_muestra - z*(desviacion_estandar_poblacion/n)
lim_sup = media_muestra + z*(desviacion_estandar_poblacion/n)

```

El intervalo de confianza con desviación de población conocida para un nivel de confianza del 90% es:

```
In [4]: print("Limite inferior:",lim_inf,"limite superior:",lim_sup)
```

```
Limite inferior: 79.44996244302611 limite superior: 80.87575184268819
```

0.2.2 Ejercicio 1 problema 2

- Suponga que no posee los datos de la población (tabla 1) y encuentre el intervalo de confianza para la media con base en la muestra de la tabla 2 (confianza 90%).

```
In [5]: lim_inf = media_muestra - z*(desviacion_standar_muestra/n)
lim_sup = media_muestra + z*(desviacion_standar_muestra/n)
```

El intervalo de confianza con desviación de población desconocida para un nivel de confianza del 90% es:

```
In [6]: print("Limite inferior:",lim_inf,"limite superior:",lim_sup)
```

```
Limite inferior: 79.68060708439687 limite superior: 80.64510720131743
```

0.2.3 Ejercicio 1 problema 3

- Con los datos de la tabla 2 obtenga el intervalo de confianza para la mediana (confianza 90%).

```
In [7]: muestra_ordenada = sorted(tabla2.iloc[:,1]) #primero ordenar la muestra
```

```

posicion_limite_inf = int(np.round(n/2 - (z*np.sqrt(n))/2)) - 1 # se encuentra la pos
posicion_limit_sup = int(np.round(1 + (n/2) + (z*np.sqrt(n))/2)) - 1 #se encuentra l

```

El intervalo de confianza para la mediana a un nivel de confianza de 90% es :

```
In [8]: print("Limite inf:",muestra_ordenada[posicion_limite_inf],"limite sup:",muestra_ordenada[posicion_limit_sup])
```

```
Limite inf: 76 limite sup: 82.7
```

0.3 Ejercicio 2

Para el ejercicio 2 se usa la tabla 3, que es una muestra de la población

0.3.1 Ejercicio 2 problema 1

1. La información presentada pertenece a población de estudiantes de Investigación científica sección C. En la tabla 3 se extrajo una muestra aleatoria de 16 estudiantes. Calcule el intervalo de confianza para la media (confianza 95%).

```
In [9]: media_muestra = tabla3.iloc[:,1].mean()
        n = tabla3.iloc[:,1].count()
        desviacion_muestra = tabla3.iloc[:,1].std()
        grados_libertad = n - 1

        confianza = 0.95
        significancia = 1 - confianza
        significancia_medios = significancia/2.0

        t = abs(stats.t.ppf(significancia_medios,grados_libertad))

        lim_inf = media_muestra - t*(desviacion_muestra/np.sqrt(n))
        lim_sup = media_muestra + t*(desviacion_muestra/np.sqrt(n))
```

El intervalo de confianza para la media a una confianza de 95% usando t-student es:

```
In [10]: print("Lim inf:",lim_inf," lim sup:",lim_sup)

Lim inf: 76.86458399124089  lim sup: 91.72291600875911
```

0.3.2 Ejercicio 2 problema 2

2. Con los datos de la tabla 3 obtenga el intervalo de confianza para la varianza (confianza 90%).

```
In [11]: n = tabla3.iloc[:,1].count()
        varianza_muestral = desviacion_muestra = tabla3.iloc[:,1].var()
        confianza = 0.9
        significancia = 1 - confianza
        significancia_medios = significancia/2.0
        grados_libertad = n - 1

        chi_cuadrado_alfa_medios = abs(stats.chi2.ppf(significancia_medios,grados_libertad))
        chi_cuadrado_uno_menos_alfa_medios = abs(stats.chi2.ppf(1- significancia_medios,grados_libertad))

        lim_inf = (grados_libertad*varianza_muestral)/chi_cuadrado_uno_menos_alfa_medios
        lim_sup = (grados_libertad*varianza_muestral)/chi_cuadrado_alfa_medios
```

El intervalo de confianza para la varianza con una confianza del 90% usando chi-cuadrado es

```
In [12]: print("lim inf:",lim_inf,"lim sup:",lim_sup)

lim inf: 116.64721773950916 lim sup: 401.5578971611225
```

0.4 Ejercicio 3

Para el ejercicio 3 usamos la tabla 4 que contiene una variable dicotómica “mayor a 75” con un valor de 1 cuando la circunferencia abdominal es mayor a 75 y 0 en caso contrario

nota matemáticamente sacar la proporción de una variable que contiene solo 0s y 1s es equivalente a sacar el promedio de esta variable.

```
In [13]: n = tabla4["No. Individuo"].count()
p = tabla4["Mayor a 75"].mean()
sp = np.sqrt(p*(1-p)/n)
confianza = 0.95
significancia = 1 - confianza
significancia_medios = significancia / 2.0

z = abs(stats.norm.ppf(significancia_medios))

limite_inferior = p - z*sp
limite_superior = p + z*sp
```

El intervalo de confianza para la proporción a una confianza del 95% es:

```
In [14]: print("lim inf:",limite_inferior,"lim sup:",limite_superior)

lim inf: 0.5378276748607178 lim sup: 0.9621723251392822
```