

Estadística II (Taller: Pruebas de hipótesis)

Juan Carlos Trejos Iglesias

2025-1

Índice

1. Ejercicio propuesto en clase	2
1.1. Datos recolectados	2
1.2. Solución	3
1.2.1. Punto A	8
1.2.2. Punto B	10
1.3. Código en python de la solución	11

1. Ejercicio propuesto en clase

El gerente de un centro comercial de la ciudad de Manizales estima que los hombres gastan más de \$40 cada vez que visitan el centro comercial y que las mujeres gastan exactamente \$25 en cada visita.

Con el fin de validar esta afirmación, se contrata una firma de investigación de mercados que recolecta datos de una muestra aleatoria de 26 hombres y 26 mujeres.

Los objetivos del estudio son:

- A. Probar si el gerente del centro comercial tiene razón o no.
- B. Calcular el margen de error del muestreo, considerando que se trabajó con una confianza del 95 % y una varianza de 4 dólares en ambas muestras.

1.1. Datos recolectados

Hombres	Mujeres
57,00	31,05
31,05	21,45
40,20	19,20
60,38	21,90
45,00	23,55
32,10	21,90
135,00	22,35
36,00	27,90
31,20	30,00
24,00	26,40
28,35	30,30
36,15	54,00
110,63	35,10
27,75	26,40
42,00	26,25
27,30	33,90
40,80	38,85
92,00	24,15
46,00	24,00
81,25	29,25
103,75	16,95
29,10	27,60
21,75	21,15
36,00	22,95
31,35	31,35
21,15	34,80

1.2. Solución

Media muestral

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Hombres

$$\bar{x}_h = \frac{1296,96}{26} = \boxed{49,88}$$

Mujeres

$$\bar{x}_m = \frac{715,65}{26} = \boxed{27,52}$$

Desviación estándar muestral de gastos en dólares de los hombres

Datos: $n = 26$, $\bar{x}_m = 49,88$

Paso 1: Calcular las diferencias de cuadrados $(x_i - \bar{x})^2$

$$\begin{aligned}(57,00 - 49,88)^2 &= 50,65 \\(31,05 - 49,88)^2 &= 354,68 \\(40,20 - 49,88)^2 &= 93,76 \\(60,38 - 49,88)^2 &= 110,19 \\(45,00 - 49,88)^2 &= 23,84 \\(32,55 - 49,88)^2 &= 300,44 \\(32,10 - 49,88)^2 &= 316,24 \\(135,00 - 49,88)^2 &= 7244,89 \\(36,00 - 49,88)^2 &= 192,74 \\(31,20 - 49,88)^2 &= 349,06 \\(28,35 - 49,88)^2 &= 463,67 \\(36,15 - 49,88)^2 &= 188,60 \\(27,75 - 49,88)^2 &= 489,87 \\(110,63 - 49,88)^2 &= 3690,19 \\(27,30 - 49,88)^2 &= 510,00 \\(42,00 - 49,88)^2 &= 62,14 \\(40,80 - 49,88)^2 &= 82,50 \\(92,00 - 49,88)^2 &= 1773,84 \\(46,00 - 49,88)^2 &= 15,08 \\(81,25 - 49,88)^2 &= 983,88 \\(103,75 - 49,88)^2 &= 2901,65 \\(29,10 - 49,88)^2 &= 431,94 \\(42,30 - 49,88)^2 &= 57,50 \\(31,35 - 49,88)^2 &= 343,47 \\(21,75 - 49,88)^2 &= 791,47 \\(36,00 - 49,88)^2 &= 192,74\end{aligned}$$

Paso 2: Sumar todas las diferencias al cuadrado:

$$\sum_{i=1}^{26} (x_i - \bar{x}_h)^2 = \boxed{22015,02}$$

Paso 3: Sustituir en la fórmula de la **varianza**:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{22015,02}{25} = \boxed{880,60}$$

Paso 4: Calcular la desviación estándar:

$$s = \sqrt{880,60} = \boxed{29,67}$$

Desviación estándar muestral de gastos en dólares de las mujeres**Datos:** $n = 26$, $\bar{x}_m = 27,52$ **Paso 1:** Calcular las diferencias al cuadrado $(x_i - \bar{x})^2$

$$(21,45 - 27,52)^2 = 36,91$$

$$(19,20 - 27,52)^2 = 69,31$$

$$(21,90 - 27,52)^2 = 31,64$$

$$(23,55 - 27,52)^2 = 15,80$$

$$(21,90 - 27,52)^2 = 31,64$$

$$(22,35 - 27,52)^2 = 26,78$$

$$(27,90 - 27,52)^2 = 0,14$$

$$(30,00 - 27,52)^2 = 6,13$$

$$(24,00 - 27,52)^2 = 12,43$$

$$(26,40 - 27,52)^2 = 1,27$$

$$(30,30 - 27,52)^2 = 7,70$$

$$(54,00 - 27,52)^2 = 700,93$$

$$(35,10 - 27,52)^2 = 57,38$$

$$(26,40 - 27,52)^2 = 1,27$$

$$(26,25 - 27,52)^2 = 1,63$$

$$(33,90 - 27,52)^2 = 40,64$$

$$(38,85 - 27,52)^2 = 128,26$$

$$(24,15 - 27,52)^2 = 11,39$$

$$(24,00 - 27,52)^2 = 12,43$$

$$(29,25 - 27,52)^2 = 2,98$$

$$(16,95 - 27,52)^2 = 111,83$$

$$(27,60 - 27,52)^2 = 0,01$$

$$(21,15 - 27,52)^2 = 40,64$$

$$(22,95 - 27,52)^2 = 20,93$$

$$(31,35 - 27,52)^2 = 14,63$$

$$(34,80 - 27,52)^2 = 52,93$$

Paso 2: Sumar todas las diferencias al cuadrado:

$$\sum_{i=1}^{26} (x_i - \bar{x}_m)^2 = \boxed{1437,58}$$

Paso 3: Sustituir en la fórmula de la **varianza**:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1437,58}{25} = \boxed{57,50}$$

Paso 4: Calcular la desviación estándar:

$$s = \sqrt{57,50} = \boxed{7,58}$$

Error estándar (de la media)

$$SE = \frac{s}{\sqrt{n}}$$

Hombres

$$SE_h = \frac{29,67}{\sqrt{26}} = \boxed{5,81}$$

Mujeres

$$SE_m = \frac{7,58}{\sqrt{26}} = \boxed{1,48}$$

1.2.1. Punto A

El objetivo es verificar si el gerente del centro comercial tiene razón respecto a los gastos promedio de hombres y mujeres:

- Que los **hombres gastan más de \$40** por visita.
- Que las **mujeres gastan exactamente \$25** por visita.

Para ello, se plantean dos pruebas de hipótesis independientes usando la distribución **t de Student**, ya que no se conoce la varianza poblacional y el tamaño de muestra es pequeño ($n = 26$).

$$t = \frac{\bar{x}_h - \mu_0}{s_h / \sqrt{n}}$$

- Prueba de hipótesis unilateral para la muestra de los hombres

Hipótesis:

$$H_0 : \mu \leq 40 \quad H_1 : \mu > 40$$

Datos muestrales:

$$\bar{x}_h = 49,88$$

$$s_h = 29,67$$

$$n = 26 \Rightarrow \text{gl} = 25$$

Estadístico de prueba:

$$t = \frac{\bar{x}_h - \mu_0}{s_h / \sqrt{n}} = \frac{49,88 - 40}{29,67 / \sqrt{26}} = \frac{9,88}{5,81} \approx \boxed{1,697}$$

Valor crítico: Para $\alpha = 0,05$, **unilateral**, derecha, con 25 gl:

$$t_{0,05,25} \approx 1,708$$

Decisión: Como $t = 1,697 < 1,708$, **Se acepta** H_0 .

Conclusión: No hay evidencia suficiente para afirmar que los hombres gasten más de \$40 por visita.

■ Prueba de hipótesis bilateral para la muestra de mujeres

Hipótesis:

$$H_0 : \mu = 25 \quad H_1 : \mu \neq 25$$

Datos muestrales:

$$\bar{x}_m = 27,52$$

$$s_m = 7,48$$

$$n = 26 \Rightarrow \text{gl} = 25$$

Estadístico de prueba:

$$t = \frac{\bar{x}_h - \mu_0}{s_h / \sqrt{n}} = \frac{27,52 - 25}{7,58 / \sqrt{26}} = \frac{2,52}{1,48} \approx \boxed{1,69}$$

Valor crítico: Para $\alpha = 0,05$, **bilateral**, con 25 gl:

$$t_{0,025,25} \approx \pm 2,060$$

Decisión: Como $|t| = 1,69 < 2,060$, **Se acepta** H_0 .

Conclusión estadística: No hay evidencia suficiente para afirmar que las mujeres un valor diferente \$25 por visita.

1.2.2. Punto B

El objetivo es determinar el **margen de error** con el que se realizó el muestreo para ambas muestras (hombres y mujeres), teniendo en cuenta que:

- Se utilizó un **nivel de confianza del 95 %**.
- Se conoce la **varianza poblacional**, la cual es igual a 4 dólares cuadrados ($\sigma^2 = 4 \Rightarrow \sigma = 2$).
- El tamaño de cada muestra es de $n = 26$.

Dado que se conoce la varianza poblacional, el cálculo del margen de error se realiza utilizando la **distribución normal estándar (Z)**, por lo que se emplea la siguiente fórmula:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Sustituimos los valores

El valor crítico $z_{\alpha/2} = 1,96$ se obtiene de la tabla de la distribución normal estándar, para un nivel de confianza del 95 %. Este valor corresponde al punto que deja un 2,5 % de probabilidad en cada cola de la curva normal, es decir:

$$P(-1,96 < Z < 1,96) = 0,95$$

Esto implica que $\alpha = 0,05 \Rightarrow \alpha/2 = 0,025$, por lo tanto:

$$z_{\alpha/2} = z_{0,025} = 1,96$$

$$E = 1,96 \cdot \frac{2}{\sqrt{26}} \approx 1,96 \cdot 0,3922 \approx \boxed{0,7687}$$

Conclusión: El margen de error con el que se realizó el muestreo fue de aproximadamente \$ $\boxed{0,77}$.

1.3. Código en python de la solución

[CLICK PARA IR AL REPOSITORIO EN GOOGLE COLAB CON LA SOLUCIÓN AL EJERCICO](#)

```
1 import numpy as np
2 from pprint import pprint
3
4 def analyze_data(data, mu_0):
5     """
6     Analiza un conjunto de datos y calcula estadísticas
7     descriptivas y un estadístico t de Student.
8
9     Parámetros:
10    - data: Lista o array de números que representan los datos de
11      la muestra.
12    - mu_0: Valor de referencia para calcular el estadístico t de
13      Student.
14
15    Retorna:
16    Un diccionario con estadísticas descriptivas y el estadístico
17      t de Student.
18    """
19
20    # Convertir los datos a un array de NumPy y aplanarlo para
21    # asegurar que sea unidimensional
22    data = np.array(data).flatten()
23
24    # Número de elementos en la muestra
25    n = len(data)
26
27    # Verificar que haya al menos dos elementos para calcular la
28    # desviación estándar
29    if n < 2:
30        raise ValueError("Los datos de entrada deben contener al
31          menos dos elementos.")
32
33    # Calcular la media de los datos
34    mean = np.mean(data)
35
36    # Calcular la suma de los datos
37    sum_data = np.sum(data)
38
39    # Calcular las diferencias al cuadrado entre cada dato y la
40    # media
41    squared_diffs = (data - mean) ** 2
42
43    # Calcular la suma de las diferencias al cuadrado
```

```
36     sum_squared_diffs = np.sum(squared_diffs)
37
38     # Calcular la varianza de la muestra (usando ddof=1 para la
        varianza muestral)
39     variance = np.var(data, ddof=1)
40
41     # Calcular la desviación estándar de la muestra
42     std_dev = np.std(data, ddof=1)
43
44     # Calcular el error estándar de la media
45     std_err = std_dev / np.sqrt(n)
46
47     # Calcular el estadístico t de Student
48     test_statistic = (mean - mu_0) / std_err
49
50     # Retornar un diccionario con los resultados
51     return {
52         "Suma de datos": sum_data,
53         "Media": mean,
54         "Diferencia de cuadrados (xi - x ) ": squared_diffs,
55         "Suma de diferencia de cuadrados": sum_squared_diffs,
56         "Varianza": variance,
57         "Desviación estándar": std_dev,
58         "Error estándar": std_err,
59         "t-student": test_statistic
60     }
61
62 # Datos de ejemplo para hombres y mujeres
63 hombresXi = [57.00, 31.05, 40.20, 60.38, 45.00, 32.55, 32.10,
        135.00, 36.00, 31.20, 28.35, 36.15, 27.75, 110.63, 27.30, 42.00,
        40.80, 92.00, 46.00, 81.25, 103.75, 29.10, 42.30, 31.35, 21.75,
        36.00]
64 hombresMu_ref = 40
65
66 mujeresXi = [21.45, 19.20, 21.90, 23.55, 21.90, 22.35, 27.90, 30.00,
        24.00, 26.40, 30.30, 54.00, 35.10, 26.40, 26.25, 33.90, 38.85,
        24.15, 24.00, 29.25, 16.95, 27.60, 21.15, 22.95, 31.35, 34.80]
67 mujeresMu_ref = 25
68
69 # Imprimir los resultados del análisis para hombres y mujeres
70 print('Datos de los hombres: ', analyze_data(hombresXi,
        hombresMu_ref))
71 print('Datos de las mujeres: ', analyze_data(mujeresXi,
        mujeresMu_ref))
```

[CLICK PARA IR AL REPOSITORIO EN GOOGLE COLAB CON LA SOLUCIÓN AL EJERCICO](#)