



Tecnológico de Monterrey

Campus Estado de México

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Tarea 4: Pruebas de hipótesis

Juan Pablo Castañeda Serrano

A01752030

Pasos para el ejercicio 1

```
# Paso 1. Datos de las latas
pesos = [11.0,11.6,10.9,12.0,11.5,12.0,12.0,11.2,10.5,12.2,11.2,10.5,12.2,11.8,12.1,11.6,11.7,11.6,11.2,12.0,11.4,10.8,11.8,10.9,11.4]

# Paso 2. Realizar la prueba t para una muestra
t_stat, p_val = ttest_1samp(pesos, 11.7)

# Mostrar el valor t y el p-valor
print(f"T-statistic: {t_stat}")
print(f"P-value: {p_val}")

# Paso 3. Verificar si debemos rechazar la hipótesis nula
alpha = 0.02 # 100% - 98% = 2%, dividido por 2 (prueba bilateral) = 1%. Dado que es bilateral, lo multiplicamos por 2 de nuevo.
if p_val < alpha:
    print("Rechazamos la hipótesis nula.")
else:
    print("Aplica la hipótesis nula.")

print(20*"")
```

```
# PASO 4. CREAR GRAFICAS
# Número de grados de libertad
df = len(pesos) - 1

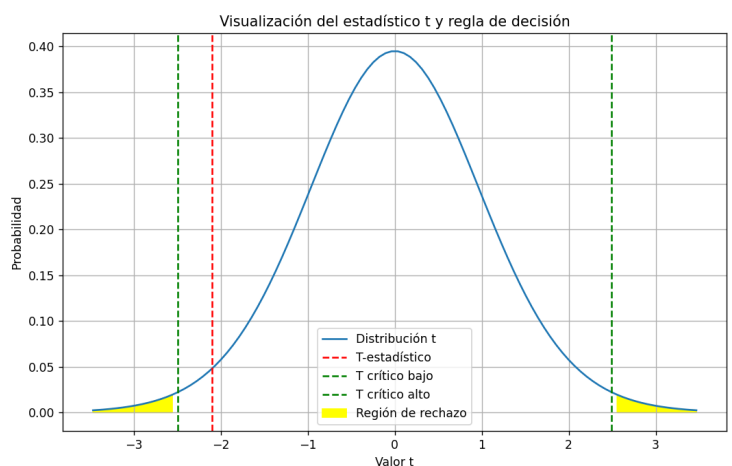
# Valores t críticos para alpha/2 y 1 - alpha/2
alpha = 0.02
t_crit_low = t.ppf(alpha / 2, df)
t_crit_high = t.ppf(1 - alpha / 2, df)

# Valores para trazar la distribución t
x = np.linspace(t.ppf(0.001, df), t.ppf(0.999, df), 100)
y = t.pdf(x, df)

# Graficar la distribución t
plt.figure(figsize=(10, 6))
plt.plot(x, y, label='Distribución t')
```

Resultados

```
PS C:\Users\sickp\Documents\estadistica> python tarea5.py
T-statistic: -2.1023453584182947
P-value: 0.04619714718347785
Aplica la hipótesis nula.
*****
```



Conclusión ejercicio 1

Si la línea roja representando el estadístico t se sitúa en las zonas marcadas en amarillo:

Esto indica que existe fundamento suficiente para descartar la suposición inicial de que el peso promedio real es 11.7. De esta manera, los responsables deberían examinar su método de empaque para asegurarse de que el peso de cada lata se alinee con sus criterios.

En caso de que el estadístico t no se encuentre en la zona de rechazo:

No se dispone de pruebas contundentes para descartar la hipótesis inicial. Esto sugiere que el método de empaque podría ser adecuado y que el peso promedio de las latas podría coincidir efectivamente con 11.7.

Pasos para el ejercicio 2

```
# Paso 1. Datos
x = np.array([17, 11, 12, 23, 20, 23, 15, 16, 23, 22, 18, 23, 25, 14, 12, 12, 20, 18,
              12, 19, 11, 11, 20, 21, 11, 18, 14, 13, 13, 19, 16, 10, 22, 18, 23])

#Paso 2. Parametros
alpha = 0.02

#Paso 3. Cálculo del valor critico
n = len(x)
t0 = stats.t.ppf(alpha/2, n-1)
print(f"t0 = {t0}")

#Paso 4. Calculo de la estadistica de prueba
m = np.mean(x)
s = np.std(x, ddof=1)
sm = s / np.sqrt(n)
te = (11.49 - m) / sm
print(f"t* = {te}")

#Paso 5. Calculo del valor p
valorp = 2 * stats.t.cdf(te, n-1)
print(f"Valor p = {valorp}")
```

```
#Paso 6. Prueba T
t_stat, p_val = stats.ttest_1samp(x, 11.7)
print(f"t statistic = {t_stat}")
print(f"p value = {p_val}")
```

```
# Por experiencias anteriores, se sabe que  $\sigma=4$  minutos. Usando un nivel de significación de 0.07, ¿está justificada la tarifa adicional?

mu_0 = 11.7
sigma = 4
n = len(x)
mean_sample = np.mean(x)
z = (mean_sample - mu_0) / (sigma / np.sqrt(n))
z_critical_positive = stats.norm.ppf(1 - 0.07/2)
z_critical_negative = stats.norm.ppf(0.07/2)

print(f"z = {z}")
print(f"z_critical_positive = {z_critical_positive}")
print(f"z_critical_negative = {z_critical_negative}")

if z > z_critical_positive or z < z_critical_negative:
    print("Rechazar H0")
else:
    print("No rechazar H0")
```

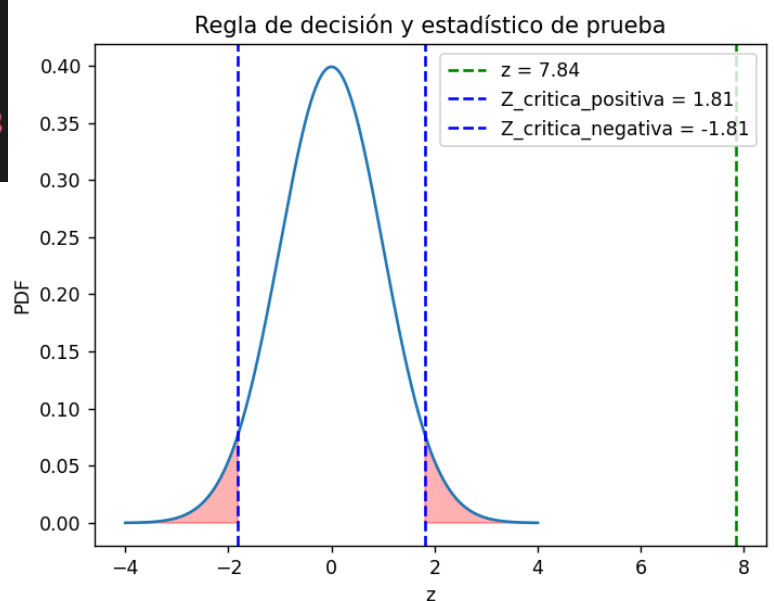
```
mu_0 = 11.7
sigma = 4
n = len(x)
mean_sample = np.mean(x)

z = (mean_sample - mu_0) / (sigma / np.sqrt(n))
z_critical_positive = stats.norm.ppf(1 - 0.07/2)
z_critical_negative = stats.norm.ppf(0.07/2)

x_values = np.linspace(-4, 4, 400)
y_values = stats.norm.pdf(x_values)
```

Resultados

```
t0 = -2.441149627906477
t* = -7.1944461825720465
Valor p = 2.5371042797624655e-08
t statistic = 6.920247689225382
p value = 5.643304419197504e-08
z = 7.838805712606993
z_critical_positive = 1.8119106729525973
z_critical_negative = -1.8119106729525978
Rechazar H0
```



Conclusión ejercicio 2

Si el estadístico de prueba z (representado por la línea verde en la gráfica) cae dentro de la región roja (área de rechazo), entonces hay suficiente evidencia para rechazar la hipótesis nula y justificar la tarifa adicional. En cambio, si z no se encuentra en esa región, la evidencia proporcionada no respalda la imposición de una tarifa adicional.