

Estimation_Theory_and_Hypothesis_testing

```
[1]: #Libraris to help with reading and manipulating data
import pandas as pd
import numpy as np

#Libraries to help with data visualization
import matplotlib.pyplot as plt
import seaborn as sns

#Library to help with statistical analysis
import scipy.stats as stats
```

0.1 Sampling Distribution

0.1.1 Q1. Suppose an automobile battery manufacturer claims that the mean lifetime of their battery 60 months with a standard deviation of 6 months. Suppose the distribution of battery life is approximately normal. Find the probability that the mean lifetime of 40 randomly sampled batteries will be less than 58 months.

```
[2]: # Given values
mu = 60 # population mean
sigma = 6 # population standard deviation
n = 40 # sample size
x_bar = 58 # sample mean we are comparing
```

```
[3]: # Standard error of the mean
SE = sigma / (n ** 0.5)
```

```
[4]: # Z-score calculation
Z = (x_bar - mu) / SE
```

```
[5]: # Probability corresponding to the Z-score
probability = stats.norm.cdf(Z)

print(f"The Z-score is : {Z}\n")
print(f"The probability that the mean lifetime of 40 randomly sampled batteries_
↪will be less than 58 months is: {probability}")
```

The Z-score is : -2.1081851067789197

The probability that the mean lifetime of 40 randomly sampled batteries will be less than 58 months is: 0.017507490509831244

0.2 Interval Estimation

0.2.1 Q2. A random sample of 40 households was selected as part of a study on electricity usage, and the number of kilowatt-hours (kWh) was recorded for each household in the sample for the first quarter of 2020. The average usage was found to be 310 kWh. In a very large study in the first quarter of the previous year, it was found that the standard deviation of the usage was 89 kWh.. Assuming the standard deviation is unchanged and that the usage is normally distributed, provide an expression for calculating a 95% confidence interval for the mean usage in the first quarter of 2019.

```
[6]: # Given values
x_bar = 310 # sample mean
sigma = 89 # population standard deviation
n = 40 # sample size
z_value = 1.96 # Z-value for 95% confidence
```

```
[7]: # Standard error of the mean
SE = sigma / (n ** 0.5)
```

```
[8]: # Confidence interval calculation
lower_bound = x_bar - z_value * SE
upper_bound = x_bar + z_value * SE
```

```
[9]: print(f"""The 95% confidence interval for the mean electricity usage in the
    ↳first quarter of 2019 is ({lower_bound} kWh,{upper_bound} kWh).
    This means we are 95% confident that the true mean usage falls within this
    ↳range.lower_bound, upper_bound""")
```

The 95% confidence interval for the mean electricity usage in the first quarter of 2019 is (282.4186142480114 kWh,337.5813857519886 kWh).

This means we are 95% confident that the true mean usage falls within this range.lower_bound, upper_bound

0.3 Hypothesis Testing

0.3.1 Q3. You are a manager of a Chinese restaurant. You want to determine whether the mean waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes. State the null and alternative hypotheses.

```
[10]: # According to Two-tailed hypothesis test

# Null Hypothesis: The mean waiting time has not changed from the previous means
    ↳of 4.5 minutes.
# mu=4.5 #(minutes)
```

```
# Alternate Hypothesis: The mean waiting time has changed from 4.5 minutes
# mu!=4.5 #(minutes)
```

0.3.2 Q4. What is the p-value in a two-tailed z-test for one sample, where the computed test statistic (z- stat) is equal to +2.00?

```
[11]: # The p-value in a two-tailed z-test can be found by looking up the probability
      ↪for the computed Z-statistic
      # and multiplying it by 2(since its two-tailed test).

      # Given Z-score for two-tailed test
      Z_stat = 2.00

      # Cumulative probability for Z = 2.00
      cumulative_prob = stats.norm.cdf(Z_stat)

      # P-value for two-tailed test
      p_value = 2 * (1 - cumulative_prob)

      print(f"Cumulative probability: {cumulative_prob}")
      print(f"P-value: {p_value}")
```

Cumulative probability: 0.9772498680518208

P-value: 0.04550026389635842

0.3.3 Q5. Samy, Product Manager of K2 Jeans, wants to launch a product line into a new market area. A Survey of a random sample of 400 households in that market showed a mean income per household of 30000 rupees. The standard deviation based on an earlier pilot study of households is 8000 rupees. Samy strongly believes the product line will be adequately profitable only in markets where the mean household income is greater than 29000 rupees. Samy wants our help in deciding whether the product line should be introduced in the new market. Perform statistical analysis and based on that draw a conclusion.

0.3.4 Assume a level of significance (α) of 5%

```
[12]: '''
      Null Hypothesis: The mean household income in the new market is less than or
      ↪equal to 29,000 rupees.
      mu<29,000

      Alternative Hypothesis: The mean household income in the new market is greater
      ↪than 29,000 rupees.
      mu>29,000
      '''

      # Given values
```

```
sample_mean = 30000 # sample mean
hypothesized_mean = 29000 # hypothesized mean ( $\mu_0$ )
sigma = 8000 # population standard deviation
n = 400 # sample size
```

```
[13]: # Standard error of the mean
SE = sigma / (n ** 0.5)

# Z-statistic calculation
Z_stat = (sample_mean - hypothesized_mean) / SE

Z_stat
```

[13]: 2.5

```
[14]: # P-value calculation for one-tailed test
p_value = 1 - stats.norm.cdf(Z_stat)

print(p_value)
print("Since the p-value 0.0062 is less than the significance level alpha=0.
↪05,we reject the null hypothesis")
```

0.006209665325776159

Since the p-value 0.0062 is less than the significance level $\alpha=0.05$,we
reject the null hypothesis

[]:

[]: