

November 25, 2024

### 0.0.1 Sampling Distribution

Q1. Suppose an automobile battery manufacturer claims that the mean lifetime of their battery is 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.

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

[2]: # Given data
mean_population = 60 # Mean lifetime of the battery
std_dev_population = 6 # Standard deviation of battery lifetime
sample_size = 40 # Number of sampled batteries
sample_mean = 58 # Sample mean to test

[7]: # Calculate the standard error of the mean
standard_error = std_dev_population / (sample_size ** 0.5)

[8]: # Calculate the z-score
z_score = (sample_mean - mean_population) / standard_error

[9]: # Find the probability using the cumulative distribution function (CDF) of the
      ↪ standard normal distribution
probability = stats.norm.cdf(z_score)

[6]: # Display the results
print("Sampling Distribution Analysis:")
print(f"Z-score: {z_score:.4f}")
print(f"Probability (mean lifetime < 58 months): {probability:.4f}")
```

```
Sampling Distribution Analysis:
Z-score: -2.1082
Probability (mean lifetime < 58 months): 0.0175
```

## 0.1 Interval Estimation

**0.1.1 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.

```
[10]: # Given data
mean_population = 60
std_dev_population = 6
sample_size = 40
sample_mean = 58

[11]: # Calculate the standard error of the mean
standard_error = std_dev_population / (sample_size ** 0.5)

[12]: # Calculate the z-score
z_score = (sample_mean - mean_population) / standard_error

[13]: # Find the probability using the cumulative distribution function (CDF) of the
      ↪ standard normal distribution
probability = stats.norm.cdf(z_score)

[14]: # Display the results
print("Sampling Distribution Analysis:")
print(f"Z-score: {z_score:.4f}")
print(f"Probability (mean lifetime < 58 months): {probability:.4f}")
```

```
Sampling Distribution Analysis:
Z-score: -2.1082
Probability (mean lifetime < 58 months): 0.0175
```

**0.1.2 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.

```
[15]: from scipy.stats import ttest_1samp

# Sample data (replace with your actual data)
waiting_times = [4.8, 5.2, 4.3, 4.6, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6] # Example
      ↪ data

# Population mean
```

```

population_mean = 4.5

# Perform a one-sample t-test
t_stat, p_value = ttest_1samp(waiting_times, population_mean)

# Display results
print("One-Sample t-Test Results:")
print(f"t-statistic: {t_stat:.4f}")
print(f"p-value: {p_value:.4f}")

# Decision based on alpha = 0.05
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: The mean waiting time has changed.")
else:
    print("Fail to reject the null hypothesis: No evidence that the mean_
    ↪waiting time has changed.")

```

One-Sample t-Test Results:

t-statistic: 1.5149

p-value: 0.1641

Fail to reject the null hypothesis: No evidence that the mean waiting time has changed.

**0.1.3 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?

```

[19]: from scipy.stats import norm

# Z-statistic
z_stat = 2.00

# Compute p-value for two-tailed test
p_value = 2 * (1 - norm.cdf(z_stat)) # CDF gives P(Z ≤ z), so 1 - CDF gives
    ↪P(Z > z)

print(f"p-value: {p_value:.4f}")

```

p-value: 0.0455

**0.1.4 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.

Assume a level of significance ( $\alpha$ ) of 5%

```
[18]: # Given data
sample_mean = 30000 # Sample mean ( $\bar{x}$ )
hypothesized_mean = 29000 # Hypothesized mean ( )
std_dev_population = 8000 # Population standard deviation ( )
sample_size = 400 # Sample size (n)
alpha = 0.05 # Significance level

# Calculate the standard error
standard_error = std_dev_population / (sample_size ** 0.5)

# Calculate the z-statistic
z_stat = (sample_mean - hypothesized_mean) / standard_error

# Calculate the p-value for one-tailed test
p_value = 1 - norm.cdf(z_stat)

# Display results
print("One-Tailed Z-Test Results:")
print(f"Z-statistic: {z_stat:.4f}")
print(f"p-value: {p_value:.4f}")

# Decision based on alpha
if p_value < alpha:
    print("Reject the null hypothesis: Launch the product line; the mean income is adequate.")
else:
    print("Fail to reject the null hypothesis: Do not launch the product line; the mean income is not adequate.")
```

One-Tailed Z-Test Results:

Z-statistic: 2.5000

p-value: 0.0062

Reject the null hypothesis: Launch the product line; the mean income is adequate.