UCSC

SCS2211 - LABORATORY II
R Lab Practical Sheet - 13

Instructions

- **Do the Activities and save in a .ipynb file**
- **File name should be <Index number>.ipynb (Eg: 2000000.ipynb) and upload to the given link.**
- **Any form of plagiarism or collusion is not allowed**

Hypothesis Testing with R

Hypothesis testing is a statistical method used to test a claim or hypothesis about a population parameter using a sample. It helps determine whether the observed data provides enough evidence to reject a given hypothesis.

Key Concepts

- Null Hypothesis (H_0): Assumes no effect or no difference.
- Alternative Hypothesis (H_A): Assumes there is an effect or difference.
- Significance Level (α): The probability of rejecting H_0 when it is true (typically 0.05).
- Test Statistic: A value calculated from the sample data to compare against a critical value.
- p-value: The probability of obtaining the observed results, assuming H_0 is true.

4 Steps of Hypothesis Testing in R

1. Step 1: Setting Null and Alternative Hypotheses
   Define the hypotheses based on the problem statement.

2. Step 2: Calculating the Test Statistic
   Use appropriate statistical tests based on data type:
   - t-test: Compare means.
   - Chi-Square test: Compare categorical variables.
   - ANOVA: Compare multiple means.
   - Shapiro-Wilk test: Check normality.

3. Step 3: p-Value and Decision
   Compare the p-value with the chosen significance level (α):

- If $p < \alpha$, reject $H0$
- If $p \geq \alpha$, fail to reject $H0$

.

4. Step 4: Conclusion
   Interpret the result in the context of the problem.

Example Activities

Activity 1: One-Sample t-Test

Scenario:

A company claims the average weight of its product is 500g. A sample of 10 products is tested.

Step 1: Define Hypotheses

- H_0: Mean weight = 500g
- H_1: Mean weight ≠ 500g

Step 2: Sample data (weights of 10 products)

```
weights <- c(502, 498, 501, 499, 503, 497, 500, 502, 496, 501)
#Perform one-sample t-test
test_result <- t.test(weights, mu = 500)
```

Step 3: View test result

```
print(test_result)
```

Step 4: Conclusion

- If p-value < 0.05, reject H_0 , meaning the product weight is significantly different from 500g.
- If p-value > 0.05, fail to reject H_0, meaning no significant difference.

Activity 2: Two-Sample t-Test

Scenario:

Compare exam scores of two different classes.

Step 1: Define Hypotheses

- H_0: No difference in mean scores
- H_1: Mean scores are different

Step 2: Sample data

```
class1 <- c(85, 78, 92, 88, 76)
class2 <- c(80, 82, 79, 85, 77)
#Perform independent t-test
test_result <- t.test(class1, class2, var.equal = TRUE)
```

Step 3: View test result

```
print(test_result)
```

Step 4: Conclusion

- If p-value < 0.05, reject H_0 (scores differ significantly).
- If p-value > 0.05, fail to reject H_0 (no significant difference).

Activity 1:

1. **Testing if Students' Average Test Score is Different from 75**

   Scenario:

   A teacher believes that students' average test score in a subject is **75**. A sample of **50 students** is taken, with a mean score of **73.5** and a standard deviation of **10**. Perform a **two-tailed hypothesis test** at **α = 0.05**.

2. **Checking if a Factory's Average Production Time is Less Than 30 Minutes**

   Scenario:

   A factory claims that it takes **30 minutes** on average to produce a product. A sample of **40 products** is taken, with a mean production time of **28.7 minutes** and a standard deviation of **3.5 minutes**. Perform a **lower tail hypothesis test** at **α = 0.05**.

Lower Tail Test of Population Mean with Known Variance

A Lower Tail Test (also called a Left-Tailed Test) is a type of hypothesis test used to check if the population mean is significantly less than a specified value. This test is performed when the population variance (σ^2) is known.

Understanding the Concept

We test whether the true mean (μ) of a population is less than a given value (μ_0).

- Null Hypothesis (H_0): The population mean is greater than or equal to the claimed value.
  $H_0 : \mu \geq \mu_0$
- Alternative Hypothesis (H_A): The population mean is less than the claimed value.
  $H_A : \mu < \mu_0$

This is called a Lower Tail Test because we focus on the left side (lower values) of the normal distribution.

Test Statistic (Z-Score)

When the population variance ($\sigma 2$) is known, we use the Z-test:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Where:

- $\bar{x}$ = Sample Mean
- μ_0 = Population Mean under H_0
- σ = Population Standard Deviation
- n = Sample Size

The Z-score tells us how far $\bar{x}$ is from μ_0 in terms of standard deviations.

p-Value and Decision

- Find the p-value from the standard normal table.
- If $p<\alpha$ (e.g., 0.05), reject H_0 (the population mean is significantly less than μ_0).
- If $p\geq\alpha$, fail to reject H_0(not enough evidence to say μ is less than μ_0).

Example: Lower Tail Test in R

Scenario:

A battery company claims that its batteries last 50 hours on average. A sample of 30 batteries has a mean lifespan of 48.5 hours, with a known standard deviation of 4 hours. Test at α = 0.05 if the batteries last significantly less than 50 hours.

Steps & R Code

Step 1: Define Hypotheses

- H_0: μ ≥ 50 (Batteries last 50 hours or more)
- H_1: μ < 50 (Batteries last less than 50 hours)

Step 2: Given values

```
x_bar <- 48.5   # Sample Mean
mu_0 <- 50      # Claimed Population Mean
sigma <- 4      # Known Population Standard Deviation
n <- 30         # Sample Size
alpha <- 0.05   # Significance Level
```

Step 3: Calculate Z-Score

```
z <- (x_bar - mu_0) / (sigma / sqrt(n))
```

Step 4: Compute p-value for lower tail test

```
p_value <- pnorm(z)
```

Step 5: Decision

```
if (p_value < alpha) {
  result <- "Reject H0: The batteries last significantly less than 50 
hours."
```

```
} else {
  result <- "Fail to reject H0: No significant evidence that batteries last less than 50 hours."
}
```

Display results

```
z
p_value
Result
```

Interpreting the Output

- If p-value < 0.05 → Reject H_0(Evidence suggests batteries last significantly less than 50 hours).
- If p-value ≥ 0.05 → Fail to reject H_0 (No strong evidence to claim batteries last less than 50 hours).

Conclusion

- A Lower Tail Test is used to check if the population mean is significantly lower than a given value.
- We use a Z-test when the population variance (σ^2) is known.
- The p-value helps determine if we should reject H_0 .

Activity 02

1. **Checking if the Average Fuel Efficiency of Cars is Less Than 25 MPG**

   Scenario:

   A car manufacturer claims that its cars have an average fuel efficiency of **25 MPG**. A sample of **45 cars** is tested, with a mean efficiency of **24.2 MPG** and a standard deviation of **2.5 MPG**. Test at **α = 0.05**.

2. **Checking if Average Blood Pressure is Lower Than 120**

   Scenario:

   A study claims that the average systolic blood pressure of adults is **120 mmHg**. A sample of **50 adults** is taken, with a mean of **118.5 mmHg** and a standard deviation of **6 mmHg**. Test at **α = 0.05**.

Upper Tail Test of Population Mean with Known Variance

An Upper Tail Test (also called a Right-Tailed Test) is a hypothesis test used to check if the population mean is significantly greater than a specified value. This test is performed when the population variance (σ^2) is known.

Understanding the Concept

We test whether the true mean (μ) of a population is greater than a given value (μ_0).

- Null Hypothesis (H_0): The population mean is less than or equal to the claimed value.
- H_0:$\mu \le \mu_0$
- Alternative Hypothesis (H_A): The population mean is greater than the claimed value.
- H_A:$\mu > \mu_0$

This is called an Upper Tail Test because we focus on the right side (higher values) of the normal distribution.

Test Statistic (Z-Score)

When the population variance (σ^2) is known, we use the Z-test:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Where:

- $\bar{x}$ = Sample Mean
- μ_0 = Population Mean under H_0
- σ = Population Standard Deviation
- n = Sample Size

The Z-score tells us how far $\bar{x}$ is from μ_0 in terms of standard deviations.

p-Value and Decision

- Find the p-value from the standard normal table (right-tailed probability).
- If $p < \alpha$ (e.g., 0.05), reject H_0(the population mean is significantly greater than μ_0).
- If $p \ge \alpha$, fail to reject H_0(not enough evidence to say μ is greater than μ_0).

Example: Upper Tail Test in R

Scenario:

A cereal company claims that the average weight of its cereal boxes is 500g. A quality control team collects a sample of 40 boxes, and the sample mean weight is 502g, with a known standard deviation of 5g. Test at α = 0.05 if the actual average weight is significantly greater than 500g.

Steps & R Code

Step 1: Define Hypotheses

- H0: μ ≤ 500 (Average weight is 500g or less)
- H1: μ > 500 (Average weight is greater than 500g)

Step 2: Given values

```
x_bar <- 502   # Sample Mean
mu_0 <- 500    # Claimed Population Mean
sigma <- 5     # Known Population Standard Deviation
n <- 40        # Sample Size
alpha <- 0.05  # Significance Level
```

Step 3: Calculate Z-Score

```
z <- (x_bar - mu_0) / (sigma / sqrt(n))
```

Step 4: Compute p-value for upper tail test

```
p_value <- 1 - pnorm(z)
```

Step 5: Decision

```
if (p_value < alpha) {
  result <- "Reject H0: The average weight is significantly greater
than 500g."
} else {
  result <- "Fail to reject H0: No significant evidence that average
weight is greater than 500g."
}

# Display results
z
p_value
Result
```

Interpreting the Output

- If p-value < 0.05 → Reject H_0 (Evidence suggests the cereal boxes weigh significantly more than 500g).
- If p-value ≥ 0.05 → Fail to reject H_0(No strong evidence to claim the weight is greater than 500g).

Conclusion

- An Upper Tail Test is used to check if the population mean is significantly greater than a given value.
- We use a Z-test when the population variance (σ^2) is known.
- The p-value helps determine if we should reject H_0

Activity 03

1. **Checking if the Average Monthly Electricity Bill is More Than $100**

   Scenario:

   A company claims that the average household electricity bill is **$100** per month. A sample of **35 households** has a mean bill of **$104** and a standard deviation of **$8**. Test at **α = 0.05**.

2. **Checking if the Average Body Temperature is More Than 98.6°F**

   Scenario:

   A medical study claims that the average human body temperature is **98.6°F**. A sample of **40 individuals** has a mean temperature of **98.9°F** and a standard deviation of **0.5°F**. Test at **α = 0.05**.

Two-Tailed Test of Population Mean with Known Variance

A Two-Tailed Test is used to check if the population mean is significantly different from a specified value (it could be greater than or less than the value). This test is performed when the population variance (σ^2) is known.

Understanding the Concept

We test whether the true mean (μ) of a population is different from a given value (μ_0), meaning it could be either larger or smaller than μ_0.

- Null Hypothesis (H_0): The population mean is equal to the claimed value.
  $H0{:}\mu=\mu_0$
- Alternative Hypothesis (H_A): The population mean is not equal to the claimed value.
  $H_A{:}\mu\neq\mu_0$

This is called a Two-Tailed Test because we check both sides (higher and lower values) of the normal distribution.

Test Statistic (Z-Score)

When the population variance (σ^2) is known, we use the Z-test:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Where:

- $\bar{x}$ = Sample Mean
- μ_0 = Population Mean under H_0
- σ = Population Standard Deviation
- n = Sample Size

The Z-score tells us how far $\bar{x}$ is from μ_0 in terms of standard deviations

p-Value and Decision

- Find the p-value by calculating the probability for both tails of the normal distribution.
- If $p<\alpha$ (e.g., 0.05), reject H_0(the population mean is significantly different from μ_0).
- If $p\geq\alpha$, fail to reject H_0(not enough evidence to say μ is different from μ_0).

For a Two-Tailed Test, the significance level α is split equally between both tails.

Example: Two-Tailed Test in R

Scenario:

A factory claims that the average weight of its product is 500g. A sample of 36 products has a mean weight of 498g, with a known standard deviation of 4g. Test at α = 0.05 if the actual average weight is different from 500g.

Steps & R Code

Step 1: Define Hypotheses

- H0: μ = 500 (Average weight is 500g)
- H1: μ ≠ 500 (Average weight is different from 500g)

Step 2: Given values

```
x_bar <- 498   # Sample Mean
mu_0 <- 500    # Claimed Population Mean
sigma <- 4     # Known Population Standard Deviation
n <- 36        # Sample Size
alpha <- 0.05  # Significance Level
```

Step 3: Calculate Z-Score

```
z <- (x_bar - mu_0) / (sigma / sqrt(n))
```

Step 4: Compute p-value for two-tailed test

```
p_value <- 2 * (1 - pnorm(abs(z)))
```

Step 5: Decision

```
if (p_value < alpha) {
  result <- "Reject H0: The average weight is significantly different
from 500g."
} else {
```

```
  result <- "Fail to reject H0: No significant evidence that average
weight is different from 500g."
}

# Display results
z
p_value
Result
```

Interpreting the Output

- If p-value < 0.05 → Reject $H0$ (Evidence suggests the weight is significantly different from 500g).
- If p-value ≥ 0.05 → Fail to reject $H0$ (No strong evidence to claim the weight is different from 500g).

Conclusion

- A Two-Tailed Test is used to check if the population mean is significantly different (either greater than or less than) a given value.
- We use a Z-test when the population variance (σ^2) is known.
- The p-value helps determine if we should reject H_0

Activity 04

1. **Checking if the Average Battery Life of a Laptop is Different from 10 Hours**

   ### Scenario:

   A laptop manufacturer claims that the average battery life of their latest model is **10 hours**. A sample of **36 laptops** is tested, and the observed **mean battery life is 9.5 hours** with a **known standard deviation of 1.2 hours**. Test at **α = 0.05** whether the battery life is significantly different from 10 hours.

2. **Checking if the Average Customer Waiting Time is Different from 15 Minutes**

   ### Scenario:

   A restaurant claims that the average waiting time for customers during peak hours is **15 minutes**. A sample of **49 customers** is observed, and their average waiting time is **16.2 minutes** with a **known standard deviation of 3.5 minutes**. Test at **α = 0.01** whether the actual waiting time is significantly different from 15 minutes.