# Session 8: Testing of Hypothesis using R

Jamuna S Murthy
Assistant Professor
Department of CSE

Academic Year - 2025-26

## 1 T-Test

A t-test compares the means of two independent groups to determine if they are significantly different.

Mathematically, the t-statistic is calculated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

### Example

```r
group1 <- c(23, 20, 19, 25, 22)
group2 <- c(30, 28, 32, 29, 27)
t.test(group1, group2)
```

**Output:**

```
t = -6.6905, df = 8, p-value = 0.00015
mean of x = 21.8, mean of y = 29.2
```

**Example 1 (Iris dataset)**

```r
data(iris)
setosa <- subset(iris, Species == "setosa")$Sepal.Length
versicolor <- subset(iris, Species == "versicolor")$
    Sepal.Length
t.test(setosa, versicolor)
```

**Example 2 (mtcars dataset)**

```r
data(mtcars)
auto <- subset(mtcars, am == 0)$mpg
manual <- subset(mtcars, am == 1)$mpg
t.test(auto, manual)
```

**Example 3 (ToothGrowth dataset)**

```r
data(ToothGrowth)
vc <- subset(ToothGrowth, supp == "VC")$len
oj <- subset(ToothGrowth, supp == "OJ")$len
t.test(vc, oj)
```

## 2 Paired T-Test

Used when comparing two related groups or paired data.

Mathematically, the paired t-statistic is:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $d$ is the difference between paired observations.

### Example

```r
before <- c(85, 70, 95, 90, 80)
after <- c(88, 74, 97, 93, 82)
t.test(before, after, paired=TRUE)
```

**Output:**

```
t = -4.3818, df = 4, p-value = 0.0117
mean difference = -2.8
```

## 3 Correlation

Correlation measures the strength and direction of a linear relationship between two variables.

Mathematically, Pearson's correlation coefficient (r) is calculated as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

### Example

```
x <- c(1, 2, 3, 4, 5)
y <- c(2, 4, 6, 8, 10)
cor.test(x, y)
```

**Output:**

```
correlation coefficient = 1, p-value < 2.2e-16
```

**Example 1 (mtcars dataset)**

```
data(mtcars)
cor.test(mtcars$mpg, mtcars$hp)
```

## 4 Chi-Square Test

Chi-Square tests the independence of categorical variables.

Mathematically:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O is observed frequency and E is expected frequency.

### Example

```
data <- matrix(c(10, 20, 20, 40), nrow=2)
chisq.test(data)
```

**Output:**

```
Chi-squared = 0, df = 1, p-value = 1
```

**Example 1 (Titanic dataset)**

```
data(Titanic)
tbl <- apply(Titanic, c(1,4), sum)
chisq.test(tbl)
```

## 5 Analysis of Variance (ANOVA)

ANOVA tests the difference between means of three or more groups.

Mathematically, ANOVA uses F-statistic:

$$F = \frac{MS_{between}}{MS_{within}}$$

## Example

```r
group1 <- c(10, 12, 14)
group2 <- c(20, 22, 24)
group3 <- c(30, 32, 34)
data <- data.frame(
  value = c(group1, group2, group3),
  group = factor(rep(c("g1", "g2", "g3"), each=3))
)
summary(aov(value ~ group, data=data))
```

**Output:**

```
F value = 150, p-value = 2.54e-06
```

**Example 1 (iris dataset)**

```r
data(iris)
summary(aov(Sepal.Width ~ Species, data=iris))
```

# 6 Correlation Analysis with Plot

Correlation analysis visualized with scatter plots.

## Example

```r
x <- c(5, 6, 7, 8, 9)
y <- c(15, 12, 18, 16, 20)
correlation <- cor(x, y)
plot(x, y, main=paste("Correlation=", round(correlation,
   2)))
abline(lm(y ~ x), col="blue")
```

**Output:** Scatter plot with regression line displaying correlation.

## Practice Questions

1. Perform a t-test comparing two groups with sample means 45 and 50.
2. Conduct a paired t-test for hypothetical pre- and post-intervention scores.
3. Compute correlation between study hours and exam scores.
4. Use Chi-Square to test independence of two categorical variables.
5. Run ANOVA for four groups with different hypothetical means.