

# Hypothesis Testing

## Hypothesis Testing

Hypothesis testing is a statistical method used to make decisions about a population parameter based on sample data. It involves making an initial assumption (the null hypothesis) and then determining whether the sample data provides enough evidence to reject this assumption in favor of an alternative hypothesis.

## Hypothesis

**Null Hypothesis ( $H_0$ ):** This is the default assumption that there is no effect or no difference. It is the hypothesis that researchers typically aim to test against.

$H_0$ : The mean weight of a population is 150 lbs.

**Alternative Hypothesis ( $H_1$  or  $H_a$ ):** This is the hypothesis that there is an effect or a difference. It is what researchers want to prove.

$H_1$ : The mean weight of a population is not 150 lbs.

## P-Value

The p-value is the probability of obtaining test results at least as extreme as the observed results, assuming the null hypothesis is true. It helps determine the significance of the results. A small p-value (typically  $< 0.05$ ) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.

## Test Statistic

A test statistic is a standardized value that is calculated from sample data during a hypothesis test. It is used to determine whether to reject the null hypothesis. The type of test statistic used depends on the type of data and the specific hypothesis test being performed (e.g., Z-test, t-test, chi-square test).

1. **Z-test:** Used for hypothesis testing with a known population sd.
2. **t-test:** Used for hypothesis testing with an unknown population sd.
3. **chi-square test:** Used for hypothesis testing with categorical data.

## Level of Significance ( $\alpha$ )

The level of significance is the threshold for rejecting the null hypothesis. It is denoted by  $\alpha$  and is usually set at 0.05 (5%). This means there is a 5% risk of concluding that a difference exists when there is no actual difference.

## Critical Region

The critical region is the set of all values of the test statistic that would lead to the rejection of the null hypothesis. It is determined by the level of significance. If the test statistic falls within the critical region, the null hypothesis is rejected.

## Critical Values

Critical values are the boundaries of the critical region. They are determined based on the level of significance and the distribution of the test statistic. For example, in a Z-test with ( $\alpha = 0.05$ ), the critical values are approx  $\pm 1.96$ .

## Steps in Hypothesis Testing

1. **State the Hypotheses:** Formulate the null and alternative hypotheses.
2. **Significance Level(alpha):** Decide the level of Significance (e.g, 0.05).
3. **Collect Data:** Gather sample data.
4. **Test Statistic:** Compute the test statistic based on the sample data.
5. **Critical Region:** Identify the critical region and critical values.
6. **Make a Decision:** Compare the test statistic to the critical values that If the test statistic falls in the critical region, reject null hypothesis. If the test statistic doesn't fall in the critical region, fail to reject the null hypothesis.
7. **Draw a Conclusion:** Interpret the result in context of research question.

## One-Sample Mean

When the Population Standard Deviation is Known:

```
one_sample_z_test <- function(sample, mu, sigma, confidence = 0.95) {  
  n <- length(sample)  
  mean <- mean(sample)  
  z <- (mean - mu) / (sigma / sqrt(n))  
  p_value <- 2 * (1 - pnorm(abs(z)))  
  return(list(z = z, p_value = p_value))  
}
```

```
sample <- c(10, 12, 14, 16, 18)  
mu <- 15  
sigma <- 2  
result <- one_sample_z_test(sample, mu, sigma)  
print(result)
```

```
## $z  
## [1] -1.118034  
##  
## $p_value  
## [1] 0.2635525
```

## When the Population Standard Deviation is Unknown

```
one_sample_t_test <- function(sample, mu, confidence = 0.95) {  
  t_test <- t.test(sample, mu = mu, conf.level = confidence)  
  return(t_test)  
}
```

```
# Example usage
```

```
result <- one_sample_t_test(sample, mu)  
print(result)
```

```
##  
## One Sample t-test  
##  
## data: sample  
## t = -0.70711, df = 4, p-value = 0.5185  
## alternative hypothesis: true mean is not equal to 15  
## 95 percent confidence interval:  
## 10.07351 17.92649  
## sample estimates:  
## mean of x  
## 14
```

## Two-Sample Mean

When the Population Standard Deviation is Known:

```
two_sample_z_test <- function(sample1, sample2, sigma1, sigma2, confidence = 0.95) {  
  n1 <- length(sample1)  
  n2 <- length(sample2)  
  mean1 <- mean(sample1)  
  mean2 <- mean(sample2)  
  z <- (mean1 - mean2) / sqrt((sigma1^2 / n1) + (sigma2^2 / n2))  
  p_value <- 2 * (1 - pnorm(abs(z)))  
  return(list(z = z, p_value = p_value))  
}
```

```
sample1 <- c(10, 12, 14, 16, 18)  
sample2 <- c(11, 13, 15, 17, 19)  
sigma1 <- 2  
sigma2 <- 2  
result <- two_sample_z_test(sample1, sample2, sigma1, sigma2)  
print(result)
```

```
## $z  
## [1] -0.7905694  
##  
## $p_value  
## [1] 0.4291953
```

## When the Population Standard Deviation is Unknown (Pooled Variance):

```
two_sample_t_test <- function(sample1, sample2, confidence = 0.95) {  
  t_test <- t.test(sample1, sample2, var.equal = TRUE, conf.level = confidence)  
  return(t_test)  
}
```

```
result <- two_sample_t_test(sample1, sample2)  
print(result)
```

```
##  
## Two Sample t-test  
##  
## data: sample1 and sample2  
## t = -0.5, df = 8, p-value = 0.6305  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.612008 3.612008  
## sample estimates:  
## mean of x mean of y  
## 14 15
```

## Dependent (Paired) Samples

```
# Example data  
sample1 <- c(10, 12, 14, 13, 15) # Before intervention  
sample2 <- c(11, 14, 13, 14, 16) # After intervention  
  
# Perform paired sample t-test  
t_test_result <- t.test(sample1, sample2, paired = TRUE)  
  
# Display the results  
print(t_test_result)
```

```
##  
## Paired t-test  
##  
## data: sample1 and sample2  
## t = -1.633, df = 4, p-value = 0.1778  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## -2.1601748 0.5601748  
## sample estimates:  
## mean difference  
## -0.8
```

## One-Sample Proportion

```
one_sample_proportion_test <- function(x, n, p, confidence = 0.95) {  
  prop_test <- prop.test(x, n, p = p, conf.level = confidence)  
  return(prop_test)  
}
```

```
x <- 40  
n <- 100  
p <- 0.5  
result <- one_sample_proportion_test(x, n, p)  
print(result)
```

```
##  
## 1-sample proportions test with continuity correction  
##  
## data: x out of n, null probability p  
## X-squared = 3.61, df = 1, p-value = 0.05743  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3047801 0.5029964  
## sample estimates:  
## p  
## 0.4
```

## Two-Sample Proportion

```
two_sample_proportion_test <- function(x1, n1, x2, n2, confidence = 0.95) {  
  prop_test <- prop.test(c(x1, x2), c(n1, n2), conf.level = confidence)  
  return(prop_test)  
}
```

```
x1 <- 40  
n1 <- 100  
x2 <- 50  
n2 <- 120  
result <- two_sample_proportion_test(x1, n1, x2, n2)  
print(result)
```

```
##  
## 2-sample test for equality of proportions with continuity correction  
##  
## data: c(x1, x2) out of c(n1, n2)  
## X-squared = 0.012692, df = 1, p-value = 0.9103  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.1562183 0.1228849  
## sample estimates:  
## prop 1 prop 2  
## 0.4000000 0.4166667
```

## One Categorical Variable (Chi-Square Goodness of Fit Test)

```
chi_square_goodness_of_fit <- function(observed, expected) {  
  chi_test <- chisq.test(observed, p = expected)  
  return(chi_test)  
}
```

```
observed <- c(50, 30, 20)  
expected <- c(0.5, 0.3, 0.2)  
result <- chi_square_goodness_of_fit(observed, expected)  
print(result)
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: observed  
## X-squared = 0, df = 2, p-value = 1
```

## Two Categorical Variables (Chi-Square Test of Independence)

```
chi_square_independence <- function(table) {  
  chi_test <- chisq.test(table)  
  return(chi_test)  
}
```

```
table <- matrix(c(10, 20, 30, 40), nrow = 2)  
result <- chi_square_independence(table)  
print(result)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: table  
## X-squared = 0.44643, df = 1, p-value = 0.504
```

## Errors and Power

Hypothesis Errors When conducting a hypothesis test, there are two types of errors that can occur:

**Type I Error(alpha):** This occurs when the null hypothesis ( $H_0$ ) is true, but we incorrectly reject it. The probability of making a Type I error is denoted by the significance level (alpha), typically set at 0.05. This means there is a 5% risk of rejecting the null hypothesis when it is actually true.

**Type II Error(beta):** This occurs when the null hypothesis is false, but we fail to reject it. The probability of making a Type II error is denoted by (beta). Unlike (alpha), (beta) is not typically set by the researcher but is influenced by factors such as sample size, effect size, and variability.

# Statistical Power

Statistical power is the probability that a test correctly rejects a false null hypothesis. It is calculated as  $(1 - \beta)$ . High power means a lower probability of making a Type II error. Power is influenced by:

1. **Sample Size**: Larger sample sizes generally increase power.
2. **Effect Size**: Larger sizes make it easier to detect a true effect of increasing power.
3. **Significance Level(alpha)**: Increasing (alpha) increases power but also increases the risk of a Type I error.
4. **Variability**: Lower variability in the data increases power.

```
set.seed(123)
sample <- rnorm(30, mean = 5.5, sd = 1)

# Null hypothesis: mean = 5
mu <- 5

# Conducting the t-test
t_test <- t.test(sample, mu = mu)
print(t_test)

##
## One Sample t-test
##
## data: sample
## t = 2.5286, df = 29, p-value = 0.01715
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
## 5.086573 5.819219
## sample estimates:
## mean of x
## 5.452896

# Type I Error (alpha)
alpha <- 0.05

# Type II Error (beta) and Power
# Using power.t.test function to calculate power
power_analysis <- power.t.test(n = length(sample),
                              delta = mean(sample) - mu,
                              sd = sd(sample),
                              sig.level = alpha,
                              type = "one.sample",
                              alternative = "two.sided")
print(power_analysis)

##
## One-sample t test power calculation
##
## n = 30
## delta = 0.4528962
## sd = 0.9810307
```

```
##          sig.level = 0.05
##          power = 0.6858779
##          alternative = two.sided

# Extracting beta and power
beta <- 1 - power_analysis$power
power <- power_analysis$power

cat("Type I Error (alpha):", alpha, "\n")

## Type I Error (alpha): 0.05

cat("Type II Error (beta):", beta, "\n")

## Type II Error (beta): 0.3141221

cat("Statistical Power:", power, "\n")

## Statistical Power: 0.6858779
```