

Hypothesis_Testing

S20426

2025-03-15

01 Proportion Testing

One Sampling

Large Sample

```
set.seed(10)
Hair_col <- c(rep("black", 1500), rep("brown", 1000), rep("blonde", 500))
sampleP <- sample(Hair_col, 1000)
Ptable <- table(sampleP)
prop.test(x = 498, n = 1000, p = 0.5, alternative = "two.sided", conf.level = 0.95, correct = FALSE)
```

```
##
## 1-sample proportions test without continuity correction
##
## data: 498 out of 1000, null probability 0.5
## X-squared = 0.016, df = 1, p-value = 0.8993
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
## 0.4670775 0.5289378
## sample estimates:
## p
## 0.498
```

Small Sample

```
set.seed(10)
sampleS <- sample(Hair_col, 10)
Stable <- table(sampleS)
Stable
```

```
## sampleS
## black blonde brown
##      5      1      4
```

```
binom.test(x = 3, n = 10, p = 0.5, alternative = "two.sided" )
```

```
##
## Exact binomial test
##
## data: 3 and 10
## number of successes = 3, number of trials = 10, p-value = 0.3438
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.06673951 0.65245285
## sample estimates:
## probability of success
## 0.3
```

Two Sampling

```
prop.test(x=c(490,400),n=c(500,500))
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: c(490, 400) out of c(500, 500)
## X-squared = 80.909, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## 0.1408536 0.2191464
## sample estimates:
## prop 1 prop 2
## 0.98 0.80
```

```
? prop.test()
```

```
## starting httpd help server ... done
```

02 Mean Testing

One Sample

One Sample Test for Mean: when population variance is known.

```
weights=c(165.1,171.5,168.1,165.6,166.8,170.0,168.8,171.1,168.8,173.6,163.5,169.9,165.4,174.
4,171.8,166.0,174.6,174.5,166.4,173.8)

library("BSDA")
```

```
## Warning: package 'BSDA' was built under R version 4.3.3
```

```
## Loading required package: lattice
```

```
##  
## Attaching package: 'BSDA'
```

```
## The following object is masked _by_ '.GlobalEnv':  
##  
##      Stable
```

```
## The following object is masked from 'package:datasets':  
##  
##      Orange
```

```
z.test(x=weights, mu = 160, sigma.x = 15, alternative = "two.sided")
```

```
##  
## One-sample z-Test  
##  
## data: weights  
## z = 2.8279, p-value = 0.004686  
## alternative hypothesis: true mean is not equal to 160  
## 95 percent confidence interval:  
## 162.9111 176.0589  
## sample estimates:  
## mean of x  
## 169.485
```

One Sample Test for Mean: when population variance is unknown.

```
weights <- c(165.1,171.5,168.1,165.6,166.8,170.0,168.8,171.1,168.8,173.6,163.5,169.9,165.4,174.4,171.8,166.0,174.6,174.5,166.4,173.8)  
t.test(weights, mu = 160, alternative = "two.sided")
```

```
##  
## One Sample t-test  
##  
## data: weights  
## t = 11.884, df = 19, p-value = 3.05e-10  
## alternative hypothesis: true mean is not equal to 160  
## 95 percent confidence interval:  
## 167.8146 171.1554  
## sample estimates:  
## mean of x  
## 169.485
```

Two Sample

Case 1 : Equal Variances

```
fat_m <- c(13.3,6.0,20.0,8.0,14.0,19.0,18.0,25.0,16.0,24.0,15.0,1.0,15.0)  
fat_w <- c(22.0,16.0,21.7,21.0,30.0,26.0,12.0,23.2,28.0,23.0)  
t.test(fat_w, fat_m, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: fat_w and fat_m
## t = 2.8, df = 21, p-value = 0.01073
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.889353 12.798339
## sample estimates:
## mean of x mean of y
## 22.29000 14.94615
```

Case 2 : Unequal Variances

```
fat_m <- c(13.3,6.0,20.0,8.0,14.0,19.0,18.0,25.0,16.0,24.0,15.0,1.0,15.0)
fat_w <- c(22.0,16.0,21.7,21.0,30.0,26.0,12.0,23.2,28.0,23.0)

t.test(fat_w,fat_m, var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: fat_w and fat_m
## t = 2.8958, df = 20.989, p-value = 0.00865
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 2.069692 12.618000
## sample estimates:
## mean of x mean of y
## 22.29000 14.94615
```

Paired Sample

The mean of paired samples refers to the mean difference between two related (dependent) samples. This is typically used in paired t-tests or before-and-after studies, where each observation in one sample has a corresponding observation in the other sample.

```
lab1 <- c(22,18,28,26,13,8,21,26,27,29,25,24,22,28,15)
lab2 <- c(25,21,31,27,11,10,25,26,29,28,26,23,22,25,17)

t.test(lab1,lab2, paired = TRUE)
```

```
##
## Paired t-test
##
## data: lab1 and lab2
## t = -1.7618, df = 14, p-value = 0.09991
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -2.0695338 0.2028671
## sample estimates:
## mean difference
## -0.9333333
```

03 Testing For Variance

One sample

```
weights <- c(165.1,171.5,168.1,165.6,166.8,170.0,168.8,  
            171.1,168.8,173.6,163.5,169.9,165.4,174.4,  
            171.8,166.0,174.6,174.5,166.4,173.8)
```

```
library(EnvStats)
```

```
## Warning: package 'EnvStats' was built under R version 4.3.3
```

```
##  
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':  
##  
##   predict, predict.lm
```

```
varTest(weights, sigma =10, alternative = "two.sided")
```

```
##  
## Results of Hypothesis Test  
## -----  
##  
## Null Hypothesis:                variance = 10  
##  
## Alternative Hypothesis:         True variance is not equal to 10  
##  
## Test Name:                     Chi-Squared Test on Variance  
##  
## Estimated Parameter(s):        variance = 12.73924  
##  
## Data:                          weights  
##  
## Test Statistic:                Chi-Squared = 24.20455  
##  
## Test Statistic Parameter:      df = 19  
##  
## P-value:                       0.3768455  
##  
## 95% Confidence Interval:       LCL = 7.367682  
##                               UCL = 27.176225
```

Two sample

```
fat_m <- c( 13.3,6.0,20.0,8.0,14.0,19.0,18.0,25.0,16.0,24.0,15.0,1.0,15.0)
fat_w <- c(22.0,16.0,21.7,21.0,30.0,26.0,12.0,23.2,28.0,23.0)

var.test(fat_m, fat_w)
```

```
##
## F test to compare two variances
##
## data: fat_m and fat_w
## F = 1.6545, num df = 12, denom df = 9, p-value = 0.4561
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.4277224 5.6846913
## sample estimates:
## ratio of variances
## 1.654525
```

04 Check the Normality Assumption

Shapiro-Wilk test

```
# shapiro.test(my_data)
```

Anderson-Darling test

```
# install.packages('nortest')
# library(nortest)
# ad.test(my_data)
```

Kolmogorov-Smirnov test

```
# ks.test(my_data, "pnorm")
```

05. Sign Test

The Sign Test is a non-parametric statistical test used to determine whether there is a significant difference between the median of a sample and a hypothesized median, or to compare two paired samples. Unlike parametric tests, it does not assume normality and is useful when dealing with ordinal or skewed data.

When to Use the Sign Test?

One-Sample Sign Test: Used to test if the median of a single sample is equal to a hypothesized value. Paired-Sample (Two-Sample) Sign Test: Used for paired data (dependent samples) to determine if there is a difference in the medians.

Case 1 : Sign Test For One Sample

```
m0 <- 130 # median in 2008
prices <- c(230.00,148.00,126.00,134.62,155.00,157.70,
           160.00,225.00,125.00,109.00,157.00,115.00,
           125.00,225.00,118.00,179.00,176.00,125.00,
           123.00,180.00,151.00,120.00,143.00,170.00,
           190.00,233.00,148.72,189.00,121.00,149.00,
           225.00,240.00)
```

```
library(BSDA)
SIGN.test(prices,md = m0, alternative="greater")
```

```
##
## One-sample Sign-Test
##
## data: prices
## s = 22, p-value = 0.02505
## alternative hypothesis: true median is greater than 130
## 95 percent confidence interval:
## 141.5795 Inf
## sample estimates:
## median of x
## 153
##
## Achieved and Interpolated Confidence Intervals:
##
##
```

	Conf.Level	L.E.pt	U.E.pt
## Lower Achieved CI	0.9449	143.0000	Inf
## Interpolated CI	0.9500	141.5795	Inf
## Upper Achieved CI	0.9749	134.6200	Inf

Case 1 :Sign Test for One Sample

```
lab1 <- c(22,18,28,26,13,8,21,26,27,29,25,24,22,28,15)
lab2 <- c(25,21,31,27,11,10,25,26,29,28,26,23,22,25,17)
```

```
library(BSDA)
SIGN.test(x = lab1,
          y = lab2,
          alternative = "two.sided",
          conf.level = 0.95)
```

```
##
## Dependent-samples Sign-Test
##
## data: lab1 and lab2
## S = 4, p-value = 0.2668
## alternative hypothesis: true median difference is not equal to 0
## 95 percent confidence interval:
## -2.8218315 0.8218315
## sample estimates:
## median of x-y
## -1
##
## Achieved and Interpolated Confidence Intervals:
##
## Conf.Level L.E.pt U.E.pt
## Lower Achieved CI 0.8815 -2.0000 0.0000
## Interpolated CI 0.9500 -2.8218 0.8218
## Upper Achieved CI 0.9648 -3.0000 1.0000
```

06. Wilcoxon Sign Rank Test

Case I: One Sample Sign Rank Test

```
m0 <- 130 # median in 2008
prices <- c(230.00,148.00,126.00,134.62,155.00,157.70,
           160.00,225.00,125.00,109.00,157.00,115.00,
           125.00,225.00,118.00,179.00,176.00,125.00,
           123.00,180.00,151.00,120.00,143.00,170.00,
           190.00,233.00,148.72,189.00,121.00,149.00,
           225.00,240.00)

wilcox.test(prices,m0, exact = FALSE, alternative = "greater")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: prices and m0
## W = 22, p-value = 0.2816
## alternative hypothesis: true location shift is greater than 0
```

Case II: Sign Rank Test for Two Paired Samples

```
lab1 <- c(22,18,28,26,13,8,21,26,27,29,25,24,22,28,15)
lab2 <- c(25,21,31,27,11,10,25,26,29,28,26,23,22,25,17)

wilcox.test(lab1, lab2, paired = TRUE, exact = FALSE)
```



```
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data: lab1 and lab2  
## V = 22, p-value = 0.1047  
## alternative hypothesis: true location shift is not equal to 0
```

Case III: Sign Rank Test for Two independent Samples

```
fat_m <- c( 13.3,6.0,20.0,8.0,14.0,19.0,18.0,25.0,16.0,24.0)  
fat_w <- c(22.0,16.0,21.7,21.0,30.0,26.0,12.0,23.2,28.0,23.0)  
wilcox.test(fat_m, fat_w, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(fat_m, fat_w, alternative = "two.sided"): cannot  
## compute exact p-value with ties
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: fat_m and fat_w  
## W = 23.5, p-value = 0.04928  
## alternative hypothesis: true location shift is not equal to 0
```