

Hypothesis Testing-1

2026-01-30

- **Test of single proportion** (One sample proportion test)
- **Test of single mean with known population variance** (One sample mean z-test)
- **Test of single mean with unknown population variance** (One sample mean t-test)

One sample proportion test

$$H_0 : P = p_0$$

$$H_1 : P > p_0 \text{ or } H_1 : P < p_0 \text{ or } H_1 : P \neq p_0$$

A built-in function:

```
prop.test(x, n, p = p_0, alternative = "two.sided", correct = FALSE)
```

x - the number of successes

n - the number of trials

p_0 - the hypothesized value of population proportion

alternative - "two.sided" or "greater" or "less"

correct = FALSE works for z-test ; TRUE works for Yate's continuity correction

Example:

```
prop.test(36, 40, p = 0.85, alternative = "greater", correct=FALSE)
```

```
##
## 1-sample proportions test without continuity correction
##
## data: 36 out of 40, null probability 0.85
## X-squared = 0.78431, df = 1, p-value = 0.1879
## alternative hypothesis: true p is greater than 0.85
## 95 percent confidence interval:
## 0.7950094 1.0000000
## sample estimates:
## p
## 0.9
```

```
ospt=prop.test(36, 40, p = 0.85, alternative = "greater", correct=FALSE) #prop.test() returns a test
ospt$statistic # Chi-square value
```

```
## X-squared
## 0.7843137
```

```
ospt$estimate # Sample proportion
```

```
## p
## 0.9
```

```
ospt$conf.int # Confidence interval
```

```
## [1] 0.7950094 1.0000000
## attr(,"conf.level")
## [1] 0.95
```

```
ospt$null.value      # Hypothesized proportion
```

```
##      p
## 0.85
```

```
ospt$p.value        # p-value
```

```
## [1] 0.1879125
```

```
ospt$parameter      # degrees of freedom
```

```
## df
## 1
```

Problem 1: 40 people were attacked by a disease and only 36 survived. At 5% LOS, test whether the survival rate attacked by this disease is more than 85 %.

$$\alpha = 0.05, p = \frac{36}{40} = 0.9, P_0 = 0.85$$

$$H_0 : P = 0.85$$

$$H_1 : P > 0.85$$

(Right tailed test) Test statistic:

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

follows a standard normal distribution.

R-code (using p-value)

Reject H_0 if

$$\text{p-value} < \alpha$$

If the conclusion is Reject H_0 , then the test results are statistically significant at α level. Else, the test results are not statistically significant at α level.

R-code (using p-value)

```
x <- 36
n <- 40
p0 <- 0.85
alpha <- 0.05
ospt=prop.test(x, n, p = p0, alternative = "greater", correct=FALSE) #prop.test() returns a test object
if (ospt$p.value<alpha) print("Reject H_0") else print("Fail to reject H_0")
```

```
## [1] "Fail to reject H_0"
```

R-code (using critical value)

Reject H_0 if

$$\text{calculated } z > z_{\alpha}$$

```
x <- 36
n <- 40
p0 <- 0.85
alpha <- 0.05
ospt=prop.test(x, n, p = p0, alternative = "greater", correct=FALSE) #prop.test() returns a test object
calc_z=sqrt(ospt$statistic) # Test statistic value
cv=qnorm(1-alpha)
if (calc_z>cv) print("Reject H_0") else print("Fail to reject H_0")

## [1] "Fail to reject H_0"
```

Same problem using reusable testing function:

```
prop_pvalue <- function(x, n, p0, alt = "two.sided") {
  prop.test(x, n, p = p0, alternative = alt, correct = FALSE)$p.value
}

# Example
prop_pvalue(36, 40, 0.85, "greater")

## [1] 0.1879125

if (prop_pvalue(36,40,0.85,"greater")<alpha) print("Reject H_0") else print("Fail to reject H_0")

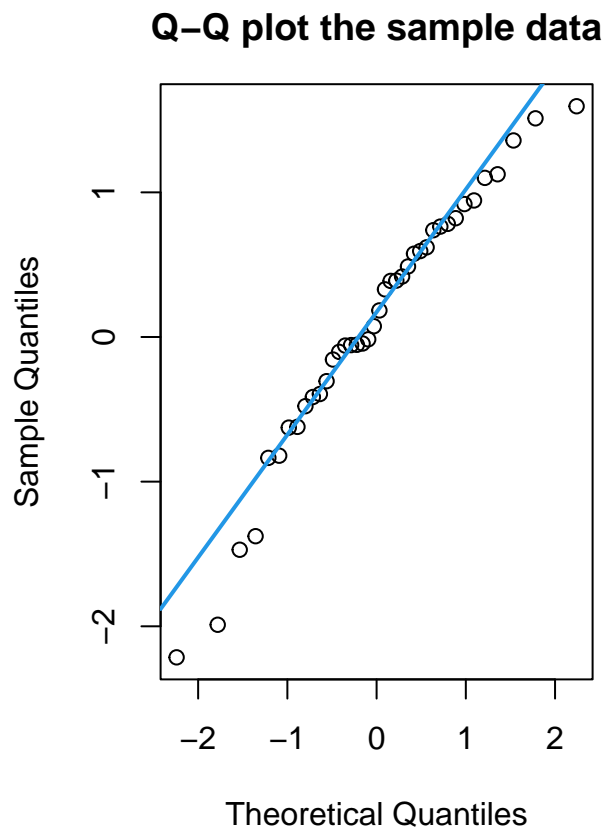
## [1] "Fail to reject H_0"
```

Assumption Checking using plots

We test the normality assumption by plotting a normal probability plot (Q-Q plot-Compares sample quantiles with theoretical normal quantiles). If the variable is normally distributed, the normal probability plot should be roughly linear.

```
par(mfrow = c(1, 2), mar = c(4, 4, 3, 1))
set.seed(1) # to maintain the consistency
x <- rnorm(40)

qqnorm(x, main = "Q-Q plot the sample data")
qqline(x, col = 4, lwd = 2)
```



Test of single mean with known population variance

Problem 2: A soft drink company claims that the mean amount of sugar in its 2-litre bottles is 50 grams. The population standard deviation of sugar content is known to be 4 grams. A random sample of 36 bottles is tested, and the sample mean sugar content is found to be 48.8 grams. At the 5% level of significance, test whether the mean sugar content is less than the claimed value.

```
xbar <- 48.8 # sample mean
mu0 <- 50    # hypothesized mean
sigma <- 4   # population standard deviation
n <- 36      # sample size
alpha=0.05
# Test statistic
z <- (xbar - mu0) / (sigma / sqrt(n))
z
```

```
## [1] -1.8
```

```
# P-value for left-tailed test
p_value <- pnorm(z)
```

```
p_value
```

```
## [1] 0.03593032
```

```
if (p_value<alpha) print("Reject H_0") else print("Fail to reject H_0")
```

```
## [1] "Reject H_0"
```

Confidence Interval for the population mean

```
xbar <- 48.8 # sample mean
mu0 <- 50    # hypothesized mean
sigma <- 4    # population standard deviation
n <- 36      # sample size
alpha=0.05
# Test statistic
z <- (xbar - mu0) / (sigma / sqrt(n))
z
```

```
## [1] -1.8
```

```
cv <- qnorm(1-alpha/2)
y=c(-cv,cv) # critical region
y
```

```
## [1] -1.959964 1.959964
```

```
xbar+(y*sigma)/sqrt(n) # Confidence interval
```

```
## [1] 47.49336 50.10664
```

Test of single mean with unknown population variance

Problem 3:

A company claims that the average weight of its packaged rice bags is 5 kg. To verify this claim, a random sample of 25 bags is selected. The sample has a mean weight of 5.12 kg and a sample standard deviation of 0.40 kg. At the 5% level of significance, test whether the mean weight of the rice bags differs from 5 kg.

```
xbar <- 5.12 # sample mean
s <- 0.40    # sample standard deviation
n <- 25      # sample size
mu0 <- 5     # hypothesized mean
alpha <- 0.05
set.seed(1)  # to maintain the consistency
x <- rnorm(n, mean = xbar, sd = s) # generate random numbers for the sample
pv=t.test(x,mu = mu0, alternative = "two.sided",conf.level=0.99)$p.value
if (pv< alpha) print("Reject H_0") else print("Fail to reject H_0")
```

```
## [1] "Reject H_0"
```