Non-Parametric Tests for Location and Dispersion

Kamesh Dubey

2024-06-06

1. Test For Location

1.1 Sign Test

The one-sample sign test is used to determine if the median of a single sample $\tilde{x}$ is equal to a specified value.

Step 1: Define the Hypothesis

$$H_0 : \tilde{x} = M_0$$
$$H_1 : \tilde{x} \neq M_0$$

Step 2: Define:

- $N^+ = \#\ x_i > M_0$.
- $N^- = \#\ x_i < M_0$.

Using above, count the number of observations above (N^+) and below (N^-) the hypothesized median M_0

Step 3: The test statistic is the smaller of N^+ and N^-, which follows a binomial distribution Binomial$(N, 0.5)$, where $N = N^+ + N^-$.

Step 4: The p-value from this binomial test, comparing N^+ and N^- to the expected distribution under H_0, determines if we reject H_0.

```
# Sample data
x = c(2, 3, -1, 4, 5, -2, 3, -3, 6, 7, 1, -1)

                    # step1
#Hypothesized median
median = 0

                  #step 2
#Calculate the differences from the hypothesized median
differences = x - median

# Count the number of positive, negative, and zero differences
positive = sum(differences > 0)
negative = sum(differences < 0)
```

```
zero = sum(differences == 0)


# Perform the sign test using the binom.test function
# Note: We exclude zero differences from the test
result = binom.test(x = c(positive, negative), alternative = "two.sided")

print(result)
```

```
##
##  Exact binomial test
##
## data:  c(positive, negative)
## number of successes = 8, number of trials = 12, p-value = 0.3877
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3488755 0.9007539
## sample estimates:
## probability of success
##               0.6666667
```

Interpreting the results

The p-value of 0.3877 suggests that there is insufficient evidence to reject the null hypothesis at a significance level of 0.05. Therefore, based on this test, we cannot conclude that the median of the sample significantly differs from the 0.

```
y = c(2, 3, 5, 0, 4, 5, 1, -2, 3, -3, 6, 7, 1, -1)

# Hypothesized median
median0 = 0

# Calculate the differences from the hypothesized median
differences = y - median0



# Count the number of positive, negative, and zero differences
positive <- sum(differences > 0)
negative <- sum(differences < 0)
zero <- sum(differences == 0)


result = binom.test(x = c(positive, negative), alternative = "greater")

print(result)
```

Lets say our data has some 0's as observation, i.e. some observations are equal to mean. What to do in that case ?

```
## 
##  Exact binomial test
## 
## data:  c(positive, negative)
## number of successes = 10, number of trials = 13, p-value = 0.04614
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.5053503 1.0000000
## sample estimates:
## probability of success
##              0.7692308
```

Interpreting the results

The p-value of 0.04614 suggests that the *observed data provides evidence against the null hypothesis.* Hence, We reject Null in favour of alternate.

Note: In the context of the sign test, $x_i = 0$ observations are disregarded because they provide no evidence for or against the alternative hypothesis. Including them would not affect the test's outcome, as the sign test focuses solely on the direction of differences rather than their magnitude. Thus, excluding $x_i = 0$ simplifies the analysis without compromising its validity.

What happens when number of observation in the sample are large ? When the number of observations is large, such as $|X| > 30$, we can still perform a one-sample sign test, but the **exact binomial test** may not be the most appropriate approach due to computational limitations and because of that the normal approximation might be more accurate in this case.

```r
y = runif(40, min = -10, max = 10)


# Hypothesized median
median0 = 0

# Calculate the differences from the hypothesized median
differences = y - median0



# Count the number of positive, negative, and zero differences
positive <- sum(differences > 0)
negative <- sum(differences < 0)
zero <- sum(differences == 0)

# Total number of observations
n <- length(y)

# Calculate the standard error
se <- sqrt(n * 0.25) #sqrt(np(1-p))

# Calculate the z-score note N+ is used as test statistics
z <- (positive + 0.5 - 0.5 * n) / se

# Calculate the p-value using the normal approximation
p_value <- 2 * pnorm(abs(z), lower.tail = FALSE)
```

```r
# Output the result
cat("Z-score:", z, "\n")
```

```
## Z-score: 1.106797
```

```r
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.2683816
```

Here, p>0.05. Then, we can't reject null. Hence, we can reject null hypothesis, ie $\tilde{x} = 0$

The assumptions of one-sample sign test are:

- The observations in the sample must be independent of each other
- Ordinal or continuous data, so at atleast ranking is possible
- No need for a symmetric distribution.