Nonparametric hypothesis tests

'Nonparametric' means 'distribution-free'. Nonparametric tests do NOT make any assumption about the distribution. Note that t test is a parametric test because it assumes that the data follow normal distribution.

Sign test

Let $X_1, X_2, \cdots, X_n$ be a random sample from a continuous distribution. Suppose, we want to test the hypothesis that the median of the distribution, denoted by m, is less than a specified value m_0. Here, we do not test hypotheses about mean because we cannot find the distribution of the test statistic. Besides, for some distributions mean does not exist. Thus,

$H_0:\ m = m_0$

$H_1:\ m < m_0$

Let

$$I_i = \begin{cases} 1, & X_i < m_0 \\ 0, & X_i \geq m_0 \end{cases}$$

Then, I_i are independent Bernoulli random variables with parameter $p = 0.5$ under the null hypothesis (because 50% values are less than median). The test-statistic is $\sum_{i=1}^{n} I_i$ that follows binomial$(n, 0.5)$ under H_0.

Let for a particular data set $\sum_{i=1}^{n} I_i = v$. Then we can calculate p-value as follows:

$$\text{p-value} = P\left(\sum_{i=1}^{n} I_i \ \leq v\right)$$

which can be calculated using the binomial distribution.

Since the value of $v = \sum_{i=1}^{n} I_i$ depends on the signs of $(X_i - m_0)$, the test is called the sign test.

For the upper-tailed test

$H_0:\ m = m_0$

$H_1:\ m > m_0$

we define

$$I_i = \begin{cases} 1, & X_i > m_0 \\ 0, & X_i \leq m_0 \end{cases}$$

$$\text{p-value} = P\left(\sum_{i=1}^{n} I_i \geq v\right)$$

Example

Consider the data: 9, 10,11, 11, 12, 13, 15, 15, 17, 19, 21. Can we conclude that the population median is less than 14?

Solution

Here, $I = 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0$. Thus, $\sum_{i=1}^{n} I_i = 6$.

Under the null, $\sum_{i=1}^{11} I_i \sim \text{binomial}(11, 0.5)$. Thus,

$$\text{p-value} = P\left(\sum_{i=1}^{11} I_i \leq 6\right) = 0.726.$$

Therefore, we cannot reject the null hypothesis.

Two-tailed sign tests

For two-tailed tests, we define

$$I_i = \begin{cases} 1, & X_i < m_0 \\ 0, & X_i \geq m_0 \end{cases}$$

$$\text{p-value} = \begin{cases} 2\,P\left(\sum_{i=1}^{n} I_i \leq v\right) & \text{if } v \leq \frac{n}{2} \\ 2\,P\left(\sum_{i=1}^{n} I_i \geq v\right) & \text{if } v \geq \frac{n}{2} \end{cases}$$

Exercise

A random sample of 20 middle-aged men was selected to test whether the median systolic blood pressure is 128. What is your conclusion if 12 men have readings above 128?

Solution

$$\sum_{i=1}^{20} I_i = 12 > \frac{20}{2}$$

p-value $= 2\, P\left(\sum_{i=1}^{20} I_i \ \geq 12\right) = 0.503$

We cannot reject the null hypothesis and conclude that the median is 128.

Signed rank test

The sign test ignores the magnitude of the differences $(X_i - m_0)$. A nonparametric test that takes the magnitude of the differences into account is the signed rank test.

We want to test the hypotheses

$H_0\colon\ m = m_0$

$H_1\colon\ m < m_0$

Let $Y_i = X_i - m_0$. We first rank the absolute differences $|Y_i|$. Let

$$I_i = \begin{cases} 1, & X_i < m_0 \\ 0, & X_i \geq m_0 \end{cases}$$

The test statistic is

$$T_{-} = \sum_{i=1}^{n} i\, I_i$$

Under the null hypothesis,

$$E(T_{-}) = \frac{n(n+1)}{4}$$

$$V(T_{-}) = \frac{n(n+1)(2n+1)}{24}$$

For moderately large n, the distribution of T_{-} is normal with the above mean and variance. We can calculate the p-value from the Z table:

p-value $= P(T_{-} \geq t)$

Example

Consider the data: 6, 10, 9, 13, 17, 20, 21. Can we conclude that the population median is less than 14?

Solution

The differences are: -8, -4, -5, -1, 3, 6, 7.
The absolute differences are: 8, 4, 5, 1, 3, 6, 7.
Ranks are: 7, 3, 4, 1, 2, 5, 6

$$T_{-} = \sum_{i=1}^{7} i\, I_i = 15$$

Do the rest part yourself.