Lecture 2: Sign Test

Caden Hewlett

2024-01-12

Lecture 2: The Sign Test

Sometimes, when the distribution of the data is unknown, we still want to compute test. These methods are referred to as *nonparametric tests.*

The first of these we will discuss is the **Sign Test**.

In order to discuss the sign test, let's get a refresher on the median.

The Median

The median is essentially the value or point θ wherein:

$$P(X > \theta) = P(X < \theta)$$

It is a measure of *center,* and can be considered the middle 50% of the data.

To calculate the median of a set of data by hand, you would do one of the following: 1. n is odd: Order the data, then take the middle most point. 2. n is even: Order the data, then take the average of the two middle most points.

Example Let $\vec{x}_1 = \{2, 3, 1, 7, 8\}$. We order the data to be $\{1, 2, 3, 7, 8\}$. Then, the middlemost point is 3. Hence, $\text{med}(\vec{x}_1) = 3$.

```
x1 = c(2,3,1,7,8)
median(x1)
```

```
## [1] 3
```

Example Let $\vec{x}_2 = \{2, 8, 4, 6\}$. We order the data to be $\{2, 4, 6, 8\}$. Then, the average of the two middlemost points is 5. Hence, $\text{med}(\vec{x}_2) = 5$.

```
x2 = c(2, 8, 4, 6)
median(x2)
```

```
## [1] 5
```

The Sign Test: Theory

Let's say that we want to test whether or not the median of a given set of data is some hypothesized value, θ.

In theory, then, half of our observations $x_1, x_2, \ldots x_n$ should be below the median, and the other half should be above.

Thus, for each of our observations, we subtract the hypothesized median, determine the sign of each element of $\vec{x}$, and determine the proportion of positives in these signs.

Let's call this proportion t. If θ is in fact the median, then $t \sim \text{binomial}(n, \frac{1}{2})$. We then can use properties of expectation and variance of the Binomial distribution to draft a test statistic the follows a standard normal distirbution (recall: the one-proportion Z Test.)

<u>**NOTE**</u> The only assumption we need is that the data are **independent.**

The Sign Test: Steps

To conduct the Sign Test, we follow this procedure:

1. Declare your Hypotheses. Our null hypothesis is that the median of X is θ.

$$H_0 : \text{med}(X) = \theta \qquad H_A : \text{med}(X) \neq \theta$$

2. Given observations $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ we define "$\vec{x}$ centered"

$$\vec{x}_\otimes = \vec{x} - \theta$$

3. Then, count the number of elements of $\vec{x}_\otimes$ greater than zero. We will denote this variable t.

$$t = \sum_{i=1}^{n} \mathbb{I}\big(\{x_\otimes\}_i > 0\big)$$

4. Then, t takes a Binomial distribution.

$$t \sim \text{bin}\big(n, 1/2\big) \qquad \mathbb{E}(t) = \frac{n}{2} \qquad \text{var}(t) = \frac{n}{4}$$

5a.: Before calculating the observed test stat, if $n < 20$, make a continuity correction. *Alternatively, use exact values from the binomial PMF.*

i. For the lower alternative $H_A : \text{med}(X) < \theta$, *add* $1/2$ to the numerator.

ii. For the lower alternative $H_A : \text{med}(X) > \theta$, *subtract* $1/2$ to the numerator.

5. Then, we have the Observed Test Statistic

$$T_{\text{obs}} = \frac{t - \dfrac{n}{2}}{\sqrt{\dfrac{n}{4}}} \sim N(0, 1)$$

6. Make conclusions - inspect the critical value calculated based on your declared α, or calculate the p-value with the standard normal distribution.

Sign Test : Worked Example

The following data are the weight gains (in g) of mice over a month obtained using a new feeding regime: $\{20, 42, 18, 21, 22, 35, 19, 18, 26, 20, 21, 32, 22, 20, 24\}$. A previous feeding method produces a weight gain over the same period with median 25 g. Test whether the new method produces the same median weight gain compared to the alternative that it produces less.

Solution

1. We test the following hypotheses at $\alpha = 0.05$:

$$H_0 : \text{med}(X) = 25, \qquad H_A : \text{med}(X) < 25$$

.

2. We subtract the hypothesized median of 25 from each of our $\vec{x}$ observations.

$$\vec{x}_\otimes = \big(\{20, 42, 18, 21, 22, 35, 19, 18, 26, 20, 21, 32, 22, 20, 24\} - 25\big)$$
$$\vec{x}_\otimes = \{-5, 17, -7, -4, -3, 10, -6, -7, 1, -5, -4, 7, -3, -5, -1\}$$

3. Then, we compute t, recalling the rules of the indicator function.

$$t = \sum_{i=1}^{n} \mathbb{I}\big(\{x_\otimes\}_i > 0\big)$$
$$t = \sum \{0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0\}$$
$$t = 4$$

4. We know that $t \sim \text{bin}(15, 1/2)$ since $n = 15$.

5. We calculated the observed test statistic with the correction. Since we have the left-tailed alternative, we add 1/2.

$$T_{\text{obs}} = \frac{t - \dfrac{n}{2} + \dfrac{1}{2}}{\sqrt{\dfrac{n}{4}}} = \frac{4 - \frac{15}{2} + \frac{1}{2}}{\sqrt{\frac{15}{4}}} = -1.549193$$

6. Then, we find the p-value by comparing with $Z = N(0, 1)$. Recall this is a left-tailed test.

$$p = P(T_{obs} < Z) = 0.061$$

Since $p > \alpha$, we fail to reject.

```
x1 = c(20, 42, 18, 21, 22, 35, 19, 18,
       26, 20, 21, 32, 22, 20, 24)

# step 2
x1prime = x1 - 25

# step 3
t = sum(as.numeric(x1prime > 0))
n = length(x1)
```

```r
# step 4 tells us that t ~ bin(n, 1/2)

# step 5 tels us to add a continuity correction
# (in this case < is HA so add 1/2)
t = t + 1/2

# then find t observed
tobs = (t - n/2)/sqrt(n/4)

# then find the p-value given alpha
alpha = 0.05
pnorm(tobs, lower.tail = TRUE)
```

Worked Example, in R

```
## [1] 0.06066763
```

Exact value:

```r
pbinom(sum(x1prime > 0), n, p = 1/2) #, alternative = "less")
```

```
## [1] 0.05923462
```