Introduction to Nonparametric Methods

December 2, 2019

Nonparametric Methods

- The methods discussed thus far are all **parametric methods**.
  - Parametric methods make a lot of assumptions about model parameters, such as distributional assumptions.
- We will introduce some **nonparametric methods**.
  - These require less restrictive assumptions.

When Are Nonparametric Methods Useful?

- The data are *nominal* or *ordinal* in nature (ordered or unordered categorical).
- The parametric assumptions for a test are not satisfied.
  - When we used t-tests, we needed to assume normality.
  - When the populations are not normally distributed, we should use a nonparametric test instead.

When Are Nonparametric Methods Useful?

- Even if a parametric method is appropriate, nonparametric methods tend to give similar results.
  - We don't always use them because they can be more complex to interpret.
  - Nonparametric methods may also have slightly lower power.

Example: The Sign Test

- A company is producing a new orange juice.
- They want to know whether people prefer their orange juice or a competitor's orange juice.
- 12 individuals were given unmarked samples of orange juice in a random order.
- Each individual identified with orange juice they preferred.
- We want to know whether preferences for the two juices are equal.

Example: Hypotheses

The hypotheses are

H_0: $p = 0.5$ No difference in preference.

H_A: $p \neq 0.5$ One product is preferred more than the other.

Example: The Data

Individual	Brand Preference
1	Tropical Orange
2	Tropical Orange
3	Citrus Valley
4	Tropical Orange
5	Tropical Orange
6	Tropical Orange
7	Tropical Orange
8	Tropical Orange
9	Citrus Valley
10	Tropical Orange
11	Tropical Orange
12	Tropical Orange

Sign Test P-Values

Under H_0, the number of + signs is distributed Binomial(n, p).

The p-value for the sign test with hypotheses

H_0: $p = 0.5$ No difference in preference.

H_A: $p \neq 0.5$ One product is preferred more than the other.

is

$$2 \times P(\text{number of + signs } \leq \text{ observed number of + signs})$$

Example: Orange Juice

Find the p-value for the orange juice example. What are the conclusions for this test?

The Sign Test

For a Binomial($n = 12$, $p = 0.5$) distribution,

Number of +	Probability	Number of +	Probability
0	0.0002	7	0.1934
1	0.0029	8	0.1208
2	0.0161	9	0.0537
3	0.0537	10	0.0161
4	0.1208	11	0.0029
5	0.1934	12	0.0002
6	0.2256		

The Sign Test

- The **sign test** is the nonparametric version of the *one-sample t-test*.
- The previous example showed it used for proportions.
- We can also use this test for measures of center.
  - For nonparametric tests, we often use the median as a measure for center.

Hypothesis Tests About a Median

$$H_0\text{: Median} = M_0$$
$$H_A\text{: Median} \neq M_0$$

- Recall: the median splits the data in half so that 50% of the values fall above and 50% fall below.
- We apply the sign test by using a + sign when the value is above the hypothesized median and a - sign when it is below.
- The computations are otherwise the same.

Large Sample Approach

- Recall: we can use a normal distribution to approximate the binomial distribution.
- For large values of n, a Binomial(n, p) is well-approximated by

$$N(\mu = np, \sigma = \sqrt{np(1-p)})$$

Example

We want to know about the median home price in St. Louis, MO.

$$H_0\text{: Median} = \$75,000$$
$$H_A\text{: Median} \neq \$75,000$$

- There is a sample of $n = 62$ sales.
- 37 had prices above $\$75,000$.
- 23 had prices below $\$75,000$.
- 2 had prices exactly equal to $\$75,000$.