統計學 (二)

Anderson's Statistics for Business & Economics (14/E)

Chapter 18(1): Nonparametric Methods

上課時間地點: 二 D56, 資訊 140306

授課教師: 吳漢銘 (國立政治大學統計學系副教授)

教學網站: http://www.hmwu.idv.tw

系級: ________ 學號: ________ 姓名: ________

Overview

1. The statistical methods for inference presented previously are ________.
2. The parametric methods begin with an ________ about the probability distribution of the ________ which is often that the population has a ________ distribution.
3. Based upon this assumption, statisticians are able to derive the ________ that can be used to make ________ about one or more parameters of the population, such as the population mean or the population standard deviation.
   (a) (Recall Chapter 9) An inference about a population mean that was based on an assumption that the population had a normal probability distribution with unknown parameters μ and σ.
   (b) Using the sample standard deviation s to estimate the population standard deviation σ.
   (c) The test statistic for making an inference about the population mean was shown to have a t distribution.

(d) The t distribution was used to compute confidence intervals and conduct hypothesis tests about the mean of a normally distributed population.

4. In this chapter we present ________________ methods which can be used to make inferences about a population without requiring an assumption about the specific form of the population's probability distribution.

   (a) (First section) how the binomial distribution uses two categories of data to make an inference about a ____________________.

   (b) (Next three sections) how _______________ data are used in nonparametric tests about two or more populations.

   (c) (Final section) use rank-ordered data to compute the __________________ for two variables.

5. For this reason, these nonparametric methods are also called __________________________.

6. The computations used in the nonparametric methods are generally done with __________________. Whenever the data are quantitative, we will transform the data into categorical data in order to conduct the nonparametric test.

18.1 Sign Test

Hypothesis Test About a Population Median

1. The ____________ provides a nonparametric procedure for testing a hypothesis about the value of a ____________________.

2. If we consider a population where ________________ is exactly equal to the median, the median is the measure of __________________ that divides the population so that ________ of the values are greater than the median and ________ of the values are less than the median.

3. Whenever a population distribution is __________, the median is often preferred over the mean as the best measure of central location for the population.

4. Example The weekly sales of Cape May Potato Chips by the Lawler Grocery Store chain.

   (a) Lawler's management made the decision to carry the new potato chip product based on the manufacture's estimate that the ______________ should be \$450 per week on a per store basis.

   (b) (Table 18.1) After carrying the product for three-months, Lawler's management requested the following hypothesis test about the population median weekly sales:

$$\begin{aligned} H_0 &: \text{Median} = 450 \\ H_a &: \text{Median} \neq 450 \end{aligned}$$

TABLE 18.1 One-Week Sales of Cape May Potato Chips at 10 Lawler Grocery Stores

Store Number	Weekly Sales (\$)	Store Number	Weekly Sales (\$)
56	485	63	474
19	562	39	662
36	415	84	380
128	860	102	515
12	426	44	721

   (c) (Table 18.2) In conducting the sign test, we compare each sample observation to the ____________________ of the population median.

      i. If the observation is greater than the hypothesized value, we record a plus sign ________

      ii. If the observation is less than the hypothesized value, we record a minus sign ________

      iii. If an observation is exactly equal to the hypothesized value, the observation is _____________ from the sample and the analysis proceeds with the smaller sample size, using only the observations where a plus sign or a minus sign has been recorded.

TABLE 18.2 Lawler Sample Data for the Sign Test About the Population Median Weekly Sales

Store Number	Weekly Sales ($)	Sign	Store Number	Weekly Sales ($)	Sign
56	485	+	63	474	+
19	562	+	39	662	+
36	415	−	84	380	−
128	860	+	102	515	+
12	426	−	44	721	+

(d) Note that there are 7 plus signs and 3 minus signs.

5. The assigning of the plus signs and minus signs has made the situation a ____________ ____________ application. The sample size __________ is the number of trials. There are two outcomes possible per trial, a ________ sign or a __________ sign, and the trials are independent. Let _____ denote the probability of a plus sign.

6. If the population median is 450, p would equal ________ as there should be 50% plus signs and 50% minus signs in the population. Thus, in terms of the binomial probability p, the sign test hypotheses about the population median are converted to the following hypotheses about the binomial probability p.

$$\begin{aligned} H_0 &: \text{Median} &= 450 \\ H_a &: \text{Median} &\neq 450 \end{aligned} \quad \Rightarrow \quad \underline{\hspace{4cm}}$$

   (a) If H_0 cannot be rejected, we cannot conclude that p is different from 0.50 and thus we cannot conclude that the population median is different from 450.

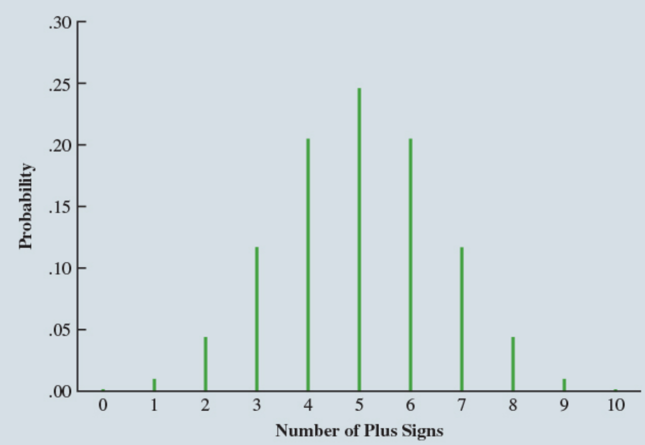
   (b) If H_0 is rejected, we can conclude that p is not equal to 0.50 and thus the population median is not equal to 450.

7. (Table 5 in Appendix B)(Table 18.3)(Figure 18.1) With $n = 10$ stores or trials and $p = 0.50$, obtain the binomial probabilities for the number of plus signs under the assumption H_0 is true. (________________________)

TABLE 18.3 Binomial Probabilities with $n = 10$ and $p = .50$

Number of Plus Signs	Probability
0	.0010
1	.0098
2	.0439
3	.1172
4	.2051
5	.2461
6	.2051
7	.1172
8	.0439
9	.0098
10	.0010

FIGURE 18.1 Binomial Sampling Distribution for the Number of Plus Signs When $n = 10$ and $p = .50$



(a) Use a 0.10 level of significance for the test.

(b) Since the observed number of plus signs for the sample data, 7, is in the upper tail of the binomial distribution, we compute the probability of obtaining 7 or more plus signs

______________________________.

(c) Since we are using a two-tailed hypothesis test, this upper tail probability is doubled to obtain the ______________________.

(d) With ____________, we cannot reject H_0. In terms of the binomial probability p, we cannot reject $H_0 : p = 0.50$, and thus we cannot reject the hypothesis that the population median is \$450.

8. The one-tailed sign tests about a population median:

   (a) Formulated the hypotheses as an ______________:

   $$H_0 : \text{Median } \leq 450$$
   $$H_a : \text{Median } > 450$$

   (b) The corresponding p-value is equal to the binomial probability that the number of plus signs is ______________________ found in the sample.

   (c) This one-tailed p-value: __________________________________.

   (d) If the example were converted to a lower tail test, the p-value would have been the probability of obtaining 7 or fewer plus signs.

   (e) The binomial probabilities provided in Table 5 of Appendix B can be used to compute the p-value when the sample size is __________.

   (f) With larger sample sizes, we rely on the ____________________________ of the binomial distribution to compute the p-value; this makes the computations quicker and easier.

Use the Normal Distribution to Approximate the Binomial Probability

1. $\boxed{\text{Example}}$ One year ago the median price of a new home was \$236,000. However, a current downturn in the economy has real estate firms using sample data on recent home sales to determine if the population median price of a new home is less today than it was a year ago.

   (a) The hypothesis test about the population median price of a new home is as follows:

   $$H_0 : \underline{\hspace{4cm}}$$
   $$H_a : \underline{\hspace{4cm}}$$

   (b) We will use a 0.05 level of significance to conduct this test. A random sample of _____ recent new home sales found _____ homes sold for more than \$236,000, _____ homes sold for less than \$236,000, and ______ home sold for \$236,000.

(c) After deleting the home that sold for the hypothesized median price of \$236,000, the sign test continues with 22 plus signs, 38 minus signs, and a sample of ____________.

(d) The null hypothesis that the population median is greater than or equal to \$236,000 is expressed by the binomial distribution hypothesis ____________.

(e) If H_0 were true as an equality, we would expect ____________ homes to have a plus sign.

(f) The sample result showing 22 plus signs is in the lower tail of the binomial distribution. Thus, the p-value is the probability of ____________ when $p = 0.50$.

(g) While it is possible to compute the exact binomial probabilities for $0, 1, 2, \cdots$ to 22 and sum these probabilities, we will use the normal distribution approximation of the binomial distribution to make this computation easier.

2. **Normal approximation of the sampling distribution of the number of plus signs when** $H_0 : p = 0.50$: For this approximation (________), the mean and standard deviation of the normal distribution are:

$$\text{Mean} : \mu = \underline{\hspace{3cm}} \qquad (18.1)$$

$$\text{Standard deviation} : \sigma = \underline{\hspace{5cm}} \qquad (18.2)$$

3. With $n = 60$ homes and $p = 0.50$, the sampling distribution of the number of plus signs can be approximated by a normal distribution with

$$\mu = 0.50n = \underline{\hspace{3cm}} \qquad \sigma = \sqrt{0.25n} = \underline{\hspace{4cm}}$$

4. The binomial probability distribution is discrete and the normal probability distribution is continuous. To account for this, the binomial probability of 22 is computed by the normal probability interval ____________. The 0.5 added to and subtracted from 22 is called the ____________ factor.

5. Thus, to compute the p-value for 22 or fewer plus signs we use the normal distribution with $\mu = 30$ and $\sigma = 3.873$ to compute the probability that the normal random variable, X, has a value less than or equal to 22.5.

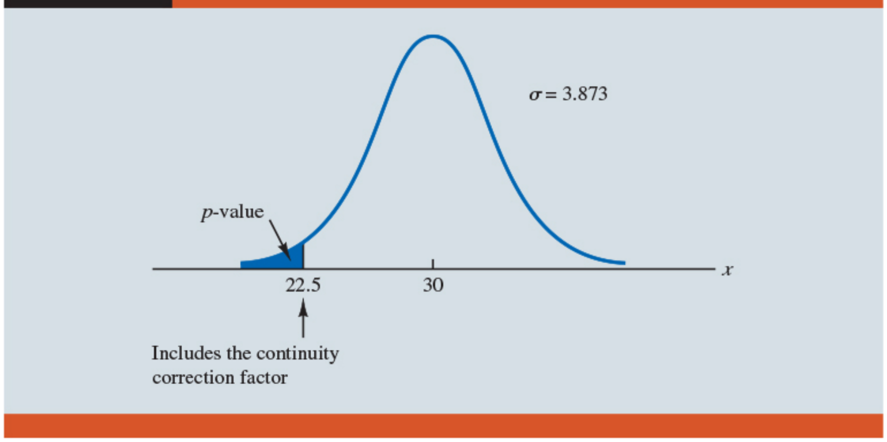
____

6. (Figure 18.2) Using this normal distribution, we compute the p-value as follows:

   p-value = ____

7. With $0.0262 < 0.05$, we ________ the null hypothesis and conclude that the median price of a new home is __________ the \$236,000 median price a year ago.

FIGURE 18.2 Normal Distribution Approximation of the p-Value for the Sign Test About the Median Price of New Homes



Hypothesis Test with Matched Samples

1. (Recall Chapter 10) Using ________________ and assuming that the differences between the pairs of matched observations were __________ distributed, the _____ distribution was used to make an inference about the difference between the means of the two populations.

2. Use the nonparametric sign test to analyze ________________ data. the sign test enables us to analyze categorical as well as quantitative data and requires no assumption about the distribution of the differences.

3. This type of matched-sample design occurs in ________________ when a sample of n potential customers is asked to compare two brands of a product such as coffee, soft drinks, or detergents. Without obtaining a quantitative measure of each individual's preference for the brands, each individual is asked to state a brand preference.

4. $\boxed{\text{Example}}$ Sun Coast Farms produces an orange juice product called Citrus Valley. The primary competition for Citrus Valley comes from the producer of an orange juice known as Tropical Orange. In a consumer preference comparison of the two brands, 14 individuals were given unmarked samples of the two orange juice products. The brand each individual tasted first was selected randomly.

   (a) If the individual selected Citrus Valley as the more preferred, a ____________ was recorded.

   (b) If the individual selected Tropical Orange as the more preferred, a ______________ was recorded.

   (c) If the individual was unable to express a difference in preference for the two products, ___________ was recorded.

5. (Table 18.4) Deleting the two individuals who could not express a preference for either brand, the data have been converted to a sign test with __________ signs and ____________ signs for the __________ individuals who could express a preference for one of the two brands.

TABLE 18.4 Preference Data for the Sun Coast Farms Taste Test

Individual	Preference	Sign	Individual	Preference	Sign
1	Tropical Orange	−	8	Tropical Orange	−
2	Tropical Orange	−	9	Tropical Orange	−
3	Citrus Valley	+	10	No Preference	
4	Tropical Orange	−	11	Tropical Orange	−
5	Tropical Orange	−	12	Citrus Valley	+
6	No Preference		13	Tropical Orange	−
7	Tropical Orange	−	14	Tropical Orange	−

6. Letting ______ indicate the proportion of the population of customers who prefer Citrus Valley orange juice, we want to test the hypotheses that there is no difference between the preferences for the two brands as follows:

$$H_0 : \_\_\_\_\_\_\_\_\_\_\_\_$$
$$H_a : \_\_\_\_\_\_\_\_\_\_\_\_$$

7. If H_0 cannot be rejected, we cannot conclude that there is a difference in preference for the two brands. However, if H_0 can be rejected, we can conclude that the consumer preferences differ for the two brands.

8. (Table 18.5) We will conduct the sign test ($\alpha = 0.05$). The sampling distribution for the number of plus signs is a __________ distribution with $p = 0.50$ and $n = 12$. (____________________)

TABLE 18.5 Binomial Probabilities with $n = 12$ and $p = .50$

Number of Plus Signs	Probability
0	.0002
1	.0029
2	.0161
3	.0537
4	.1208
5	.1934
6	.2256
7	.1934
8	.1208
9	.0537
10	.0161
11	.0029
12	.0002

9. Under the assumption H_0 is true, we would expect ___________________ plus signs. With only two plus signs in the sample, the results are in the __________ of the binomial distribution.

10. To compute the p-value for this two-tailed test, we first compute the probability of 2 or fewer plus signs and then ________ this value. Using the binomial probabilities of 0, 1, and 2 shown in Table 18.5, the p-value is

$$p\text{-value} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

11. We reject H_0. The taste test provides evidence that consumer preference _________________ for the two brands of orange juice. We would advise Sun Coast Farms of this result and conclude that the competitor's Tropical Orange product is the more preferred. Sun Coast Farms can then pursue a strategy to address this issue.

12. Similar to other uses of the sign test, one-tailed tests may be used depending upon the application.

13. As the sample size becomes large, the _____________________________ of the binomial distribution will ease the computation.

14. While the Sun Coast Farms sign test for matched samples used categorical preference data, the sign test for matched samples can be used with ______________ data as well.

    (a) This would be particularly helpful if the ___________________ are _______________ distributed and are __________.

    (b) In this case a positive difference is assigned a plus sign, a negative difference is assigned a negative sign, and a zero difference is removed from the sample.

    (c) The sign test computations proceed as before.

😍 **EXERCISES 18.1**: 1, 3, 6, 9