STAT 518 --- Section 3.4: The Sign Test

• The <u>sign test</u>, as we will typically use it, is a method for analyzing <u>paired</u> data.

Examples of Paired Data:

• Similar subjects are paired off and one of two treatments is given to each subject in the pair.

or

• We could have two observations on the same subject.

The key: With paired data, the pairings cannot be switched around without affecting the analysis.

• We might label one of the variables X_i and the other variable Y_i.

• Our entire bivariate data set for n' individual pairs is:

• The bivariate random vectors are assumed to be independent across observations.

• The goal may be to determine whether the X variable tends to be larger than or smaller than the corresponding Y variable.

• Assuming the data are at least ordinal, we could classify each pair as "+" if $X_i < Y_i$ or "–" if $X_i > Y_i$.

• If $X_i = Y_i$ then the pair is classified as "0" or "tie".

• We further assume internal consistency: If P(+) > P(–) for one pair, then P(+) > P(–) for all pairs, and same holds for P(+) < P(–) and P(+) = P(–).

Test Statistic:

• The null distribution of *T* is

where *n* =

• The hypotheses of the sign test can be stated in a variety of ways.

• Most generally, we can test any one of:

H_0:	H_0:	H_0:
H_1:	H_1:	H_1:

• These could be stated in terms of comparing the population medians of *X* and *Y*:

H_0:	H_0:	H_0:
H_1:	H_1:	H_1:

• The corresponding rejection rules in each case are:

• The P-values for each case are:

where $T \sim \text{Binom}(n, 0.5)$.

• The exact critical region and P-value are found using Table A3 (or the computer) similarly to the other binomial-based tests.

Example 1: Six students are given two tests, one after being fed, and one on an empty stomach. Is there evidence that students perform better on a full stomach? (Use $\alpha = .05$.)

	Student					
Scores	1	2	3	4	5	6
X (with food)	74	71	82	77	72	81
Y (without food)	68	71	86	70	67	80

Example 2: 18 boy/girl sets of twins were scored for "empathy" on a personality test. In ten sets, the girl scored higher; in 7 sets, the boy scored higher, and in one set, the scores were equal. Can we conclude a difference in median empathy between the boy and girl populations? (Use α = .05.)

Some Notes

• One way to view the sign test is simply as the binomial test, where p^* = 0.5.

• We classify each trial as "+" or "–" and determine whether the probability of "+" is different from/greater than/ less than 0.5.

• Note that performing the quantile test about the median is essentially performing the sign test, where the second variable is simply the constant number x^*.

• The sign test is appropriate for any data measured on an ordinal or stronger scale.

• If the paired differences $Y_i - X_i$ are continuous with a symmetric distribution, the <u>Wilcoxon signed-rank test</u> (we will see it in Chapter 5) may be more powerful than the sign test.

• If the paired differences $Y_i - X_i$ have a normal distribution, the <u>paired t-test</u> is the most powerful option.

Efficiency of the Sign Test

Population	A.R.E.(sign vs. signed-rank)	A.R.E.(sign vs. paired-t)
Normal		
Uniform (light tails)		
Double exponential (heavy tails)		

Sec. 3.5 --- Variations of the Sign Test

• Tests based on the sign test can be used to answer a variety of questions.

McNemar's Test

• Consider two paired <u>binary</u> variables (nominal).

• Both *X* and *Y* can only take the values 0 and 1, say.

• This type of data often arises from "before vs. after" experiments.

• $X = 1$ might represent having some condition _______ a treatment is applied and $Y = 1$ having it ___________ the treatment is applied.

• <u>Question</u>: Is the probability of having the condition the same before and after the treatment is applied?

• Or does the treatment change the probability of having it?

Null and Alternative hypotheses:

H_0: **vs. H_1:**

$\leftrightarrow$

• The data from such a study can be summarized with a 2×2 table:

• Consider the ($X_i = 0$, $Y_i = 1$) entries to be the "+" observations.
• Let the ($X_i = 1$, $Y_i = 0$) entries be the "–" observations.

• Then we can use the ___________ test to test H_0.

• Note the a and d entries are treated as ___________.

• The test statistic is simply $T_2 =$

• The null distribution of T_2 is

Using Table A3 with $n = b + c$ and $p = 0.5$, reject H_0 if

where t is the value corresponding to a probability of $\approx \alpha/2$.

The P-value is

Example 1: Suppose 200 subjects were asked last month and again this month whether they approved of the president's job performance. 90 said "yes" both times; 90 said "no" both times; 12 said "yes" the first month and "no" the second, and 8 said "no" the first month and "yes" the second. At $\alpha = 0.05$, has the president's approval rating significantly changed?

Contingency Table:

Hypotheses:

Test Statistic: T_2 =

P-value:

Conclusion:

• For large samples (n > 20), we can use the test statistic

T_1 =

which has a null distribution.

Why is this?

Cox-Stuart Test for Trend

• An ordered sequence of numbers exhibits <u>trend</u> if the later numbers in the sequence tend to be greater than the earlier numbers (____________ trend) or if the later numbers in the sequence tend to be less than the earlier numbers (____________ trend).

• In the arranged data, we essentially pair points to the left of the middle ordered value with points to the right of the middle ordered value, and perform a sign test.

• We assume the data $X_1, X_2, \ldots, X_{n'}$ are at least ordinal in scale.

• Pair the data as $(X_1, X_{1+c}), (X_2, X_{2+c}), \ldots, (X_{n'-c}, X_{n'})$ where $c = n'/2$ if n' is even; $c = (n' + 1)/2$ if n' is odd.

• Note that if n' is odd, the middle value is ignored.

• If the first element in a pair is less than the second, we write a "+" for that pair.
• If the first element in a pair is greater than the second, we write a "–" for that pair.
• If the first element in a pair equals the second, we ignore that pair.

Null hypothesis: H_0:

3 possible alternatives:

Test Statistic: T =

Null distribution of T is ______________ with p =
and n = the number of <u>untied</u> pairs.

• The decision rule and p-value are obtained in the same way as the sign test.

<u>Example</u>: NASA data give the average global temperatures for the last 13 decades, from the 1880s to the 2000s: `-0.493, -0.457, -0.466, -0.497, -0.315, -0.077, 0.063, -0.036, -0.025, -0.002, 0.317, 0.563, 0.923.` **(Temperatures given in degrees F, centered by subtracting from 1951-1980 mean). Does Cox-Stuart test find evidence (at $\alpha = 0.05$) of an increasing trend?**

Hypotheses:

***n′* =** **and *c* =**

Pairs:

***T* =**

Look at binomial table with *p* = 0.5 and *n* =

P-value =

Conclusion:

• In order to test for a specified type of trend other than increasing or decreasing (such as periodic, alternating, etc.), the data must first be reordered to reflect the expected ordering according to the specified trend. Then the Cox-Stuart test can be implemented on the reordered data.