The Procedure of Sign Test and SAS Implementation

Weixing Song

I. Sign Test

1. Data

The data consist of observations on a bivariate random sample

\[(X_1, Y_1), \cdots, (X_{n'}, Y_{n'})\]

where there are \(n'\) pairs of observations. There should be some natural basis for pairing the observations; otherwise the \(X\)’s and \(Y\)’s are independent, and the more powerful Mann-Whitney test is more appropriate.

Within each pair \((X_i, Y_i)\) a comparison is made, and the pair is classified as “plus” or “+” if \(X_i < Y_i\), as “minus” or “−” if \(X_i > Y_i\), or as “tie” or “0” if \(X_i = Y_i\). Thus the measurement scale needs only to be ordinal.

2. Assumptions

- The bivariate random variables \((X_i, Y_i), i = 1, \cdots, n'\), are mutually independent.
- The measurement scale is at least ordinal within each pair. That is, each pair \((X_i, Y_i)\) may be determined to be a “plus”, “minus”, or “tie”.
- The pair \((X_i, Y_i)\) are internally consistent, in that if \(P(+) > P(−)\) for one pair, then it is also for all pairs. The same is true for \(P(+) < P(−)\) and \(P(+) = P(−)\).

3. Test Statistic

\[T = \text{total number of } +\text{’s.}\]

4. Null Distribution

The null distribution of \(T\) is Bin\((n, 1/2)\). Where \(n\) is the number of nontied pairs.

5. Hypotheses Testing

(i) Two-Tailed Test:

\[H_0 : P(+) = P(−) \text{ v.s. } H_1 : P(+) \neq P(−)\]
Let $t$ be the observed $T$. Then

$$p \text{ value} = 2[P(T \leq t) \land P(T \geq t)].$$

if $n > 20$, using normal approximation, then

$$p \text{ value} = 2 \left[ P \left( Z \leq \frac{2t - n + 1}{\sqrt{n}} \right) \land P \left( Z \geq \frac{2t - n - 1}{\sqrt{n}} \right) \right], \quad Z \sim N(0, 1).$$

(ii) **Lower-Tailed Test:**

$H_0 : \quad P(+) \geq P(-) \quad v.s. \quad H_1 : \quad P(+) < P(-)$

Let $t$ be the observed $T$, then

$$p \text{ value} = P(T \leq t), \quad T \sim \text{Bin}(n, 1/2).$$

if $n > 20$, using normal approximation, then

$$p \text{ value} = P \left( Z \leq \frac{2t - n + 1}{\sqrt{n}} \right), \quad Z \sim N(0, 1).$$

(iii) **Upper-Tailed Test:**

$H_0 : \quad P(+) \leq P(-) \quad v.s. \quad H_1 : \quad P(+) > P(-)$

Let $t$ be the observed $T$, then

$$p \text{ value} = P(T \geq t), \quad T \sim \text{Bin}(n, 1/2).$$

if $n > 20$, using normal approximation, then

$$p \text{ value} = P \left( Z \geq \frac{2t - n + 1}{\sqrt{n}} \right), \quad Z \sim N(0, 1).$$

6. **Example**

An item A is manufactured using a certain process. Item B serves the same function as A but is manufactured using a new process. The manufacturer wishes to determine whether B is preferred to A by the consumer, so she selects a random sample consisting of 10 consumers, gives each of them one A and one B and asks them to use the items for some period of time. The sign test (one-tailed) will be sued to test

$H_0 : \quad P(+) \leq P(-) \quad v.s. \quad H_1 : \quad P(+) > P(-)$

The incorporated 1 in the p-value formula is a “correction of continuity” that improves the normal approximation to the binomial.
where “+” represents the event “item B is preferred over item A”, and “-” represents the event “item A is preferred over item B”.

At the end of the allotted period of time the consumers report their preferences to the manufacturer. 8 consumers preferred B to A, 1 preferred A to B, and 1 reported "no preference". Therefore

\[ T = 8, \quad n = 8 + 1 = 9, \]

and the p-value is

\[ p \text{ value} = P(T \geq t) = P(T \geq 8) = 0.0195. \]

7. Computer Assistance

(i) Minitab

First, obtain the difference of the preference between B and A; Second, go to Stat→Nonparametrics and choose 1-Sample Sign...; Third, in the 1-Sample Sign dialog box, choose Test Median, don’t change the default value 0, then select greater than as the alternative. The result is

<table>
<thead>
<tr>
<th>N Below Equal Above</th>
<th>P Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

and the p-value is 0.0195.

(ii) SAS

PROC UNIVARIATE in SAS reports the sign statistic, the codes are in the following, in which we just made up some data.

```sas
DATA ITEM;
   INPUT DIFF @@;
DATALINES;
   1 1 1 1 1 1 1 1 -1 0
;
RUN;
PROC UNIVARIATE DATA=ITEM;
   VAR DIFF;
RUN;
```

The result is

Tests for Location: Mu0=0

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>-----p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
The testing statistic is SAS for the sign test is $M = T - n/2$ and the default hypothesis in SAS is two-tail. So the p-value for our example should be $0.0391/2 = 0.01951$.

II. Some Variations of The Sign Test

A. McNemar Test for Significance of Changes

1. **Background:** Suppose now that the data are not ordinal as in the sign test but nominal, with two categories that we will call “0” and “1”. That is, each $X_i$ is either 0 or 1, and similarly for each $Y_i$. Then a question sometimes asked is, “Can we detect a difference between the probability of (0,1) and the probability of (1,0)?” such a question arises when $X_i$ in the pair $(X_i, Y_i)$ represents the condition (or state) of the subject before the experiment and $Y_i$ represents the condition of the same subject after the experiment.

2. **Data:** The data consist of observations on $n'$ independent bivariate random variables $(X_i, Y_i), i = 1, 2, \ldots, n'$. The measurement scale for the $X_i$ and the $Y_i$ is nominal with two categories, which we call 0 and 1; that is, the possible values of $(X_i, Y_i)$ are $(0, 0), (0, 1), (1, 0)$, and $(1, 1)$. In the McNemar test the data are usually summarized in a $2 \times 2$ contingency table, as follows.

<table>
<thead>
<tr>
<th>Classification of $X$</th>
<th>Classification of $Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = 0$</td>
<td>$Y = 0$</td>
</tr>
<tr>
<td>$X = 1$</td>
<td>$Y = 1$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. **Assumptions**
   - The pairs $(X_i, Y_i)$ are mutually independent;
   - The measurement scale is nominal with two categories for all $X_i$ and $Y_i$;
   - The difference $P(X_i = 0, Y_i = 1) - P(X_i = 1, Y_i = 0)$ is negative for all $i$, or zero for all $i$, or positive for all $i$.

4. **Test Statistic and Null Distribution:**
   The test statistic for the McNemar test is usually written as
   $$ T_1 = \frac{(b - c)^2}{b + c}, \quad T_1 \sim \chi^2_1. $$
However, for \( b + c \leq 20 \), the following test statistic is preferred,
\[
T_2 = b, \quad T_2 \sim \text{Bin}(b + c, 1/2).
\]

5. Hypotheses:

\[
H_0 : \quad P(X_i = 0, Y_i = 1) = P(X_i = 1, Y_i = 0) \quad \text{for all } i
\]

versus
\[
H_1 : \quad P(X_i = 0, Y_i = 1) \neq P(X_i = 1, Y_i = 0) \quad \text{for all } i.
\]

The above hypothesis is equivalent to
\[
H_0 : \quad P(X_i = 0) = P(Y_i = 0) \quad \text{for all } i \text{ v.s. } H_1 : \quad P(X_i = 0) \neq P(Y_i = 0) \quad \text{for all } i,
\]
or
\[
H_0 : \quad P(X_i = 1) = P(Y_i = 1) \quad \text{for all } i \text{ v.s. } H_1 : \quad P(X_i = 1) \neq P(Y_i = 1) \quad \text{for all } i.
\]

6. Example: Prior to a nationally televised debate between the two presidential candidates, a random sample of 100 persons stated their choice of candidates as follows. 84 persons favored the Democratic candidate, and the remaining 16 favored the Republican. After the debate the same 100 people expressed their preference again. Of the persons who formerly favored the Democrat, exactly 1/4 of them changed their minds, and also 1/4 of the people formerly favoring the Republican switched to the Democratic side. The results are summarized in the following 2 \( \times \) 2 contingency table.

<table>
<thead>
<tr>
<th>Before</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrat</td>
<td>63</td>
<td>4</td>
</tr>
<tr>
<td>Republican</td>
<td>21</td>
<td>12</td>
</tr>
</tbody>
</table>

7. SAS implementation:

PROC FREQ in SAS has the McNemar test. Here is the SAS codes for the above example.

```sas
DATA PREF;
INPUT BEFORE $ AFTER $ COUNT;
DATALINES;
DEMO DEMO 63
DEMO REPU 21
REPU DEMO 4
REPU REPU 12
```

5
The first FREQ procedure only reports the asymptotic p-value,

McNemar's Test

<table>
<thead>
<tr>
<th>Statistic (S)</th>
<th>11.5600</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Pr &gt; S</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

and the second FREQ procedure also reports the exact p-value,

McNemar's Test

<table>
<thead>
<tr>
<th>Statistic (S)</th>
<th>11.5600</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Asymptotic Pr &gt; S</td>
<td>0.0007</td>
</tr>
<tr>
<td>Exact Pr &gt;= S</td>
<td>9.105E-04</td>
</tr>
</tbody>
</table>

**B. Cox and Stuart Test for Trend**

1. **Background:** This modification is introduced in 1955 and it is used to test for the presence of trend. A sequence of numbers is said to have trend if the later numbers in the sequence tend to be greater than the earlier numbers (upward trend) or less than the earlier numbers (downward trend). This test involves pairing the later numbers with the earlier numbers and then performing a sign test on the pairs thus formed. If there is a trend, one member of each pair will have a tendency to be higher or lower than the other member. On the other hand, if there is no trend and the sequence of numbers actually represents observations on independent and identically distributed random variables, there will be no tendency for one particular member of each pair to exceed the other one.
2. **Data:** The data consist of observations on sequence of random variables \( X_1, X_2, \ldots, x_{n'} \) arranged in a particular order, such as the order in which the random variables are observed. It is desired to see if a trend exist in the sequence. Group the random variables into pairs \((X_1, X_{1+m}), (X_2, X_{2+m}), \ldots, (X_{n'-m}, X_{n'})\), where

\[
m = \begin{cases} 
    n'/2 & \text{if } n' \text{ is even,} \\
    (n' + 1)/2 & \text{if } n' \text{ is odd,}
\end{cases}
\]

Note that the middle random variable is eliminated using this scheme if \( n' \) is odd. Replace each pair \((X_i, X_{i+m})\) with a “+” if \( X_i < X_{i+m} \), or a “-” if \( X_i > X_{i+m} \), eliminating ties. The number of untied pairs is called \( n \).

This test may be used to detect any specified type of nonrandom pattern, such as a sine wave or other periodic pattern. The sequence of random variables is merely rearranged so that the smallest numbers, as predicted, will be near the beginning of the sequence and the larger numbers near the end. Then the presence of an upward trend in the rearranged sequence is evidence that the predicted pattern is present in the original arrangement of the sequence.

3. **Assumptions**

- The random variables \( X_1, X_2, \ldots, X_{n'} \) are mutually independent;
- The measurement scale of \( X_i \)’s is at least ordinal;
- Either the \( X_i \)’s are identically distributed or there is a trend; that is, the later random variables are more likely to be greater than instead of less than the earlier random variables or vice versa.

4. **Test Statistic and Null Distribution:**

\[
T = \text{Total number of “+”'s}, \quad T \sim \text{Bin}(n, 1/2).
\]

5. **Hypotheses:**

The null hypothesis is that no trend is present. An upper-tailed test is used to detect an upward trend. A lower-tailed test is used to detect an downward trend. The two-tailed test is used if the alternative hypothesis is that any type of trend exists.

6. **Example:** Total annual precipitation is recorded yearly for 19 years. This record is examined to see if the amount of precipitation is tending to increase or decrease. The precipitation in inches was 45.25, 45.83, 41.77, 36.26, 45.37, 52.25, 35.37, 57.16, 35.37, 58.32, 41.05, 33.72, 45.73, 37.90, 41.72, 36.07, 49.83, 36.24 and 39.90. Because \( n' = 19 \) is odd, the middle number 58.32 is omitted. The remaining numbers are paired.

\[(45.25, 41.05), (45.83, 33.72), (41.77, 45.73), (36.26, 37.90), (45.37, 41.72), \]

\[\text{Example: Total annual precipitation is recorded yearly for 19 years. This record is examined to see if the amount of precipitation is tending to increase or decrease. The precipitation in inches was 45.25, 45.83, 41.77, 36.26, 45.37, 52.25, 35.37, 57.16, 35.37, 58.32, 41.05, 33.72, 45.73, 37.90, 41.72, 36.07, 49.83, 36.24 and 39.90. Because } n' = 19 \text{ is odd, the middle number 58.32 is omitted. The remaining numbers are paired.}\]
There are no ties, so \( n = 9 \), and \( T = 4 \).

7. SAS implementation:

We will use the PROC UNIVARIATE in SAS to implement the sign test for the above example. Here is the SAS codes.

```
DATA PRECI;
  INPUT X Y @@;
  DIFF = Y - X;
DATALINES;
45.25 41.05 45.83 33.72 41.77 45.73
36.26 37.90 45.37 41.72 52.25 36.07
35.37 49.83 57.16 36.24 35.37 39.90
;RUN;
PROC UNIVARIATE DATA=PRECI;
  VAR DIFF;
RUN;
```

The first FREQ procedure only reports the asymptotic p-value,

Tests for Location: \( \mu_0 = 0 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>-----p Value-----</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t</td>
<td>t -0.96427</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M -0.5</td>
<td>Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S -6.5</td>
<td>Pr &gt;=</td>
</tr>
</tbody>
</table>

From above result, we see the p-value is 1.000, therefore, then null hypothesis “no trend exists” can not be rejected.

Remark: Daniels (1950) proposed the use of Spearman’s \( \rho \) to test for trend by pairing measurements, called \( X_i \)’s with the time (or order) at which the measurements were taken. The assumption is that the \( X_i \)’s are mutually independent, and the null hypothesis is that they are identically distributed. The alternative hypothesis is that the distribution of the \( X_i \)’s is related to time so that as time goes on, the \( X \) measurements tend to become larger (or smaller).

The Spearman’s \( \rho \) is defined as

\[
\rho = \frac{\sum_{i=1}^{n} R(X_i) R(Y_i) - n \left( \frac{n+1}{2} \right)^2}{\left( \sum_{i=1}^{n} R(X_i)^2 - n \left( \frac{n+1}{2} \right)^2 \right)^{1/2} \left( \sum_{i=1}^{n} R(Y_i)^2 - n \left( \frac{n+1}{2} \right)^2 \right)^{1/2}}
\]
which is simply Pearson’s $r$ computed on the ranks and average ranks. If there is no ties, an equivalent but computationally easier form is given by

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} [(R(X_i) - R(Y_i))^2]}{n(n^2 - 1)}.$$ 

For the precipitation example, $\rho = -0.24748$.

The two-tailed test for trend involves rejection of the null hypothesis of no trend if $\rho$ is too large or too small. Exact quantiles of $\rho$ when $X$ and $Y$ are independent are given in some special table for $n \leq 30$ and no ties, see attached table. For larger $n$, we can use the following procedure.

(A). Two-Tailed Test

$H_0$: The $X_i$ and $Y_i$ are mutually independent, v.s. the alternative $H_1$: Either (a) there is a tendency for the larger values of $X$ to be paired with the larger values of $Y$ or (b) there is a tendency for the smaller values of $X$ to be paired with the larger values of $Y$.

The corresponding approximate p-value is:

$$p\text{-value} = 2P(Z \geq |\rho|\sqrt{n - 1}).$$

(B). Lower-Tailed Test for Negative Correlation

$H_0$: The $X_i$ and $Y_i$ are mutually independent, v.s. the alternative $H_1$: There is a tendency for the smaller values of $X$ to be paired with the larger values of $Y$.

The corresponding approximate p-value is:

$$p\text{-value} = P(Z < \rho\sqrt{n - 1}).$$

(C). Upper-Tailed Test for Positive Correlation

$H_0$: The $X_i$ and $Y_i$ are mutually independent, v.s. the alternative $H_1$: There is a tendency for the larger values of $X$ to be paired with the larger values of $Y$.

The corresponding p-value is:

$$p\text{-value} = P(Z > \rho\sqrt{n - 1}).$$

For the Two-tailed test, the approximate p-value for our example is $2P(Z \geq 0.2475\sqrt{19 - 1}) = 0.3070$. So we can not reject the null hypothesis.

The following SAS codes can be used to obtain Spearman’s $\rho$.

```
DATA PRECP;
  INPUT X YEAR;
  DATALINES;
  45.25 1950 45.83 1951 41.77 1952 36.26 1953
```
which gives the following result

Spearman Correlation Coefficients, N = 19
Prob > |r| under H0: Rho=0

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>1.00000</td>
<td>-0.24748</td>
</tr>
<tr>
<td>X</td>
<td>-0.24748</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

References
