3 Nonparametric tests with fuzzy data

Ranking data is an important concept on nonparametric tests. By using Equ. (2.4) above, we can rank continuous fuzzy numbers simultaneously and easily. Traditional nonparametric tests could not deal with fuzzy data. To success for this, the ranking technique will be applied. We provide sign test and Wlicoxon signed-ranks test latter and some empirical examples will be given to show for this.

3.1 One-sample sign test with fuzzy data

The oldest of all the nonparametric procedures may be sign test. Its use was found as early as 1710 by Arbuthnott [12]. The sign test uses the direction of differences between two observations without concerning quantitative measures.

Assumption

The *n* fuzzy samples which are designed by $\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_n$ are random samples of independent measurements form a population with unknown median *M*.

Hypothesis

$$\text{A. (Two-sided)} \quad H_0 : M = M_0, H_1 : M \neq M_0$$

Select a level of significance α

Test statistic

Record the sign of the difference obtained by subtracting the hypothesized median M_0 from each sample value after defuzzificaton; that is, record the sign of the *n* difference, $Q(\tilde{A}_i) - Q(M_0), i = 1, 2, \cdots, n.$ Let the total numbers of plus sign is *C*. If H_0 is true, we expect a random fuzzy sample from the population have about as many plus signs as minus signs when the *n* differences $Q(\tilde{A}_i) - Q(M_0)$ haven been computed. If we observe a sufficient small number of either plus or minus signs, we reject H_0 in two-sided case.

Decision rule

Reject H_0 at the α level of significance if

$$P(T \le C) = \sum_{i=0}^{c} C_i^{n'} (\frac{1}{2})^{n'} < \frac{\alpha}{2}$$

or

$$P(T \ge C) = \sum_{i=C}^{n'} C_i^{n'} (\frac{1}{2})^{n'} > \frac{\alpha}{2}$$

where *n'* is the total sample numbers for $Q(\tilde{A}_i) \ne Q(M_0)$.

Example 3.1 *A survey showed the median of the time which adults spend on exercise per week is roughly 2.5~3.5 hours, i.e.* $M_0 = [2, 2.5, 3.5, 4]$. *In order to check it up. We select 10 adults, survey their time spending on exercise per week, the calculations for obtaining the test statistics are summarized in Table 2. At the* $\alpha = 0.05$ *level of significance does the median of the population equals to the random samples, i.e.* $M = M_0$*?*

Table 2 Adults spend on exercise per week

Adult	Hours $(\tilde{A}_i)$	$Q(\tilde{A}_i)$	Sign
1	[2.5,3,3.5]	3	0
2	[1.5,2.3,3.5]	2.43	-
3	[2.5,3.5,5.5]	3.83	+
4	[2.5,3,3.5]	3	0
5	[4,5,6]	5	+
6	[2.7,3.9]	3.3	+
7	[2,2.5,3,3.5]	2.75	-
8	[3,3.5,4.5,5]	4	+

9	[1.5,2,2.5,3]	2.25	-
10	[3,3.5]	3.25	-

$H_0: \ M = [2, 2.5, 3.5, 4] \quad V.S \quad M \neq [2, 2.5, 3.5, 4]$

$n' = 10 - 2 = 8$

$P(T \leq 4) = \sum_{i=0}^{4} C_i^8 (\frac{1}{2})^8 = 0.527$

Because $0.527 > \frac{0.05}{2} = 0.025$,

hence we accept H_0.

At the $\alpha = 0.05$ level of significance the median of the population equals to the random samples.

3.2 Wilcoxon signed-ranks test with fuzzy data

As mentioned above, the sign test with fuzzy data utilizes only the sign of the difference between $Q(\tilde{A}_i)$ and $Q(M_0)$. There is another procedure with fuzzy data, is called the Wilcoxon signed-ranks test. In this test procedure the magnitude of the differences are concerned. Because the Wlicoxon signed-ranks test uses more data information than the sign test, it's often a more powerful test.

Assumption

A. Two fuzzy observations each has *n* subjects. Let *i* denote the particular subject that is being referred to and the first fuzzy observation measured on subject *i* be denoted by $\tilde{A}_i$ and second fuzzy observation by $\tilde{B}_i$.

Hypotheses

$$
\begin{array}{llll}
A. & H_0 : M = M_0 & vs & H_1 : M \neq M_0 \\
B. & H_0 : M \geq M_0 & vs & H_1 : M < M_0 \\
C. & H_0 : M \leq M_0 & vs & H_1 : M > M_0
\end{array}
$$

Select a significance level α.

Test statistic

The Wlicoxon signed rank statistic W^+ is computed by ordering $d(\tilde{A}_i, \tilde{B}_i)$ from smallest to largest, the rank of each $d(\tilde{A}_i, \tilde{B}_i)$ is given a rank of r_i. There may be two or more $d(\tilde{A}_i, \tilde{B}_i)$ equal. In this situation assign each tied value the mean of the rank positions. For example, if three smallest are all equal, rank them 1, 2, and 3, but assign each a rank of $\frac{1+2+3}{3} = 2$.

Denote $I_i = \begin{cases} 1, & \tilde{B}_i - \tilde{A}_i > 0 \\ 0, & \text{otherwise} \end{cases}$, $I'_i = \begin{cases} 1, & \tilde{B}_i - \tilde{A}_i < 0 \\ 0, & \text{otherwise} \end{cases}$.

The Wlicoxon signed ranked statistic W^+ is defined as $W^+ = \sum_{i=1}^{n} I_i r_i$ and W^- is defined as $W^- = \sum_{i=1}^{n} I'_i r_i$. To simplify notation, we call the smaller of the two *T*, i.e. $T = \min\{W^+, W^-\}$.

Decision rule

Decision rules for each set of hypotheses listed above are as follows.

A. Reject H_0 at a α level significance if $T \leq W_{\frac{\alpha}{2}}$.

B. Reject H_0 at a α level significance if $W^+ \leq W_{\frac{\alpha}{2}}$.

C. Reject H_0 at a α level significance if $W^- \leq W_{\frac{\alpha}{2}}$.

Example 3.2 *A survey showed female spends more time on housework than male of each spouse per week. In order to check it up. 11 couples are invited to do a fuzzy questionnaire about the time spending on housework per week. The calculations for obtaining the test statistics are summarized in Table 3. At the* $\alpha = 0.05$ *level of significance what can we say about this survey?*

Table 3 Housework time

Couple	Husband $(\tilde{A}_i)$	Wife $(\tilde{B}_i)$	$Q(\tilde{A}_i)$	$Q(\tilde{B}_i)$	$d(\tilde{A}_i, \tilde{B}_i)$	$\tilde{B}_i - \tilde{A}_i$	r_i
1	[0,1,2]	[5,6.5,8]	1	6.5	5.5	+	10
2	[5,6,7]	[6,7,10,11]	6	8.5	2.5	+	5
3	[1,2,3]	[4,5,7]	2	5.33	3.33	+	8
4	[4,6,7]	[6,7,10,11]	5.67	8.5	2.83	+	6
5	[4,6]	[5,7]	5	6	1	+	2
6	[6,7,9]	[8,9,10]	7.33	9	1.67	+	3
7	[3,4,5,5.5]	[6,7,9]	4.36	7.33	2.97	+	7
8	[6,7,9,9.5]	[4,6,7]	7.86	5.67	2.19	-	4
9	[6,7,8]	[5,6,7,8]	7	7.67	0.67	+	1
10	[7,9,11]	[3,5,6]	9	4.67	4.33	-	9
11	[0,2]	[6,8]	1	7	6	+	11

$W^+ = \sum_{i=1}^{11} I_i r_i = 53$, $W^- = \sum_{i=1}^{11} I'_i r_i = 13$, N=11.

M = *the median of the time which female spend on housework per week.*

M_0 = *the median of the time which male spend on housework per week.*

$H_0 : M = M_0$ *vs* $M > M_0$.

Because $W_{0.05} = 14 > W^- = 13$, hence we reject H_0.

At the $\alpha = 0.05$ level of significance, female spend more time on housework than male of each spouse per week.