**Example 9.1 (Act 9.1: Application of Sign Test)**

Data in Table below (**Ref: Problems 9.9-9.11: Hospital data, page 365**) refers to a sample from a larger dataset of people in a hospital on the usage of antibiotic on their recent visit to the hospital along with other related information. It appears that the distribution of the white-blood-cell (WBC) count in the population is not normally distributed. We wish to test vs where is the median white-blood cell count in the population.

1. What test should you use for the test?
2. Implement above test in SAS.
3. Report value of the test statistic and p-value for the test. What is your conclusion about the test on the basis of the observed p-value?

| **wbc** | **Antibio** | **Bact\_cul** | **Sex** | **Service** | **Dur\_stay** | **Age** | **Temp** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | No | No | Female | med. | 5 | 30 | 99.0 |
| 5 | No | Yes | Female | med. | 10 | 73 | 98.0 |
| 12 | No | No | Female | surg. | 6 | 40 | 99.0 |
| 4 | No | No | Female | surg. | 11 | 47 | 98.2 |
| 11 | No | No | Female | surg. | 5 | 25 | 98.5 |
| 6 | Yes | No | Male | surg. | 14 | 82 | 96.8 |
| 8 | Yes | Yes | Male | med. | 30 | 60 | 99.5 |
| 7 | No | No | Female | med. | 11 | 56 | 98.6 |
| 7 | No | No | Female | med. | 17 | 43 | 98.0 |
| 12 | No | Yes | Male | surg. | 3 | 50 | 98.0 |
| 7 | No | Yes | Female | med. | 9 | 59 | 97.6 |
| 3 | No | No | Male | surg. | 3 | 4 | 97.8 |
| 11 | Yes | No | Female | surg. | 8 | 22 | 99.5 |
| 14 | Yes | Yes | Female | surg. | 8 | 33 | 98.4 |
| 11 | No | Yes | Female | surg. | 5 | 20 | 98.4 |
| 9 | No | No | Male | surg. | 5 | 32 | 99.0 |
| 6 | Yes | No | Male | surg. | 7 | 36 | 99.2 |
| 6 | No | No | Male | surg. | 4 | 69 | 98.0 |
| 5 | Yes | No | Male | med. | 3 | 47 | 97.0 |
| 6 | No | No | Male | surg. | 7 | 22 | 98.2 |
| 10 | No | No | Male | surg. | 9 | 11 | 98.2 |
| 14 | Yes | No | Male | surg. | 11 | 19 | 98.6 |
| 4 | No | No | Female | med. | 11 | 67 | 97.6 |
| 5 | No | No | Female | surg. | 9 | 43 | 98.6 |
| 5 | No | No | Female | med. | 4 | 41 | 98.0 |

**Solution**

/\*Activity 9.1(i)\*\*\*\*\*\*\*\*/

**data** act9pt1;

input wbc Antibio $ Bact\_cul $ Sex $ Service $ Dur\_stay Age Temp;

cards;

8 No No Female med. 5 30 99.0

5 No Yes Female med. 10 73 98.0

12 No No Female surg. 6 40 99.0

4 No No Female surg. 11 47 98.2

11 No No Female surg. 5 25 98.5

6 Yes No Male surg. 14 82 96.8

8 Yes Yes Male med. 30 60 99.5

7 No No Female med. 11 56 98.6

7 No No Female med. 17 43 98.0

12 No Yes Male surg. 3 50 98.0

7 No Yes Female med. 9 59 97.6

3 No No Male surg. 3 4 97.8

11 Yes No Female surg. 8 22 99.5

14 Yes Yes Female surg. 8 33 98.4

11 No Yes Female surg. 5 20 98.4

9 No No Male surg. 5 32 99.0

6 Yes No Male surg. 7 36 99.2

6 No No Male surg. 4 69 98.0

5 Yes No Male med. 3 47 97.0

6 No No Male surg. 7 22 98.2

10 No No Male surg. 9 11 98.2

14 Yes No Male surg. 11 19 98.6

4 No No Female med. 11 67 97.6

5 No No Female surg. 9 43 98.6

5 No No Female med. 4 41 98.0

;

**run**;

/\*proc print noobs; run;\*/

/\*Activity 9.1(ii) & (iii)\*\*\*\*\*\*\*\*/

**proc** **univariate** data=act9pt1 mu0=**6**;

ods select TestsForLocation;

var wbc;

title "Result of act 9.1: sign test";

**run**;

|  |
| --- |
| Result of act 9.1: sign test |

The UNIVARIATE Procedure

Variable: wbc (white-blood cell count)

| **Tests for Location: Mu0=6** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 2.865687 | **Pr > |t|** | 0.0085 |
| **Sign** | **M** | 3.5 | **Pr >= |M|** | 0.1892 |
| **Signed Rank** | **S** | 68 | **Pr >= |S|** | 0.0134 |

The value of the sign test statistic is 3.5 and the p-value is 0.1892 (marked in red in the table).

Since p-value (0.1892)>, we accept the null hypothesis that the median WBC in the population is indeed 6.

**Example 9.2 (Act 9.2: Application of Signed Rank Test)**

A common symptom of **otitis media** in young children is the prolonged presence of fluid in the middle ear, known as middle-ear effusion, which may result in temporary hearing loss and interfere with normal learning skills in the first 2 years of life. One hypothesis is that babies who are breastfed for at least 1 month build up some immunity against the infection and prolonged effusion than bottled-fed babies. In order to test the hypothesis, 24 pairs of babies are considered in a study, where one member is a breastfed baby and the other a bottled-fed baby, having episode of middle-ear effusion. The data below refers to the outcome variables, **duration of middle-ear effusion** after the first episode of otitis media, of breastfed and bottle-fed babies.

| **Obs** | **breast\_fed** | **bottle\_fed** |
| --- | --- | --- |
| **1** | 20 | 18 |
| **2** | 11 | 35 |
| **3** | 3 | 7 |
| **4** | 24 | 182 |
| **5** | 7 | 6 |
| **6** | 28 | 33 |
| **7** | 58 | 223 |
| **8** | 7 | 7 |
| **9** | 39 | 57 |
| **10** | 17 | 76 |
| **11** | 17 | 186 |
| **12** | 12 | 29 |
| **13** | 52 | 39 |
| **14** | 14 | 15 |
| **15** | 12 | 21 |
| **16** | 30 | 28 |
| **17** | 7 | 8 |
| **18** | 15 | 27 |
| **19** | 65 | 77 |
| **20** | 10 | 12 |
| **21** | 7 | 8 |
| **22** | 19 | 16 |
| **23** | 34 | 28 |
| **24** | 25 | 20 |

1. Read these data into SAS.

(b) Use SAS to implement the signed rank test for vs the alternative hypothesis , where and be the median duration of middle-ear effusion in breastfed and bottled-fed babies in the population. Note that alternative follows from the fact that babies who are breastfed for at least 1 month build up some immunity against the infection and prolonged effusion to lead to the shorter duration of middle-ear effusion than the bottled-fed babies.

(c) Implement the test in SAS, and report value of the test statistic and p-value.

(d) What conclusion can you make regarding the test at 5% level of significance?

Solution

/\*\*\*\*\*Act 9.2, (a), Ref: problem 9.15, page 366\*\*\*\*\*/

**data** act9pt2;

input breast\_fed bottle\_fed;

diff=breast\_fed-bottle\_fed;

cards;

20 18

11 35

3 7

24 182

7 6

28 33

58 223

7 7

39 57

17 76

17 186

12 29

52 39

14 15

12 21

30 28

7 8

15 27

65 77

10 12

7 8

19 16

34 28

25 20

;

**run**;

/\*proc print data=act9pt2;run;

proc univariate data=act9pt2 normal;

var breast\_fed bottle\_fed;

run;\*/

\*proc univariate performs Wilcoxon signed-rank test;

**proc** **univariate** data=act9pt2;

ods select TestsForLocation;

title "Result (b) & (c) of act 9.2: signed rank test";

var diff;

**run**;

|  |
| --- |
| Result (b) & (c) of act 9.2: signed rank test |

The UNIVARIATE Procedure

Variable: diff

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | -2.3157 | **Pr > |t|** | 0.0298 |
| **Sign** | **M** | -4.5 | **Pr >= |M|** | 0.0931 |
| **Signed Rank** | **S** | -77 | **Pr >= |S|** | 0.0154 |

The value of the signed rank test statistic is -77 and left-tailed p-value is 0.0154/2=0.0077, since the value of the test statistic is negative.

**Conclusion:** Since the left-tailed p-value (0.0077), we reject the null hypothesis of no median difference in the duration of middle –ear effusion in breast-fed and bottled-fed babies in the population against the left-tailed alternative (i.e., in favor of the left-tailed alternative), at 5% level of significance.

**Example 9.3 (Act 9.3: Application of Wilcoxon Rank-sum Test)**

Ref: **Smoke** sas data available in SAS Studio in BioStat course. Use nonparametric methods to test whether there is a difference between the median number of days abstinent from smoking by male vs female populations. Let and be median abstinence from smoking in the male and female populations.

1. Use SAS to carry out the Wilcoxon rank-sum test for vs .
2. Report the value of the test statistic and the p-value.
3. What is the conclusion about the test at 5% level of significance?
4. What other tests can you use for this test?

Solution

/\*\*\*\*\*Act 9.3, Ref: problem 9.15, page 366\*\*\*\*\*/

libname lib "C:\Users\kislam\Desktop\Courses\Stat 468 568 BioStat";

ods select WilcoxonTest;

**proc** **npar1way** data=lib.smoke wilcoxon;

title "Result of act 9.3: signed sum test";

class gender;

var Day\_abs;

**run**;

|  |
| --- |
| Result of act 9.3: signed sum test |

The NPAR1WAY Procedure

| **Wilcoxon Two-Sample Test** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Statistic** | **Z** | **Pr > Z** | **Pr > |Z|** | **t Approximation** | |
| **Pr > Z** | **Pr > |Z|** |
| 13271.00 | 0.6701 | 0.2514 | 0.5028 | 0.2517 | 0.5035 |
| **Z includes a continuity correction of 0.5.** | | | | | |

1. The SAS code is printed above.
2. The value of the rank sum test statistic is 13271, with approximated value 0.6701. The two-sided p-value (pT)=0.5028>alpha=0.05.
3. **Conclusion**: Accept null hypothesis that the median abstinence from smoking in the male and female populations are the same at 5% level of significance.
4. We can use sign test, signed rank test for the same problem.

**Example 9.4 (Act 9.4: Permutation Test for Rank-sum Test)**

**Ref: Pilot study data on problem #2, Page 359**

A pilot study is planned to test the efficacy of vitamin E supplementation as a possible preventive agent for Alzheimer’s disease. Twenty subjects age 65+ are randomized to either a supplement of vitamin E of 400 IU/day (group 1, n=10) or placebo (group 2, n=10). It is important to compare the total vitamin E intake (from food and supplements) of the two groups at baseline. The baseline intake of each group in IU/day is as follows:

| **grp1** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7.5 | 12.6 | 3.8 | 20.2 | 6.8 | 403.3 | 2.9 | 7.2 | 10.5 | 205.4 |

| **grp2** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8.2 | 13.3 | 102 | 12.7 | 6.3 | 4.8 | 19.5 | 8.3 | 407.1 | 10. |

Let and be location parameters (i.e., medians) of the baseline intake of vitamin E in the populations with vitamin E supplement group and placebo group, respectively.

(a) Test the or against all possible alternatives using the permutation rank sum test.

(b) Report the observed value of the test statistic and p-value from the null distribution of the test statistic obtained from the permutation test. What is the conclusion about the test?

(c) What other non-parametric tests can be used for this test?

**Solution**

There is no default test in SAS for a permutation test as I know of. We will write code for the rank-sum permutation test in proc iml (a good way to introduce proc iml; just for your information iml refers to **interactive matrix language,** an environment in SAS to perform any operation in matrix setting, like MatLab or R).

/\*Act 9.4: Pemutation rank-sum test\*/

**proc** **iml**;

/\*define two groups as vectors:

grp1 and grp2\*/

grp1 = {**7.5**,**12.6**,**3.8**,**20.2**,**6.8**,**403.3**,**2.9**,**7.2**,**10.5**,**205.4**};

grp2 = {**8.2**,**13.3**,**102.0**,**12.7**,**6.3**,**4.8**,**19.5**,**8.3**,**407.1**,**10.2**};

call randseed(**123456**); /\* set random number seed \*/

/\* stack two grps into into a single grp vector\*/

grp = grp1//grp2; /\* stack data into a single vector\*/

n1 = nrow(grp1);

n = n1 + nrow(grp2);

/\*Assign ranks to the combined vector grp\*/

r=rank(grp);

/\*observed value of the rank-sum statistics due to grp1\*/

robs=sum(r[**1**:n1]);

/\*generate k=1000 samples WOR for 1000 permutations\*/

k= **1000**;

/\* define nulldist, a result holder vector\*/

nulldist = j(k,**1**);

do i = **1** to k;

rs = sample(r, n, "wor");

nulldist[i] = sum(rs[**1**:n1]); /\* sum of ranks due to x1 sample \*/

end;

pL=mean(nulldist<=robs);

pR=mean(nulldist>=robs);

pT=**2**\*min(pL,pR,**0.5**);

title "Act 9.4: Obs rank-sum statistic:";

print robs;

print "p-values of rank-sum permutation test:", pL pR pT;

**quit**;

|  |
| --- |
| Act 9.4: Obs rank-sum statistic: |

| **robs** |
| --- |
| 97 |

|  |
| --- |
| p-values of rank-sum permutation test: |

| **pL** | **pR** | **pT** |
| --- | --- | --- |
| 0.294 | 0.732 | 0.588 |

1. See the SAS code above for implementing the rank-sum permutation test.
2. The observed value of the test statistic (robs) due to grp1 is 97 and the two-sided p-value of the permutation test is 0.588. Conclusion: Since the p-value>0.05, we fail to reject the null hypothesis at 5% level of significance against the three possible alternative hypotheses.
3. One could use a sign test, signed test or rank-sum test for this problem.