**Homework 9**

Copy and paste or type answers from the SAS output in answering questions as appropriate.

**Last name: Zuo ; First name: Lian**

**HW 9.1**

The data in table below (**Ref: Table 2.15 Page 37**) are a sample of cholesterol levels taken from 24 hospital employees who were on a standard American diet and who agreed to adopt a vegetarian diet for 1 month. Serum-cholesterol measurements were made before adopting the diet and 1 month after. We wish to test if there is any significant difference in the serum-cholesterol levels before and after the vegetarian diet. In other words, we wish to test if there is any effect of the vegetarian diet on serum-cholesterol levels.

1. What test should you use?
2. How should you set up the null and alternative hypothesis for the test?
3. Implement the test in SAS, and report value of the test statistic and p-value from your sas output.
4. What conclusion can you make regarding the test at 5% level of significance?

| **Before** | **After** |
| --- | --- |
| 195 | 146 |
| 145 | 155 |
| 205 | 178 |
| 159 | 146 |
| 244 | 208 |
| 166 | 147 |
| 250 | 202 |
| 236 | 215 |
| 192 | 184 |
| 224 | 208 |
| 238 | 206 |
| 197 | 169 |
| 169 | 182 |
| 158 | 127 |
| 151 | 149 |
| 197 | 178 |
| 180 | 161 |
| 222 | 187 |
| 168 | 176 |
| 168 | 145 |
| 167 | 154 |
| 161 | 153 |
| 178 | 137 |
| 137 | 125 |

Solution:

1. Sign Test
2. we wish to test or , or in general, , where and are location parameters, preferably medians, of the paired population, and .

data hw9pt1;

input bf af;

dif=bf-af;

datalines;

195 146

145 155

205 178

159 146

244 208

166 147

250 202

236 215

192 184

224 208

238 206

197 169

169 182

158 127

151 149

197 178

180 161

222 187

168 176

168 145

167 154

161 153

178 137

137 125

;

run;

title "Result of hw9pt1";

proc univariate data=hw9pt1;

ods select testsforlocation;

var dif;

run;

title;

**Result of hw9pt1**

**The UNIVARIATE Procedure**

**Variable: dif**

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 5.696519 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 9 | **Pr >= |M|** | 0.0003 |
| **Signed Rank** | **S** | 134 | **Pr >= |S|** | <.0001 |

**(**d)

The Value of the sign test statistics is 9 and the P-Value is 0.0003.

Since P-Value 0.0003 < , we reject the null hypothesis that the median Dur\_stay in the population is indeed 7.

**HW 9.2**

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. We wish to test if the median duration of hospital stay (Dur\_stay) is 7 days.

Suppose refers to the unknown median duration of hospital stay of patients in the population.

1. How do you specify the null and alternative hypothesis?
2. What test should you use for the test?
3. Implement above test in SAS.
4. Report value of the test statistic and p-value for the test.
5. What is your conclusion about the test on the basis of the observed p-value at 5% level of significance?

The dataset for this problem appears in the activity 9 SAScode in canvas or SAS Studio act9.sas.

| **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:**

(a) we wish to test or

**(**b) Sign Test

**(c)**

data hw9;

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 univariate data=hw9 mu0=7;

ods select TestsForLocation;

var Dur\_stay;

title "Result of hw9: sign test"

run;

(d)

**Result of hw9: sign test run**

**The UNIVARIATE Procedure**

**Variable: Dur\_stay**

| **Tests for Location: Mu0=7** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 1.399708 | **Pr > |t|** | 0.1744 |
| **Sign** | **M** | 1.5 | **Pr >= |M|** | 0.6776 |
| **Signed Rank** | **S** | 33 | **Pr >= |S|** | 0.3229 |

**(**d)

The Value of the sign test statistics is 1.5 and the P-Value is 0.6776.

Since P-Value (0.6776 > , we accept the null hypothesis that the median Dur\_stay in the population is indeed 7.

**HW 9.3**

Using the hospital data in HW 9.2, we wish to test if there is any significant difference in the median WBC counts due to medical and surgical services.

1. Set up the null and alternative hypothesis for the test.
2. What test should you use for the test? Why?
3. Implement the test in SAS.
4. Report value of the test statistic and p-value for the test.
5. What is your conclusion about the test on the basis of the observed p-value?

Solution:

data hw9;  
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;  
ods select WilcoxonTest;  
proc npar1way data=hw9 wilcoxon;  
title "Result of hw 9.3: signed sum test";  
class Service;  
var wbc;  
run;

**Result of hw 9.3: signed sum test**

**The NPAR1WAY Procedure**

| **Wilcoxon Two-Sample Test** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Statistic** | **Z** | **Pr < Z** | **Pr > |Z|** | **t Approximation** | |
| **Pr < Z** | **Pr > |Z|** |
| **Z includes a continuity correction of 0.5.** | | | | | |
| 89.0000 | -1.5666 | 0.0586 | 0.1172 | 0.0652 | 0.1303 |

1. vs .
2. Use the Wilcoxon rank-sum test ,the two underlying populations are not normal.
3. The SAS code is printed above.
4. The value of the rank sum test statistic is 89, with approximated value -1.5666. The two-sided p-value (pT)=0.1172<alpha=0.05.
5. **Conclusion**: Reject null hypothesis that the median WBC counts due to medical and surgical services are the same at 5% level of significance.

**HW 9.4**

The desirable levels of an HDL (high-density lipoprotein, known as good cholesterol) for healthy men 20 or over is 40 mg/dL or higher, while the desirable levels for women 20 or over is 50mg/dL or higher. Two random samples from male and female patients treated in a health facility are considered to test if there is any significant difference in the median of HDL levels in the male and female populations. It appears that in the distribution of HDL levels in the population of treated patients is not normal.

1. Set up a suitable null and alternative hypothesis for the test.
2. Use a permutation test applied to sign, signed rank or rank sum statistics, as you wish.
3. Find the value of the test statistic for the given sample.
4. Generate 9000 permutation samples and compute a p-value by comparing the null distribution of the test statistic due to the 9000 permutations to the observed value of the test statistic in (c).
5. What is your conclusion about the test at 5% level of significance?

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Male | 61.9 | 56.9 | 57.9 | 56.4 | 58.7 | 59.5 | 60.0 | 63.4 | 64.4 | 58.1 |  |  |
| Female | 64.3 | 67.1 | 69.0 | 69.3 | 72.0 | 67.4 | 70.6 | 72.0 | 66.8 | 70.9 | 68.8 | 69.1 |

Solution:

/\*hw 9.4: Pemutation rank-sum test\*/  
proc iml;  
/\*define two groups as vectors:  
grp1 and grp2\*/  
Male = {61.9,56.9,57.9,56.4,58.7,59.5,60.0,63.4,64.4,58.1};  
Female = {64.3,67.1,69.0,69.3,72.0,67.4,70.6,72.0,66.8,70.9,68.8,69.1};  
call randseed(123456); /\* set random number seed \*/  
/\* stack two grps into into a single grp vector\*/  
grp = Male//Female; /\* stack data into a single vector\*/  
n1 = nrow(Male);   
n = n1 + nrow(Female);  
/\*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=9000 samples WOR for 9000 permutations\*/  
k= 9000;   
/\* 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 "hw 9.4: Obs rank-sum statistic:";  
print robs;  
print "p-values of rank-sum permutation test:", pL pR pT;  
quit;

**hw 9.4: Obs rank-sum statistic:**

| **robs** |
| --- |
| 56 |

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

| **pL** | **pR** | **pT** |
| --- | --- | --- |
| 0 | 1 | 0 |

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

(b)See the SAS code above for implementing the rank-sum permutation test.

(c)

(e)The observed value of the test statistic (robs) due to grp1 is 26 and the two-sided p-value of the permutation test is 0. Conclusion: Since the p-value< 0.05, we reject the null hypothesis at 5% level of significance against the three possible alternative hypotheses.