WST221 Practical Notes - 2014

HYPOTHESIS TESTING IN SAS

REFERENCES

1. Cody R.P. and Smith J.K. , (1997). *Applied Statistics and the SAS Programming Language*, Prentice Hall.
2. Steyn, A.G.W., Smit, C.F., Du Toit, S.H.C. and Strasheim, C., (1996). *Modern Statistics in Practice* (2nd revised impression), J.L. van Schaik.
3. Wackerly, D.D., Mendenhall, W. and Scheaffer, R.L., (2008). *Mathematical Statistics with Applications* (Seventh Edition), Thomson.

1. ONE SAMPLE CASE (PARAMETRIC TESTS)

1.1 Testing H_0: $\mu = \mu_0$

Example: Steyn, Smit, Du Toit and Strasheim, page 423, Example 12.9
A firm's board of directors has to decide whether newly appointed representatives would take the traditional course in sales techniques or, would instead, change to a new course offered by a consultant. Suppose the first-year sales figures of ten representatives selected at random, who completed the new course are as follows:
R287900 R419400 R338300 R287500 R310850
R292600 R390050 R369850 R430400 R338450
The mean first year sales of representatives who took the traditional course is R300000.

Solution
SAS Program

```
data ex_12_9;
input sales @@;
diff=sales-300000;
cards;
287900 419400 338300 287500 310850
292600 390050 369850 430400 338450
;
proc means mean std stderr t prt;
var diff;
run;
```

SAS Output

```
          Analysis Variable : DIFF
                  Mean       Std Dev     Std Error            T  Prob>|T|
          ---------------------------------------------------------------
              46530.00      53767.58      17002.80    2.7366080    0.0230
          ---------------------------------------------------------------
```

PROC MEANS Step	Comment
PROC MEANS; or PROC MEANS DATA=EX_12_9; or PROC MEANS MEAN STD STDERR T PRT;	The last data set is used to calculate descriptive statistics unless the name of the data set is specified in the DATA option of PROC MEANS. Only the mean, standard deviation and standard error are given. T gives the *t*-value and PRT the *p*-value for the *t*-test.
VAR DIFF;	The VAR statement specifies variables for which statistics must be calculated. If the statement is omitted, statistics will be calculated for all numerical variables.

Hypothesis test from SAS Output
$H_0 : \mu = 300000$ $\quad H_1 : \mu > 300000$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob>|T|) = 0.0230/2 = 0.0115 H_0 is rejected.
$\therefore$ The mean first-year sales of the ten representatives who attended the new course is significantly higher than R300 000.

Note: The *p*-value (Prob>|T|) for a two-sided hypothesis is given in the output. In the case of a one sided hypothesis the Prob>|T| value must be divided by 2.

2. ANALYSIS OF CATEGORICAL DATA

2.1 The (I x J) independence test

Example: Wackerly, Mendenhall and Scheaffer, page 724, Example 14.3
A survey was conducted to evaluate the effectiveness of a new flu vaccine that had been administered in a small community. The vaccine was provided free of charge in a two-shot sequence over a period of 2 weeks to those wishing to avail themselves of it. Some people received the two-shot sequence, some appeared only for the first shot, and the others received neither.
A survey of 1000 local inhabitants in the following spring provided the information shown in the table below. Do the data present sufficient evidence to indicate a dependence between the two classifications – vaccine category and occurrence or nonoccurrence of flu?

Table of data tabulation

Status	No Vaccine (0)	One Shot (1)	Two Shots (2)	Total
Flu (1)	24	9	13	46
No flu (0)	289	100	565	954
Total	313	109	578	1000

Solution
SAS Program

```
proc format;
value vac 0='No Vaccine'
          1='One Shot'
          2='Two Shots';
value status 0='No Flu'
             1='Flu';

data ex14_3w;
input vaccine status f @@;
format vaccine vac. status status.;
cards;
0 1 24   1 1 9   2 1 13
0 0 289  1 0 100 2 0 565
;
proc freq;
tables status*vaccine/chisq expected;
weight f;
run;
```

Alternative data step

```
data ex14_3w;
format vaccine vac. status status.;
 do vaccine = 0 to 2;
  do status = 0 to 1;
   input f @@;
   output;
  end;
 end;
cards;
289 24 100 9 565 13
;
```

SAS Output

The FREQ Procedure

Table of status by vaccine

status Frequency Expected Percent Row Pct Col Pct	vaccine No Vaccine	One Shot	Two Shots	Total
No Flu	289 298.6 28.90 30.29 92.33	100 103.99 10.00 10.48 91.74	565 551.41 56.50 59.22 97.75	954 95.40
Flu	24 14.398 2.40 52.17 7.67	9 5.014 0.90 19.57 8.26	13 26.588 1.30 28.26 2.25	46 4.60
Total	313 31.30	109 10.90	578 57.80	1000 100.00

Statistics for Table of status by vaccine

Statistic	DF	Value	Prob
Chi-Square	2	17.3130	0.0002
Likelihood Ratio Chi-Square	2	17.2519	0.0002
Mantel-Haenszel Chi-Square	1	14.9155	0.0001
Phi Coefficient		0.1316	
Contingency Coefficient		0.1305	
Cramer's V		0.1316	

Sample Size = 1000

Hypothesis test from SAS Output

H_0 : The occurrence or nonoccurence of flu is independent of vaccine category

H_1 : The occurrence or nonoccurence of flu is dependent of vaccine category

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (Prob) = 0.0002 H_0 is rejected.

$\therefore$ The occurrence or nonoccurence of flu depends on vaccine category.

3. TWO SAMPLE CASE (PARAMETRIC TESTS)

3.1 Testing $H_0 : \sigma^2_1 = \sigma^2_2$ and H_0: $\mu_1 - \mu_2 = 0$ (Two Independent Samples)

Example: Cody and Smith.
Students are randomly assigned to a control or treatment group (where a drug is administered). Their response time to a stimulus is then measured. The times are as follows:

Control	Treatment
(response time in millisec)	
80	100
93	103
83	104
89	99
98	102

Do the treatment scores come from a population whose mean is different from the mean of the population from which the control scores were drawn?

Solution
SAS Program
```
data response;
input group$ time @@;
cards;
c  80   c  93   c  83   c 89   c  98
t 100   t 103   t 104   t 99   t 102
;
proc ttest;
 class group;
 var time;
run;
```

SAS Output
```
                                    The TTEST Procedure
                                        Statistics
                              Lower CL         Upper CL  Lower CL           Upper CL
 Variable  group           N      Mean    Mean     Mean   Std Dev  Std Dev   Std Dev  Std Err
 time      c               5    79.535    88.6   97.665    4.3741   7.3007    20.979    3.265
 time      t               5    99.025   101.6   104.17    1.2424   2.0736    5.9587   0.9274
 time      Diff (1-2)            -20.83     -13   -5.173    3.6249   5.3666    10.281   3.3941

                                          T-Tests
              Variable    Method           Variances      DF    t Value    Pr > |t|
              time        Pooled           Equal           8      -3.83      0.0050
              time        Satterthwaite    Unequal      4.64      -3.83      0.0141

                                  Equality of Variances
                  Variable    Method       Num DF    Den DF    F Value    Pr > F
                  time        Folded F          4         4      12.40    0.0318
```

Note:
The information in the bottom line of the output above is used to test the hypothesis of equal variances. If the *p*-value is small (say the Prob>F’ value is less than 0.05) then the null hypothesis of equal variances is rejected. The *t*-value and *p*-value for unequal variances are used. If the Prob>F’ value is greater than 0.05 the *t*-value and *p*-value for equal variances are used. In this example Prob>F’ = 0.0318. The two samples come from populations with variances that differ significantly.

PROC TTEST Step	Comment
PROC TTEST; or PROC TTEST DATA=RESPONSE;	PROC TTEST computes a *t* statistic for testing the hypothesis that the means of two groups of observations in a SAS data set are equal. The last data set is used unless the name of the data set is specified in the DATA option of the PROC TTEST statement.
CLASS GROUP;	This statement identifies the independent variable; the variable that identifies the two groups of subjects.
VAR TIME;	Identifies the dependent variable. When more than one dependent variable is listed, a separate *t*-test is computed for each dependent variable in the list.

Hypothesis test from SAS Output

$H_0 : \sigma^2_{\text{control}} = \sigma^2_{\text{treatment}}$
$H_1 : \sigma^2_{\text{control}} \neq \sigma^2_{\text{treatment}}$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob>F') = 0.0318 H_0 is rejected.
$\therefore$Population variances differ significantly.

$H_0 : \mu_{\text{control}} = \mu_{\text{treatment}}$
$H_1 : \mu_{\text{control}} \neq \mu_{\text{treatment}}$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob>|T|) = 0.0145 H_0 is rejected.
$\therefore$The average response times for the two groups differ significantly.

Note:
The p-value for a two-sided hypothesis is given in the output. In the case of a one sided hypothesis the Pr > |t| value must be divided by 2.

Example: Steyn, Smit, Du Toit and Strasheim, page 433, Example 12.16

Two risk factors that have a bearing on the condition of the heart are fitness and cholesterol level. In a research project on this subject the amounts of triglycerides (unsaturated fats) in the blood samples of nine coronary patients and 21 marathon athletes were measured. The observations in millimol per litre are as follows:

Coronary patients:	3.80	2.71	1.60	1.62	1.93	1.32	1.09	2.28	0.65
Marathon athletes:	0.86	0.84	1.15	1.12	0.72	1.62	1.23	1.22	1.13
	0.98	0.62	0.38	0.86	1.25	0.90	0.56	0.66	0.73
	0.73	0.50	0.92						

Test the hypothesis that, compared to marathon athletes, coronary patients have a significantly higher population mean triglyceride level.

Solution

SAS Program

```
data response;
input group$ triglyc @@;
cards;
c 3.80  c 2.71  c 1.60  c 1.62  c 1.93  c 1.32  c 1.09  c 2.28  c 0.65
m 0.86  m 0.84  m 1.15  m 1.12  m 0.72  m 1.62  m 1.23  m 1.22  m 1.13
m 0.98  m 0.62  m 0.38  m 0.86  m 1.25  m 0.90  m 0.56  m 0.66  m 0.73
m 0.73  m 0.50  m 0.92
;
proc ttest;
class group;
var triglyc; run;
```

SAS Output

```
                                       The TTEST Procedure
                                           Statistics
                                Lower CL            Upper CL  Lower CL            Upper CL
 Variable  group            N       Mean    Mean        Mean   Std Dev  Std Dev    Std Dev  Std Err
 triglyc   c                9      1.163  1.8889      2.6147    0.6378   0.9443     1.8091   0.3148
 triglyc   m               21     0.7672  0.9038      1.0404    0.2296   0.3001     0.4334   0.0655
 triglyc   Diff (1-2)             0.5241  0.9851      1.4461    0.4483   0.5649      0.764   0.2251

                                            T-Tests
              Variable    Method           Variances      DF    t Value    Pr > |t|
              triglyc     Pooled           Equal          28       4.38      0.0002
              triglyc     Satterthwaite    Unequal       8.7       3.06      0.0140

                                     Equality of Variances
                  Variable    Method       Num DF    Den DF    F Value    Pr > F
                  triglyc     Folded F          8        20       9.90    <.0001
```

Hypothesis test from SAS Output

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (Prob>F') < 0.0001 H_0 is rejected.

$\therefore$ Population variances differ significantly.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

Use $\alpha = 0.01$. Reject H_0 if p-value < 0.01.

Since p-value (Pr > |t|) = 0.0141/2 = 0.00705 H_0 is rejected.

$\therefore$ The mean amount of triglycerides for coronary patients is significantly higher than for marathon athletes.

3.2 Testing H_0: $\mu_1 - \mu_2 = 0$ (Two Dependent Samples)

Example: Wackerly, Mendenhall and Scheaffer, page 646, Example 12.2

We wish to compare two methods for determining the percentage of iron ore in ore samples. Because inherent differences in the ore samples would be likely to contribute unwanted variability in the measurements that we observe, a matched-pairs experiment was created by splitting each of 12 ore samples into two parts. One-half of each sample was randomly selected and subjected to method 1; the other half was subjected to method 2. The results are presented in the table below. Do the data provide sufficient evidence that method 2 yield a higher average percentage than method 1? Test using $\alpha = 0.05$.

Table: Percentage of iron ore in ore samples

Sample	Method 1	Method 2		Sample	Method 1	Method 2
1	38.25	38.27		7	35.42	35.46
2	31.68	31.71		8	38.41	38.39
3	26.24	26.22		9	42.68	42.72
4	41.29	41.33		10	46.71	46.76
5	44.81	44.80		11	29.20	29.18
6	46.37	46.39		12	30.76	30.79

Solution

SAS Program

```
data ex12_2w;
input method1 method2;
diff=method1-method2;
cards;
38.25      38.27
31.68      31.71
26.24      26.22
41.29      41.33
44.81      44.80
46.37      46.39
35.42      35.46
38.41      38.39
42.68      42.72
46.71      46.76
29.20      29.18
30.76      30.79
;
proc means n mean stderr t prt;
var diff;
run;
```

SAS Output

```
                          The MEANS Procedure
                       Analysis Variable : diff
  N             Mean       Std Error     t Value    Pr > |t|
 ------------------------------------------------------------
 12       -0.0166667       0.0077198       -2.16      0.0538
 ------------------------------------------------------------
```

Hypothesis test from SAS Output

$H_0 : \mu_{\text{method1}} = \mu_{\text{method2}}$

$H_1 : \mu_{\text{method1}} < \mu_{\text{method2}}$

Reject H_0 if p-value < 0.05.

Since p-value (Prob>|T|) = 0.0538/2 = 0.0269 H_0 is rejected.

$\therefore$ Method 2 yields a significantly higher average percentage of iron ore than does method 1.

Note:

The p-value for a two-sided hypothesis is given in the output. In the case of a one sided hypothesis the Pr > |t| value must be divided by 2.

Example: Steyn, Smit, Du Toit and Strasheim, page 437, Example 12.18
The time (in minutes) it takes operators to fit a certain part before and after completing a training programme appears in the table below. Determine whether the training programme significantly decreased the mean fitting time.

Table: Time required by eight operators before and after training

Operator	Before training	After training
1	23	17
2	17	14
3	16	12
4	15	13
5	19	12
6	21	20
7	13	14
8	20	15

Solution
SAS Program

```
data ex_12_18;
input before after @@;
diff=before-after;
cards;
23 17  17 14  16 12  15 13
19 12  21 20  13 14  20 15
;
proc means n mean stderr t prt;
var diff;
run;
```

SAS Output

```
Analysis Variable : DIFF
N          Mean     Std Error            T  Prob>|T|
-----------------------------------------------------
8     3.3750000     0.9437293     3.5762374    0.0090
-----------------------------------------------------
```

Hypothesis test from SAS Output
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$
Use $\alpha = 0.01$. Reject H_0 if p-value < 0.01.
Since p-value (Prob>|T|) = 0.0090/2 = 0.0045 H_0 is rejected.
$\therefore$ The training programme significantly decreased the mean fitting time.

4. TWO SAMPLE CASE (NONPARAMETRIC TESTS)

4.1 Testing Testing H_0: $\eta_1 - \eta_2 = 0$ (Two Independent Samples. Wilcoxon Rank Sum Test)

Example: Wackerly, Mendenhall and Scheaffer, page 756, Example 15.4
The bacteria counts per unit volume are shown in the table below for two types of cultures, I and II. Four observations were made for each culture. Do these data present sufficient evidence to indicate a difference in the locations of the population distributions for cultures I and II?

Table: Bacteria counts for different cultures

Culture I	27	31	26	25
Culture II	32	29	35	28

Solution
SAS Program
```
data ex15_4w;
input culture obs @@;
cards;
1 27  1 31  1 26  1  25
2 32  2 29  2 35  2  28
;
proc npar1way wilcoxon;
class culture;
var obs;
exact wilcoxon;  run;
```

SAS Output
```
                          The NPAR1WAY Procedure
                Wilcoxon Scores (Rank Sums) for Variable obs
                       Classified by Variable culture
                          Sum of      Expected       Std Dev          Mean
culture          N        Scores      Under H0      Under H0         Score
--------------------------------------------------------------------------
1                4          12.0          18.0      3.464102           3.0
2                4          24.0          18.0      3.464102           6.0

                       Wilcoxon Two-Sample Test
                     Statistic (S)              12.0000
                     Normal Approximation
                     Z                          -1.5877
                     One-Sided Pr <  Z           0.0562
                     Two-Sided Pr > |Z|          0.1124

                     t Approximation
                     One-Sided Pr <  Z           0.0782
                     Two-Sided Pr > |Z|          0.1564

                     Exact Test
                     One-Sided Pr <=  S          0.0571
                     Two-Sided Pr >= |S - Mean|  0.1143
                   Z includes a continuity correction of 0.5.

                          Kruskal-Wallis Test
                          Chi-Square         3.0000
                          DF                      1
                          Pr > Chi-Square    0.0833
```

Hypothesis test from SAS Output
$H_0 : \eta_1 = \eta_2$
$H_1 : \eta_1 \neq \eta_2$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob >= |S - Mean|) = 0.1143 H_0 is not rejected.
$\therefore$ We do not have sufficient evidence to reject the hypothesis that the population distributions of bacteria counts for the two cultures are identical.

Example: Steyn, Smit, Du Toit and Strasheim, page 596, Example 16.4

After an inter-university tennis tournament in which four universities took part, a joint ranking of the twenty best male players is compiled. The positions occupied by the players from two of the universities are given in the table below.

Is there a difference in the quality of the male tennis players from the two universities?

University	Individual	Position on list
1	1	1
	2	3
	3	10
	4	15
	5	16
2	6	2
	7	4
	8	5
	9	6
	10	8
	11	9

Solution

SAS Program

```
data ex_16_4;
input sample position @@;
cards;
1 1  1 3  1 10  1 15  1 16
2 2  2 4  2  5  2  6  2  8  2 9
;
proc npar1way wilcoxon;
class sample;
var position;
exact wilcoxon;
run;
```

SAS Output

```
                          The NPAR1WAY Procedure

               Wilcoxon Scores (Rank Sums) for Variable position
                          Classified by Variable sample

                              Sum of      Expected       Std Dev          Mean
sample          N             Scores      Under H0       Under H0         Score
------------------------------------------------------------------------------
1               5               34.0          30.0       5.477226      6.800000
2               6               32.0          36.0       5.477226      5.333333


                        Wilcoxon Two-Sample Test

                   Statistic (S)               34.0000

                   Normal Approximation
                   Z                            0.6390
                   One-Sided Pr >  Z            0.2614
                   Two-Sided Pr > |Z|           0.5228

                   t Approximation
                   One-Sided Pr >  Z            0.2686
                   Two-Sided Pr > |Z|           0.5372

                   Exact Test
                   One-Sided Pr >=  S           0.2684
                   Two-Sided Pr >= |S - Mean|   0.5368

              Z includes a continuity correction of 0.5.

                        Kruskal-Wallis Test
                     Chi-Square         0.5333
                     DF                      1
                     Pr > Chi-Square    0.4652
```

Hypothesis test from SAS Output

$H_0 : \eta_1 = \eta_2$
$H_1 : \eta_1 \neq \eta_2$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr >= |S - Mean|) = 0.5368 H_0 is not rejected.
$\therefore$ The quality of the male players from the two universities therefore does not appear to differ.

5. CORRELATION

5.1 Testing H_0: $\rho = 0$ against H_1: $\rho \neq 0$, (Pearson's correlation)

Example: Steyn, Smit, Du Toit and Strasheim, page 493, Example 13.9

In the television series *Beyond 2000* it was alleged that there is a negative linear correlation between income and the number of hours that a person sets aside for sleep. To investigate this statement, a random sample of twenty people, all working for the same company, was taken. The following data was obtained:

Monthly income and average number of hours slept per day by twenty employees

Observation	Number of hours slept	Monthly Income (R1000)	Observation	Number of hours slept	Monthly Income (R1000)
1	7.0	0.844	11	8.1	5.440
2	9.3	1.708	12	7.5	6.065
3	7.4	1.728	13	7.4	6.444
4	8.9	2.909	14	7.3	7.236
5	7.7	2.676	15	7.6	7.349
6	8.9	3.440	16	7.1	8.235
7	7.3	3.616	17	7.4	8.379
8	6.3	4.096	18	5.9	9.131
9	7.7	4.420	19	6.0	9.746
10	7.6	5.808	20	6.7	9.983

Solution

SAS Program

```
data ex_13_9;
input sleep income @@;
cards;
7.0 0.844  9.3 1.708  7.4 1.728  8.9 2.909  7.7 2.676
8.9 3.440  7.3 3.616  6.3 4.096  7.7 4.420  7.6 5.808
8.1 5.440  7.5 6.065  7.4 6.444  7.3 7.236  7.6 7.349
7.1 8.235  7.4 8.379  5.9 9.131  6.0 9.746  6.7 9.983
;
proc corr nosimple;
var sleep income;
run;
```

SAS Output

```
                              Correlation Analysis
                         2 'VAR' Variables:  SLEEP    INCOME
          Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 20
                                        SLEEP            INCOME
                       SLEEP           1.00000          -0.56374
                                        0.0               0.0096

                       INCOME         -0.56374           1.00000
                                     0.0096            0.0
```

Hypothesis test from SAS Output

$H_0 : \rho = 0$

$H_1 : \rho < 0$

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (Prob > |r|) = 0.0096/2 = 0.0048, H_0 is rejected.

$\therefore$ Income and number of hours slept therefore do show a negative linear correlation.

5.2 Testing H_0: $\rho = 0$ against H_1: $\rho \neq 0$, (Spearman's correlation)

Example: Wackerly, Mendenhall and Scheaffer, page 784, Example 15.12
Suppose that eight elementary-science teachers have been ranked by a judge according to their teaching ability (low rank means good teaching ability), and all have taken a national teachers' examination. The data are given in the table below. Do the data suggest agreement between the judge's ranking and the examination score? Alternatively, we might express this question by asking whether a correlation exists between the judge's ranking and the ranks of examination scores.

Table: Judge's ranking and examination score for teachers

Teacher	Judge's Rank	Exam Score
1	7	44
2	4	72
3	2	69
4	6	70
5	1	93
6	3	82
7	8	67
8	5	80

Solution
SAS Program
```
data ex15_12w;
input rank score;
cards;
7 44
4 72
2 69
6 70
1 93
3 82
8 67
5 80
;
proc corr spearman;
var rank score;
run;
```

SAS Output
```
                                  The CORR Procedure
                          2  Variables:    rank     score

                                  Simple Statistics
Variable            N          Mean       Std Dev        Median       Minimum       Maximum
rank                8       4.50000       2.44949       4.50000       1.00000       8.00000
score               8      72.12500      14.27723      71.00000      44.00000      93.00000

                        Spearman Correlation Coefficients, N = 8
                               Prob > |r| under H0: Rho=0
                                            rank         score
                            rank         1.00000      -0.71429
                                                        0.0465

                            score       -0.71429       1.00000
                                          0.0465
```

Hypothesis test from SAS Output
$H_0 : \rho = 0$
$H_1 : \rho \neq 0$
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob > |R|) = 0.0465, H_0 is rejected.
$\therefore$ Judge's rank and examination score is significantly correlated.

6. ANALYSIS OF VARIANCE (PARAMETRIC TESTS)

6.1 One-way Analysis of Variance (One-way ANOVA)

Example: Wackerly, Mendenhall and Scheaffer, page 670, Example 13.2
Four groups of students were subjected to different teaching techniques and tested at the end of aspecified period of time. As a result of dropouts from the experimental groups (due to sickness, transfer, etc.), the number of students varied from group to group. Do the data shown in the table below present sufficient evidence to indicate a difference in mean achievement for the four teaching techniques?

Percentages scored by 23 students in an examination

Group	Percentage					
1	65	87	73	79	81	69
2	75	69	83	81	72	79	90
3	59	78	67	62	83	76
4	94	89	80	88			

Solution
SAS Program

```
data ex_13_2w;
input group percentage @@;
cards;
1 65 1 87 1 73 1 79 1 81 1 69
2 75 2 69 2 83 2 81 2 72 2 79 2 90
3 59 3 78 3 67 3 62 3 83 3 76
4 94 4 89 4 80 4 88
;
proc glm;
class group;
model percentage=group;
means group/hovtest scheffe;
run;
```

SAS Output

```
                                      The GLM Procedure

                                   Class Level Information
                               Class         Levels    Values
                               group              4    1 2 3 4

                           Number of Observations Read          23
                           Number of Observations Used          23

Dependent Variable: percentage
                                               Sum of
  Source                      DF          Squares     Mean Square    F Value    Pr > F
  Model                        3       712.586439      237.528813       3.77    0.0280
  Error                       19      1196.630952       62.980576
  Corrected Total             22      1909.217391

              R-Square     Coeff Var      Root MSE    percentage Mean
              0.373235      10.26019      7.936030           77.34783

  Source                      DF        Type I SS     Mean Square    F Value    Pr > F
  group                        3      712.5864389     237.5288130       3.77    0.0280

  Source                      DF      Type III SS     Mean Square    F Value    Pr > F
  group                        3      712.5864389     237.5288130       3.77    0.0280

               Levene's Test for Homogeneity of percentage Variance
                   ANOVA of Squared Deviations from Group Means

                                   Sum of        Mean
             Source          DF     Squares      Square    F Value    Pr > F
             group            3      7064.0      2354.7       0.94    0.4404
             Error           19     47548.9      2502.6
```

Hypothesis test from SAS Output

H_0 : Samples come from distributions with equal variances.
H_1 : Variances of the samples differ.
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr > F) = 0.4404 H_0 is not rejected.
$\therefore$ Sample variances do not differ significantly.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1 : One of the μ's differ
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr > F) = 0.0280 H_0 is rejected.
$\therefore$ There is sufficient evidence to indicate a difference in mean achievement among the four teaching procedures.

From the Scheffe pairwise comparisons on the next page, the percentages of Groups 3 (70.83%) and 4 (87.75%) differs significantly from each other.

```
                           The GLM Procedure
                     Scheffe's Test for percentage
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type
                 II error rate than Tukey's for all pairwise comparisons.

                    Alpha                        0.05
                    Error Degrees of Freedom       19
                    Error Mean Square        62.98058
                    Critical Value of F       3.12735

          Comparisons significant at the 0.05 level are indicated by ***.

                         Difference
                group       Between     Simultaneous 95%
              Comparison      Means    Confidence Limits
                4 - 2         9.321     -5.915   24.557
                4 - 1        12.083     -3.608   27.774
                4 - 3        16.917      1.226   32.608  ***
                2 - 4        -9.321    -24.557    5.915
                2 - 1         2.762    -10.762   16.286
                2 - 3         7.595     -5.929   21.119
                1 - 4       -12.083    -27.774    3.608
                1 - 2        -2.762    -16.286   10.762
                1 - 3         4.833     -9.201   18.868
                3 - 4       -16.917    -32.608   -1.226  ***
                3 - 2        -7.595    -21.119    5.929
                3 - 1        -4.833    -18.868    9.201
```

The display of the results of pairwise comparisons will be different by adding the option 'lines' to the means statement: `means group/hovtest scheffe lines;`

```
                     Scheffe's Test for percentage
          NOTE: This test controls the Type I experimentwise error rate.
                   Alpha                              0.05
                   Error Degrees of Freedom             19
                   Error Mean Square              62.98058
                   Critical Value of F             3.12735
                   Minimum Significant Difference   14.647
                   Harmonic Mean of Cell Sizes    5.508197
                       NOTE: Cell sizes are not equal.


           Means with the same letter are not significantly different.
            Scheffe Grouping          Mean      N    group

                        A           87.750      4    4
                        A
                   B    A           78.429      7    2
                   B    A
                   B    A           75.667      6    1
                   B
                   B                70.833      6    3
```

6.2 Two-way Analysis of Variance

Example: Wackerly, Mendenhall and Scheaffer, page 689, Example 13.5
A stimulus-response experiment involving three treatments was laid out in a randomized block design using four subjects. The response was the length of time until reaction, measured in seconds. The data, arranged in blocks, are shown in the table below. Do the data present sufficient evidence to indicate a difference in the mean responses for stimuli (treatments)? Subjects? Use $\alpha = 0.05$.

	Subject			
Treatment	**1**	**2**	**3**	**4**
1	1.7	1.5	0.1	0.6
2	3.4	2.6	2.3	2.2
3	2.3	2.1	0.8	1.6

Solution
SAS Program
```
data ex13_5w;
input treatment subject time @@;
cards;
1 1 1.7 1 2 1.5 1 3 0.1 1 4 0.6
2 1 3.4 2 2 2.6 2 3 2.3 2 4 2.2
3 1 2.3 3 2 2.1 3 3 0.8 3 4 1.6
;
proc glm;
class treatment subject;
model time= treatment subject;
run;
```

SAS Output
```
                         The GLM Procedure
                      Class Level Information
                   Class          Levels    Values
                   treatment           3    1 2 3
                   subject             4    1 2 3 4

                Number of Observations Read          12
                Number of Observations Used          12

Dependent Variable: time
                                     Sum of
  Source                 DF         Squares     Mean Square    F Value    Pr > F
  Model                   5      8.95166667      1.79033333      23.61    0.0007
  Error                   6      0.45500000      0.07583333
  Corrected Total        11      9.40666667

              R-Square     Coeff Var      Root MSE     time Mean
              0.951630      15.58746      0.275379      1.766667

  Source                 DF       Type I SS     Mean Square    F Value    Pr > F
  treatment               2      5.47166667      2.73583333      36.08    0.0005
  subject                 3      3.48000000      1.16000000      15.30    0.0032

  Source                 DF     Type III SS     Mean Square    F Value    Pr > F
  treatment               2      5.47166667      2.73583333      36.08    0.0005
  subject                 3      3.48000000      1.16000000      15.30    0.0032
```

Hypothesis test from SAS Output

(a)

H_0 : The average times are the same for different treatments.

H_1 : At least one of the average times for treatments differ from the rest.

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (Pr > F) = 0.0005 H_0 is not rejected.

$\therefore$ The mean stimulus-response times for different treatments do differ significantly from one another.

(b)

H_0 : The average times are the same for different subjects.

H_1 : At least one of the average times for subjects differ from the rest.

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (Pr > F) = 0.0032 H_0 is rejected.

$\therefore$ The mean stimulus-response times for different subjects do differ significantly from one another.

6.3 Factorial Design

Factorial design
(r x k) factorial design with h repetitions per cell

Factor A: r levels
Factor B: k levels

rkh units are randomly assigned to rk cells so that every combination of treatments is applied to h units.

Interaction
There is interaction between Factor A and Factor B if the relationship between the mean response and the different levels of one factor depends upon the level of the other factor.

Example: Steyn, Smit, Du Toit and Strasheim, page 530, Example 14.7
Two of the factors that might affect the number of fungi counted on the leaves of an avocado pear tree are
- Factor A: Concentration of the fungicide used to spray the leaves, namely, a 10% or 30% solution of copper sulphate.
- Factor B: The position (N, S, E, W) of the section of the tree from which leaves are selected for examination.

The table below reflects the data obtained after random choice of four trees for each of the eight combinations and fungus counts.

Number of fungi counted on the leaves of 32 avocado pear trees

Fungicide (Factor A)	Position (Factor B)			
	North	South	East	West
10% soln	93	69	84	73
	78	64	65	76
	85	61	79	71
	81	58	81	77
30% soln	103	49	67	80
	108	44	73	81
	107	38	84	85
	111	41	82	78

Solution
SAS Program
```
proc format;
value a 1='10% solution'  2='30% solution';
value b 1='North'  2='South'  3='East'  4='West';
data ex_14_7;
input a b number @@;
format a a. b b.;
cards;
1 1  93 1 1  78 1 1  85 1 1  81
1 2  69 1 2  64 1 2  61 1 2  58
1 3  84 1 3  65 1 3  79 1 3  81
1 4  73 1 4  76 1 4  71 1 4  77
2 1 103 2 1 108 2 1 107 2 1 111
2 2  49 2 2  44 2 2  38 2 2  41
2 3  67 2 3  73 2 3  84 2 3  82
2 4  80 2 4  81 2 4  85 2 4  78
;
proc glm;
class a b;
model number=a b a*b;  run;
```

SAS Output

```
                         General Linear Models Procedure
                             Class Level Information
                   Class    Levels    Values
                   A             2    10% solution 30% solution
                   B             4    East North South West
                     Number of observations in data set = 32
Dependent Variable: NUMBER
Source                    DF        Sum of Squares          Mean Square   F Value     Pr > F
Model                      7         9328.87500000        1332.69642857     43.11     0.0001
Error                     24          742.00000000          30.91666667
Corrected Total           31        10070.87500000

                    R-Square                  C.V.             Root MSE          NUMBER Mean
                    0.926322              7.334247           5.56027577          75.81250000

Source                    DF             Type I SS          Mean Square   F Value     Pr > F
A                          1           40.50000000          40.50000000      1.31     0.2637
B                          3         7378.62500000        2459.54166667     79.55     0.0001
A*B                        3         1909.75000000         636.58333333     20.59     0.0001

Source                    DF           Type III SS          Mean Square   F Value     Pr > F
A                          1           40.50000000          40.50000000      1.31     0.2637
B                          3         7378.62500000        2459.54166667     79.55     0.0001
A*B                        3         1909.75000000         636.58333333     20.59     0.0001
```

Table with mean number of fungicide for different concentrations and positions
(These results can be obtained from a Proc Means procedure)

Position	**Concentration 10%**	**Concentration 30%**
North	84.25	107.25
South	63.00	43.00
East	77.25	76.50
West	74.25	81.00

Hypothesis test from SAS Output

$H_0(a)$: The effects of the levels of factor A are the same.
∴The number of fungi counted is the same for different concentrations of fungicide.
$H_1(a)$: The effects of the levels of factor A differ.
∴The number of fungi counted differ for different concentrations of fungicide.
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr > F) = 0.2637 H_0 is not rejected.

$H_0(b)$: The effects of the levels of factor B are the same.
∴The number of fungi counted is the same for different positions (north, south, east or west) of the tree.
$H_1(b)$: The effects of the levels of factor B differ.
∴The number of fungi counted is the same for different positions (north, south, east or west) of the tree.
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr > F) = 0.0001 H_0 is rejected.

$H_0(ab)$: There is no interaction between factors A and B.
∴ There is no interaction between concentration of fungicide used and the position on the tree.
$H_1(ab)$: There is interaction between factors A and B.
∴ There is interaction between concentration of fungicide used and the position on the tree.
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Pr > F) = 0.0001 H_0 is rejected.

From the above hypothesis tests it follows that both position and interaction affect the number of fungicide counted. From the data one can conclude that trees on the northern side that were treated with a 10% fungicide solution had lower fungicide counts than the ones that were treated with a 30% fungicide solution. We arrive at the opposite conclusion when comparing the averages on the southern side.

7. NONPARAMETRIC TESTS FOR MORE THAN TWO MEDIANS

7.1 Kruskal-Wallis Test

Example: Wackerly, Mendenhall and Scheaffer, page 767, Example 15.7
A quality control engineer has selected independent samples from the output of three assembly lines in an electronics plant. For each line, the output of ten randomly selected hours of production was examined for defects. Do the data in the table below provide evidence that the probability distributions of the number of defects per hour of output differ in location for at least two of the lines? Use $\alpha = 0.05$.

Number of defects from three independent assembly lines

Line	Percentage									
1	6	38	3	17	11	30	15	16	25	5
2	34	28	42	13	40	31	9	32	39	27
3	13	35	19	4	29	0	7	33	18	24

Solution
SAS Program
```
data ex15_7w;
input line defects @@;
cards;
1  6 1 38 1  3 1 17 1 11 1 30 1 15 1 16 1 25 1  5
2 34 2 28 2 42 2 13 2 40 2 31 2  9 2 32 2 39 2 27
3 13 3 35 3 19 3  4 3 29 3  0 3  7 3 33 3 18 3 24
;
proc npar1way wilcoxon;
class line;
var defects;
run;
```

SAS Output
```
                        The NPAR1WAY Procedure
              Wilcoxon Scores (Rank Sums) for Variable defects

                        Classified by Variable line
                        Sum of      Expected        Std Dev          Mean
line          N         Scores      Under H0       Under H0         Score
---------------------------------------------------------------------------
1            10         120.00         155.0      22.727774        12.000
2            10         210.50         155.0      22.727774        21.050
3            10         134.50         155.0      22.727774        13.450

                     Average scores were used for ties.

                            Kruskal-Wallis Test
                         Chi-Square         6.0988
                         DF                      2
                         Pr > Chi-Square    0.0474
```

Hypothesis test from SAS Output
$H_0 : \eta_1 = \eta_2 = \eta_3$
H_1 : At least two medians differ
Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.
Since p-value (Prob > CHISQ) = 0.0474 H_0 is rejected.
$\therefore$ At least one of the three lines produce a significant different median number of defects than the other lines.

7.2 Friedman Test

Example: Steyn, Smit, Du Toit and Strasheim, page 610, Example 16.9
Ten wine tasters each had to judge different types of semi-sweet wines and arrange them in order of preference. The data is given below. Test whether there is a significant difference in the popularity of the four wines among the wine tasters.

Orders of preference of ten wine tasters for four types of wine

	Wine taster									
Type of wine	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	2	2	1	4	4	2	3	4	1	4
B	3	3	4	3	1	3	2	2	2	1
C	1	1	2	1	2	1	1	1	3	2
D	4	4	3	2	3	4	4	3	4	3

Solution
SAS Program

```
data ex_16_9;
input taster1-taster10;
cards;
2 2 1 4 4 2 3 4 1 4
3 3 4 3 1 3 2 2 2 1
1 1 2 1 2 1 1 1 3 2
4 4 3 2 3 4 4 3 4 3
;
proc transpose prefix=wine;

proc means noprint sum;
var wine1-wine4;
output out=sumsq1 sum=sum1-sum4;
proc print;

data sumsq2;
set sumsq1;
t2=sum1**2+sum2**2+sum3**2+sum4**2;

data final;
set sumsq2;
h=10; c=4; n=h*c; df=c-1;
q=12*t2/(n*(c+1))-3*h*(c+1);
p_value=1-probchi(q,df);
proc print;
var c h n t2 q p_value;

proc print data=sumsq1;
title1 'Data Set sumsq1';
proc print data=sumsq2;
title1 'Data Set sumsq2';
proc print data=final;
title1 'Data Set final';

run;
```

SAS Output

```
Data Set final
OBS    C     H     N     T2       Q       P_VALUE
1      4     10    40   2686     11.16    0.010891

Data Set sumsq1
OBS    _TYPE_    _FREQ_    SUM1    SUM2    SUM3    SUM4
1        0        10        27      24      15      34

Data Set sumsq2
OBS    _TYPE_    _FREQ_    SUM1    SUM2    SUM3    SUM4     T2
1        0        10        27      24      15      34     2686

Data Set final
OBS  _TYPE_  _FREQ_  SUM1  SUM2  SUM3  SUM4   T2    H   C   N   DF    Q      P_VALUE
1      0       10     27    24    15    34    2686  10  4   40  3    11.16   0.010891
```

Solution (Alternative)
SAS Program

```
data ex_16_9;
input wine$ taster order @@;
cards;
A 1 2 A 2 2 A 3 1 A 4 4 A 5 4 A 6 2 A 7 3 A 8 4 A 9 1 A 10 4
B 1 3 B 2 3 B 3 4 B 4 3 B 5 1 B 6 3 B 7 2 B 8 2 B 9 2 B 10 1
C 1 1 C 2 1 C 3 2 C 4 1 C 5 2 C 6 1 C 7 1 C 8 1 C 9 3 C 10 2
D 1 4 D 2 4 D 3 3 D 4 2 D 5 3 D 6 4 D 7 4 D 8 3 D 9 4 D 10 3
;
proc means noprint sum;
var order;
by wine;
output out=sumsq1 sum=sumrank;

proc means noprint uss;
var sumrank;
output out=sumsq2 uss=t2;

data final;
set sumsq2;
h=10; c=4; n=h*c; df=c-1;
q=12*t2/(n*(c+1))-3*h*(c+1);
p_value=1-probchi(q,df);
proc print;
var c h n t2 q p_value;

proc print data=sumsq1;
title1 'Data Set sumsq1';
proc print data=sumsq2;
title1 'Data Set sumsq2';
proc print data=final;
title1 'Data Set final';

run;
```

SAS Output

```
Data Set final
OBS    C    H    N    T2     Q      P_VALUE
1      4    10   40   2686   11.16  0.010891

Data Set sumsq1
OBS   WINE   _TYPE_   _FREQ_   SUMRANK
1     A       0        10        27
2     B       0        10        24
3     C       0        10        15
4     D       0        10        34

Data Set sumsq2
OBS   _TYPE_   _FREQ_    T2
1       0        4      2686

Data Set final
OBS   _TYPE_   _FREQ_    T2     H    C    N    DF    Q       P_VALUE
1       0        4       2686   10   4    40   3     11.16   0.010891
```

Hypothesis test from SAS Output

H_0 : The four types of wine are equally popular

H_1 : The four types of wine are not equally popular

Use $\alpha = 0.05$. Reject H_0 if p-value < 0.05.

Since p-value (P_VALUE) = 0.010891 H_0 is rejected.

$\therefore$ The four types of wine are not equally popular among the wine tasters.

Example: Steyn, Smit, Du Toit and Strasheim, page 608, Example 16.8

Write a SAS program to give the ranks of the building valuations, as obtained by arranging the valuations blockwise (see 4.2).

Solution

SAS Program

```
data ex_16_8;
input block1 - block11;
cards;
320 160 195 165 300 315 120 170 120 300 210
430 405 170 160 206 190 165 215 159 150 320
515 185  95 185 220 160 205 240 305 240 235
;
proc rank;
proc print;
run;
```

SAS Output

OBS	BLOCK1	BLOCK2	BLOCK3	BLOCK4	BLOCK5	BLOCK6	BLOCK7	BLOCK8	BLOCK9	BLOCK10	BLOCK11
1	1	1	3	2	3	3	1	1	1	3	1
2	2	3	2	1	1	2	2	2	2	1	3
3	3	2	1	3	2	1	3	3	3	2	2