THE UNIVERSITY OF SYDNEY
STAT2012 STATISTICAL TESTS

Semester 2	**Assignment 2**	2015

Important points

- In each hypothesis test, you should state the hypotheses, test statistic, assumption, p-value and conclusion.
- You should show working. Computer output will not be accepted as working.
- You are encouraged to include the R outputs as checks for your answers. Note that you can enter observations in a vector and then create the necessary factor labels. For example, in Q1, you can enter data as

  ```
  y=c(29,24,20,26,32,31,32,26,21,23,31,25,18,20,25,28,22,30)
  ```

  and then create factor labels for the 2 levels of gender and 3 levels of software.

1. A professor studied which of the three statistical softwares was the best for his students. He believed that the time (in hours) it took a student to master a particular software might be influenced by gender. A 3×2 factorial experiment with three replicates was designed, as shown below:

Software	Excel	SAS	R
Male	29	32	18
	24	26	20
	20	21	25
Female	26	23	28
	32	31	22
	31	25	30

   (a) Complete the ANOVA table.

   (b) Test at the 5% level of significance if the times students take to master a software differ across types of software.

   (c) Test at the 5% level of significance if the factors of software type and gender of students interact.

2. Eighteen business people who fly frequently were asked to evaluate four airlines that they have used in the last trip based on 20 questions on a 5-point Likert scale. The results are shown below:

Airlines

	A	B	C	D
	36	48	57	84
	18	30	60	71
	24	42	78	66
	39	48	60	57
		52		60
Mean	29.25	44.00	63.75	67.60
Var.	98.25	74.00	92.25	113.30

Test at the 5% level of significance if the mean scores differ across airlines using

(a) the ANOVA method by first filling up an ANOVA table,

(b) the Bonferroni multiple comparison method if the result in (a) is significant, and

(c) the Kruskal-Wallis test.

Compare the three results.

3. Five business people who fly frequently were asked to evaluate four airlines that they have used before based on 20 questions on 5-point Likert scale. The results are shown below:

Airlines

Person	A	B	C	D
1	89	84	78	76
2	77	67	52	81
3	85	77	75	69
4	65	72	62	73
5	58	47	52	62

(a) Test at the 5% level of significance if the variance of scores for airline D is less than 150. Provide a 95% lower-sided confidence interval for the true variance.

(b) Test at the 5% level of significance if the variances of scores for airlines B and D differ. Provide a 95% 2-sided confidence interval for the ratio of true variances.

(c) Test at the 5% level of significance if the mean scores differ across airlines using the Friedman test.

4. A study was carried out to find a relationship between the number of customers, X, and weekly sales, Y, (in thousand dollars) of a drug store. A random sample of 10 weekly sales gives the data as follows:

X	907	926	506	741	789	889	874	510	529	410
Y	112.0	110.5	68.4	92.1	94.2	100.8	94.5	67.3	72.4	61.2

The following summary statistics are given:

$$\sum_{i=1}^{10} x_i = 7,081;\ \sum_{i=1}^{10} y_i = 873.4;\ \sum_{i=1}^{10} x_i^2 = 5,370,001;\ \sum_{i=1}^{10} y_i^2 = 79,396.24;\ \sum_{i=1}^{10} x_i y_i = 651,006.1.$$

(a) Fit a linear regression line, $Y = \alpha + \beta X$, to the data and interpret the result. Plot the data and the regression line.

(b) Test, at the 5% level of significance, if the linear regression model of Y on X is significant.

(c) Construct a 95% Prediction Interval of the sales of a particular week if 750 customers visit the store during the week?

(d) Construct a 95% Estimation Interval of the expected weekly sales if 750 customers visit the store during the week?

(e) According to the linear regression line obtained in (a), what is the expected sales of the week if the drug store was closed on that week for refurbishment. Is the estimate reasonable? Explain.

(f) Compute the sample correlation coefficient r and the coefficient of determination r^2 between the weekly sales and the number of customers. Interpret the value of r^2.