

**Question 1 [7 marks]**

In an effort to determine if men do tend to shy away from marrying women smarter than themselves, a sample of 10 newly-wed couples was selected and an IQ test given to each person. The (fictional) data for each couple are given in Appendix A, along with various graphs and analyses.

- a) Identify the most appropriate test, giving reasons for your choice, including references to specific graphs.

[2] -The data are matched pairs as the man and woman are paired.  
 - Since the data are paired, we look only at the boxplot of differences which suggests a paired t-test since the differences appear to come from a normal distribution. No marks if they look at the other boxplots given their observation of paired data. If they think the data are independent, then they should look at the two boxplots for the two samples.

- b) Regardless of your answer in part (a), complete the most appropriate parametric test for testing whether the data support the view that men tend to marry women not as smart as themselves. Use the .05 level of significance.

[4] Assuming we calculate the difference by taking the husband's IQ score and subtracting the wife's IQ score from it we get the following hypotheses:

$$H_0 : \mu_d = 0$$

$$H_a : \mu_d > 0$$

Difference 9 1.77778 2.72845 0.90948

95% lower bound for mean difference: 0.08655

T-Test of mean difference = 0 (vs > 0): T-Value = 1.95 P-Value = 0.043

The test statistic is  $1.78/.909 = 1.95$ . The critical value is 1.86 (*there are only 9 pairs, not 10 as the question states, and 8 df*). We reject the null hypothesis and conclude that there is evidence that men tend to marry women with a lower IQ.

- one mark for hypotheses
- one mark for test-statistic
- one mark for critical value and critical region
- one mark for decision to reject null H and conclusion

If a 2-sample t-test is completed, then I would suggest full marks if they suggested the two "samples" are independent in (a) since they have already lost that mark.

- c) Why might there be a problem to make a claim about the relative intelligence of men and women?

[1] The two samples are not random samples representative of the larger populations of men and women. They are linked by being married to each other. Thus we cannot conclude anything about the relative IQ of men and women.

***No marks for commenting on the small sample size.***

**Question 2 [ 6 marks]**

In another study of men and women, a sample of ten men and a sample of ten women were enlisted in a study in effort to determine whether women are truly more concerned about hygiene than men. Each participant in the study was followed for a week and the number of times they washed their hands was recorded. The (fictional) data are given below, with various graphs and analyses in Appendix B:

- a) Identify the most appropriate test, giving precise reasons, with reference to specific graphs and the manner in which the data were collected.

[2] -The data are not paired in this case  
-the boxplots suggest that there are outliers.

This suggests the non-parametric Mann-Whitney U test. I would give only 1 mark out of the 2 for the suggestion of a two sample t-test....

- b) Perform this *most* appropriate test to determine if women do wash their hands more frequently than men. Use a 5% significance level.

[3] The hypotheses are

$$H_0 : M_M - M_W = 0$$

$$H_a : M_M - M_W < 0$$

(where M is the population median)

The test statistic is  $W = 62.5$

which is significant at any significance level greater than 0.0234. Thus there is evidence that women wash their hands more frequently.

Full marks for someone who puts a two sample t-test down for part (a) and then does a two sample t-test here in part (b).

- c) Explain whether this experiment would allow you to make a general statement about the difference in hygiene between men and women.

[1] If you believe hand washing is a good indicator of hygiene then this experiment does allow you to make a more general statement about the difference in hygiene between men and women. If you do not believe it is a good indicator then you won't accept the generalization. ***The "fact" that these people were being followed for a week means that it is possible women appeared to wash their hands more often because women are more prone to changing their behaviour than men, when conscious of being watched.***

***No marks for commenting on the small sample sizes.***

**Question 3. [ 19 marks ]**

An agricultural scientist conducted an experiment to determine the effects of two fertilizers and four wheat types on crop yields. Appendix C gives the data and various analyses. Complete all hypothesis tests at the .05 level of significance.

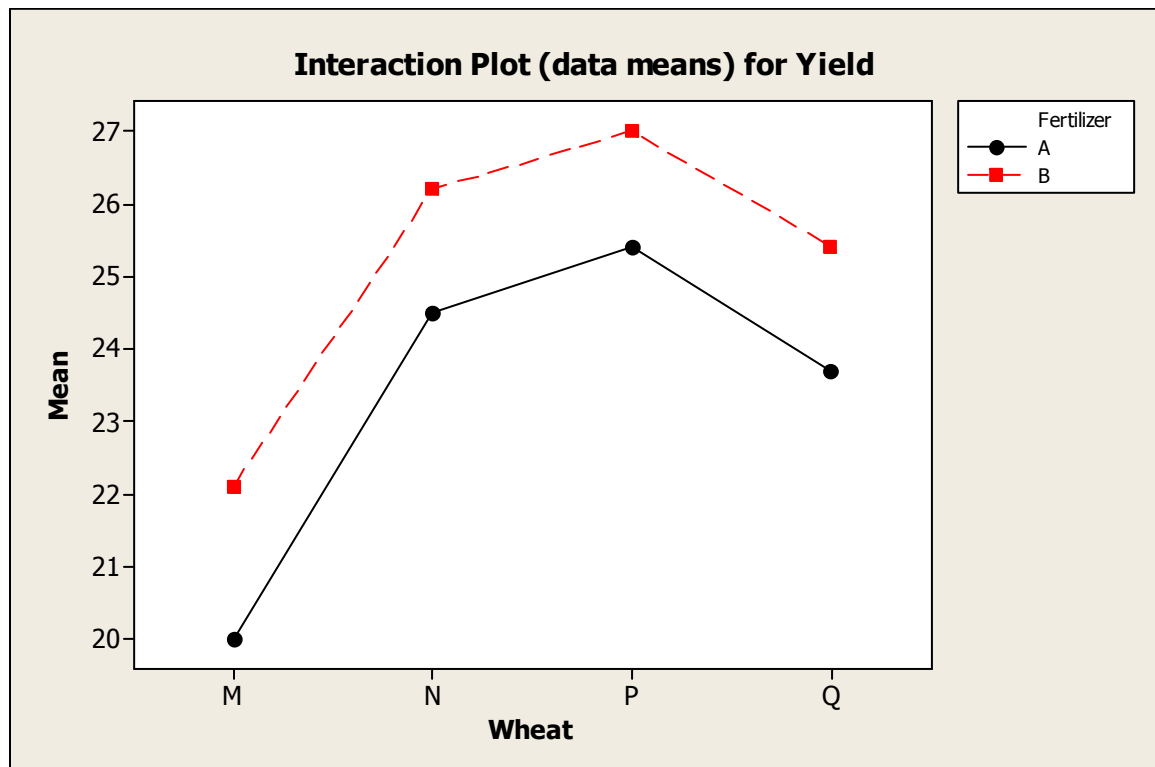
a) Comment on whether the model assumptions are warranted.

[2] -The normality plot and the histogram suggest that the normality assumption is warranted. There are also no points that are more than two SD away from zero in the residual plot so outliers are not an issue.

-The “even” scatter in the residual plot suggest that the equal variance assumption is also warranted.

b) Sketch the interaction plot, using appropriate labels. What does this suggest to you about the possibility of interaction and of main effects. Explain briefly.

[5] The parallel line segments suggests that there little interaction between the two factors. It also suggests a main effect due to fertilizer and a main effect due to wheat.



-one mark for labelling the y-axis properly

-one mark for labelling the two “response curves” by the fertilizers

-one mark for the plots

-one mark for comment on no interaction

-one mark for comments on main effects

c) Fill in the missing terms in the ANOVA table below:

[3]

Source	DF	SS	MS	F	P
Fert	1	18.904	18.9037	48.63	0.000
Wheat	3	92.021	30.6738	78.90	0.000
Interaction	3	0.221	0.0738	0.19	0.902
Error	16	6.22	0.3887		
Total	23	<b>117.366</b>			

S = 0.6235    R-Sq = 94.70%    R-Sq(adj) = 92.38%

-one mark for the DF

-one mark for the SS and MS (given the DF)

-one mark for the F-statistics, given their MS values

d) Test whether the effect of Fertilizer on the yield depends on the Wheat Type.  
Copy the F-statistic from the above table, but complete the test using the p-value.

[3] -one mark for hypotheses, Ho: no interaction, Ha: some interaction

-The F statistic for an interaction is 0.19 with a p-value of 0.902.

-Thus there is no evidence that the effect of fertilizer is dependent on wheat type or that there is interaction between wheat and fertilizer in their effects on yield.

e) Test for differences in mean yields among the Wheat Types.

Copy the F-statistic from the above table, but complete the test using the p-value.

[3] -one mark for hypotheses, Ho: no wheat effect, Ha: some wheat effect

-The main effects of fertilizer has an F-statistic of 78.90 with a p-value < 0.001.

-Thus there is strong evidence of a difference in yield based on wheat type.

f) After examining the individual confidence intervals for Wheat yield means, you decide to test the difference in means between Wheat Types M and Q. Calculate the appropriate confidence interval. What do you conclude?

[3] Since the test is done after examining the means (and implicitly doing the comparisons), we must do a Tukey paired comparison.

The  $q(.05)$  value is 4.046, based on num 4 and denom 16 d.f.

The interval is  $3.5 \pm (4.046/\sqrt{2}) \cdot .6235 \cdot \sqrt{(2/6)}$  or  $3.5 \pm 2.86 \cdot .6235 \cdot \sqrt{(1/3)} =$

$3.5 \pm 1.03$  or  $3.5 \pm 1.0$ .

If they use a t-interval, then the  $t(\alpha/2)$  value is 2.12 based on 16 d.f., in which case the interval is  $3.5 \pm 2.12 \cdot .6235 \cdot \sqrt{(1/3)} = 3.5 \pm .76 = 3.5 \pm .8$ .

-one mark for observed difference of 3.5 and standard error of  $.6235 \cdot \sqrt{(2/6)}$

-one mark for  $q(.05)/\sqrt{2}$  and **no mark for t-value**.

-one mark for concluding there is a true difference in mean yields between M and Q.

**Question 4. [9 marks]**

A polling firm collected data on two environmentally conscious cities. The people who were actively involved (token participation in Earth Day was strictly symbolic!) in protecting the environment included those who used CFLs (Compact Fluorescent Lights) and hybrid cars on a regular basis to reduce their “carbon footprint” on the environment.

City#1	
Sample Size:	200
People With Active Involvement:	68

City#2	
Sample Size:	300
People With Active Involvement:	93

- (a) What is the 95% confidence interval for the proportion of people in City#1 who are actively involved?

$$\begin{aligned}
 [3] \text{ 95\% CI: } \hat{p} \pm SE(\hat{p}) &= \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.34 \pm 1.96 \sqrt{\frac{0.34(0.66)}{200}} \\
 &= 0.34 \pm 1.96 * 0.0335 \\
 &= 0.34 \pm 0.0657 \\
 &= (0.2743, 0.4057)
 \end{aligned}$$

- 1.5 marks for  $\hat{p}$  &  $SE(\hat{p})$
- 0.5 mark for  $z_{\alpha/2}$ ,
- 1.0 mark for correct CI.

If 1.645 is used as  $z_{\alpha/2}$ , then 95% CI: (0.2849, 0.3951) and it is not correct. Reduce 1.5 marks. (TQ—let's reduce by .5 for wrong z-value) I would not expect a 1-sided CI based on  $z(\alpha)=1.645$ , but it is possible.

- (b) What sample size would be required to have a 95% confidence interval in part (a) that has a margin of error of  $\pm 3\%$ ? Assume the true population proportion is unknown.

$$[2] n_{\max} = \frac{z_{\alpha/2}^2 pq}{E^2} = \frac{1.96^2 (0.5)(0.5)}{(.03)^2} = \left( \frac{1.96}{2(0.03)} \right)^2 = 1067.1111 \approx 1068 \text{ (Round up)}$$

- 1 mark for correct formula and assuming  $p = 0.5$  or using  $p\text{-hat} = .34$
- 1 mark for correct value of  $n$
- Reduce 0.5 mark if  $z_{\alpha/2} = 1.645$

- (c) Test the hypothesis that the proportions of actively involved people in City#1 and City#2 are different. Use a 5% level of significance.

$$S1: H_0 : p_1 = p_2, H_a : p_1 \neq p_2$$

(Two-Tailed Test)

$$S2: z_{Calc} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{SE(\hat{p}_1 - \hat{p}_2)_{Pooled}} = \frac{(0.34 - 0.31) - 0}{\sqrt{0.322 * (0.678) \left( \frac{1}{200} + \frac{1}{300} \right)}} = \frac{0.03}{0.0427} = 0.7033$$

S3:  $z_{Crit} = z_{\alpha/2} = 1.96$

S4: Since  $\{|z_{Calc}| = 0.7033\} < \{z_{Crit} = 1.96\} \rightarrow$  Do NOT Reject Ho.

There is no difference in proportion of people in the two cities who are actively involved.

[4]

- 1 Mark each for the 4 steps or their other forms (such as 6 steps or 7 steps!).
- Reduce 1 mark for each of the wrong steps without penalizing twice. For example, if  $SE(\hat{p}_1 - \hat{p}_2)$  is wrong but other steps are right, reduce 1 mark only.

**Question 5. [5 marks]**

A business school wishes to examine the grade distributions in some undergraduate courses. It collected the following data broken down by ‘Letter Grade’ and the ‘Course’:

Course	Letter Grade					Totals
	A	B	C	D	F	
Statistics:	22	35	50	35	23	165
BusDecision:	36	40	44	20	10	150
Accounting:	30	35	45	30	12	152
HResources:	40	35	35	30	5	145
Totals	128	145	174	115	50	612

Partial Minitab output for the analysis was obtained:

	A	B	C	D	F	Total
1	22	35	50	35	23	165
	34.51	39.09	46.91	31.00	13.48	
	4.535	0.429	0.203	0.515	6.723	
2	36	40	44	20	10	150
	31.37	35.54	42.65	28.19	12.25	
	0.683	0.560	0.043	2.378	0.415	
3	30	35	45	30	12	152
	31.79	36.01	43.22	28.56	12.42	
	0.101	0.028	0.074	0.072	0.014	
4	40	35	35	30	5	145
	30.33	34.35	41.23	27.25	11.85	
	3.085	0.012	0.940	0.278	3.957	
Total	128	145	174	115	50	612

Chi-Sq = 25.044, DF = 12, P-Value = 0.015

Complete the test of hypothesis, using the .05 level of significance. What is the approximate p-value?

S1:  $H_0$ : Letter Grade and Course are (Statistically) Independent Variables

$H_a$ : Letter Grade and Course are (Statistically) NOT Independent Variables

Alternate wording which is acceptable is:

Courses have the same Grade Distribution, etc.

$$S2: \chi^2_{Calc} = \sum_{i,j} \chi^2_{ij} = 25.044$$

$$S3: \chi^2_{Crit} = \chi^2_{\alpha=0.05} (df = (r-1)(c-1)) = \chi^2_{\alpha=0.05} (12) = 21.0261$$

S4: Since  $\{\chi^2_{Calc} = 25.044\} > \{\chi^2_{Crit} = 21.0261\} \rightarrow$  Reject  $H_0$ .

The "Letter Grade" and "Course" variables are statistically NOT Independent.

$$S5: p\text{-Val} = P[\chi^2(12) > \chi^2_{Calc} = 25.044] \times 1$$

Thus, from the table,  $0.01 < p\text{-Val} < 0.05$  (*TQ—table provided yields between .01 and .02*)

**{The actual value is 0.015}**

- 1 Mark each for S2, S3, S4, and S5.
- Do not penalize repeatedly for errors which are carried over. However, if various  $E_{ij}$  and  $O_{ij}$  values are wrong, then reduce two marks or more.

**Question 6.**

- (a) Examine the scatter and residual plots and comment on whether the regression model assumptions are warranted or not in each case.

[2]

- no particular problem with assumption of normally dist. Errors as the relative number of outlying residuals is consistent with the empirical rule
- no particular problem with assumption of constant variance (one could observe a slight case of increasing variance in model 1 and perhaps the opposite in model 2).

For the remaining questions, **refer only to Model 1**, and unless otherwise stated, **perform all tests at the .05 level of significance.**

- (b) Test whether the model is useful for predicting the real estate taxes.

- [3] -Ho:  $\beta_1=0, \beta_2=0, \dots, \beta_6 = 0$ , Ha: one  $\beta_j$  is nonzero ( $j=1, \dots, 6$ )  
 - $F= 772866/12420 = 62$ , rejection region is  $F > 2.2$  (7 and 52 d.f.)  
 -decide to reject null H, conclude model is useful.

- (c) Calculate the R-square value.

[1] -  $5410065/6055922 = .893$

- (d) Test whether the **Price** variable is a useful predictor of real estate taxes, given the other predictors in the model.

- [3] -Ho:  $\beta(\text{price}) = 0$ , Ha:  $\beta(\text{price})$  not zero  
 - $t = .6122/.1378 = 4.44$ , rejection region is  $t > 2.01$  (52 d.f.) or  $t > 1.96$   
 -decide to reject null H, conclude Price is useful predictor of taxes, given the other predictors.

- (e) Give a precise interpretation for the numerical value of the (estimated) coefficient for the **Price** variable, using the appropriate units of measurement.

[2] For an increase of one unit (\$100) in the price of a house, the **estimated** increase in taxes is \$.61.

-one mark for idea of slope .61 (unit increase in price resulting in .61 increase in taxes)

-one mark for proper units

*We will not insist on the adjective "estimated".*

- (f) What is the meaning of the estimated coefficient corresponding to the NE variable?

[1] Houses in the NE area of the city pay, on average, \$18.32 in taxes more than houses not in the NE area.

(g) Calculate a **99% interval** for the average taxes for 2774 sq. ft. houses selling for \$180,000 that are two years old, have four features, located in the NE area, are not custom built, and are not on a corner lot.

[3] This is a confidence interval since it refers to average taxes for a population of houses meeting the criteria stated.

The CI is "Fit"  $\pm$  t(alpha/2) \* "SE Fit"

or  $\$1501.6 \pm 2.68 * (\$46) = \mathbf{\$1501.6 \pm \$123}$  ( \$118 if z-value of 2.576 used)

= ( \$1379, \$1625 )

-one mark for Fit value and SE Fit value

-one mark for t of 2.68 or 2.57 or 2.58

-one mark calculating  $1800 \pm 123$  or lower and upper limits

One mark for the 95% interval as given in the Minitab output since this is a recognition of the need for a confidence interval Give full marks for taking the 95% CI and adjusting it for the 99% CI by using the t(.05/2) value to calculate the SE of 46.

(h) Calculate a **99% interval** for the actual taxes for a house with the characteristics listed in part (f).

[2] **This is a prediction interval since it refers to an actual house.**

-one mark for standard error of  $\text{sqrt}(46^2 + 111.45^2) = \$120.57$

-one mark for interval  $1501.6 \pm 2.68 * 120.57 = \mathbf{\$1501.6 \pm 323}$  ( $\pm 311$  if z used).

(let's not worry about the dollar units) = (\$1179, \$1825 ).

No mark for the 95% interval as given in the Minitab output.

If they take the 95% prediction interval in output, divide the ME by 2.01 to obtain the SE, and then multiply by 2.68 to obtain the ME for the 99% PI, then give full marks.

Only one mark for dividing the ME by 1.96 and multiplying by 2.576, as Minitab uses the t-values and not the z-values.