ADM 2304 Fall 2008
Final Exam Solutions

1
(a)

Let p be the proportion supporting an election (ignoring the undecided).

-Ho: p=.5: Ha: p>.5
-p-hat = 430/825 = .5212
-z = .0212 / sqrt(.5*.5/825) = 1.22
-z(.05)= 1.645
-do not reject null H, and conclude there were not more people supporting an election over coalition

Optional alternative approach to compare two sample proportions from the same sample.

If Ha: p(election) > p(coalition), then must calculate standard error accounting for *non-independent sample proportions:*

-p1hat - p2hat = .5212 - .4788 = .0424
-std err is sqrt(.52*.48 / 825 + .52*.48 / 825 + 2 * .52*.48 / 825) and z = 1.22

We do not need to ignore the undecideds if we compare the non-independent proportions:
-p1hat = 430/992 = .433, p2hat = 395/992 = .398,
-std err is sqrt(p1hat * q1hat /992 + p2hat*q2hat/992 + 2*p1hat*p2hat/992) = .028933
-z = (.433468 - .398185) / .028933 = 1.219457 = 1.22.

For some strange reason, many students calculated std err as
Sqrt(p-bar * q-bar * (1/430 + 1/395). This is obviously incorrect.

(b) Ho: p(W) = .30, p(Ont)=.39, p(Que)=.24, p(Mar)=.07

	Observed Freq	Exp.Freq = 992*p(i)	Cont. To chi-square
West	301	297.6	.0388
Ont	390	386.88	.0252
Que	240	238.08	.0155
Mar	61	69.44	1.0258
Total	992	992	1.1

Df. = 3
Critical value is 7.81
We do not reject the null H since 1.1 not > 7.81
Conclude the observed distribution is consistent with the population distribution.

(c)

Test and CI for Two Proportions

Sample	X	N	Sample p
1	39	135	0.288889
2	152	390	0.389744

Difference = p (1) - p (2)
Estimate for difference: -0.100855
95% CI for difference: (-0.191344, -0.0103652)
Test for difference = 0 (vs not = 0): Z = -2.10 P-Value = 0.036

Ho: p1=p2; Ha: p1 ≠ p2
P1-hat = 39/135 = .29, p2-hat = 152/390 = .39
Pooled proportion p-hat is (39+152)/(135+390) = .36
Std err is sqrt(.36*.64*(1/135 + 1/390)) = .048
Z = -.10 / .048 = -2.1 or 2.1
Reject null H since |z| > 1.96 for alpha = .05
Conclude that there is a difference between the proportions of BC voters and Ontario voters who supported the coalition.

(d)

Chi-Square Test: C1, C2, C3, C4, C5, C6

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	C1	C2	C3	C4	C5	C6	Total
1	73	75	35	168	60	19	430
	58.52	45.95	26.01	169.05	104.03	26.44	
	3.584	18.370	3.109	0.007	18.637	2.094	
2	39	19	17	152	139	29	395
	53.76	42.21	23.89	155.29	95.56	24.29	
	4.050	12.761	1.988	0.070	19.742	0.914	
3	23	12	8	70	41	13	167
	22.73	17.84	10.10	65.66	40.40	10.27	
	0.003	1.914	0.437	0.288	0.009	0.726	
Total	135	106	60	390	240	61	992

Chi-Sq = 88.701, DF = 10, P-Value = 0.000

Table above gives the missing values.

Ho: distribution of responses same across region (response and region are independent)
Ha: disribution differs by region (response and region are associated)

Chi-square stat is 88.7, critical value is 18.31 based on alpha of .05 and 10 d.f.
Reject null H and conclude that the distribution of responses does differe by region
p-value is prob (chi-square > 88.7) which is less than .0000005
(chi-square table gave critical values for .0000005 as 48.51 for 10 d.f.)

2.

(a) The two samples are independent since they are each random samples from distinct populations. Just because both advisors are new or come from the same firm does not make the samples paired.

(b) Since the samples are independent, the sample of differences is not meaningful and we must ignore the boxplot of differences. To decide between the parametric and non-parametric tests, we look at the two boxplots for advisor1 and advisor2-- they are both symmetric and there are no outliers, making it reasonable to assume that each sample is from a normally distributed population. Thus the most appropriate test is the 2-sample t-test.

(c)

Two-Sample T-Test and CI: Advisor1, Advisor2

```
Two-sample T for Advisor1 vs Advisor2

           N    Mean  StDev  SE Mean
Advisor1  15  -29.57   2.82     0.73
Advisor2  15  -32.25   2.42     0.63

Difference = mu (Advisor1) - mu (Advisor2)
Estimate for difference:  2.68000
95% CI for difference:  (0.71386, 4.64614)
T-Test of difference = 0 (vs not =): T-Value = 2.79  P-Value = 0.009  DF = 28
Both use Pooled StDev = 2.6286
```

Since the question advised that the tests can be selected from the tests in the output, we will calculate the confidence interval assuming equal population variances.

The observed difference is 2.68 or -2.68.
Pooling the sample variances we have the pooled variance as
$[14 * 2.82^2 + 14 * 2.42^2]/28 = 6.9$ and pooled stdev = 2.63

The critical value is 2.05 (t (.05/2) baed on 28 d.f.)

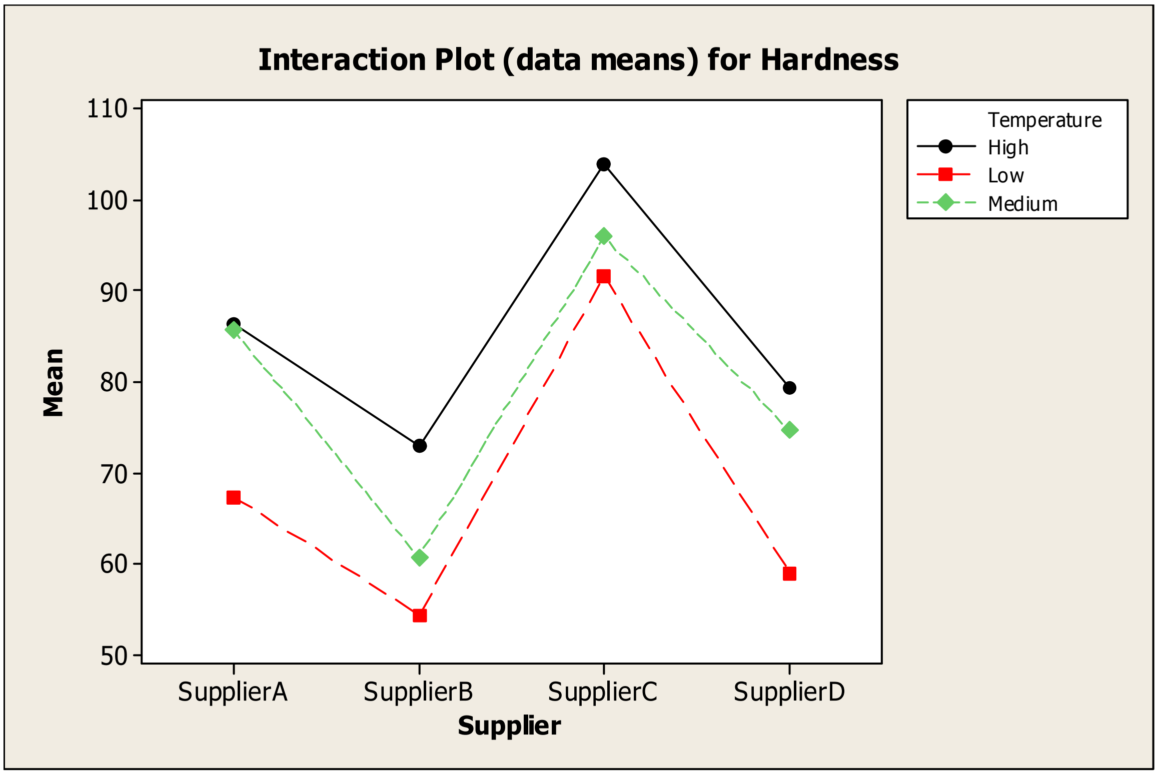
The 95% CI is 2.68 +/- 2.05 * 2.63 sqrt(1/15+1/15)
= 2.68 +/- 2.05 * .96 = 2.68 +/- 1.97 = (.71%, 4.65%)

(d)
Ho: mu1=mu2; Ha: mu1 not = mu2.
Since the 95% CI above excludes the value zero, we reject the null H at the .05 level. Conclude that the 2.68% difference is significantly different than zero.

(e)
The corresponding non-parametric test is the Mann-Whitney test. It leads to the same conclusion as (d) because the p-value of .0279 is less than alpha.

(f) The W statistic is the sums of the ranks for the advisor1 samples. It is not the sum of ranks for positive or negative differences as we do not calculate differences when the samples are independent.

3.
(a)



The Minitab plot indicates that the means are measures of "Hardness". In a manual plot, the vertical axis should be labelled "Hardness" as the means are average hardnesses based on the responses or measures on each windshield.

(b)
Plot shows there is some slight interaction between supplier and temperature in terms of how they affect the hardness.
There appears to be a supplier main effect on the hardness (Supplier C appears to resuklt in the hardest windshields) and a temperature main effect on the hardness (with high temperatures resulting in the hardest windshields).

(c)

There are no problems with:
-the assumption of constant error variance, since the residuals have a constant vertical spread;
-the assumption of normally distributed errors, since there are no residuals beyond +/- 2 standard errors.

(d) If the effect of temperature on hardness depends on the supplier, then this means that temperature and supplier interact.

Ho: no interaction between temperature and supplier;
Ha: some interaction between temperature and supplier.

F = MS int / MSE = 42.27 / 50.47 = .84
F(.05) = 2.51 based on 6 num and 24 den d.f.

Do not reject the null H, conclude no interaction.

(e)
The Tukey criterion T is the margin of error for comparing the observed differences.

T = [q(.05)/ sqrt(2)] * sqrt(MSE * (1/k + 1/k)

Where q is based on num 4 groups and den 24 d.f.
MSE = 50.47, and k = 9 since each supplier mean is based on 9 observations.

T = [3.9 / sqrt (2)] * sqrt (50.47 * (2/9) = 3.9 * sqrt (50.47 / 9) = 9.24

The supplier (column) means are
C - 97.2; B - 62.67; D - 71.0; A - 79.78
The supplier C mean is more than 9.24 higher than each of the other three means; therefore, the supplier C mean is significantly higher than each of the other supplier means.

4.
(a)
The main problem is that there are two outliers--an extreme standardized residual at 3.59 and one at 2.63; making it difficult to assume normally distributed errors. Without these outliers, the residuals have constant spread and therefore it is reasonable to assume constant error variance without these outliers.

(b)
To improve the fit, one should omit at least one if not two observations that do not fit the "mold" of the other observations.

(c)
To go from Model 1 to Model 2, we dropped the 2 outlying cases mentioned in (a) and dropped the "assets" variable as well.

(d)
Model 2 is considered better than Model 1 because:
-the standard error of estimate is lower (which means the adjusted R-square is higher);
-the coefficient of multiple determination (R-square) is higher;

-it is a simpler model;
-the residual plot does not suffer from extreme outliers.

(e)
We could drop 2 more outliers.
Also some of the companies have a market value of 0, possibly because these are privately held companies. These cases make it more difficult to find a relationship between market value and profits. Clearly these companies have profits even though their market value is listed as zero.

(f)
Ho: beta1=0, … , beta4=0; Ha: one betaj is nonzero
F = 106 and F(.05) = 2.61, based on 4,43 d.f.
We reject Ho and conclude that the model is useful

(g)
Ho: beta(MV) = 0; Ha: beta(MV) not = 0.
-t=8.38 and t(.005) = 2.70 based on 43 d.f. (z(.005)=2.575 is also acceptable based on d.f.>30
-we reject Ho, and conclude that the Market Value variable is useful, given the other variables in the model.

(h)
The coefficient of .0177 means that for every $1 million increase in Market Value, the average profits is estimated to increase by $.0177 millions = $17,700, provided that the other variables are kept constant.

(i)
This should be a confidence interval which is 4074 +/- 2.7 (136.1) = 4074 +/- 367 millions of dollars; 2.575 may be used instead of 2.7.

(j)
For the prediction interval, the standard error is sqrt($136.1^2 + 571.5^2$) = 588
The 99% P.I. is 4074 +/- 2.7 * 588 = 4074 +/- 1588 = (2486, 5662) million $

The actual profit for company 13 which has the same characteristics is $6310 millions. This value is not in the 99% P.I. because it is an outlier.