ADM 2304 **Winter 2013**

Solutions and Marking Guide to Assignment 1.

Problem 1. [2 marks]

2 marks: done by MyStatLab

Problem 2. [12 marks]

(a) -Ho: $p = .3462$ or $p \geq .3462$ Ha: $p < .3462$
n = 1500 with p-hat = .31
-z = (.31 - .3462) / sqrt(.3462*.6538/1500) = -2.95
-At the .01 significance level, we reject Ho since -2.95 < -2.32 or -2.33
-Conclude there is a potential decrease in Conservative support.

```
Test of p = 0.3462 vs p < 0.3462
                                    99%
                                  Upper
Sample     X     N  Sample p      Bound  Z-Value  P-Value
1        465  1500  0.310000  0.337780    -2.95    0.002
```

4 marks:
1 for hypotheses (deduct .5 for 2-sided alternative)
1 for z-statistic (deduct .5 if z is based on SE= sqrt(.31.69/1500) instead of SD based on p=.3462.) CI would be based on the SE.*
1 for rejection region z < -2.32,
1 for decision and conclusion

(b)

$$n = z^2 pq/ M^2 = 2.575^2 \; .31*.69 / .01^2$$
$$= 14183$$

3 marks: 1 mark for using p-hat of .31, 1 mark for z=2.575, 1 mark for calculation

(c)
Ho: p=.3962, Ha: p < .3962
n=16 is not sufficient to assume a normal distribution for p-hat since np=16*.3962 = 6.3 is not greater or equal to 10 or since X=3 is not greater or equal to 10.
Use the binomial (n=16, p=.3962) to calculate
the p-value = $P(X \leq 3) = P(0) + P(1) + P(2) + P(3) = 0.069$
(see calculations below)
Since p-value not < .05, we do not reject the null H.
Conclude that there is insufficient evidence to show that Conservative support among students is lower than .3962.

Cumulative Distribution Function

```
Binomial with n = 16 and p = 0.3962

x  P( X <= x )
3    0.0690932
```

Note that Minitab handles this easily if you do not assume a normal distribution for p-hat.

Test and CI for One Proportion

```
Test of p = 0.3962 vs p < 0.3962

                              99%
                            Upper    Exact
Sample   X   N  Sample p    Bound  P-Value
1        3  16  0.187500  0.502936   0.069
```

If normal approximation is used, then z = (.1875 - .3962)/sqrt(.3962*.6538/16)
= -.2087/ .1272 = -1.64 which would not be in the rejection region z < -2.33.
This would not lead to a rejection of the null hypothesis.
Note p-value of P(z < -1.64) = 0.0505

Another possibility is to use the adjustment p-hat = (x+2)/(n+4) = 5/20= .25
and to calculate the CI based on this value:

at most .25 + 2.33 * sqrt(.25*.75/16) = .25 + 2.33 * 0.1083 = .25 + .252 = .502.

Since this CI covers the value .3962, we do not reject the null hypothesis.

Note that, using .25, z is (.25 - .3962)/.1272 = -1.15 with a p-value of P(z < -1.15) = 1251. This is not a good approximation.

4 marks:
1 for hypotheses
1 for some recognition that the normal approximation is not appropriate
1 for showing p-value of .069 by calculating the binomial prob'y or using Minitab.
1 for decision and conclusion

If students calculate the z-statistic manually based on the faulty assumption that the sample proportion is normally distributed, then give 2 marks out of 4
If they use the adjustment (x+2)/(n+4), then give them 3 marks out of 4.

We cannot infer anything about the national level of support since the sample of U of Ottawa students is not a sample from that population.
(1 mark)

Problem 3. [11 marks]

(a) Using Minitab, we can calculate the population mean avginc value as $37828.

```
Variable     N  N*    Mean  SE Mean  StDev  Minimum     Q1  Median     Q3
avginc     849   7   37828      870  25341    11202  25651   30493  40390
```

*1 mark for mean of **37828***

(b) 1 mark for both a boxplot and a histogram

1 mark for explaining that the extreme skewness of the population data means that a sample mean of 31 observations will not be normally distributed.

(c) 2 marks for evidence of 20 CIs.

(d) 1 mark for any graph of data (note the graphs should differ from student to student)

2 marks for any reasonable comment re data distribution— that the large sample probably did (or did not) come from a population that is extremely skewed and how this means the sample mean is not (or is) normally distributed and the CI is not (or is) valid.

1 mark for manual calculation of CI

(e) 1 mark for count

(f) 1 mark for commenting that one should not expect that 95% of all intervals will cover the true value, since the sample mean is not normally distributed.

Problem 4. [10 marks]

(a)

Ho: $\mu3 = \mu5$, Ha: $\mu3 \neq \mu5$

t = (37.85 - 42.73) / sqrt(7.93^2/31 + 7.06^2/44) = -2.74

Reject Ho at 0.01 level since |t| > 2.575 (using z approx.) or 2.66 (based on 59 df)

Conclude there is a difference in average final exam marks.

```
Two-sample T for FX_3 vs FX_5

       N   Mean  StDev  SE Mean
FX_3  31  37.85   7.93      1.4
FX_5  44  42.73   7.06      1.1

Difference = mu (FX_3) - mu (FX_5)
Estimate for difference:  -4.87680
99% CI for difference:  (-9.61080, -0.14279)
T-Test of difference = 0 (vs not =): T-Value = -2.74  P-Value = 0.008  DF = 59
```

4 marks:
1 for hypotheses
1 for t-statistic (must show manual calculation)
1 for decision based on critical value
1 for conclusion

(b) 2-sided CI is:

4.88 ± 2.576 * sqrt(7.93^2/31 + 7.06^2/44) = 4.88 ± 2.576 * 1.78
= 4.88± 4.59 = (0.29, 9.47) if we use the z critical value

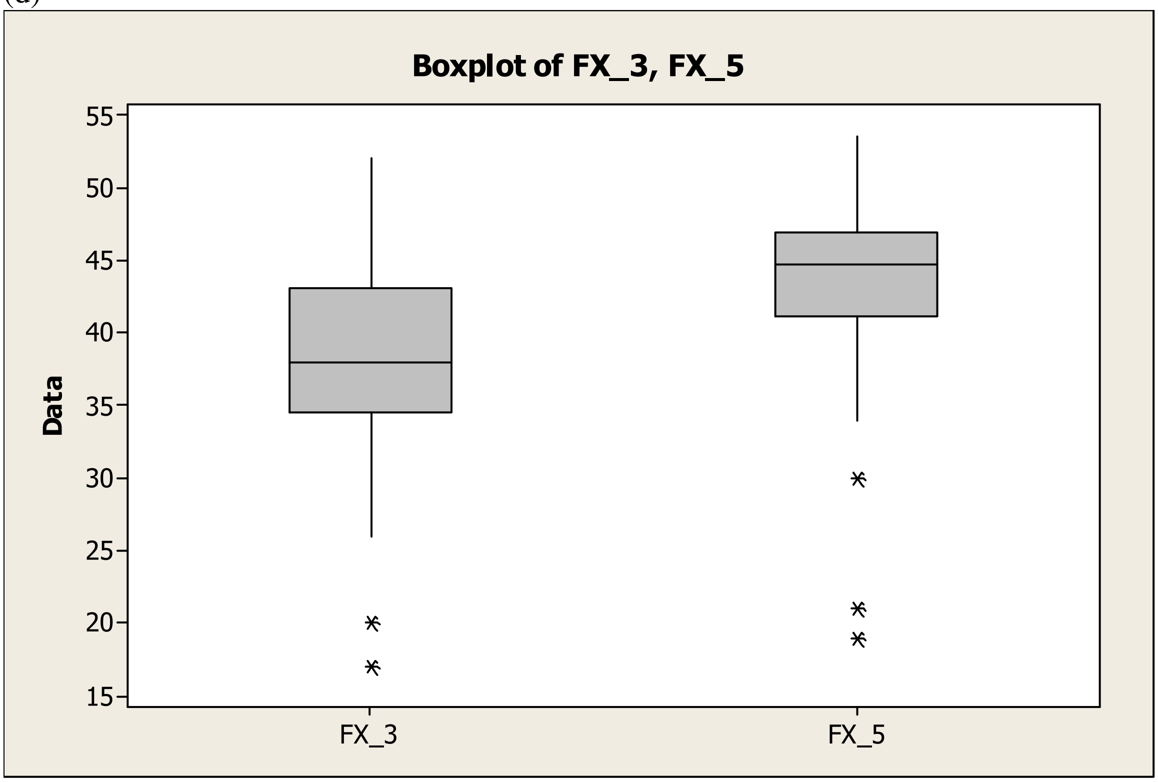
4.88 ± 2.66 * 1.78 = 4.88 ± 4.73 = (0.15, 9.61) or (-9.61, -0.15)

2 marks for showing manual calculation of CI (do not worry about accuracy)

(c) We reject the null H again the CI does not cover zero.

1 mark for noting the same decision and conclusion as before

(d)



Accept any reasonable comment. Either that the data do not come from extremely skewed distributions and therefore the sample means are normally distributed (This is more likely). Or that the data do come from extremely skewed distributions and therefore the sample means are not normally distributed.

1 mark for any graph of the two samples
1 mark for any reasonable comment

(e) We could use the Mann-Whitney test a.k.a. Wilcoxon Rank Sum test.

1 mark for identification of the alternative non-parametric test