**ADM 2304**
**Fall 2012**
**Solutions and Marking Guide to Assignment 1.**

**Problem 1. [ 3 marks ]**

*3 marks: done by MyStatLab*

**Problem 2. [ 10 marks ]**

(a) Ho: p=.306, Ha: p > .306
n = 1836 with p-hat = .36
z = (.36 - .306) / sqrt(.306*.694/1836) = 5.0
At the .05 significance level, we reject Ho since 5.45 > 1.645
Conclude there has been an increase in NDP support.

```
Test of p = 0.306 vs p > 0.306
                                    95%
                                  Lower
Sample    X     N  Sample p       Bound  Z-Value  P-Value
1       661  1836  0.360022    0.341595     5.02    0.000
```

*4 marks:*
*1 for hypotheses (deduct .5 for 2-sided alternative)*
*1 for z-statistic or CI or p-value (deduct .5 if z is based on SE= sqrt(.36*.64/1836) instead of SDbased on p=.306.)*
*1 for decision to reject the null H.*
*1 for conclusion*

(b)

$$n = z^2 pq / M^2 = 1.96^2 \; .36*.64 / .01^2 = 8851$$

*2 marks: 1 mark for using p-hat of .36, 1 mark for answer.*

(c)
Ho: p=.306, Ha: p > .306
n=17 is not sufficient to assume a normal distribution for p-hat since np=17*.306 = 5.2 is not greater or equal to 10 or since X=8 is not greater or equal to 10.
Use the binomial (n=17, p=.306) to calculate
the p-value = $P(X \geq 8) = 1 – P(X<= 7) = 1 – P(0) – P(1) - \ldots - P(7)$ = 1 - 0.885 = .115
(see calculations below)
Since p-value not < .05, we do not reject the null H. To use the critical value approach, we reject the null hypothesis if $X \geq 9$ since $P(X \geq 9) = 1 – P(X \leq 8)$ = 0..05. Since X = 8 is not in the rejection region, we do not reject the null hypothesis.
Conclude that there is insufficient evidence to suggest that NDP support among students exceeds .306

**Cumulative Distribution Function**

```
Binomial with n = 17 and p = 0.306

x  P( X <= x )
7     0.884737
```

| x | binomial prob'y |
| --- | --- |
| 0 | 0.002010 |
| 1 | 0.015063 |
| 2 | 0.053134 |
| 3 | 0.117140 |
| 4 | 0.180774 |
| 5 | 0.207239 |
| 6 | 0.182752 |
| 7 | 0.126625 |
| Total | 0.884737 |

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 = 0.306 vs p > 0.306

                                 95%
                               Lower    Exact
Sample   X   N  Sample p       Bound  P-Value
1        8  17  0.470588  0.260114    0.115
```

If normal approximation is used, then z = (.47 - .306)/sqrt(.306*.694/17)
= ,164/ .112=1.46, which has a p-value of P(z>1.46) = .07. Clearly this is not close to correct value of 0.115.

Another possibility is to use the adjustment p-hat = (x+2)/(n+4) = 10/21= .476
and to calculate the CI based on this value:

at least .476 – 1.645 * sqrt(.476*.524/21) = .476 – 1.645 * 0.109 = .476 - ,179 = .297.

Since this CI covers the value .306, we do not reject the null hypothesis.

Note that, using .476, z is (.476 - .306)/.109 = 1.56 with a p-value of P(z > 1.56) = 0.06.
this is still 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 .115 or rejection region of* $X \geq 9$ *by calculating the binomial prob'y or using Minitab.*
*1 for decision and conclusion (if the answer uses the z of 1.46, then the decision would be to not reject the null H since z not > 1.645, and we cannot conclude that NDP support among students is > .306*

*If students calculate the z-statistic manually based on the assumption that the sample proportion is normally distributed, then give them 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 for the NDP since the sample of U of Ottawa students from Quebec is not a sample from that population.
*(no mark)*

**Problem 3.** [ 10 marks ]

(a) Using Minitab, we can calculate the population mean of the male BMIs as 26.244

```
Variable    N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
bmi_m     639   0  26.244    0.153  3.873   17.100  23.600  25.600  28.500
```

*1 mark for mean of* ***26.244***

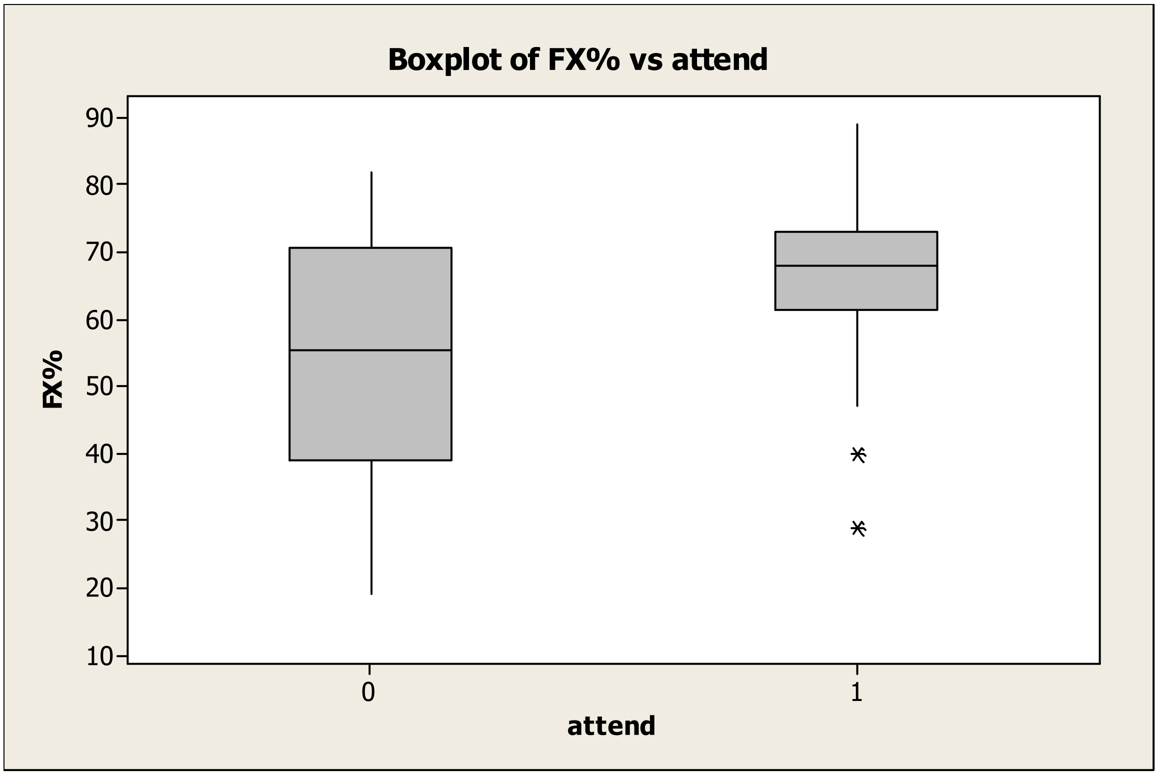
*(b) 3 marks for evidence of 20 CIs.*
*(c) 1 mark for any graph of data*
*2 marks for any reasonable comment re data distribution— that the small sample can or cannot be assumed to come from a normal distribution or that the large sample came from a population that is not extremely skewed and how this means the CI is or is not valid.*
*1 mark for manual calculation of CI*

*(d) 1 mark for count*
*(e) 1 mark for one of the binomial probabilities below.*

| *x* | *prob* |
|---|---|
| *20* | *0.358486* |
| *19* | *0.377354* |
| *18* | *0.188677* |
| *17* | *0.059582* |
| *16* | *0.013328* |
| *15* | *0.002245* |
| *14* | *0.000295* |
| *13* | *0.000031* |

**Problem 4. [ 12 marks ]**

(a)




Since first sample is small, we need to assume the population of marks is normally distributed. This is a reasonable assumption since the boxplot is symmetric and there are no outliers.
Since the second sample is large, we only need to assume the population is not extremely skewed. This is a valid assumption given the boxplot distribution.
The 2-sample t test is valid.

*2 marks: 1 for a relevant comment on the assumptions for each population based on the relevant sample.*

(b)

**Two-Sample T-Test and CI: FX%, attend**

```
Two-sample T for FX%

attend   N  Mean  StDev  SE Mean
0       28  54.4   17.8      3.4
1       48  66.4   10.8      1.6


Difference = mu (0) - mu (1)
Estimate for difference:  -11.9256
99% CI for difference:  (-21.9768, -1.8744)
T-Test of difference = 0 (vs not =): T-Value = -3.22  P-Value = 0.003  DF = 38
```

Ho: mu(0) = mu(1); Ha: mu(0) ≠ mu(1)

t = (66.4 – 54.4)/ sqrt(17.8^2/28 + 10.8^2/48) = 3.24 or -3.24

We reject the null hypothesis at the .01 level of significance since 3.22 > 2.57 or 2.58 which is the approximate critical value based on z approximation. The exact t critical value based on 38 d.f. is 2.71.
Conclude there is a difference in average marks based on attendance.

**Inverse Cumulative Distribution Function**

```
Student's t distribution with 38 DF

P( X <= x )          x
      0.005   -2.71156
```

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

(c) p-value is 2 * Prob (t > 3.22) or 2 * Prob (t < -3.22)
Using the normal approximation, this is 2 * 0.0006 = 0.0012

To get the exact p-value, use Minitab to find
2 * P(t < -3.22) = 2 * 0.0013 = 0.0026 = 0.003.

```
Student's t distribution with 38 DF
    x  P( X <= x )
-3.22    0.0013131
```

*2 marks: 1 for showing p-value is double the tail probability of P(t<-3.24) or P(t>3.24)*
*1 for a reasonable attempt at showing how the tail prob of .0013 can be found.*

(d) 2-sided CI is:
12 ± 2.576 * sqrt(17.8^2/28 + 10.8^2/48) = 12 ± 2.576 * 3.71
= 12 ± 9.56 = (2.44, 21.56) if we use the z critical value
Using the t-distribution, we have:
12 ± 2.71 * 3.71 = 12 ± 10 = ( 2, 22 ) or ( -22, -2 )

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

(e) We reject the null H again since the p-value of 0.003 is less than .01 or because the CI does not cover zero.

*1 mark for noting the same decision and conclusion as before*

(f) We could use the Mann-Whitney test a.k.a. Wilcoxon Rank Sum test.
*1 mark for identification of the alternative non-parametric test*