ADM 2304 -- ASSIGNMENT 1 SOLUTIONS

1. ***[10 marks]***

a. A recent poll of 1000 respondents found the three major parties in a virtual tie, with the Greens lagging behind at 4.4%. Test whether there is evidence at the 5% level of significance to show that the support for the Greens has dropped. Show your manual calculations.

Test and CI for One Proportion

```
Test of p = 0.0678 vs p < 0.0678


Sample   X     N  Sample p  95% Upper Bound  Z-Value  P-Value
1       44  1000  0.044000         0.054668    -2.99    0.001

Using the normal approximation.
```

Ho: p=.0678, Ha: p< .0678
Z = (.044-.0678)/sqrt(.0678*.9322/1000)=-0.0238/0.00795=-2.99
Reject Ho if Z<-1.645
Decide to reject Ho, conclude there is evidence support for Green party has dropped.

4 marks:
1 for hypotheses Ho: p=.0678, Ha: p< .0678 (stated separately from Minitab output)
1 for showing how the z-statistic is calculated
1 for showing p-value (from Minitab or as P(z < -2.99)) or rejection region of < -1.645.
1 for showing decision to reject Ho and conclusion support has dropped (0.5 each)

b. What sample size would be required to obtain a 99% 2-sided confidence interval for the true proportion of Green Party support with a margin of error of 1%?

M = 0.01, z = 2.575 (accept range 2.57 to 2.58) for 99% CI, p-hat =.044, q-hat =.956

Based on the sample results,
n = pq $(z / M)^2$ = .044*.956 * $(2.575/0.01)^2$ = 2789 (accept range from 2778 to 2799)

Based on the 2011 results,
n = .0678*.9322*$(2.575/0.01)^2$ = 4190 (the range here is plus-or-minus 15).

Since we rejected the null hypothesis above, assuming p=0.0678 is not reasonable.

Based on the very conservative but very unlikely value of p=0.5,
n = 16576 (here the range is plus-or-minus 65).

2 marks:
-1 mark for z-value (2.57 to 2.58) and M = 0.01
-1 mark for use of proper formula and calculation

If the solution uses p=q=0.5, then only give maximum of 1 mark.
If the solution uses p=0.0678, then give maximum of 1.5 marks.

c. Suppose that, in a random sample of 17 University of Ottawa students, only 4 indicated a preference for the Conservatives. Test whether this is sufficient evidence to indicate that the level of support for the Conservatives among U of O students is lower than the 39.62% share of the popular vote in 2011. Use the .01 level of significance and explain how you would calculate the p-value for this test. Why does this not allow us to infer anything about the national level of support for the Conservatives?

Test and CI for One Proportion

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


                                99%
                              Upper
Sample  X   N  Sample p       Bound  Z-Value  P-Value
1       4  17  0.235294    0.474627    -1.36    0.087
```

**** NOTE * The normal approximation may be inaccurate for small samples.***

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       4  17  0.235294    0.543388    0.133
```

Cumulative Distribution Function

```
Binomial with n = 17 and p = 0.3962

x  P( X <= x )
4     0.132676
```

Ho: p=.3962, Ha: p<.3962
Use of z-statistic requires np at least 10, but np=17*.3962 < 10, or recognize that X=4 < 10.
Must use binomial probability to calculate p-value = P(X $\leq$ 4) = 0.133
Do not reject Ho since p-value not < 0.05
Conclude insufficient evidence to conclude support lower than 0.3962.

4 marks:

0.5 for hypotheses
1 for recognizing normal approximation not appropriate
1 for calculating p-value of 0.133 (whether using Minitab or binomial calculation)
1 for decision not to reject Ho and conclusion (0.5 for each)
0.5 for comment that a student sample is not representative of the Canadian electorate.

Note that if the solution uses the z-statistic and the normal approximation to find the p-value, then deduct 2 marks for not recognizing the inappropriateness of the normal approximation and not finding the p-value using the binomial probability. Could still get the other 2 marks.

Question 2. [9 marks]

A file in the assignments area on Blackboard Learn called **incomes.mtw** contains data on the median and average incomes for neighbourhoods in Ottawa-Gatineau.

a. Treating the **average** incomes as the population, use Minitab to calculate the population mean. Set aside all population information until part d.

```
Variable    N  N*   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3  Maximum
avginc    268   0  47994      815  13341    22511  39813   46433  55273   135608
```

The mean of the average incomes is **$47994.**
If solution is based on median incomes, then mark is zero.

1 mark

b. Now use Minitab (Calc Menu – Random Data – Sample from Columns) to draw twenty samples of size n = 30. This procedure must be replicated twenty times (note that if you open up the same sampling dialog box each time from the menu, then you only have to replace the last destination column with the next one). Using each sample, use Minitab to calculate a 90% confidence interval estimate for the population mean, assuming you do not know the population standard deviation (this interval estimation can be done in one operation on all twenty columns).

One-Sample T: sample1

```
Variable   N   Mean  StDev  SE Mean       95% CI
sample1   30  48640  12169     2222  (44096, 53184)
```

This has to be done twenty times.

-3 marks for showing results of twenty CIs, whether based on population of median incomes or population of average incomes

c. For the first sample, confirm the Minitab generated interval by calculating the interval manually. Display the sample data graphically and comment on whether the relevant assumption regarding the population distribution is warranted (state clearly the assumption needed to justify the interval estimation).

The boxplot shows that the population income data are not extremely skewed; in general, the sample will be somewhat skewed.

Since the sample is almost large (n = 30), we can safely assume that the sample mean has a sampling distribution which is normally distributed. Accept comment that the sample is still small (not greater than 30) requiring the sample comes from a normal population (this may or may not be reasonable given the sample.

3 marks
-1 mark for graph

-2 mark for commenting sample size not quite large and need to assume normality of population distribution, or for commenting on large sample size and reasonableness of assumption that the distribution is not extremely skewed.

d. Count the number of intervals that contain the true value of the population mean from part a.

1 mark for the count (no need to verify)

e. What proportion of students would you expect to count 18 of their 20 intervals that contain the population mean?

The expected proportion is equal to the probability of counting 18 out of 20 using the binomial probability, based on n=20 and p=0.9
Minitab calculates it as

Probability Density Function

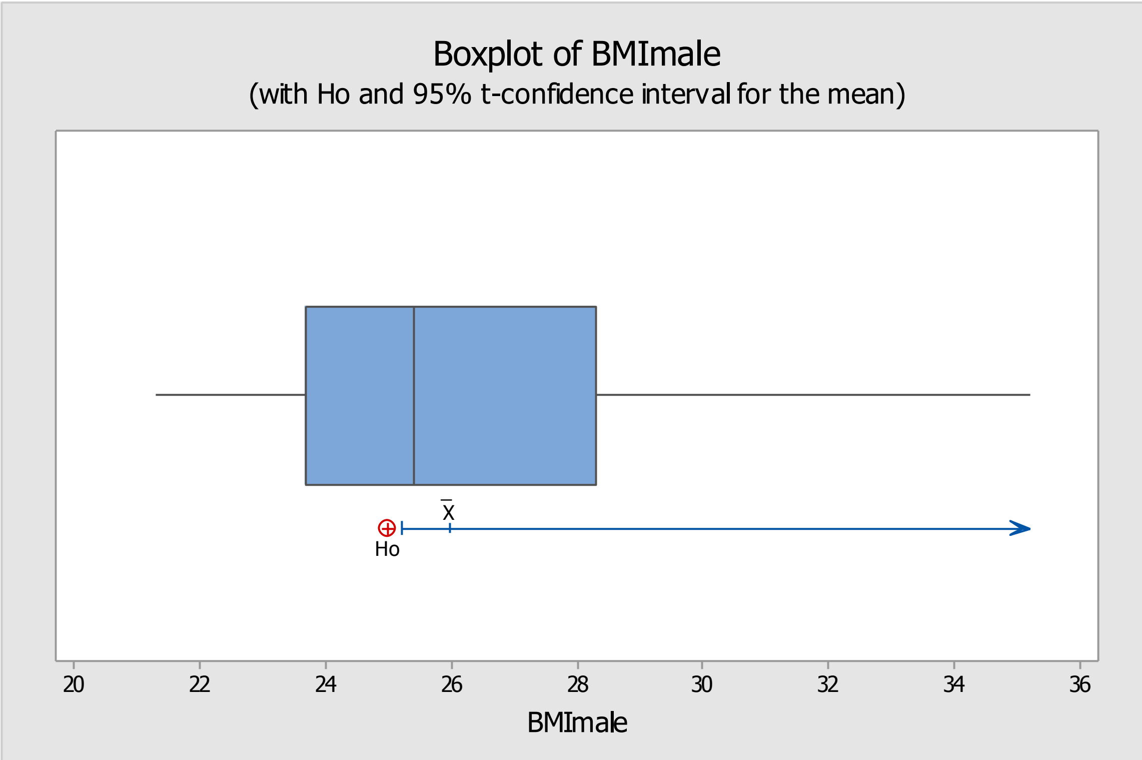
```
Binomial with n = 20 and p = 0.9

 x  P( X = x )
18    0.285180
```

1 mark for the value 0.285, which is lower than one might think.

2. ***[6 marks]***

The file **BMIsamples.mtw** contains two samples of BMI values from the male and female populations. Test at the 0.05 level of significance whether there is sufficient evidence here to show that the average **male** BMI (in the population) exceeds 25. Explain whether your test satisfies the underlying assumptions, with reference to graphical evidence, and show your manual calculations.



The sample is skewed but not extremely so. Therefore we can assume the sample mean is normally distributed since the sample size is > 30. Do not accept the reason that the sample does not seem to come from a normal population since this is unnecessary because this is not a necessary assumption because of the CLT.

One-Sample T: BMImale

```
Test of μ = 25 vs > 25


Variable   N    Mean  StDev  SE Mean  95% Lower Bound     T      P
BMImale   49  25.973  3.183    0.455           25.211  2.14  0.019
```

Ho: $\mu = 25$ Ha: $\mu > 25$

T = (25.973 – 25)/0.455 = 2.14

Reject Ho since p-value < 0.05 or t = 2.14 > 1.645.

Conclude average male BMI exceeds 25.

6 marks

-1 mark for any graph (usually boxplot or histogram or dotplot), the CI underneath is not necessary.

-1 mark for comment on underlying distribution, given large sample size

-1 mark for hypotheses, separately from Minitab

-1 mark for showing calculation of t-statistic

-1 mark for application of rejection region

-1 mark for decision to reject Ho and conclusion that average male BMI is exceeds 25(0.5 mark each)

3. ***[10 marks]***

Two of the columns, **OWmale** and **OWfemale**, in the same dataset code the BMI values as:

0 - if BMI $\leq$ 25.4 (these are considered "not overweight");
1 - if BMI $\geq$ 25.5 (these are considered "overweight").

a. Test whether there is sufficient evidence to show that the proportion of overweight males (proportion of males who are overweight) is different than the proportion of overweight females in the population. Use the critical value approach and the 0.05 level of significance. Perform the test manually after using Minitab to summarize the data (note that the mean coded value in each sample is the sample proportion).
b. Now find the p-value for your sample result and explain how you would find the p-value if you did not have statistical software to perform the test for you.
c. Finally calculate manually the 95% 2-sided confidence interval for the true difference between the proportions of overweight males and overweight females.
d. Explain how the results in parts b and c are consistent with your conclusion in part a.

(a)

Test and CI for Two Proportions: OWfemale, OWmale

```
Event = 1


Variable   X   N  Sample p
OWfemale  10  39  0.256410
OWmale    24  49  0.489796


Difference = p (OWfemale) - p (OWmale)
Estimate for difference:  -0.233386
95% CI for difference:  (-0.429272, -0.0374997)
Test for difference = 0 (vs ≠ 0):  Z = -2.23  P-Value = 0.026

Fisher's exact test: P-Value = 0.030
```

Ho: p1-p2 = 0, Ha: p1 – p2 ≠ 0
p-bar = (10+24)/(39+49)=34/88

z-statistic = (0.233)/sqrt((34*54/(88*88)*(1/39+1/49))
= 0.233 / 0.1045 = ± 2.23

Reject Ho if |z| > 1.96
Decision to reject Ho, conclude number of overweight males is different from number of overweight females.

5 marks:
-1 mark for statement of hypotheses separately from Minitab output
-1 mark for pooling proportions
-1 mark for manual calculation of z
-1 mark for rejection region of absolute value of z greater than 1.96
-1 mark for decision and conclusion that number of overweight males is different from number of overweight females (0.5 mark each)

(b)

p-value is Prob(|Z| > 2.23) = 2 * P(Z > 2.23) = 2 * P(Z < -2.23) = 2 * 0.0129 = 0.026

2 marks:
-1 mark for showing one tail probability
-1 mark for multiplying this by 2.

(c)
0.4898 – 0.2564 ± 1.96 * sqrt(0.2564 *.7436 / 39 + .4898 * .5102 / 49)

= 0.233 ± 1.96 * 0.0999 = 0.233 ± .196 = (0.037, 0.429)

2 marks
-1 mark for use of appropriate formula
-1 mark for correct z-value and correct SE calculation of 0.10.

(d)

Since the p-value < 0.05 and the 95% CI does not cover zero, we reject the null hypothesis and conclude that the proportions of overweight males and of overweight females are different.

1 mark:
0.5 for mentioning each decision rule