ADM 2304 -- ASSIGNMENT 1 SOLUTIONS

Note that Minitab output does not substitute for the different elements of a hypothesis test—it should be viewed strictly as a calculator. The hypotheses, decision and conclusion must be written or typed and shown separately from any computer output for marks. No marks are given for a solution that only provides Minitab output.

Problem 1.

[10 marks]

(a)

Test and CI for One Proportion

```
Test of p = 0.3189 vs p < 0.3189

Sample    X     N  Sample p  95% Upper Bound  Z-Value  P-Value
1       439  1500  0.292667         0.311990    -2.18    0.015

Test of p = 0.3189 vs p < 0.3189

Sample    X     N  Sample p  95% Upper Bound  Z-Value  P-Value
1       440  1500  0.293333         0.312669    -2.12    0.017

Using the normal approximation.
```

4 marks:
1 for hypotheses (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 < -2.12)) or rejection region of < -1.645
1 for showing decision to reject Ho and conclusion support has dropped (0.5 each)

(b)

M = 0.01, z = 1.96 for 95% CI, p-hat =.293, q-hat =.707

$n = pq\,(z / M)^2 = .293*.707 * (1.96/0.01)^2 = 7958$ (accept range from 7478 to 8287)

2 marks:
-1 mark for z-value and M value (accept z=1.9 to 2.0))
-1 mark for use of proper formula and calculation, if using p=.293, but 0 mark if using p=0.50)

(c)

Test and CI for One Proportion

```
The normal approximation may be inaccurate for small samples.

Test of p = 0.3189 vs p < 0.3189
                                               Exact
Sample  X   N  Sample p  95% Upper Bound  P-Value
1       3  23  0.130435         0.303638    0.036

Test of p = 0.3189 vs p < 0.3189

Sample  X   N  Sample p  95% Upper Bound  Z-Value  P-Value
1       3  23  0.130435         0.245943    -1.94    0.026

Using the normal approximation.
```

The second test uses the normal approximation which is inappropriate, and the first is the exact calculation using the binomial distribution.

4 marks:

1 for hypotheses of Ho: p=.3189, Ha: p<.3189
1 for recognizing normal approximation not appropriate since we only observe 3 out of 23 or np = 23.3189 < 10.*
1 for calculating p-value of 0.036 (whether using Minitab or binomial calculation)
1 for ***decision to reject*** *Ho and conclusion that there is sufficient evidence to conclude support lower than .3189.*

Note that if the solution uses the z-statistic and the normal approximation to find the p-value, then give the 2 marks only for hypotheses and decision/conclusion.

Some students may make the adjustment from section 11.6, using p-hat = (3 + 2)/(23+4)=0.185

The z-statistic would be (0.185- .3189)/sqrt(.3189*.6811/27) = -.1339/.0897= -1.49, with a p-value of P(z < -1.49) = 0.068.

A 95% CI would be an upper bound of 0.185 + 1.645 * sqrt(.185* .815/ 27) or 0.185 + 1.645*0.075 or 0.185 + 0.123 = 0.308, which is closer to the binomial calculation above.

This solution can get one more mark for a total of 3 out of 4, if the confidence interval estimate is given without a p-value.

Problem 2.

[10 marks]

Test and CI for Two Proportions: OW_male, OW-female

```
Event = 1


Variable    X   N  Sample p
OW_male    21  31  0.677419
OW-female  11  35  0.314286


Difference = p (OW_male) - p (OW-female)
Estimate for difference:  0.363134
95% CI for difference:  (0.137895, 0.588372)
Test for difference = 0 (vs not = 0):  Z = 2.95  P-Value = 0.003

Ho: p1-p2=0 vs Ha: p1-p2 ≠ 0
p-bar = (32)/66 = 0.485

z-statistic = (0.363)/sqrt((0.485*0.515)(1/31+1/35))
                                 = 0.363 / 0.123 = ± 2.95
Reject the null hypothesis since 2.95 is > 1.96
Conclude there is a difference between the proportion of overweight males vs females.


-Decision to reject null H
-There is sufficient evidence to show a difference.
```

5 marks:
-1 mark for statement of hypotheses, separately from Minitab output
-1 mark for pooling proportions
-1 mark for z-statistic (must show manual calculation)
-1 mark for decision to reject null H
-1 mark for conclusion that there is evidence to show a difference.

Some may do this test as a 2-sample test with similar calculated values:

```
Two-sample T for OW_male vs OW-female

            N   Mean  StDev  SE Mean
OW_male    31  0.677  0.475    0.085
OW-female  35  0.314  0.471    0.080


Difference = mu (OW_male) - mu (OW-female)
Estimate for difference:  0.363134
95% CI for difference:  (0.130094, 0.596174)
T-Test of difference = 0 (vs not =): T-Value = 3.11  P-Value = 0.003  DF = 64
Both use Pooled StDev = 0.4730
```

Using the pooled stdev of 0.473, the standard error for the t-statistic is
0.473 sqrt(1/31 +1/35) = 0.117 with t = 0.363/0.117 =3.10

```
Two-sample T for OW_male vs OW-female

            N   Mean  StDev  SE Mean
OW_male    31  0.677  0.475    0.085
OW-female  35  0.314  0.471    0.080

Difference = mu (OW_male) - mu (OW-female)
Estimate for difference:  0.363134
95% CI for difference:  (0.129822, 0.596446)
T-Test of difference = 0 (vs not =): T-Value = 3.11  P-Value = 0.003  DF = 62
```

Here the standard error is sqrt(0.475^2/31 + 0.471^2/35) = 0.117
and t = 0.363/0.117 = ± 3.10

If the 2-sample t-test is done, then the hypotheses
Ho: μ1 - μ2 = 0 vs Ha: μ1 - μ2 ≠ 0 does not get a mark,
and the manual calculation of t does not get a mark.
However, if they get the decision and conclusion consistent with their t-statistic, then they can get 2 marks for these.

(For the 2-sample t-test, the exact t-critical values are closer to ± 1.99 whereas the normal approximation has ± 1.96 but both are acceptable).

(b)

For the 2-proportions test, the p-value is
Prob(|Z| > 2.95) = 2 * P(Z < -2.95) = 2 * P(Z > 2.95) = 2*0.0016 = 0.003

For the 2-sample test, however, the p-value is
2 P(t > 3.11) = 2 * 0.0014 = 0.0028 = 0.003

Cumulative Distribution Function

```
Student's t distribution with 64 DF

    x  P( X <= x )
-3.11    0.0013964
```

1 mark for any calculation close to any of the above.

(c)

For the 2-proportions test, the CI is
0.363 ± 1.96 * sqrt(0.363 *.637/ 31 + .314 * .686 / 35)
= 0.363 ± 1.96 * 0.117 = 0.363 ± 0.229= (0.134, 0.592)

For the 2-sample t-test, the CI can be:

0.363 ± 1.99 * 0.117 = 0.363 ± 0.233 or 0.363 ± 1.96 * 0.117 = 0.363 ± 0.229

These CIs do not cover zero.

2 marks for any CI like the above.

(d) Depending on the p-value calculated, and the CI calculated, the decisions should be consistent with (a).

1 mark for comparison of p-value to alpha;
1 mark for check of whether the CI covers zero.

Problem 3.

[5 marks]

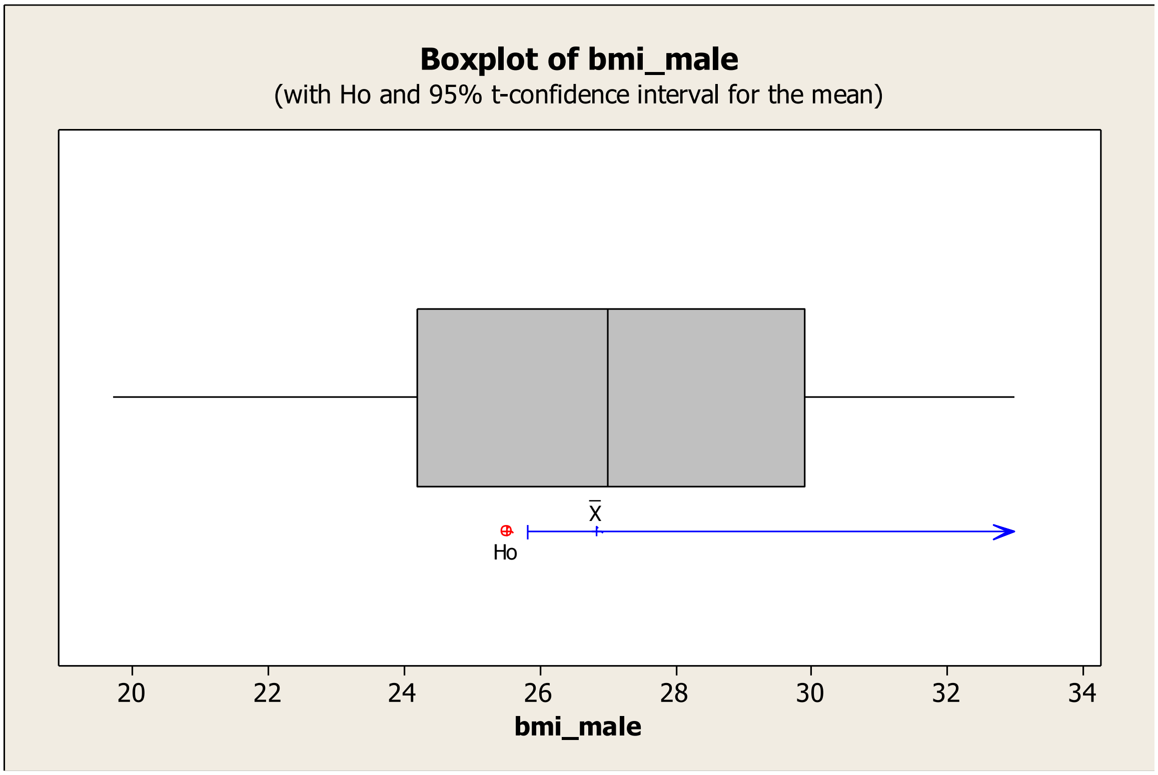
One-Sample T: bmi_male

Test of mu = 25.5 vs > 25.5

Variable	N	Mean	StDev	SE Mean	95% Lower Bound	T	P
bmi_male	31	26.8355	3.3080	0.5941	25.8271	2.25	0.016

From the boxplot below, the sample is relatively symmetric and there are no outliers. With the sample size of 31, we can safely assume the sample mean is normally distributed since a large sample only requires that the population is not extremely skewed (in fact, it is safe to assume the population is normally distributed which guarantees the sample mean is normally distributed). However, the normality assumption is not a minimal requirement.

5 marks
-1 mark for comment on population distribution of data or sampling distribution of sample mean
-1 mark for hypotheses, separately from Minitab
-1 mark for showing calculation of t-statistic
-1 mark for decision to reject null H
-1 mark for conclusion that average male BMI exceeds 25



Problem 4.

[10 marks]

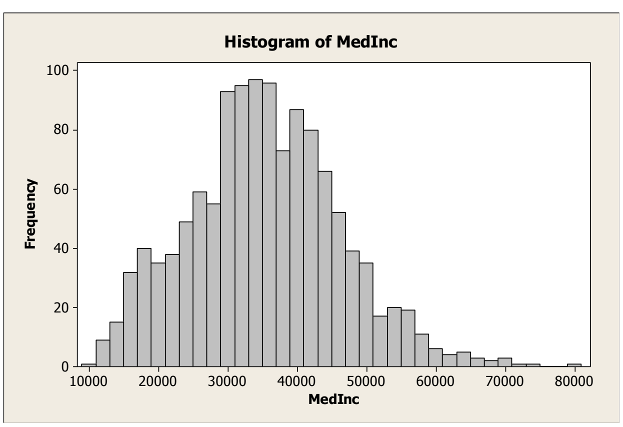
(a)

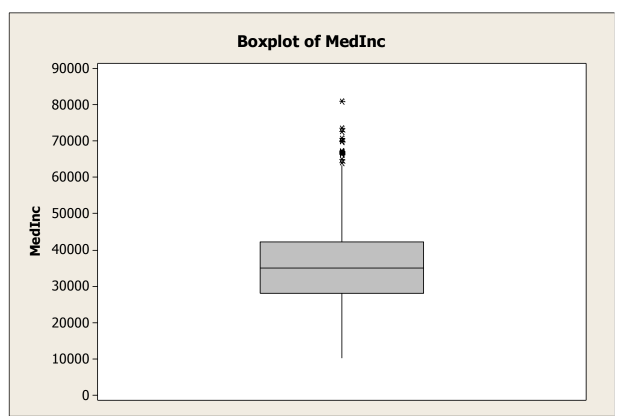
Descriptive Statistics: MedInc

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
MedInc	1239	0	35392	312	10974	10073	28128	35044	42274

1 mark for the population mean of $35392.

(b)





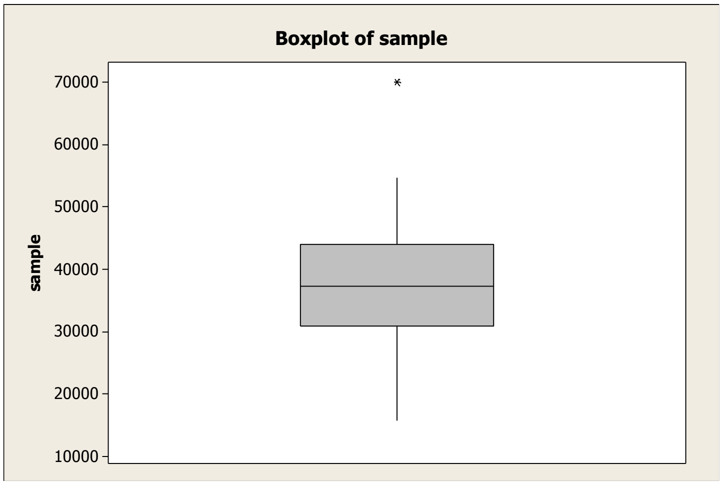
The population data is positively skewed. However, since it is not extremely skewed, it is reasonable to assume that samples of size 40 would result in means that are normally distributed.

1 mark for the two graphs and 1 mark for the explanation

(c)

3 marks for showing the twenty CIs.

(d)



This boxplot of a sample shows that the population income data can be assumed to be ***not extremely skewed***; but in general, the sample will be somewhat skewed.

Since the sample is considered large (n > 30), we can safely assume that the sample mean has a sampling distribution which is normally distributed.

One-Sample T: sample

```
Variable   N     Mean    StDev  SE Mean        95% CI
sample    40  37484.0  10656.9   1685.0  (34075.7, 40892.3)
```

3 marks
-1 mark for any graph
-1 mark for manual calculations
-1 mark for commenting on large sample size and reasonableness of assumption that the distribution is not extremely skewed (no marks for comment that requires the population to be approximately normal since sample is large).

(e)

1 mark for the count