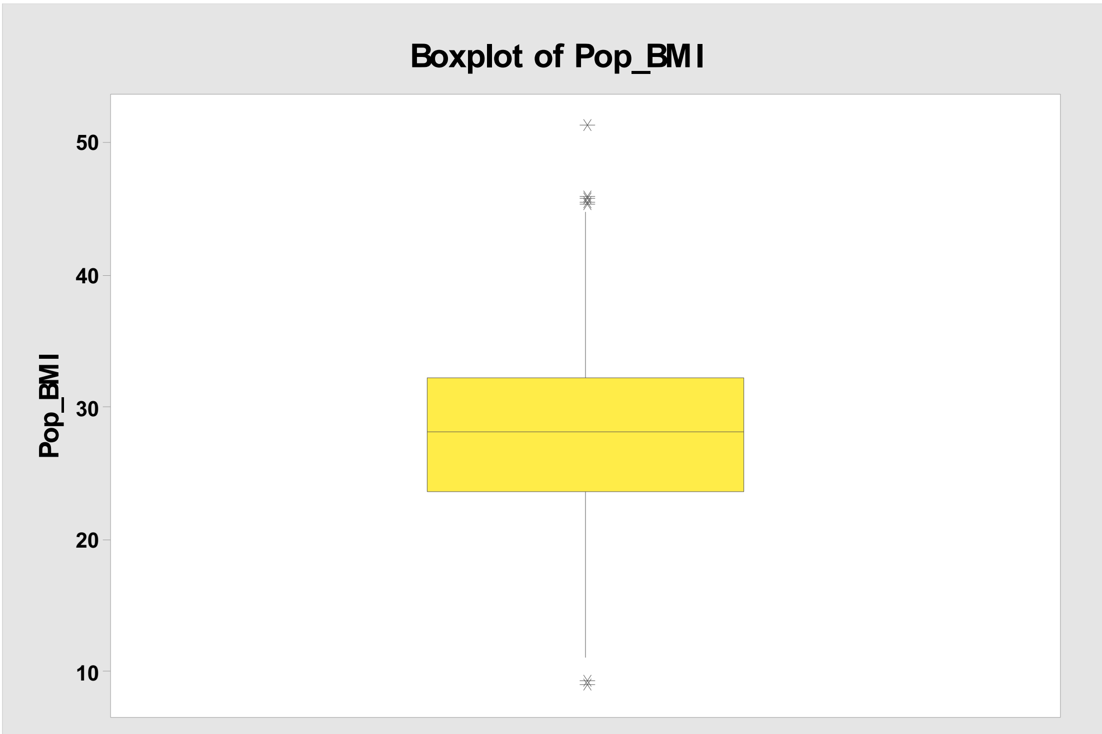
ADM2304X – ASSIGNMENT 1

Professor Dr. Suren Phansalker

University of Ottawa
ADM2304 X

Question 1

Part a)



Descriptive Statistics: Pop_BMI (Copy-pasted from Minitab)

```
Variable      N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
Maximum
Pop_BMI    3000   0  27.951    0.111  6.067    9.070  23.703  28.170  32.168
51.150
```

We can see here that the population data is quite normal and has outliers on both sides.

Part b)

Descriptive Statistics: Pop_Coded

```
Variable      N  N*     Mean  SE Mean    StDev  Minimum       Q1   Median
Q3  Maximum
Pop_Coded  3000   0  0.25933  0.00800  0.43834  0.00000  0.00000  0.00000
1.00000  1.00000
```

Tally for Discrete Variables: Pop_Coded

```
Pop_Coded  Count
        0   2222
        1    778
       N=   3000
Population proportion of highly obese males
```

$$p = \frac{x}{n} = \frac{778}{3000} = 0.25933$$

Part c)

Descriptive Statistics: Sample 1

```
Variable     N  N*     Mean   SE Mean    StDev   Minimum       Q1   Median       Q3
Maximum
Sample1    100   0  0.2700    0.0446   0.4462    0.0000   0.0000   0.0000   1.0000
1.0000
```

Tally for Discrete Variables: Sample1

```
Sample1  Count
      0     73
      1     27
     N=    100
```

The sample proportion can be calculated as follows:

$$\hat{p} = \frac{x}{n} = \frac{27}{100} = 0.27$$

$$SE(\hat{p}) = \sqrt{\frac{0.27(1-0.27)}{100}} = 0.0444$$

The Confidence interval can be calculated as follows:

$$\hat{p} \pm Z^{*} SE(\hat{p})$$

$$0.27 \pm 1.96(0.0444)$$

$$0.27 \pm 0.0870$$

$$(0.1830, 0.3570)$$

Test and CI for One Proportion: Sample 1

```
Event = 1


Variable    X    N  Sample p         95% CI
Sample1    27  100  0.270000  (0.182986, 0.357014)

Using the normal approximation.
```

From the calculation above, our confidence interval is (0.1830, 0.357014) and I double checked in Minitab indeed I get the same number.

Part d)

Test and CI for One Proportion: Sample 1, Sample 2, Sample 3, Sample 4, Sample 5, Sample 6, Sample 20.

```
Event = 1


Variable    X    N  Sample p         95% CI
Sample1    27  100  0.270000  (0.182986, 0.357014)
Sample2    28  100  0.280000  (0.191998, 0.368002)
Sample3    29  100  0.290000  (0.201064, 0.378936)
Sample4    23  100  0.230000  (0.147518, 0.312482)
Sample5    23  100  0.230000  (0.147518, 0.312482)
Sample6    27  100  0.270000  (0.182986, 0.357014)
Sample7    31  100  0.310000  (0.219353, 0.400647)
Sample8    23  100  0.230000  (0.147518, 0.312482)
Sample9    27  100  0.270000  (0.182986, 0.357014)
Sample10   21  100  0.210000  (0.130169, 0.289831)
Sample11   30  100  0.300000  (0.210183, 0.389817)
Sample12   26  100  0.260000  (0.174029, 0.345971)
Sample13   26  100  0.260000  (0.174029, 0.345971)
Sample14   28  100  0.280000  (0.191998, 0.368002)
Sample15   28  100  0.280000  (0.191998, 0.368002)
Sample16   31  100  0.310000  (0.219353, 0.400647)
Sample17   26  100  0.260000  (0.174029, 0.345971)
Sample18   26  100  0.260000  (0.174029, 0.345971)
Sample19   18  100  0.180000  (0.104701, 0.255299)
Sample20   25  100  0.250000  (0.165131, 0.334869)

Using the normal approximation.
```

When I randomly sampled 19 times in Minitab, I found that only Sample 19 did not contain the population proportion 0.25933.

Part e)

We should expect that 95% of the 20 random samples would somewhat contain the population proportion which is 19 of them.

From our 20 random samples, there is 19 of them that contained the population proportion. This is indeed consistent with what we expected

Part f) To calculate the probability of the exact number of Confidence Intervals we obtained among the 20, which we found in part e, we are going to use the Binomial distribution

The probability of 19 confidence intervals among 20 contain population proportion

$P(X=x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$

$P(X=19) = \frac{20!}{(20-19)!19!} 0.95^{19} 0.05^1 = 0.3774$

Question 2

Part a)

Null Hypothesis Ho: p=0.175

Alternative Hypothesis Ha: p$\neq$ 0.175

$SE(\hat{p}) = \sqrt{\frac{0.175(1-0.175)}{100}} = 0.0380; Z_{calc} = \frac{0.27-0.175}{0.0380} = 2.50 \; ; Z_{crit} = 1.96$

We Reject Ho

if $|Z_{calc}| > 1.96$

We Reject Ho since 2.50 > 1.96

- "H_obese" is not 17.5%.
- P-value = P (Z >$Z_{calc}) * 2 = (1 - 0.9938) * 2 = 0.0124$

Part b)

$\hat{p} = \frac{x}{n} = \frac{27}{100} = 0.27 \; ; SE(\hat{p}) = \sqrt{\frac{0.27(1-0.27)}{100}} = 0.0444$

Let's calculate 95% Confidence interval:

$$\hat{p} \pm Z^{*}SE(\hat{p})$$

$$0.27 \pm 1.96(0.0444)$$

$$0.27 \pm 0.0870$$

$$(0.1830, 0.3570)$$

We are 95% confident that the true proportion is between 0.1830 and 0.3570

The Hypothesis value I found is 0.175 is not covered by the confidence interval, we reject Ho.

It is consistent with the conclusion in part a.

Part c) Below I am confirming my results in part ''a''

Test and CI for One Proportion: Sample1

```
Test of p = 0.175 vs p ≠ 0.175

Event = 1


Variable   X    N Sample p         95% CI         Z-Value P-Value
Sample1    27  100  0.270000  (0.182986, 0.357014)     2.50
0.012

Using the normal approximation.
```

- The Z-value found my Minitab output is 2.50
- The P-value is 0.012 which is supported by the calculation we did above.
- Since P-value is 0.012, that's less than our significance level
- Therefore, we reject Ho.

Part d)

$n = \left(\frac{1.96}{0.04}\right)^2 0.27 * 073 = 473.2371$, This is approximately 474.

We need a 474-sample size to obtain a margin of error of ± 4%.

Part e)

Ho: p=0.36

Ha: p<0.36

$SE(\hat{p}) = \sqrt{\frac{0.36*0.64}{100}} = 0.0480$; $Z_{calc} = \frac{0.27-0.36}{0.0480} = -1.875$; $Z_{crit} = -1.645$

Reject Ho if $Z_{calc} < -1.645$

Reject Ho since -1.875 < - 1.645

The hypothesis that population proportion, the "H_obese" is less than 36%.

Part f) $\hat{p} = \frac{x}{n} = \frac{27}{100} = 0.27$; $SE(\hat{p}) = \sqrt{\frac{0.27(1-0.27)}{100}} = 0.0444$

Upper bound

$$\hat{p} + Z^{*}SE(\hat{p})$$

$$0.27 + 1.645(0.0444) = 0.3430$$

Confidence interval

$$(0, 0.3430)$$

The hypothesized value 0.36 is not covered by the confidence interval, so we reject Ho which is consistent with the conclusion in part e.

Question 3

np =50* 0.03 =1.5 <10

So, we cannot use the normal approximation since np is less than 10

Ho: p=0.03

Ha: p>0.03

Exact p-value =P($x \geq 4$)= 1- P(x=0) –P(x=1)-P(x=2)-P(x=3) = 0.062761

```
Test of p = 0.03 vs p > 0.03


                                                Exact
Sample  X   N  Sample p  95% Lower Bound  P-Value
1       4  50  0.080000         0.027788    0.063
```

Question 4

a. Ho: $p_{females}$

Ha: $p_{females} \neq p_{males}$

$$\widehat{p_{females}} = \frac{48}{150} = 0.32$$

$$\widehat{p_{males}} = \frac{25}{125} = 0.2$$

$$\widehat{p_{pool}} = \frac{48 + 25}{150 + 125} = 0.2655$$

$$SE\left(\widehat{p_{females}} - \widehat{p_{males}}\right) = \sqrt{0.2655 * (1 - 0.2655)(\frac{1}{150} + \frac{1}{125})} = 0.0535$$

$$Z_{calc} = \frac{0.32 - 0.20}{0.0535} = 2.2430$$

P-value = P(Z >Z_{calc}) $* 2 = (1 - 0.9875) * 2 = 0.0125 * 2 = 0.0250$

i. Reject Ho since 0.025 < 0.05.

ii. Do not reject Ho since 0.025 > than 0.05.

iii. Based on the observations in ''I'' and ''ii'' we can say that the difference between the proportion of obese male and female are significant at level of 5% and not significant at LS 1%.

Part b) Confidence interval

$$\widehat{p_{females}} = \frac{48}{150} = 0.32$$

$$\widehat{p_{males}} = \frac{25}{125} = 0.2$$

$$SE\left(\widehat{p_{females}} - \widehat{p_{males}}\right) = \sqrt{\frac{0.32 * 0.68}{150} + \frac{0.2 * 0.8}{125}} = 0.0523$$

Confidence interval

$$0.32 - 0.2 \pm 1.96\ (0.0523)$$
$$(0.0176, 0.2224)$$

Test and CI for Two Proportions

```
Sample   X    N  Sample p
1       48  150  0.320000
2       25  125  0.200000


Difference = p (1) - p (2)
Estimate for difference:  0.12
95% CI for difference:  (0.0175806, 0.222419)
Test for difference = 0 (vs ≠ 0):  Z = 2.24  P-Value = 0.025

Fisher's exact test: P-Value = 0.028
```

Part c)

Ho: $p_{females} = p_m$

Ha: $p_{females} > p_{males}$

$$\widehat{p_{females}} = \frac{48}{150} = 0.32$$

$$\widehat{p_{males}} = \frac{25}{125} = 0.2$$

$$\widehat{p_{pooled}} = \frac{48+25}{150+125} = 0.2655$$

$$SE\left(\widehat{p_{females}} - \widehat{p_{males}}\right) = \sqrt{0.2655 * (1 - 0.2655)(\frac{1}{150} + \frac{1}{125})} = 0.0535$$

$$Z_{calc} = \frac{0.32 - 0.20}{0.0535} = 2.2430$$

$$Z_{crit} = 1.645$$

Reject Ho if $Z_{calc} > 1.645$

Reject Ho since 2.2430 > 1.645

The proportion of obese females is greater than obese males.

Part d)

Upper bound

$$\widehat{p_{females}} = \frac{48}{150} = 0.32$$

$$\widehat{p_{males}} = \frac{25}{125} = 0.2$$

$$SE\left(\widehat{p_{females}} - \widehat{p_{males}}\right) = \sqrt{\frac{0.32 * 0.68}{150} + \frac{0.2 * 0.8}{125}} = 0.0523$$

Lower bound

$$0.32 - 0.2 - 1.645\ (0.0523) = 0.0340$$

Confidence interval
(0.0340, 1)

Since zero is not covered by the confidence interval, we reject Ho.
This is consistent with the conclusion in Part c)