ASSIGNMENT 1
JULIA ZANON
ADM 2304 B
6464514

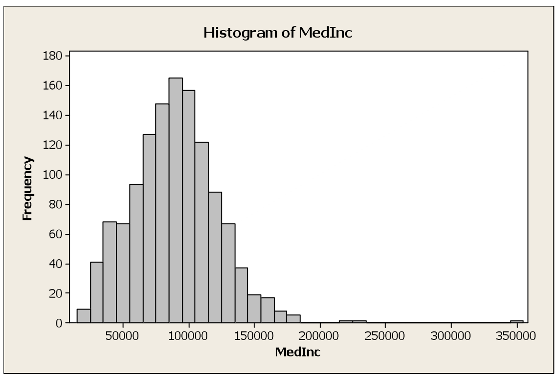
Question 1

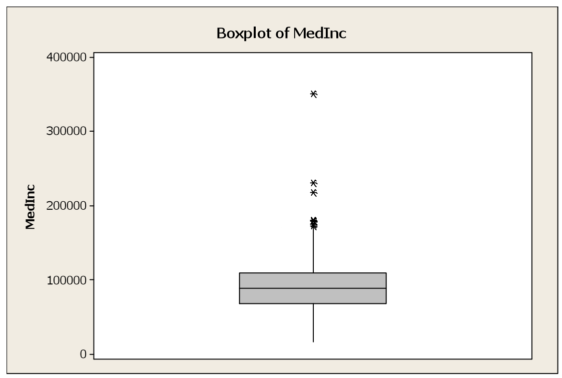
a)

Descriptive Statistics: MedInc

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
MedInc	1241	0	**89368**	918	32337	15947	67725	89103	109436

Treating the median family incomes as the population, the population mean is 89368.





b) **One-Sample T: C4, C5, C6, C7, C8, C9, C10, C11, ...**

Variable	N	Mean	StDev	SE Mean	95% CI
C4	31	91071	31811	5713	(79402, 102739)
C5	31	84535	26936	4838	(74654, 94415)
C6	31	89754	26768	4808	(79935, 99572)
C7	31	95831	26436	4748	(86134, 105528)
C8	31	80415	29035	5215	(69765, 91065)
C9	31	91486	28986	5206	(80854, 102119)
C10	31	81167	26272	4719	(71530, 90803)
C11	31	84551	29019	5212	(73907, 95196)
C12	31	81502	35856	6440	(68350, 94654)
C13	31	84261	26236	4712	(74637, 93884)
C14	31	93014	30605	5497	(81788, 104240)
C15	31	90737	26969	4844	(80845, 100629)
C16	31	92972	33754	6062	(80591, 105354)
C17	31	94515	54454	9780	(74541, 114489)
C18	31	88981	35529	6381	(75949, 102013)
C19	31	87860	32870	5904	(75804, 99917)
C20	31	92669	30170	5419	(81602, 103735)
C21	31	84638	26710	4797	(74841, 94436)
C22	31	90925	37277	6695	(77252, 104599)
C23	31	89691	32937	5916	(77610, 101772)

c) **Descriptive Statistics: C4**

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum

```
C4          31    0  91071      5713  31811     25460  70378   88174  117037   156830
```

95% Confidence Interval

$$\overline{X} \pm T_{\alpha/2}\frac{s}{\sqrt{n}}$$

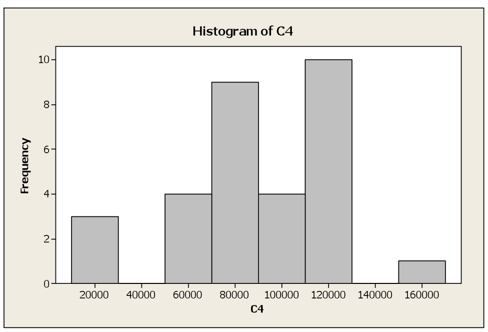
$$= 91071 \pm 1.645 \times \frac{31811}{\sqrt{31}}$$

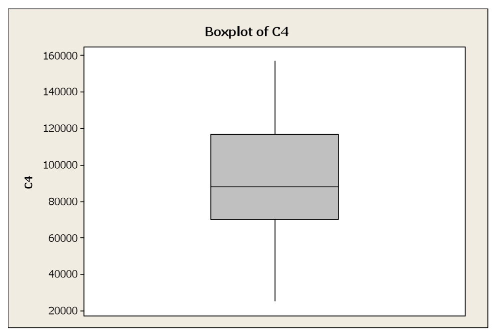
$$= = 91071 \pm 1.645 \times 5713.42$$

$$= 91071 \pm 9398.58$$

=[81 672.42, 100 469.58]

The large sample probably came from a population that is extremely skewed and this means the sample is not normally distributed and the CI is not valid. (But it is close)





d) Twenty intervals out of my twenty samples include the mean of the population! Therefore, the mean is within 100% of the intervals.

<u>Question 2</u>

a)

p = 0.3962-0.05
= 0.3462

$\hat{p}$ = 0.304

n = 1000

q = 1-p
= 0.6538

H_o : p=0.3462

H_a : p<0.3462

$$z_{stat} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.304 - 0.3462}{\sqrt{\frac{(0.3462)(0.6538)}{1000}}}$$

$$= \frac{-0.0422}{0.01504} = -2.81$$

$P(z \rightarrow -2.81) = 0.0025$

Since $0.0025<0.01$, $p<\alpha$, therefore we reject H_O on the 0.01 level of significance.
Conclude there is potential decrease in Conservative support.

b)

$$n = pq\left[\frac{z_{\alpha/2}}{ME}\right]^2$$

$$= (0.304)(0.696)\left[\frac{2.575}{0.01}\right]^2$$

$$= 14030$$

c)

p = 0.3962
p̂ = 3/17
= 0.1765
n =17
q =1-p
=0.6038

H_O: p=0.3962
H_a: p<0.3962

n=17 is not sufficient to assume a normal distribution for p-hat since np=17*0.3962=6.7 is not greater or equal to 10 or since X=3 is not greater or equal to 10.

We proceed with the binomial test in order to calculate the P value.

$$P(X \le 3) = P(0) + P(1) + P(2) + P(3)$$

$$=$$

$$SD(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.3962 \times 0.6038}{17}} = 0.1186$$

$$z_0 = \frac{\hat{p} - p}{SD(p)} = \frac{0.1765 - 0.3962}{0.1186} = -1.8524 \approx -1.85$$

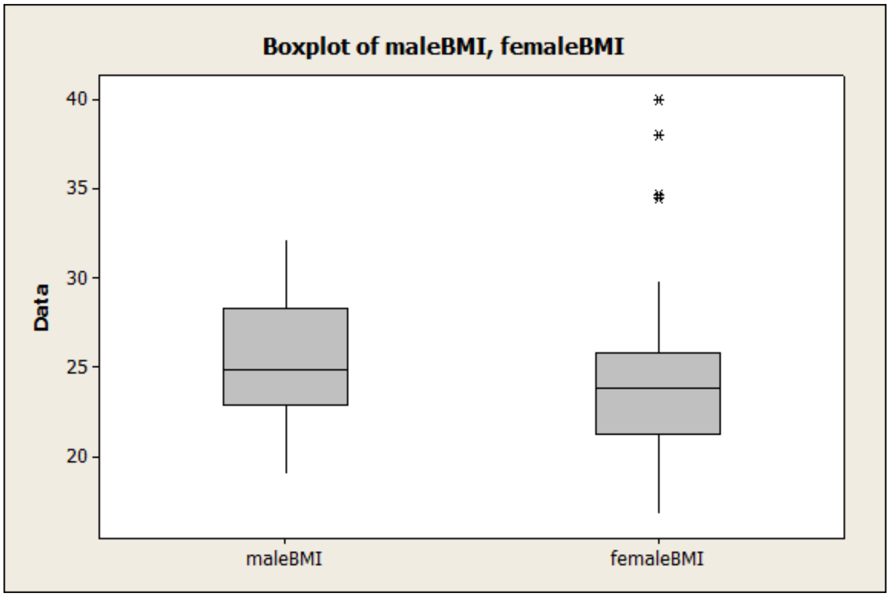
$P(z \rightarrow -1.85) = 0.0322$

Since $0.0322<0.05$, $p<\alpha$, therefore reject H_O on 0.05 level of significance.

*When proportion p=0.3962, there is a 3.22% chance that the sample proportion p̂=0.304.

<u>Question 3</u>

a) Yes, the 2-sample T test is valid after examining the distributions because the independent samples are both approximately normally distributed, according to their boxplots below. The data do not come from extremely skewed distributions and therefore the sample means are approximately normally distributed.



b)

Two-Sample T-Test and CI: maleBMI, femaleBMI

```
Two-sample T for maleBMI vs femaleBMI

             N   Mean  StDev  SE Mean
maleBMI     39  25.75   3.48     0.56
femaleBMI   34  24.71   5.42     0.93


Difference = mu (maleBMI) - mu (femaleBMI)
Estimate for difference:  1.04
95% CI for difference:  (-1.13, 3.22)
T-Test of difference = 0 (vs not =): T-Value = 0.96  P-Value = 0.340  DF = 54
```

$Ho: \mu_1 - \mu_2 = 0$
$Ha: \mu_1 - \mu_2 \neq 0$
$difference : 1.04$

2 Sample T-test: (manually)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{25.75 - 24.71}{\sqrt{\frac{3.48^2}{39} + \frac{5.42^2}{34}}} = 0.96$$

Do not reject Ho at 0.05 significance level since t < 2.000 (based on 54 df)
Conclude there is no difference in average male BMI from the average female BMI.

c) P(at t(df=54)=0.96)=0.340
Since 0.340>0.05, then p> α, fail to reject Ho on a 0.05 level of significance

d) Confidence Interval 95% : (manually calculated)

$$(\bar{X}_1 - \bar{X}_2) \pm T \times \sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}$$

$$= 1.04 \pm 2.000 \times \sqrt{\frac{3.48^2}{39} + \frac{5.42^2}{34}}$$

$$= 1.04 \pm 2.000 \times 1.08$$

$$= 1.04 \pm 2.16$$

=[-1.12, 3.2]

Therefore the 95% confidence interval for the true difference is [-1.12, 3.2].

e) We fail to reject the null H again the CI covers zero.

f) If the 2 sample T test was not valid, we could use the Mann-Whitney test.