V506 Connect Homework Exercise 3 – Spring 2017

Group 4 – Damodar Panigrahi and Carlos Sathler

Point and Interval Estimation

Page 288, Problem 2 [2 pts]

$n = 81, \bar{x} = 40, \sigma = 5, CI = 95\% \;\Rightarrow z = 1.96$

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}} = 1.96 * \frac{5}{\sqrt{81}} = 40 \pm 1.0889 = (\mathbf{38.9111, 41.0889})$$

Page 289, Problem 5 [2 pts]

$n = 49, \bar{x} = 20, \sigma = 5$

a. **It is 20. It indicates that if we're taking one statistic from the sample, in this case the mean of the sample, it is the best estimate of the corresponding population parameter. In this case the best estimate is 20 which is the sample mean.**

b. $CI = 95\% \;\Rightarrow z = 1.96$

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}} = 20 \pm 1.96 * \frac{5}{\sqrt{49}} = 20 \pm 1.4 = (\mathbf{18.6, 21.4})$$

It indicates that if we take many samples of this size from this population we will likely see that 95% of the time the sample means falls in the above interval.

Page 289, Problem 7 [2 pts]

$n = 60, \bar{x} = 8.6, \sigma = 2.3$

a. **8.6. It indicates that if we're taking one statistic from the sample, in this case the mean of the sample, it is the best estimate of the corresponding population parameter. In this case the best estimate is 8.6 which is the sample mean.**

b. $CI = 99\% \;\Rightarrow z = 2.58$

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}} = 8.6 \;\pm 2.58 * \frac{2.3}{\sqrt{60}} = 8.6 \;\pm 0.7661 = (\mathbf{7.8339, 9.3661})$$

c. **If we take many samples of this size from this population we will likely see that 99% of the time the sample mean falls in the above interval**

Page 296, Problem 9

a. $CI = 95\%, n = 12 \;\Rightarrow 11\; degrees\ of\ freedom.$
$t = 2.201$

b. $CI = 90\%, n = 20 \;\Rightarrow 19\; degrees\ of\ freedom.$
$t = 1.729$

c. $CI = 99\%, n = 8 \;\Rightarrow 7\; degrees\ of\ freedom.$

$t = 3.499$

Page 297, Problem 13

x	$\bar{x}$	$(x-\bar{x})^2$
107	98.6	70.56
92	98.6	43.56
97	98.6	2.56
95	98.6	12.96
105	98.6	40.96
101	98.6	5.76
91	98.6	57.76
99	98.6	0.16
95	98.6	12.96
104	98.6	29.16
		276.4

$$n = 10, \bar{x} = 98.6, \sum (x - \bar{x})^2 = 276.4$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{276.4}{9}} = 5.5418$$

$$CI = 90\%\,, degrees\ of\ freedom = 9 \Rightarrow t = 1.833$$

$$\bar{x} \pm t * \frac{s}{\sqrt{n}} = 98.6 \pm 1.833 * \frac{5.5418}{\sqrt{10}} = 98.6 \pm 3.2123 = (\mathbf{95.3877}, \mathbf{101.8123})$$

This implies that when collecting multiple samples from this population we expect to see that 90% of the time the sample mean will lie between $\mathbf{95.3877}$ and $\mathbf{101.8123}$.

Page 297, Problem 14

x	$\bar{x}$	$(x-\bar{x})^2$
29	35.0667	36.8044
40	35.0667	24.3378
38	35.0667	8.6044
37	35.0667	3.7378
38	35.0667	8.6044
37	35.0667	3.7378
33	35.0667	4.2711
42	35.0667	48.0711
38	35.0667	8.6044
30	35.0667	25.6711
21	35.0667	197.8711
29	35.0667	36.8044
45	35.0667	98.6711
35	35.0667	0.0044
34	35.0667	1.1378
		506.9333

$n = 15, \bar{x} = 35.0667, \sum (x - \bar{x})^2 = 506.9333$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{506.9333}{14}} = 6.0174$$

$CI = 98\%\ , degrees\ of\ freedom = 14 \Rightarrow t = 2.624$

$$\bar{x} \pm t * \frac{s}{\sqrt{n}} = 35.0667\ \pm 2.624 * \frac{6.0174}{\sqrt{15}} = 35.07\ \pm 4.0769 = (\mathbf{30.9898}, \mathbf{39.1436})$$

This implies that when collecting multiple samples from this population we expect to see that 98% of the time the sample mean will lie between $\mathbf{30.9898}$ and $\mathbf{39.1436}$.

Page 300, Problem 16

a. $n = 400, X = 300$
$$\pi = p = \frac{X}{n} = \frac{300}{400} = \mathbf{0.75}$$

b. $n * p = 400 * .75 = 300$
$n * (1 - p) = 400 * (1 - .75) = 400 *\ .25 = 100$
Both $n\pi > 5$ and $n(1 - \pi) > 5$ thus we can use CLT and the standard normal distribution.

$CI = 99\%\ \Rightarrow z = 2.58$
$$p\ \pm z * \sqrt{\frac{p * (1 - p)}{n}} = 0.75 \pm 2.58 * \sqrt{\frac{0.75 * (1 - 0.75)}{400}} = 0.75 \pm 0.0559 = (\mathbf{0.6941}, \mathbf{0.8059})$$

c. **There is a 99% probability that the mean of other samples of 300 voters taken from this population will show support for this candidate with a proportion falling in the above range.**

Page 300, Problem 18

a. $n = 300, X = 15$
$$\pi = p = \frac{X}{n} = \frac{15}{300} = \mathbf{0.05}$$

b. $n * p = 300 * .05 = 15$
$n * (1 - p) = 300 * (1 - .05) = 285$
Both $n\pi > 5$ and $n(1 - \pi) > 5$ thus we can use CLT and the standard normal distribution.

$CI = 95\%\ \Rightarrow z = 1.96$
$$p\ \pm z * \sqrt{\frac{p * (1 - p)}{n}} = 0.05 \pm 1.96 * \sqrt{\frac{0.05 * (1 - 0.05)}{300}} = 0.05 \pm 0.0247 = (\mathbf{0.0253}, \mathbf{0.0747})$$

c. **No. Zach can say that he is reasonably sure (with 95% confidence) that the average proportion of defects in this lot is no greater than 7.47% (0.05 + 0.0247).**

Page 304, Problem 24

$\sigma = 0.5, E = 0.2, CI = 95\% \;\Rightarrow z = 1.96$

$$n = \left(\frac{z * \sigma}{E}\right)^2 = \left(\frac{1.96 * 0.5}{0.2}\right)^2 = 24.01 \;\Rightarrow \boldsymbol{Sample\ size = \ 25\ boxes}$$

Page 309, Problem 48

$n = 500, p = 0.65, CI = 95\% \;\Rightarrow z = 1.96$

$$p \;\pm z * \sqrt{\frac{p * (1-p)}{n}} = 0.65 \;\pm 1.96 * \sqrt{\frac{0.65 * (1-0.65)}{500}} = 0.65 \;\pm 0.0418 = (\mathbf{0.6082, 0.6918})$$

This implies that when collecting multiple samples from this population we expect to see that 95% of the time the proportion will lie between $\mathbf{0.6082}$ and $\mathbf{0.6918}$.

Page 310, Problem 53

$\pi = 0.6, E = 0.05, CI = 95\% \;\Rightarrow z = 1.96$

$$n = \pi * (1-\pi)\left(\frac{z}{E}\right)^2 = 0.6 * (0.4) * \left(\frac{1.96}{0.05}\right)^2 = 368.7936 \Rightarrow \boldsymbol{Sample\ size = \ 369\ households}$$

<u>Hypothesis Testing, Population Means, and Proportions</u>

Page 330, Problem 4

$n = 64, \bar{X} = 215, \sigma = 15,$

Step 1 $H_o: \mu \geq 220$
$H_1: \mu < 220$

Step 2 $\alpha = 0.025$

Step 3 Z statistic since sigma is known

Step 4 Reject H_o if $Z < -Z_\alpha$

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{215 - 220}{15/\sqrt{64}} = -2.67$$

$-Z_\alpha = -Z_{0.025} = -1.96$ {Z_α for 0.5 – 0.025 = 0.475 is 1.96}
-2.67 is less than -1.96

Step 5 Because -2.67 falls in the rejection region we **reject** the **null hypothesis**. Based on the sample data we have evidence to suggest (at 0.025 confidence level) that the population mean is less than 220.

Page 330, Problem 5

$n = 48 \quad \bar{X} = 59{,}500 \quad \sigma = 5{,}000$

a. $H_o: \mu = 60000$
$H_1: \mu \neq 60000$

$\alpha = 0.05$

b. Reject H_o if $|Z| > Z_{\alpha/2}$

c. $$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{59500 - 60000}{5000/\sqrt{48}} = -0.6928$$

d. $Z_{\alpha/2} = Z_{0.05/2} = 1.96$
0.6928 is not greater than 1.96, so we **do not reject the null hypothesis**.

e. $p = 2 * (0.5 - 0.2549) = 0.4902$ {Corresponding area under Z=0.69 from the z-table is 0.2549. This a two-tail test we should double}

Our p-value suggests that if the null hypothesis were true, there would be 49.02% chance of selecting a sample with the mean we found here (59,500).

Also, p-value 0.4902 (computed) > 0.05 (critical). Thus, the ***null hypothesis can't be rejected***.

Page 331, Problem 8

$n = 35 \quad \bar{X} = 84.85 \quad \sigma = 9.95$

a. $H_o: \mu \leq 80$
$H_1: \mu > 80$

$\alpha = 0.01$

b. Reject H_o if $Z > Z_\alpha$

c. $$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{84.85 - 80}{9.95/\sqrt{35}} = 2.8837$$

d. $Z_\alpha = Z_{0.01} = 2.33$
2.8837 is greater than 2.33, so we **reject the null hypothesis** {Z_α for 0.5 – 0.01 = 0.49 is 2.33. There is no Z_α for 0.49. The closest values is either 0.4901 or 0.4898. We would like to be conservative. We don't want to reject the null hypothesis now and then take on additional effort to prove that this is not the case in future. Or in other words, we have larger not-to reject area for us to be conservative (which is same as having smaller area to reject). We would like to use the 0.4901 (the z-value of which is 2.33, which is our critical value). This way the area of not to reject is larger compared to that of area if we were to consider 0.4898.}

e. $p = 0.5 - 0.4980 = 0.002$

Our p-value suggests that if the null hypothesis were true, there would be 0.2% chance of selecting a sample with the mean we found here (84.85).

Also 0.002 (computed) < 0.01 (critical). Thus, the ***null hypothesis can be rejected***.

{Corresponding area under for Z = 2.884 from the z-table could be either 0.4980 (which corresponds to 2.88) or 0.4981 (which corresponds to 2.89). The p-value corresponding to these are 0.002 (=0.5 – 0.4980) or 0.0019 (=0.5-0.4981) respectively. Thus, we chose 0.002, because larger (0.002 > 0.0019) is more conservative.}

Page 339, Problem 19

n = 12, degrees of freedom = 11

Teenager	x	$\bar{x}$	$(x - \bar{x})^2$
1	51	82.5	992.25
2	175	82.5	8556.25
3	47	82.5	1260.25
4	49	82.5	1122.25
5	44	82.5	1482.25
6	54	82.5	812.25
7	145	82.5	3906.25
8	203	82.5	14520.25
9	21	82.5	3782.25
10	59	82.5	552.25
11	42	82.5	1640.25
12	100	82.5	306.25
		Total:	38933

$$\bar{X} = \frac{\sum x}{n} = \frac{990}{12} = 82.50 \qquad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{38933}{11}}$$

Step 1 $H_o: \mu \leq 50$
$H_1: \mu > 50$

Step 2 $\alpha = 0.05$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t > t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{82.5 - 50}{\sqrt{\frac{38933}{11}}/\sqrt{12}} = 1.8923$$

$t_{\alpha,n-1} = t_{0.05,11} = 1.796$

1.8923 is greater than 1.796

Step 5 Because 1.8923 falls in the rejection region we **reject the null hypothesis**. Based on the sample data we have evidence to suggest that (at the 0.05 confidence level) the population mean is greater than 50.

p-value = 0.05 (percentage for t=1.796, which is to the left of 1.8923 for 11 degrees of freedom)
Our p-value suggests that if the null hypothesis were true, there would be a 5% chance of selecting a sample with the mean we found here (82.5).

Page 339, Problem 20

n = 15, degrees of freedom = 14

Agent	x	$\bar{x}$	$(x - \bar{x})^2$
1	53	56.4	11.56
2	57	56.4	0.36
3	50	56.4	40.96
4	55	56.4	1.96
5	58	56.4	2.56
6	54	56.4	5.76
7	60	56.4	12.96
8	52	56.4	19.36
9	59	56.4	6.76
10	62	56.4	31.36
11	60	56.4	12.96
12	60	56.4	12.96
13	51	56.4	29.16
14	59	56.4	6.76
15	56	56.4	0.16
		Total:	195.6

$$\bar{X} = \frac{\sum x}{n} = \frac{846}{15} = 56.40 \qquad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{195.6}{14}}$$

Step 1 $H_o: \mu \leq 53$
$H_1: \mu > 53$

Step 2 $\alpha = 0.05$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t > t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{56.4 - 53}{\sqrt{\frac{195.6}{14}}/\sqrt{15}} = 3.5229$$

$t_{\alpha,n-1} = t_{0.05,14} = 1.761$
3.5229 is greater than 1.761

Step 5 Because 3.5229 fall in the rejection region we **reject the null hypothesis**.

Based on the sample data we have evidence to suggest (at the 0.05 confidence level) that the population mean is greater than 53.

p-value = 0.005 (percentage for t=2.977, which is to the left of 3.5229 for 14 degrees of freedom) Our p-value suggests that if the null hypothesis were true, there would be less than 5% chance of selecting a sample with the mean we found here (56.4).

Also, 0.005 (computed) ≤ 0.05 (critical value). Thus. the ***null hypothesis can be rejected***.

Page 344, Problem 23

$n = 120$ $\bar{X} = 45{,}500$ $\sigma = 3{,}000$

Step 1 $H_o: \mu = 45000$
$H_1: \mu \neq 45000$

Step 2 $\alpha = 0.1$

Step 3 Z statistic since sigma is known

Step 4 Reject H_o if $|Z| > Z_{\alpha/2}$

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{45500 - 45000}{3000/\sqrt{120}} = 1.8257$$

$Z_{\alpha/2} = Z_{0.1/2} = 1.64$ (There is no z-score for 0.45. The closest values are either 0.4495 for z=1.64 or 0.4505 for z=1.65. We would like to be conservative, so we pick the smaller z-score to get a larger rejection reason)
1.8257 is greater than 1.64

Step 5 Because 1.8257 falls in the rejection region we **reject the null hypothesis**. Based on the sample data we have evidence to suggest (at the 0.1 confidence level) that the population mean is different than 45,000.

p-value = 2 * (0.5 -0.4656) = 0.0688 (used percentage for z-score=1.82 to be conservative) Our p-value suggests that if the null hypothesis were true, there would be less than 10% chance of selecting a sample with the mean we found here (45,500).

Also, p-value 0.0688 (computed) < 0.10 (critical). Thus, the ***null*** hypothesis can be ***rejected***.

Page 344, Problem 27

$n = 50$ $\bar{X} = 6.8$ $s = 0.9$ $degrees\ of\ freedom = 49$

Step 1 $H_o: \mu \geq 7$
$H_1: \mu < 7$

Step 2 $\alpha = 0.05$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t < -t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{6.8 - 7}{0.9/\sqrt{50}} = -1.5713$$
$$-t_{\alpha,n-1} = -t_{0.05,49} = -1.677$$
-1.5713 is not less than -1.677

Step 5 Because -1.5713 does not fall in the rejection region we **do not reject the null hypothesis**. Based on the sample data we have evidence to suggest that (at the 0.05 confidence level) the population mean is not less than 7.

p-value = 0.1 (percentage for t=1.299, which is to the left of 1.5713 for 49 degrees of freedom)
Our p-value suggests that if the null hypothesis were true, there would be a 10% chance of selecting a sample with the mean we found here (6.8).

p-value, 0.10 (computed) is not less than 0.05 (critical value). Thus. the ***null hypothesis can't be rejected***.

Page 345, Problem 34

n = 15, degrees of freedom = 14

Day	x	$\bar{x}$	$(x - \bar{x})^2$
1	25	26.06666667	1.137777778
2	27	26.06666667	0.871111111
3	25	26.06666667	1.137777778
4	26	26.06666667	0.004444444
5	25	26.06666667	1.137777778
6	28	26.06666667	3.737777778
7	28	26.06666667	3.737777778
8	27	26.06666667	0.871111111
9	24	26.06666667	4.271111111
10	26	26.06666667	0.004444444
11	25	26.06666667	1.137777778
12	29	26.06666667	8.604444444
13	25	26.06666667	1.137777778
14	27	26.06666667	0.871111111
15	24	26.06666667	4.271111111
		Total:	32.933333333

$$\bar{X} = \frac{\sum x}{n} = \frac{391}{15} = 26.067 \qquad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{32.933333333}{14}}$$

Step 1 $H_o: \mu \le 25$
$H_1: \mu > 25$

Step 2 $\alpha = 0.01$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t > t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{26.06666667 - 25}{\sqrt{\frac{32.933333333}{14}}/\sqrt{15}} = 2.6935$$

$t_{\alpha,n-1} = t_{0.01,14} = 2.624$

2.6935 is greater than 2.624

Step 5 Because 2.6935 fall in the rejection region we **reject the null hypothesis**. Based on the sample data we have evidence to suggest that (at the 0.01 confidence level) the population mean is greater than 25.

p-value = 0.01 (percentage for t=2.624, which is to the left of 2.6953 for 14 degrees of freedom) Our p-value suggests that if the null hypothesis were true, there would be a 1% chance of selecting a sample with the mean we found here (26.067).

Page 346, Problem 44

$n = 36$ $\bar{X} = 64$ $s = 8.8$ $degrees\ of\ freedom = 35$

Step 1 $H_o: \mu \geq 69$

$H_1: \mu < 69$

Step 2 $\alpha = 0.1$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t < -t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{64 - 69}{8.8/\sqrt{36}} = -3.4091$$

$-t_{\alpha,n-1} = -t_{0.1,35} = -1.306$

-3.4091 is less than -1.306

Step 5 Because -3.4091 falls in the rejection region we **reject the null hypothesis**. So the answer is yes. Based on the sample data we have evidence to suggest that (at the 0.1 significance level) the residents in Legacy Ranch use less water on average.

Page 347, Problem 48

$n = 200$ $\bar{X} = 4800$ $s = 1300$ $degrees\ of\ freedom = 199$

Step 1 $H_o: \mu \geq 5000$

$H_1: \mu < 5000$

Step 2 $\alpha = 0.05$

Step 3 t statistic since we don't know sigma

Step 4 Reject H_o if $t < -t_{\alpha,n-1}$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{4800 - 5000}{1300/\sqrt{200}} = -2.1757$$
$$-t_{\alpha,n-1} = -t_{0.05,199} = -1.653$$
-2.1757 is less than -1.653

Step 5 Because -2.1757 falls in the rejection region we **reject the null hypothesis**. So the answer is yes. Based on the sample data we have evidence to suggest that (at the 0.05 significance level) the prevention plans were effective in reducing the mean amount of a claim.

Part II:

1. Data Set Exercise 69 (Parts a-c), on page 311 of the textbook. This problem uses the Goodyear, Arizona Real Estate data (REAL-ESTATE-2003.csv) that is documented on page 716 of the textbook and is available via Canvas. [6 pts]
   a. Develop a 95% confidence interval for the mean selling price of the homes.

```
TITLE "Exercise 1a - 69a from the book";
PROC MEANS DATA=hwe3.Realestate2003 ALPHA=0.05 MEAN LCLM UCLM;
VAR Price;
RUN;
```

The MEANS Procedure

Analysis Variable : Price		
Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
221.1028571	211.9868004	230.2189139

   b. Develop a 95% confidence interval for the mean distance the home is from the center of the city.

```
TITLE "Exercise 1b - 69b from the book";
PROC MEANS DATA=hwe3.Realestate2003 ALPHA=0.05 MEAN LCLM UCLM;
VAR Distance;
RUN;
```

The MEANS Procedure

Analysis Variable : Distance		
Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
14.6285714	13.6853508	15.5717921

   c. Develop a 95% confidence interval for the proportion of homes with an attached garage.

```
TITLE "Exercise 1c - 69c from the book";
PROC FREQ DATA=hwe3.Realestate2003;
        TABLES Garage / BINOMIAL(Level=2) ALPHA=0.05;
RUN;
```

Binomial Proportion	
Garage = 1	
Proportion	0.6762
ASE	0.0457
95% Lower Conf Limit	0.5867

Binomial Proportion	
Garage = 1	
95% Upper Conf Limit	0.7657
Exact Conf Limits	
95% Lower Conf Limit	0.5779
95% Upper Conf Limit	0.7643

2. Computer Data Exercise 50, on page 347 of the textbook. This problem also uses the REAL-ESTATE-2003.csv dataset. [12 pts]
   a. A recent article in the Arizona Republic indicated that the mean selling price of the homes in the area is more than $220,000. Can we conclude that the mean selling price in the Goodyear, AZ, area is more than $220,000? Use the .01 significance level. What is the p-value?

      Ho: Mu <= 220000, Ha: Mu >22000, Alpha = 0.01
      p-value = 0.4054.
      The answer to the question is no. p-value is greater than alpha we do not reject Ho; based on this sample we **do not** have evidence to suggest that Mu is more than 220000 at the 0.01 confidence level.

```
TITLE "Exercise 2a - 50a from the book";
PROC TTEST DATA=hwe3.Realestate2003 H0=220 ALPHA=0.01 SIDES=U PLOTS=none;
VAR Price;
RUN;
```

Variable: Price

N	Mean	Std Dev	Std Err	Minimum	Maximum
105	221.1	47.1054	4.5970	125.0	345.3

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
221.1	210.2	Infty	47.1054	39.9086	57.1696

DF	t Value	Pr > t
104	0.24	0.4054

   b. The same article reported the mean size was more than 2,100 square feet. Can we conclude that the mean size of homes sold in the Goodyear, AZ, area is more than 2,100 square feet? Use the .01 significance level. What is the p-value?

Ho: Mu <= 2100, Ha: Mu >2100, Alpha = 0.01
p-value < 0.0001.
The answer to the question is yes. p-value is less than alpha so we reject Ho; based on this sample, we **have evidence** to suggest Mu is greater than 2100 at the 0.01 confidence level.

```
TITLE "Exercise 2b - 50b from the book";
PROC TTEST DATA=hwe3.Realestate2003 H0=2100 ALPHA=0.01 SIDES=U PLOTS=none;;
VAR Size;
RUN;
```

Variable: Size

N	Mean	Std Dev	Std Err	Minimum	Maximum
105	2223.8	248.7	24.2667	1600.0	2900.0

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
2223.8	2166.5	Infty	248.7	210.7	301.8

DF	t Value	Pr > t
104	5.10	<.0001

c. Determine the proportion of homes that have an attached garage. At the .05 significance level, can we conclude that more than 60 percent of the homes sold in Goodyear, AZ, area had an attached garage? What is the p-value?

Ho: Mu <= 60%, Ha: Mu > 60%, Alpha = 0.05
p = 0.6762 and z-score = 1.5936
p-value 0.0555
Probability of p=0.6762 if Mu=0.6 where true is 0.0555 (one tail p-value), which is not a significant result at the desired significance level. Based on this sample we do not have evidence to suggests Mu > 60% at the 0.5 significance level and we do not reject Ho.
The answer to the problem question is no.

```
TITLE "Exercise 2c";
PROC FREQ DATA=hwe3.Realestate2003;
	TABLES Garage / BINOMIAL(p=0.6 Level=2) ALPHA=0.05;
RUN;
```

Garage	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	34	32.38	34	32.38
1	71	67.62	105	100.00

Binomial Proportion	
Garage = 1	
Proportion	0.6762
ASE	0.0457
95% Lower Conf Limit	0.5867
95% Upper Conf Limit	0.7657
Exact Conf Limits	
95% Lower Conf Limit	0.5779
95% Upper Conf Limit	0.7643

Test of H0: Proportion = 0.6	
ASE under H0	0.0478
Z	1.5936
One-sided Pr > Z	0.0555
Two-sided Pr > |Z|	0.1110

d. Determine the proportion of homes that have a pool. At the .05 significance level, can we conclude that more than 60 percent of homes sold in the Goodyear, AZ, area had a pool? What is the p-value?

Ho: Mu <= 60%, Ha: Mu > 60%, Alpha = 0.05
p = 0.3619 and z-score z-score = -4.9801
p-value < 0.0001
Probability of p=0.3619 if Mu=0.60 were true is less than 0.0001 (one tail p-value).
So, based on this sample we have evidence to suggest Mu < 0.60 (since z-score is negative) at the 0.05 significance level. Conversely, probability of p=0.3619 if Mu <= 0.60 were true is more than 0.9999 (1-0.0001). Therefore we cannot reject the null hypothesis at the 0.05 significance level. The answer to the problem question is no.

```
TITLE "Exercise 2d";
PROC FREQ DATA=hwe3.Realestate2003;
	TABLES Pool / BINOMIAL(p=0.6 Level=2) ALPHA=0.05;
RUN;
```

Pool	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	67	63.81	67	63.81
1	38	36.19	105	100.00

Binomial Proportion	
Pool = 1	
Proportion	0.3619
ASE	0.0469
95% Lower Conf Limit	0.2700
95% Upper Conf Limit	0.4538
Exact Conf Limits	
95% Lower Conf Limit	0.2704
95% Upper Conf Limit	0.4615

Test of H0: Proportion = 0.6	
ASE under H0	0.0478
Z	-4.9801
One-sided Pr < Z	<.0001
Two-sided Pr > |Z|	<.0001

3. Data Set Exercise 52, page 347 of the textbook. This problem uses the Buena School District Bus data (Buena School District.csv) that is documented on page 721 of the textbook and is available via Canvas. [12 pts]
   a. Select the variable for the number of miles traveled last month. Conduct a test of hypothesis to determine whether the mean number of miles traveled is equal to 840. Use the .01 significance level. Find the p-value and explain what it means.

      Ho: Mu = 840, Ha: Mu =/= 840, Alpha = 0.01
      p-value = 0.0393.
      The answer to the problem is yes. p-value is greater than alpha so we do not reject Ho; based on this sample, we **do not** have evidence to suggest Mu is different than 840 at the 0.01 confidence level.
      The p-value tells us that if Mu were 840 we would have a 3.93% chance of getting a sample mean equal to 830.1. It's a small chance, but not small enough for us to accept the suggestion that Mu =/= 840 given the 1% significance level for our test.

```
TITLE "Exercise 3a - 52a from the book";
PROC TTEST DATA=hwe3.Buena HO=840 ALPHA=0.01 SIDES=2 PLOTS=none;;
VAR Miles;
RUN;
```

Variable: Miles

N	Mean	Std Dev	Std Err	Minimum	Maximum
80	830.1	42.1882	4.7168	741.0	1008.0

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
830.1	817.7	842.6	42.1882	34.9491	52.8315

DF	t Value	Pr > |t|
79	-2.10	0.0393

b. Using the maintenance cost variable, conduct a test of hypothesis to determine whether the mean maintenance cost is less than $500 at the .05 significance level. Determine the p-value and interpret the result.

Ho: Mu >= 500, Ha: Mu < 500, Alpha = 0.05
p-value < 0.0001.
The answer to the problem question is yes. p-value is less than alpha so we reject H0; based on this sample, we **have evidence** to suggest Mu is less than 500 at the 0.05 confidence level.

```
TITLE "Exercise 3b - 52b from the book";
PROC TTEST DATA=hwe3.Buena HO=500 ALPHA=0.05 SIDES=L PLOTS=none;;
VAR Maintenance;
RUN;
```

Variable: Maintenance

N	Mean	Std Dev	Std Err	Minimum	Maximum
80	450.3	53.6860	6.0023	329.0	570.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
450.3	-Infty	460.3	53.6860	46.4627	63.5895

DF	t Value	Pr < t
79	-8.28	<.0001

c. Suppose we consider a bus "old" if it is more than eight years old. At the .01 significance level, can we conclude that less than 40 percent of the buses are old? Report the p-value.

Ho: Mu >= 40%, Ha: Mu < 40%, Alpha = 0.01
p = 0.035 and z-score z-score = -0.9129
p-value = 0.01807

Probability of p=0.35 if Mu=0.40 were true is 0.01807 (one tail p-value) which is not a significant result at the desired confidence level. Based on this sample we do not have evidence to suggest that Mu < 40% at the 0.01 significance level and we do not reject Ho. The answer to the problem question is no.

```
TITLE "Exercise 3c";
* Create new dataset with variable "old";
DATA hwe3.Buena2;
        SET hwe3.Buena;
        IF Age > 8 THEN Old = 1;
        ELSE Old = 0;
RUN;
PROC FREQ DATA=hwe3.Buena2;
        TABLE Old / BINOMIAL(p=0.40 Level=2) ALPHA=0.01;
RUN;
```

Old	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	52	65.00	52	65.00
1	28	35.00	80	100.00

Binomial Proportion	
Old = 1	
Proportion	0.3500
ASE	0.0533
99% Lower Conf Limit	0.2126
99% Upper Conf Limit	0.4874
Exact Conf Limits	
99% Lower Conf Limit	0.2190
99% Upper Conf Limit	0.4994

Test of H0: Proportion = 0.4	
ASE under H0	0.0548
Z	-0.9129
One-sided Pr < Z	0.1807
Two-sided Pr > |Z|	0.3613

4. Carry out the following tasks with the data from the Baltimore Longitudinal Study of Aging. This data set is named: BLSA.DAT, which is a text file available from Canvas. [10 pts]
   a. Construct a 95% confidence interval for the mean diastolic blood pressure of all males. Interpret the results.

```
TITLE "Exercise 4a";
PROC MEANS DATA=hwe3.Blsa ALPHA=0.05 MEAN LCLM UCLM;
WHERE sex="M";
VAR dbp;
RUN;
```

The 95% confidence interval suggests that if we select multiple samples from this population we expect to see that 95% of the time the sample mean will lie between the values below. The first two confidence limit values are based on the binomial approximation the “exact” values are based on the binomial distribution.

Analysis Variable : dbp		
Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
79.9944444	78.8620660	81.1268228

   b. Construct a 99% confidence interval for the proportion of males that smoke. Interpret the results.

```
TITLE "Exercise 4b";
DATA hwe3.Blsa2;
    SET hwe3.Blsa;
    IF smoker = "Y" THEN smoke = 1;
    ELSE smoke = 0;
RUN;

PROC FREQ DATA=hwe3.Blsa2;
    TABLE smoke / BINOMIAL(Level=2) ALPHA=0.01;
    WHERE sex="M";
RUN;
```

The 99% confidence interval suggests that if we select multiple samples from this population we expect to see that 99% of the time the sample proportion will lie between the values below. The first two confidence limit values are based on the binomial approximation the “exact” values are based on the binomial distribution.

Binomial Proportion
smoke = 1

Binomial Proportion	
smoke = 1	
Proportion	0.3583
ASE	0.0253
99% Lower Conf Limit	0.2932
99% Upper Conf Limit	0.4234
Exact Conf Limits	
99% Lower Conf Limit	0.2941
99% Upper Conf Limit	0.4264

c. Construct a 90% confidence interval for the mean systolic blood pressure of all females who are over 30 years old. Interpret the results.

```
TITLE "Exercise 4c";
PROC MEANS DATA=hwe3.Blsa ALPHA=0.1 MEAN LCLM UCLM;
WHERE sex="F" and Age > 30;
VAR sbp;
RUN;
```

The 90% confidence interval suggests that if we select multiple samples from this population we expect to see that 90% of the time the sample proportion will lie between the values below.

Analysis Variable : sbp		
Mean	Lower 90% CL for Mean	Upper 90% CL for Mean
125.9124579	123.9393899	127.8855259

d. Test the claim, at the .05 level of significance, that males who are 50 and older have a mean systolic blood pressure that is greater than the mean systolic blood pressure for all men. Use the p-value method.

Ho: Mu <= 130.8416667, Ha: Mu > 130.8416667, Alpha = 0.05
p-value < 0.0001
The answer to the question is yes. p-value is less than alpha so we do reject Ho; based on this sample, we **have evidence** to suggest at the 0.05 confidence level that males who are 50 and older have a mean systolic blood pressure that is greater than the mean pressure for all men.

```
TITLE "Exercise 4d";
PROC MEANS DATA=hwe3.Blsa MEAN;
WHERE sex="M";
VAR sbp;
```

```
RUN;
* got 130.8416667;
PROC TTEST DATA=hwe3.Blsa HO=130.8416667 ALPHA=0.05 SIDES=U PLOTS=none;;
WHERE Age>=50 and sex=”M”;
VAR sbp;
RUN;
```

Variable: sbp

N	Mean	Std Dev	Std Err	Minimum	Maximum
180	139.1	23.3233	1.7384	92.0000	220.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
139.1	136.2	Infty	23.3233	21.1372	26.0177

DF	t Value	Pr > t
179	4.76	<.0001

e. Test the claim, at the .01 level, that females who have lower weight than the mean weight of all females also have a lower mean systolic blood pressure than the average female. Use the p-value method.

Ho: Mu >= 122.9416667, Ha: Mu < 122.9416667, Alpha = 0.01
p-value = 0.0828
The answer to the question is no. p-value is greater than alpha so we do not reject Ho; based on this sample, we **do not** have evidence to suggest at the 0.01 confidence level that females who have lower weight than the mean weight of all females also have a lower mean systolic blood pressure than the average female.

```
TITLE "Exercise 4e";
PROC MEANS DATA=hwe3.Blsa MEAN;
WHERE sex="F";
VAR weight sbp;
RUN;
/* found
weight  62.2766667
sbp 122.9416667 */
PROC TTEST DATA=hwe3.Blsa H0=122.9416667 ALPHA=0.01 SIDES=L PLOTS=none;
WHERE sex="F" and weight < 62.2766667;
VAR sbp;
RUN;
```

Variable: sbp

N	Mean	Std Dev	Std Err	Minimum	Maximum

N	Mean	Std Dev	Std Err	Minimum	Maximum
209	121.1	19.5293	1.3509	84.0000	200.0

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
121.1	-Infty	124.2	19.5293	17.3254	22.3213

DF	t Value	Pr < t
208	-1.39	0.0828