Project 2: Confidence Intervals and Hypothesis Tests

Name:

Part 1: Using an Input statement for data.

Questions that you must answer are highlighted in yellow. For this exercise, we will be using an input statement to read data into SAS.

First, create a temporary dataset named “seven” using the following code:

```
data seven;
input sbp @@;
cards;
172 148 123 140 108 152
123 129 133 130 137 128
115 161 142
;
run;
```

Notes:

- This will create a data set “seven” in your work library.
- If you leave SAS, you will need to complete this step again.
- This data set is created from a sample of 15 participants who were taking a medication to lower systolic blood pressure.

Calculate the mean and standard deviation for sbp. Include your SAS code and output here:

```
proc means data=seven mean std;
var sbp;
run;
```

The SAS System

The MEANS Procedure

Analysis Variable : sbp	
Mean	Std Dev
136.0666667	17.1024922

Calculate a 99% confidence interval for the mean. Include your SAS code and output here:

```
proc means data=seven mean std lclm uclm alpha=0.01;
var sbp;
run;
```

Analysis Variable : sbp			
Mean	Std Dev	Lower 99% CL for Mean	Upper 99% CL for Mean
136.0666667	17.1024922	122.9213921	149.2119412

Based on this result, if we were to conduct a hypothesis test to test whether the mean is equal (or not equal) to 130 at α=0.01, what would you expect to find?
I would expect to find that this a fair estimate of the mean since 130 is within our interval of 99% Confidence Level.

Create the hypotheses to test whether the mean is equal to 130. Use α=0.10.

Include your SAS code here:

```
proc ttest data=seven h0=130;
var sbp;
run;
```

Write your hypotheses:
H_0: $\mu = 130$
H_1: $\mu \neq 130$
We fail to reject the H_0 if p-val is less than .1

What is your p-value? 0.1911

What is your conclusion? Be sure to write both parts (reject or fail to reject and the conclusion in words).
Since our p-value is 0.1911 and this is greater than 0.1, we fail to reject the H_0. There is not enough sufficient evidence to say that the mean of sbp is not 130.

Run the hypothesis test again. This time test whether the mean is greater than 140 by adding sides=U to your statement. Use α=0.05. State your hypotheses, p-value, and conclusion.

H_0: $\mu = 140$
H_1: $\mu > 140$
We fail to reject the H_0 if p-val is less than 0.05
p-val = 0.8059
Since our p-value is 0.8059 and this is greater than 0.05, we fail to reject the H_0. There is not enough sufficient evidence to say that the mean of sbp is not greater than 140.

<u>Part 2: Using a .sas7bdat file for data.</u>

Please complete the following:

1. Download p075.sas7bdat from eLearning.
3. Save p075.sas7bdat in STA5990Data folder. File name should be exact.

Run the following SAS code to create a work data set named “eight” from your p075.sas7bdat file. Note: you will be using a similar file for this week’s project. This will be included in the project instructions.

```
libname in 'G:\My Drive\STA5990Data';

data eight;
set in.p075;
```

```
run;
```

Use the following code to explore the variables in the data set.

```
proc print data=eight (obs=10);
run;
```

Choose three variables in the data set that are continuous measures (we used weight and BMI for demo) and find their means and standard deviations using the following SAS code:

```
proc means data=eight n mean std;
var age dbp u;
run;
```

Paste your output and edited SAS code here:

```
proc means data=eight n mean std;
var age dbp u;
run;
```

Variable	Label	N	Mean	Std Dev
age	Age in Years	250	54.8244873	11.4683755
dbp	Diastolic Blood Pressure (mmHg)	250	76.7676672	8.4448878
u		250	0.0492539	0.0286188

Customize the code above in Part 1 to create a 95% confidence interval for each of your chosen variables. Paste your output and SAS code here.

```
proc means data=eight n mean std uclm lclm alpha=.05;
var age dbp u;
run;
```

The SAS System

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Upper 95% CL for Mean	Lower 95% CL for Mean
age	Age in Years	250	54.8244873	11.4683755	56.2530392	53.3959355
dbp	Diastolic Blood Pressure (mmHg)	250	76.7676672	8.4448878	77.8196000	75.7157344
u		250	0.0492539	0.0286188	0.0528188	0.0456890

Suppose we are interested in the number of participants with hypertension (HTN). Customize the SAS code from the practice to determine the number of hypertensive participants:

Paste your output and SAS code here.

```
proc freq data=eight;
table HTN / missing;
run;
```

Hypertension Status				
HTN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	103	41.20	103	41.20
1	147	58.80	250	100.00

Suppose we are interested in the number hypertensive participants by sex. Run the following code:

```
proc freq data=eight;
table male*HTN / missing;
run;
```

Paste your output here.

Frequency Percent Row Pct Col Pct	Table of male by HTN		
	HTN(Hypertension Status)		
male(Male Indicator)	0	1	Total
0	65 26.00 38.24 63.11	105 42.00 61.76 71.43	170 68.00
1	38 15.20 47.50 36.89	42 16.80 52.50 28.57	80 32.00
Total	103 41.20	147 58.80	250 100.00

How many people males are hypertensive? 42

What percentage of females are not hypertensive? 26%

Customize the following code to create a 95% confidence interval for the proportion of participants with hypertension:

```
data nine;
r=147; n=250; alpha=0.05; p=r/n; z=probit(1-alpha/2);
*standard error; se=(sqrt(n*p*(1-p)))/n;
*continuity correction; cc=1/(2*n);
LL=p-(z*se+cc);
UL=p+(z*se+cc);
run;

proc print data=nine;
var p LL UL;
run;
```

Paste your output here:

Obs	p	LL	UL
1	0.588	0.52499	0.65101

Run a hypothesis test to determine whether the proportion of participants with hypertension is equal (or not equal) to 40%. Report your hypotheses, p-value, and conclusion.

```
ods select BinomialTest;
proc freq data=eight;
tables HTN / binomial(p=0.4);
run;
```

H_0: p = .4
H_1: p ≠ .4
We reject H_0 if p-val < .05
p-val = .6985
Since our p-value is 0.6985 and this is greater than 0.05, we fail to reject the H_0. There is not enough sufficient evidence to say that the percentage of people with hypertension is not 40%.