Hypothesis Testing on one-sample proportions, two-sample proportions, one-sample mean, two-sample mean, and paired means

Please note that these Examples were prepared for the asynchronous version of BIOS 5200. Excuse the overlap in materials if you are a current or former student in the asynchronous section of the course.

35. It is assumed that the proportion of women in the United States ages 40 years and older that received a mammogram within the past two years is 0.80. You believe the actual proportion is less than this. Taking a random sample of 100 counties in the United States you get a proportion of 0.75. Evaluate your result using the steps of hypothesis testing. Because of the large sample size, use the z test.

We translate the question into the proper information. N = 100, $\hat{p} = 0.75$, X = 75. X translate into count of mammogram. The question then becomes hypothesis testing (as stated in the question) using Z-test for one proportion. Why Z-test for one proportion, we are given a hypothesized proportion for the population (p = 0.80), and an observed proportion (0.75). We did a one-sided test. Why? You believe that the actual proportion is less than what you are being told (0.8). That translate into one-sided hypothesis.

H0: p ≥ 0.80
H1: p < 0.80. α = 0.05

Test Statistic: $$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

DF = NA

Decision Rule: Reject H0 if computed (one-sided) p-value ≤ 0.05 or if Z-statistic ≥ 1.645 (The Z-critical value for one-sided α = 0.05). The information from example 35 gets translated into SAS code as below:

```
data bios5200.ex5_35;
input mammogram $ count;
datalines;
Yes 75
No 25
;
run;
```

We use two variables to input aggregate data into SAS for one sample with categorical (including binary) responses. One variable captures the categorical (including binomial) variable and one variable captures the count/frequency for each level of the categorical (including binomial variable). See code above.

We always use a data step to create new data in SAS from scratch. The data step always begin with a data statement followed by a library name (if we want permanent data) and a data name. If we want to save the data permanently, we have to give SAS the name of a library to save our newly created data. The name of the library and the name of the dataset is always separated by a period (.).

Performing Z-test on binomial proportion requires the use of the FREQ procedure. We always call a SAS procedure with PROC followed by the name of the procedure. For frequency tables, we use PROC FREQ. All SAS procedures require the name of the dataset to be analyzed which must be called in the PROC statement with DATA = library.dataname. All SAS statements at any steps always end with a semicolon (;).

In PROC FREQ, when we use aggregate data, we always call the WEIGHT statement to assign the aggregate/count/frequency to each level of the categorical variable. (see PROC FREQ code below)!

All PROC FREQ requires a TABLE statement. After the TABLE statement, we must provide the variable to be analyzed. If we need additional output to be generated, we must put a / after the variable followed by the name of the OPTIONS needed.

For a BINOMIAL proportion, we use the option BIN/BINOMIAL. To control the output, we use BINOMIAL (LEVEL = "**put level here")** statement to select the level of interest. For hypothesis testing, we include hypothesized value as BINOMIAL (LEVEL = "put level here", p = **p0**). By default, SAS uses p = 0.5 as the test value. To accurately do hypothesis test, we have to specify the correct hypothesized value under the null hypothesis. For Confidence interval using one proportion, we use BINOMIAL (LEVEL = "put level here") alpha = **alpha**. (see Code below).

Note that if you use both p and alpha together, the CI will be for p-p0 and not for p.

This is the code to solve question 35 from chapter 5

```
proc freq data = bios5200.ex5_35;
weight count; *use always with aggregate data;
tables mammogram /binomial (level = 'Yes' p = 0.8) alpha = 0.05;
run;
```

What is you do not have aggregate data? No weight statement is needed. See below for a cross-tabulation table also known as a two-way table for individual level data.

Test of H0: Proportion = 0.8	
ASE under H0	0.0400
Z	-1.2500
One-sided Pr < Z	0.1056

Conclusion: **Since computed p = 0.1056 > 0.05, we fail to reject the null hypothesis and conclude that the proportion of women 40 and older receiving mammogram in those 100 counties does not seem to be different from what is expected for all women 40 and older in the USA.**

46. Using the data in Table 5.2, assess whether the proportion of CPAP use is similar between males and females. Apply the six steps of hypothesis testing.

Male and female are mutually exclusive groups. So, we have two-sample proportion and will use the Z-test for two proportions.

H_0: $P_{Female} = P_{Male}$.
H_1: $P_{Female} \neq P_{Male}$, α = 0.05

Test statistic: $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

Decision Rule: Reject H0 if computed p-value ≤ 0.05 or if Z-statistic ≥ 1.96.

Why use a two-sided test? The hypothesis uses the word similar. So, that is a non-directional test.

```
proc freq data = bios5200.table5_2;
table CPAP*Sex/chisq; *no weight statement needed for individual level data;
run;
```

33. We found that at baseline there was no difference in the level of restless sleep patterns between 100 adult participants in an intervention group and 100 adult participants in a control group. Of interest was whether an intervention could lower the percentage experiencing restless sleep. After a six-week intervention, 18% of those in the intervention group and 30% of those in the control group experienced restless sleep problems. Apply the steps of hypothesis testing using the z test.

This is an example of a two-sample proportion. Why? We have an intervention group and a control group and those two groups are mutually exclusive. We have a binary response restless sleep. We also know that 18% in the intervention and 30% in the control experienced restless sleep at the end of a 6-week intervention. Other information given to use include N in each group. We have 100 participants for each group. We need to create the SAS dataset to analyze the data.

$H_0: P_I \geq P_C$.

$H_1: P_I < P_C$, α = 0.05

Test Statistic: $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

Decision Rule: Reject Null is computed p ≤ 0.05 or if Z-statistic ≥ 1.645.

Why are we doing a one-sided hypothesis. The key word is found in the statement of interest: "Of interest was whether an intervention could **lower** the percentage experiencing restless sleep."

```
*Creating data for two proportions using aggregate information;

data bios5200.restless;
input intervention $ restless $ count;
datalines;
Yes Yes 18
Yes No 82
No Yes 30
No No 70
;
run;


*Hypothesis Testing for two proportions;
proc freq data = bios5200.restless;
weight count; *This example uses aggregate data.  Therefore, we use weight
statement;
table intervention*restless /chisq;
run;
```

What if we wanted to get the confidence interval for the difference in proportions with restless sleep between the intervention and the control? We would use the SAS procedure PROC FREQ. But, the chisq option would get substituted with the riskdiff option, to find the difference in risk of restless sleep between intervention and control. We add order = data to keep control the way the table is presented. That way, since yes/yes is first in the data, the 2-way table maintains that order.

```
data bios5200.restless;
input intervention $ restless $ count;
```

```
datalines;
Yes Yes 18
Yes No 82
No Yes 30
No No 70
;
run;


*confidence interval for two proportions;
proc freq data = bios5200.restless order = data;
weight count; *This example uses aggregate data.  Therefore, we use weight
statement;
table intervention*restless /riskdiff; *use riskdiff to get confidence
interval for two proportions;

run;
```

40. One COVID-19 testing site says that the average time for individuals to get their test results is 20 hours, with a standard deviation of 5 hours. To test these claims, a random sample of 10 tests was identified, with the wait times 27, 20, 29, 18, 33, 22, 42, 21, 35, and 27 hours. Apply the six steps of hypothesis testing to evaluate whether or not the mean time to get test results is 20 hours.

What types of data do we have? What information are given in the question? We have a quantitative measurement and a hypothesized mean of 20 hours for a COVID-19 test. This example amounts to a one-sample t-test since we are testing hypotheses on one population mean.

$H_0: \mu = 20$.

$H_1: \mu \neq 20$, α = 0.05.

N = 10

DF = 10-1 = 9

T = 2.262

We have one variable. We will create the data using a data step in SAS and permanently save the data in a folder using a SAS library.

```
* One Sample T-test;

data bios5200.covid_test;
input hours;
datalines;
27
20
29
18
```

```
3
22
42
21
35
27
;
run;
```

We will use the SAS procedure PROC UNIVARIATE to run a one-sample t-test. By default, PROC UNIVARIATE uses a hypothesized mean **mu0= 0** by default. To change that, we use **mu0=mean** option in the PROC UNIVARIATE statement.

```
proc univariate data = bios5200.covid_test mu0=20;
var hours;
run;
```

PROC UNIVARIATE also provides summary statistic for all variables specified in the VAR statement. We use PROC UNIVARIATE for histogram and uses the HISTOGRAM statement.

```
proc univariate data = bios5200.covid_test;
var hours;
histogram;
run;
```

To get a box-and-whisker plot in SAS, we use PROC SGPLOT. The VBOX and HBOX statements are used to display the boxplot either vertically or horizontally. SAS will identify outliers in the plot. To not use show outliers, use the EXTREME option in the VBOX/HBOX statement.

```
proc sgplot data = bios5200.covid_test;
vbox hours;
run;

proc sgplot data = bios5200.covid_test;
vbox hours/extreme;
run;
```

To get confidence interval for **one mean**, we use the SAS procedure PROC MEANS. By default, SAS will display n, mean, std, min, and max. Of course, we can change that to get other summary statistic, including lower confidence limit, upper confidence limit, lower quartile, upper quartile, median, skewness, kurtosis, coefficient of variation, etc. By default, if we request the confidence interval, SAS provides 95% CI. We can change that with the alpha = alpha option. See below!

```
proc means data = bios5200.covid_test n mean std stderr min q1 median q3 max
cv skewness kurtosis lclm uclm alpha = 0.05;
var hours;
run;
```

Analysis on paired quantitative data

To do paired T-test, we need to measure a continuous outcome at two time points on the same group of people.

Example:

A new drug is proposed to lower total cholesterol and a study is designed to evaluate the efficacy of the drug in **lowering** cholesterol. Fifteen (15) patients agree to participate in the study and each is asked to take the new drug for 6 weeks. Before starting the treatment, each patient's total cholesterol level is measured. The measurements before and after the trial are given in the table below. Using a significance level of 0.005, do you have sufficient evidence to support the claim of the drug lowering cholesterol level? Provide a 95% confidence interval for the mean difference in cholesterol level.

Subject	Cholesterol Before	Cholesterol After
1	215	205
2	190	156
3	230	190
4	220	180
5	214	201
6	240	227
7	210	197
8	193	173
9	210	204
10	230	217
11	180	142
12	260	262
13	210	207
14	190	184
15	200	193

We start by creating the SAS dataset for analysis using a SAS data step.

```
data bios5200.cholesterol;
input subject choles_B choles_A;
choles_diff = choles_B-choles_A;  *creating the difference/change in
cholesterol level while inputting the data;
datalines;
1     215   205
2     190   156
3     230   190
4     220   180
5     214   201
6     240   227
7     210   197
8     193   173
9     210   204
10    230   217
11    180   142
12    260   262
13    210   207
```

```
14     190    184
15     200    193
;
run;
```

Using a significance level of 0.005, do you have sufficient evidence to support the claim of the drug lowering cholesterol level?

- Step 1: Set up the hypothesis and significance level:
  - $H_0: \mu_d \leq 0$ (because the difference was computed as before-after, check the SAS code above)
  - $H_1: \mu_d > 0$
  - Significance level ~ α = 0.005
- Step 2: Compute the appropriate degree of freedom: DF = 15-1 = 14
- Step 3: Select and compute the appropriate test statistic: $TS = \frac{\bar{x}_d - 0}{\frac{s_d}{\sqrt{n}}}$
- Step 4: Set up the decision rule (based on a distribution table, here the T-table).
  - If p-value ≤ 0.01 or TS > 2.977, we reject the null hypothesis.
  - If p-value > 0.01 or TS < 2.977, we do not reject the null

We use the SAS procedure PROC TTEST with the PAIRED statement for paired t-test. The result of the paired t-test is provided along with a 95% confidence interval for the difference in means.

```
proc ttest data = bios5200.cholesterol;
paired choles_B *choles_A;
run;
```

SAS OUTPUT for PAIRED T-TEST.

The TTEST Procedure

Difference: choles_B - choles_A

N	Mean	Std Dev	Std Err	Minimum	Maximum
15	16.9333	14.1647	3.6573	-2.0000	40.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
16.9333	9.0892	24.7775	14.1647	10.3703	22.3391

DF	t Value	Pr > |t|
14	4.63	0.0004

Since this is a one-sided test, the actual one-sided p-value would be pr>1|t| = 0.0004/2 = 0.0002

When we run PROC TTEST with PAIRED statement, we have three tables of output (see result under paired T-test).

The first table provides the summary statistic for the score difference. That means, before the difference between the paired values is being summarized. The second table provides confidence interval for the mean difference. The last table provides hypothesis test for paired means.

Things to note in the output of PROC TTEST with PAIRED statement. Not only does SAS gives the result of the paired t-test, it also gives the 95% confidence interval for the mean difference. This is only possible when we have both measures (before and after).

If you only have access to the difference values, you can still do paired t-test in SAS and confidence interval for the mean difference. However, we have to use two separate SAS procedures: PROC UNIVARIATE for the paired t-test on the difference value and PROC MEANS for the confidence interval on the difference values.

```
* paired t-test using difference values;
proc univariate data = bios5200.triglyceride mu0 = 0;
var trig_Diff;
run;

*Confidence interval using for paired means using difference values;

proc means data = bios5200.triglyceride alpha = 0.05 n mean std stderr lclm
uclm;
var trig_Diff;
run;
```

Hypothesis testing and confidence interval for two population means

44. Assess whether the mean component physical quality of life score significantly differs between those with a voice disorder versus those without a voice disorder using the data in Table 5.2. Apply the six steps of hypothesis testing.

Two things to identify in the question before we start. 1. Physical quality of life (QOL) score is a quantitative variable and 2. The subjects either have voice problems or they did not. We are asked to test whether physical QOL score differs between those with

vs. those without a voice problem. Since we have two mutually groups, those with voce problems cannot also have voice problems. We must be comparing means between two groups. Also, since we are asked for a difference, we need to do a two-sided hypothesis test.

- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_1 \neq \mu_2$, α = 0.05

- Step 2: Compute the appropriate degree of freedom: DF = n1+n2-2 = 92
- Step 3: Select and compute the appropriate test statistic: $TS = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$
- Step 4: Set up the decision rule (based on a distribution table, here the T-table).
  - If p-value ≤ 0.05 or TS ≥ 1.99, we reject the null hypothesis.
  - If p-value > 0.05 or TS < 1.99, we do not reject the null hypothesis
- Step 5: Draw your conclusion, either reject the null or fail to reject the null based on the decision rule or computed p-value.

To do two-sample t-test, SAS uses the procedure PROC TTEST. However, contrary to paired t-test, we must use the CLASS statement instead of the PAIRED statement. Then, we must identify the variable to be analyzed using a VAR statement. We can list multiple variables after the VAR statement, but we can only have one CLASS statement.

```
proc ttest data = bios5200.table5_2;
class VP;
var physical;
run;
```

45. Continuing with Exercise 44, calculate the 95% confidence interval for the difference in means, interpret it, and say what it tells us about statistical significance.

The same PROC TTEST code for two sample t-test also gives the 95% confidence interval for the difference in means.

SAS OUTPUT from the above code:

The TTEST Procedure

Variable: Physical

VP	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
No		57	69.0907	20.1819	2.6732	4.5500	95.4500
Yes		37	49.3116	24.2995	3.9948	11.3600	90.9100
Diff (1-2)	Pooled		19.7791	21.8856	4.6204		
Diff (1-2)	Satterthwaite		19.7791		4.8067		

VP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		69.0907	63.7357	74.4457	20.1819	17.0387	24.7581
Yes		49.3116	41.2098	57.4135	24.2995	19.7606	31.5641
Diff (1-2)	Pooled	19.7791	10.6025	28.9557	21.8856	19.1289	25.5779
Diff (1-2)	Satterthwaite	19.7791	10.1845	29.3737			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	92	4.28	<.0001
Satterthwaite	Unequal	66.842	4.11	0.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	36	56	1.45	0.2086

The Last (4th) table provide the test for equality of variance which will determine if the result of the pooled (equal) variances should be reported or the result of the Satterthwaite (unequal) variances should be reported. When the value under Pr > F is ≤ 0.05, we should report the Satterthwaite (unequal variances) result for both Hypothesis testing and confidence interval.