

# ANLT5030 – Unit 4

## Assignment 1 Tutorial

SAS Studio



# Instructions

For this assignment, you will continue practicing with confidence intervals in a more ambiguous context. For this assignment, read Case Problem 2, Ethical Behavior of Business Students at Bayview University, on pages 452–454 of your text, and download the accompanying Bayview data set from CengageBrain. Use SAS to create the following elements in your report:

- Use descriptive statistics to summarize the data and comment on your findings.
- Develop 95% confidence intervals for the proportion of all students, the proportion of male students, and the proportion of female students who were involved in some type of cheating.
- Conduct a hypothesis test to determine if the proportion of business students at Bayview University who were involved in some type of cheating is less than that of business students at other institutions, as reported by the *Chronicle of Higher Education*.
- Conduct a hypothesis test to determine if the proportion of business students at Bayview University who were involved in some type of cheating is less than that of nonbusiness students at other institutions, as reported by the *Chronicle of Higher Education*.
- Describe the advice you would give to the dean based on your analysis of the data related to cheating at Bayview University.

Remember, to download the Bayview data set Web file from the Chapter 9 content, access the free content at CengageBrain.



# Case Problem

## Case Problem 2 Ethical Behavior of Business Students at Bayview University

During the global recession of 2008 and 2009, there were many accusations of unethical behavior by Wall Street executives, financial managers, and other corporate officers. At that time, an article appeared that suggested that part of the reason for such unethical business behavior may stem from the fact that cheating has become more prevalent among business students (*Chronicle of Higher Education*, February 10, 2009). The article reported that 56% of business students admitted to cheating at some time during their academic career as compared to 47% of nonbusiness students.

Cheating has been a concern of the dean of the College of Business at Bayview University for several years. Some faculty members in the college believe that cheating is more widespread at Bayview than at other universities, whereas other faculty members think that cheating is not a major problem in the college. To resolve some of these issues, the dean commissioned a study to assess the current ethical behavior of business students at Bayview. As part of this study, an anonymous exit survey was administered to a sample of 90 business students from this year's graduating class. Responses to the following questions were used to obtain data regarding three types of cheating.

During your time at Bayview, did you ever present work copied off the Internet as your own?

Yes\_\_\_ No\_\_\_

During your time at Bayview, did you ever copy answers off another student's exam?

Yes\_\_\_ No\_\_\_

During your time at Bayview, did you ever collaborate with other students on projects that were supposed to be completed individually?

Yes\_\_\_ No\_\_\_

Any student who answered Yes to one or more of these questions was considered to have been involved in some type of cheating. A portion of the data collected follows. The complete data set is in the DATAfile named Bayview.

DATAfile

Bayview

Student	Copied from Internet	Copied on Exam	Collaborated on Individual Project	Gender
1	No	No	No	Female
2	No	No	No	Male
3	Yes	No	Yes	Male
4	Yes	Yes	No	Male



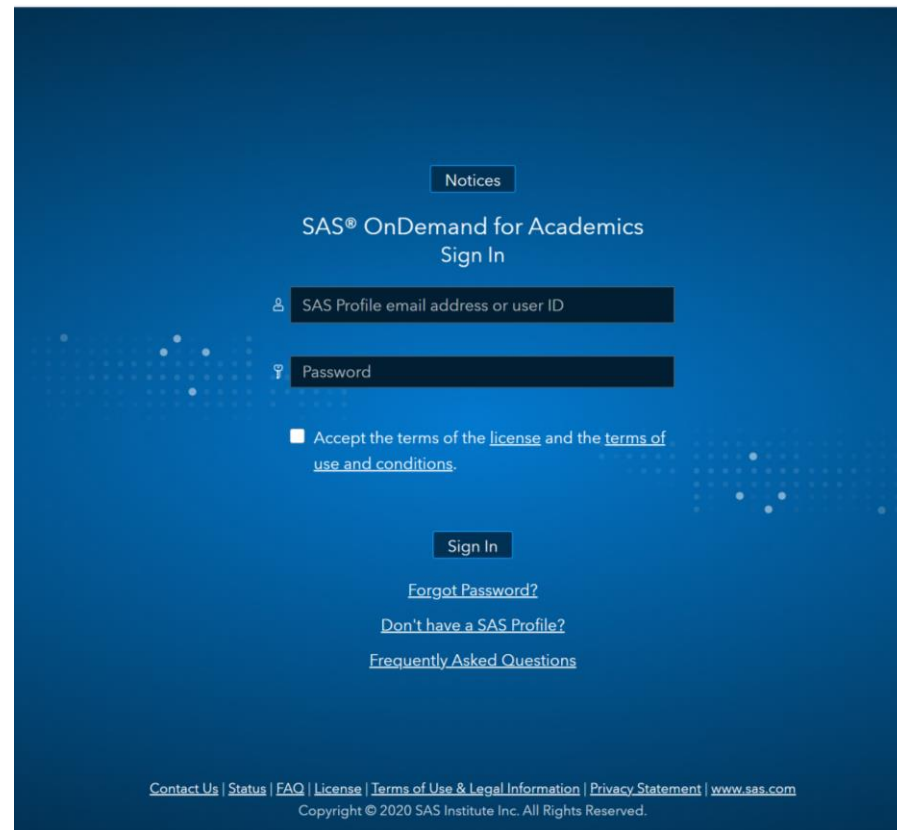
# Dataset

- Download the BayView.xlsx file from the course datasets or from the Unit 4 Welcome announcement in the course announcements.



# Access the SAS OnDemand for Academics Control Center

<https://odamid.oda.sas.com/SASODAControlCenter>



The screenshot shows the SAS OnDemand for Academics Sign In page. The background is a dark blue gradient with a subtle pattern of white dots. At the top, there is a 'Notices' button. Below it, the text 'SAS® OnDemand for Academics' and 'Sign In' are centered. There are two input fields: 'SAS Profile email address or user ID' and 'Password'. Below the password field is a checkbox labeled 'Accept the terms of the [license](#) and the [terms of use and conditions](#)'. A 'Sign In' button is centered below the checkbox. At the bottom, there are three links: 'Forgot Password?', 'Don't have a SAS Profile?', and 'Frequently Asked Questions'. The footer contains a row of links: 'Contact Us', 'Status', 'FAQ', 'License', 'Terms of Use & Legal Information', 'Privacy Statement', and 'www.sas.com', followed by the copyright notice 'Copyright © 2020 SAS Institute Inc. All Rights Reserved.'



# SAS OnDemand for Academics (SODA) Control Center

The screenshot displays the SAS OnDemand for Academics (SODA) Control Center dashboard. At the top, the SAS logo is on the left, and the user's location (United States) and name (Stefanie Reay) are on the right. The main heading is "SAS® OnDemand for Academics Dashboard". Below this, there are tabs for "Planned Events" and "Notices". A navigation bar includes "Applications", "Enrollments", and "Courses". The "Applications" tab is active, showing a list of SAS products:

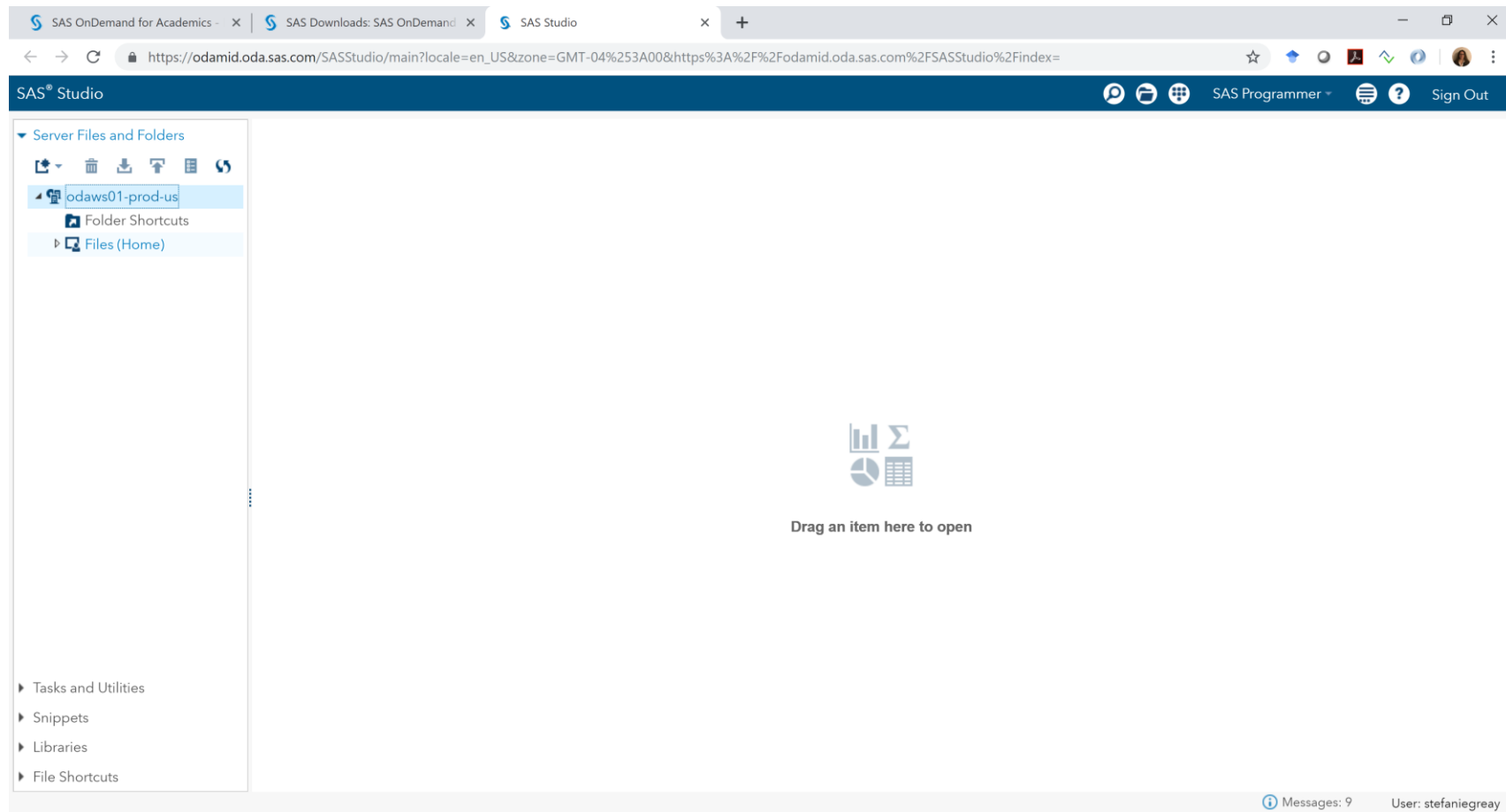
- SAS® Studio**: Write and run SAS code with a Web-based SAS development environment. *Actions: [Clear my saved tabs.](#)*
- SAS® Enterprise Guide®**: Deliver the power of SAS from an easy-to-use, point-and-click interface. ([Download Required](#))
- SAS® Enterprise Miner™**: Reveal valuable insights with powerful data mining software. ([Configuration Steps Required](#)) *Actions: [Clear my project locks.](#)*
- SAS® Forecast Studio**: Generate large numbers of high-quality forecasts automatically. ([Configuration Steps Required](#)) *Actions: [Manage your personal environment.](#)*
- JMP® Software access to SAS® hosted servers**: Statistical discovery software. Users must have a copy of JMP® software. ([Configuration Steps Required](#))

On the right side, there is a "Reference" section with links to the [Support Site](#), [Step-by-Step Reference Guides](#), and [Frequently Asked Questions](#). Below this is a "Quotas (learn more)" section showing progress bars for "Home Directory (46.5MB/5120MB)" at 1% and "Course Directory (207.0MB/3072MB)" at 7%.

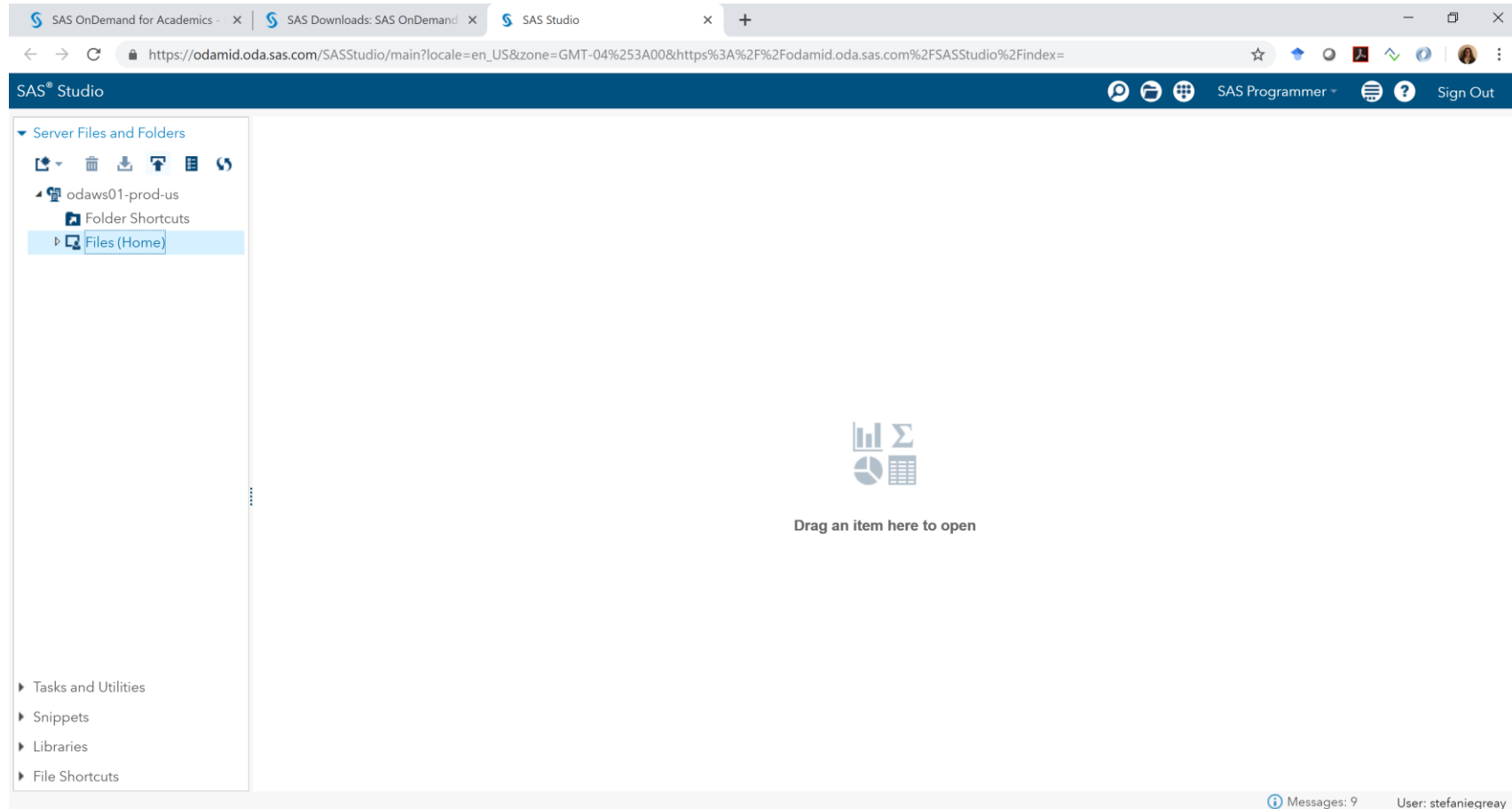
At the bottom, there is a link for "Other Ways to Access SAS® OnDemand for Academics Resources".



# Click on Files(Home)

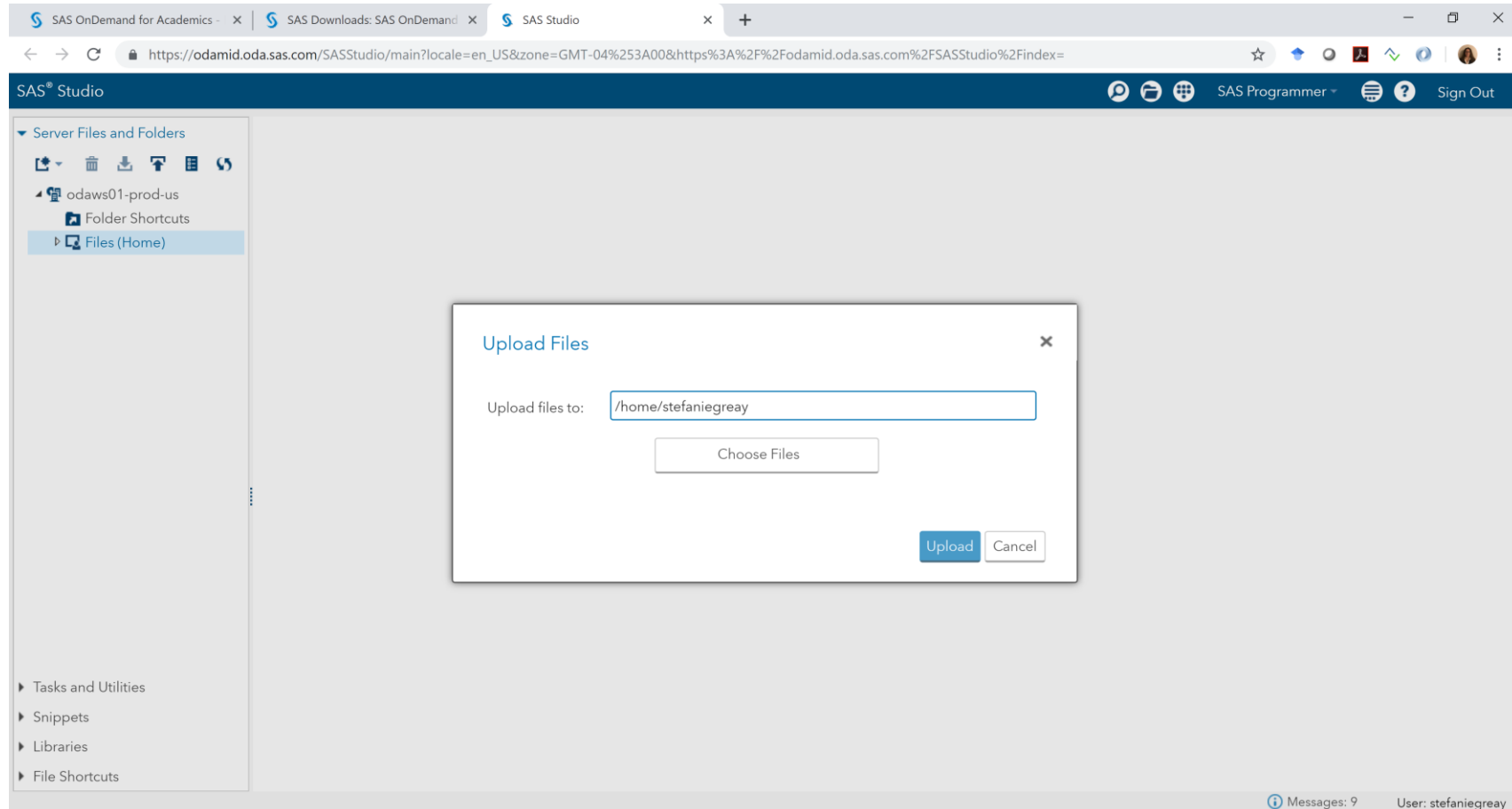


# The Upload button will display in dark blue





You can create a folder at this point, if you wish, or simply upload to your home directory.



Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”

Upload Files

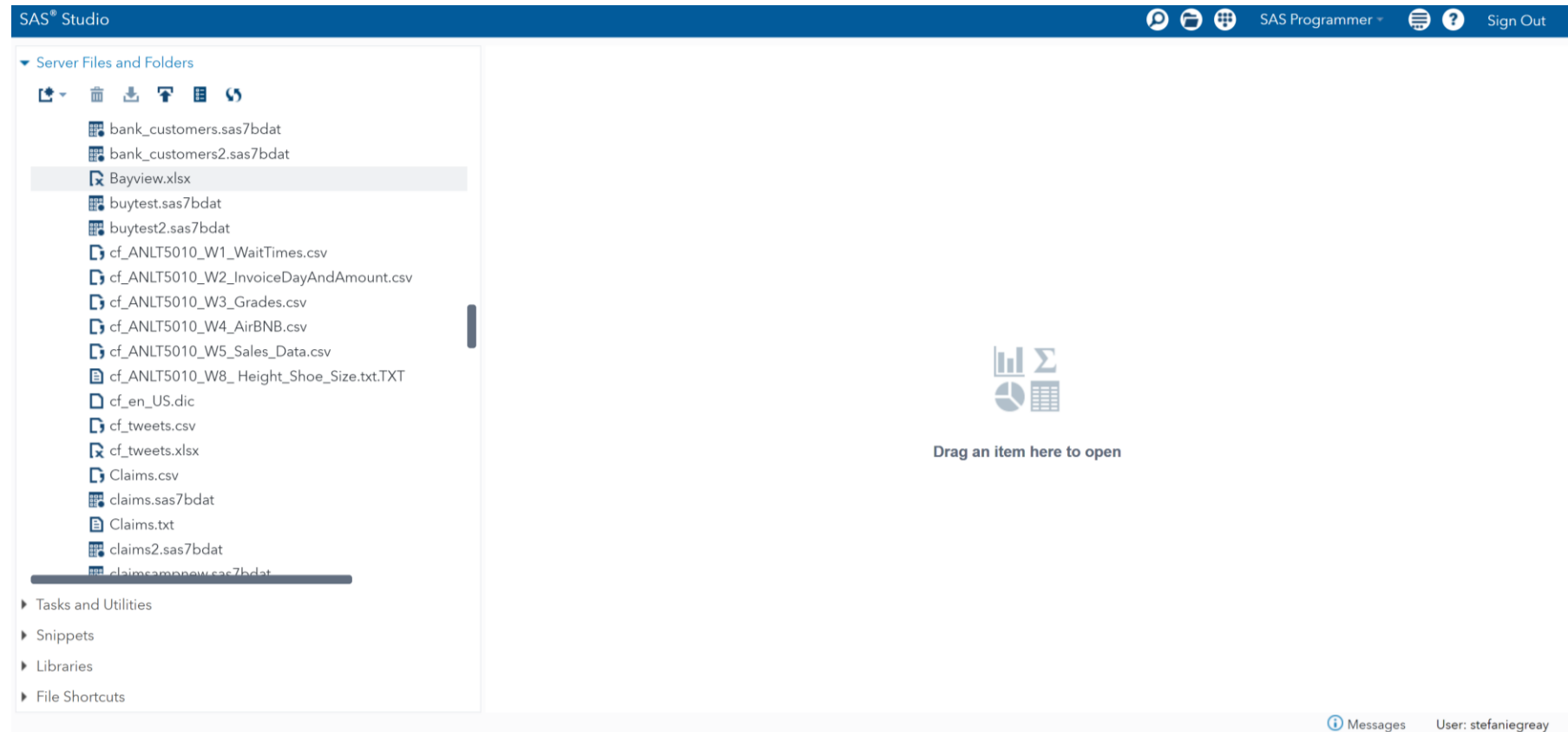
Upload files to:

**Selected files:**

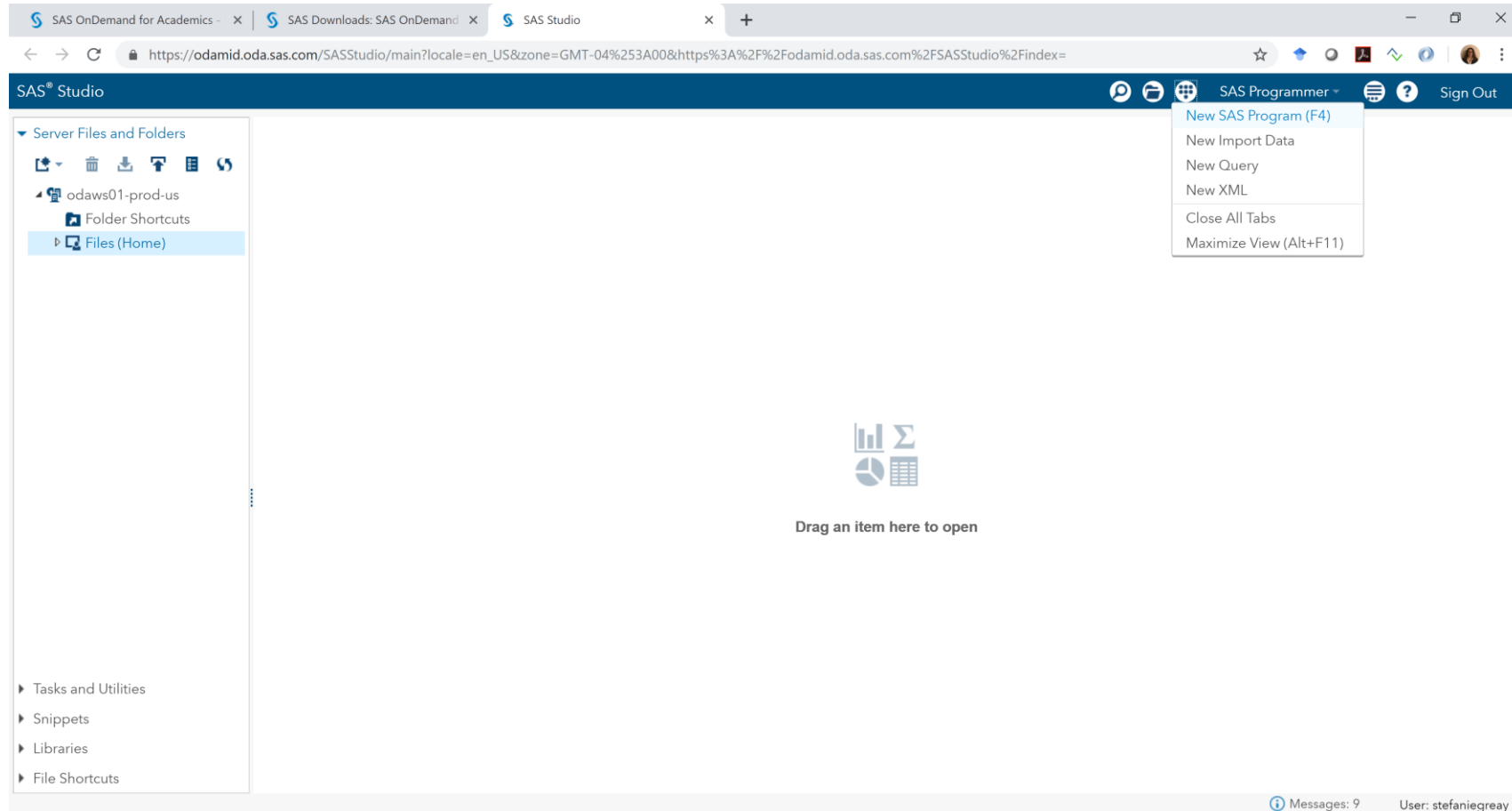
1	XLSX	Bayview.xlsx	10.9 kb
---	------	--------------	---------



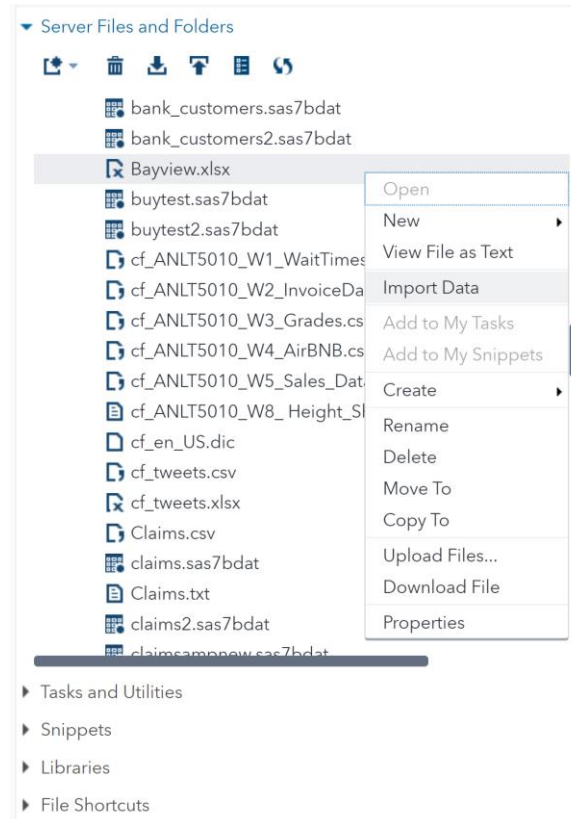
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



# To get started with the SAS portion of the Unit 3 Assignment 1 assignment, start a new SAS program.



# Import the dataset into a SAS dataset format (from the current xlsx format)



The Proc Import code will be written for you (save this as a template to use for future imports!)

SAS® Studio

SAS Programmer

Sign Out

Server Files and Folders

APA.rtf

APA1.rtf

APA2.rtf

APA\_Styl.sas

APA\_Style\_outputtesting.sas

APA\_Style\_test.sas

APA\_tst.rtf

bank\_customers.sas7bdat

bank\_customers2.sas7bdat

Bayview.xlsx

buytest.sas7bdat

buytest2.sas7bdat

cf\_ANLT5010\_W1\_WaitTimes.csv

cf\_ANLT5010\_W2\_InvoiceDayAndAmount.csv

cf\_ANLT5010\_W3\_Grades.csv

cf\_ANLT5010\_W4\_AirBNB.csv

cf\_ANLT5010\_W5\_Sales\_Data.csv

cf\_ANLT5010\_W8\_Height\_Shoe\_Size.txt.TXT

cf\_en\_US.dic

Tasks and Utilities

Snippets

Libraries

File Shortcuts

\*Bayview

SettingsCode/ResultsSplit

LogCode

FILE INFORMATION

SOURCE FILE

File name: Bayview.xlsx

Source location: /home/stefaniegreay

Worksheet name: First worksheet

OUTPUT DATA

SAS server: SASApp

Data set name: IMPORT

Library: WORK

CODE

LOG

RESULTS

OUTPUT DATA

Table of Contents

The CONTENTS Procedure

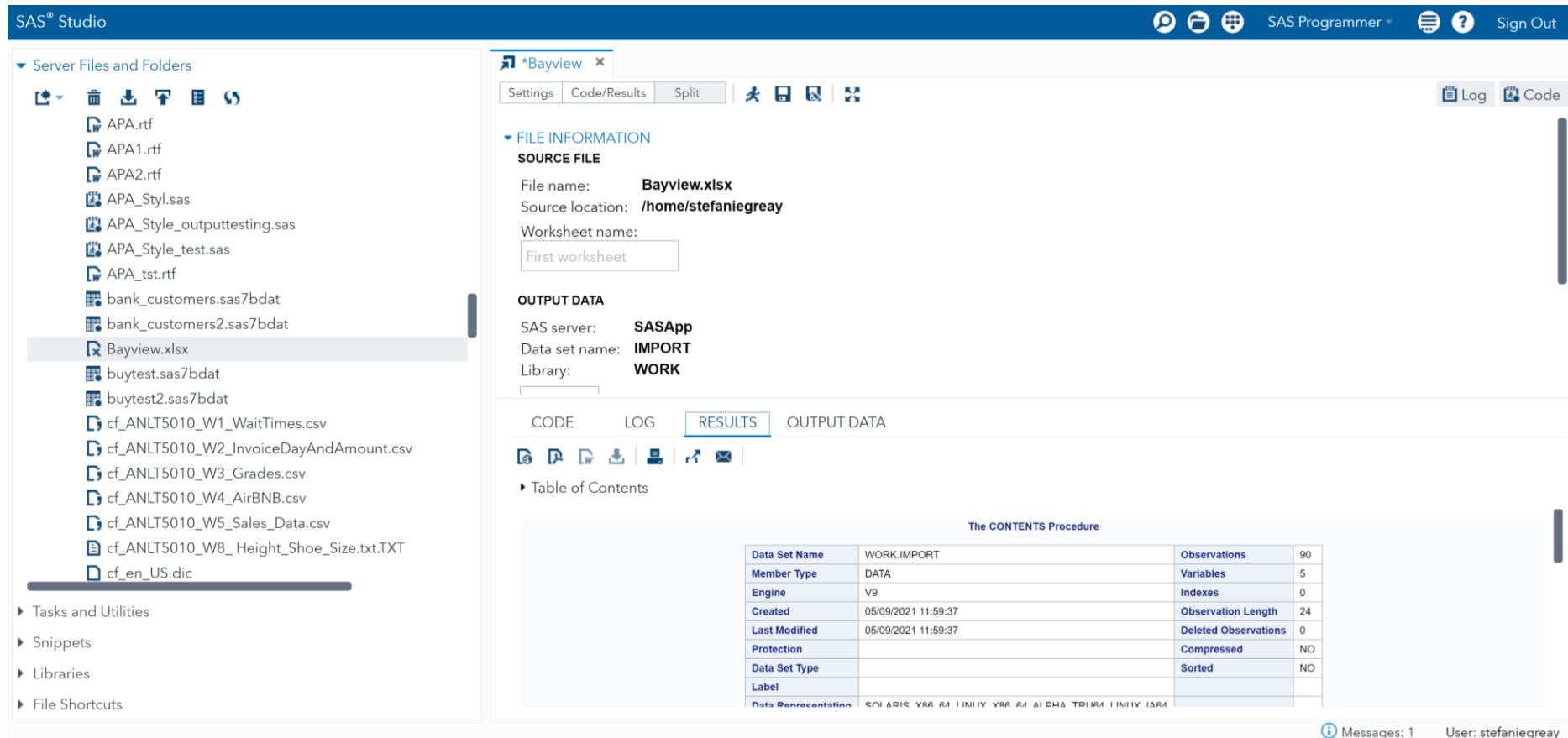
Data Set Name	WORK.IMPORT	Observations	90
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	05/09/2021 11:59:37	Observation Length	24
Last Modified	05/09/2021 11:59:37	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLAPIS YR6 64 I INIUY YR6 64 ALPHA TRIIRA I INIUY IARA		

Messages: 1User: stefaniegreay

NOTE: you have to run this code for the data to actually import.



To run the code, click the icon that looks like a guy running.



SAS® Studio

Server Files and Folders

- APA.rtf
- APA1.rtf
- APA2.rtf
- APA\_Style.sas
- APA\_Style\_outputtesting.sas
- APA\_Style\_test.sas
- APA\_tst.rtf
- bank\_customers.sas7bdat
- bank\_customers2.sas7bdat
- Bayview.xlsx**
- buytest.sas7bdat
- buytest2.sas7bdat
- cf\_ANLT5010\_W1\_WaitTimes.csv
- cf\_ANLT5010\_W2\_InvoiceDayAndAmount.csv
- cf\_ANLT5010\_W3\_Grades.csv
- cf\_ANLT5010\_W4\_AirBNB.csv
- cf\_ANLT5010\_W5\_Sales\_Data.csv
- cf\_ANLT5010\_W8\_HeightShoe\_Size.txt.TXT
- cf\_en\_US.dic

Tasks and Utilities

Snippets

Libraries

File Shortcuts

\*Bayview x

Settings Code/Results Split

Log Code

FILE INFORMATION

SOURCE FILE

File name: Bayview.xlsx

Source location: /home/stefaniegreay

Worksheet name: First worksheet

OUTPUT DATA

SAS server: SASApp

Data set name: IMPORT

Library: WORK

CODE LOG RESULTS OUTPUT DATA

Table of Contents

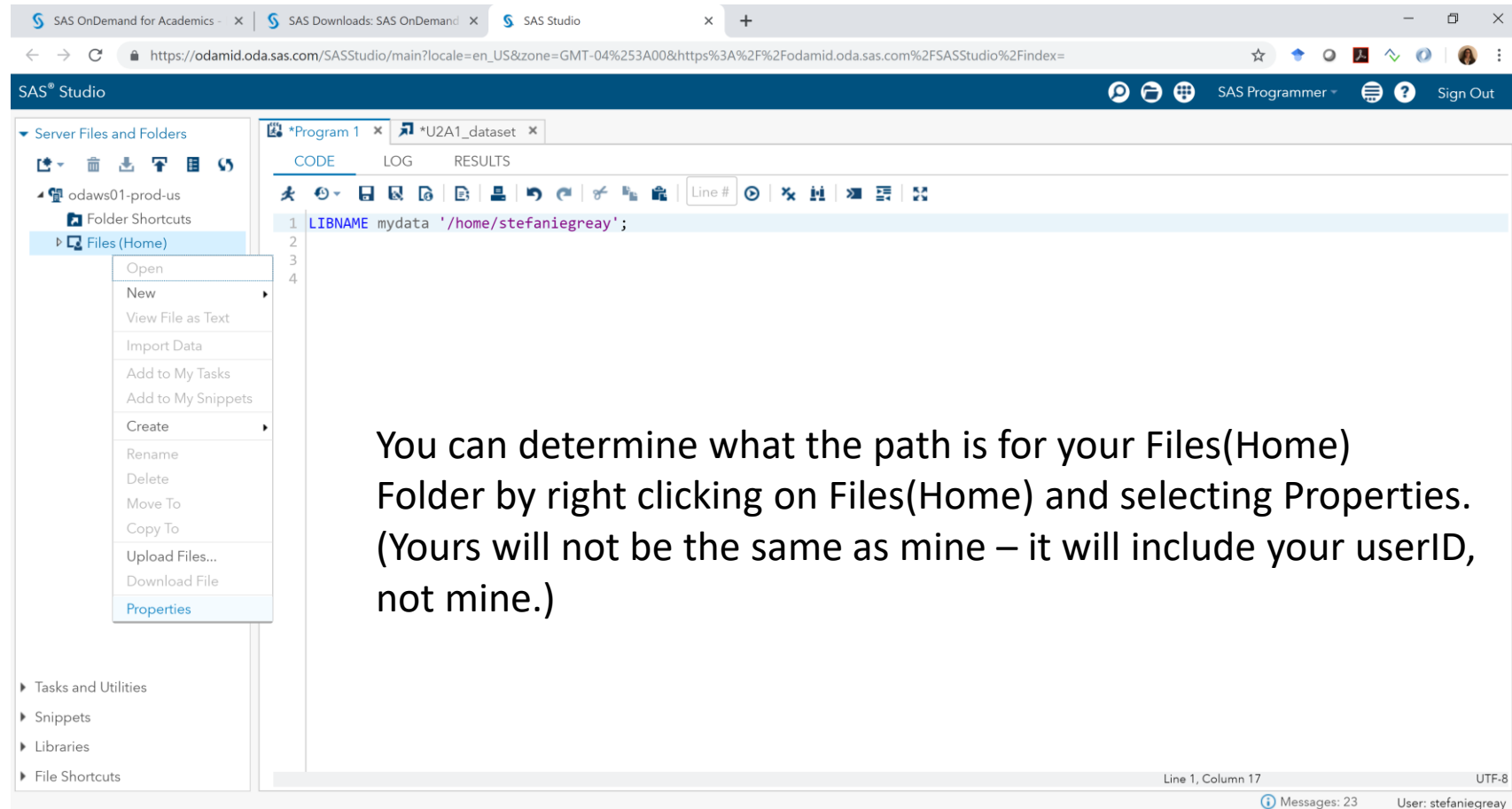
The CONTENTS Procedure

Data Set Name	Member Type	Engine	Created	Last Modified	Protection	Data Set Type	Label	Observations	Variables	Indexes	Observation Length	Deleted Observations	Compressed	Sorted
WORK.IMPORT	DATA	V9	05/09/2021 11:59:37	05/09/2021 11:59:37				90	5	0	24	0	NO	NO

Messages: 1 User: stefaniegreay

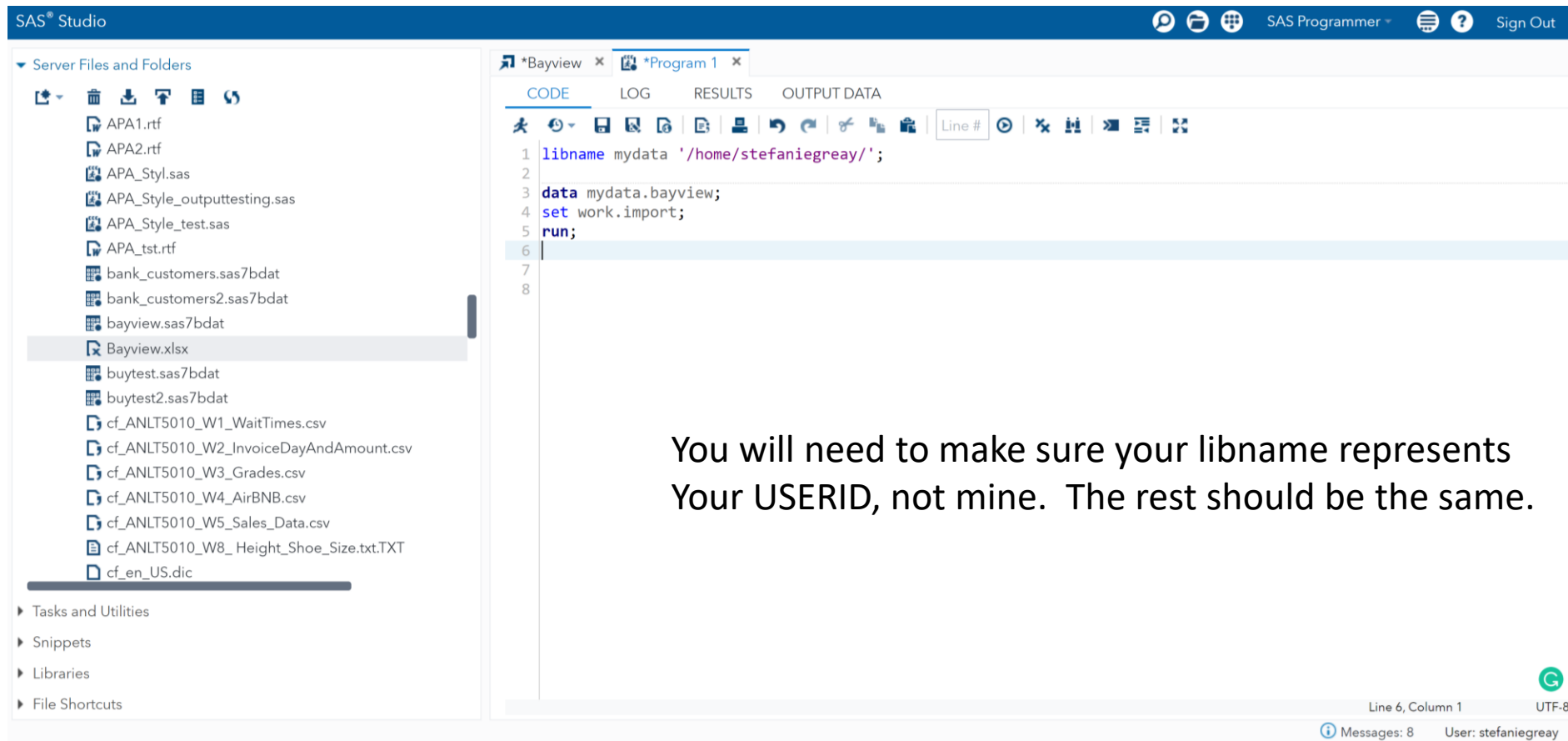


# To create a SAS Library for your Files(Home) folder, you need to use a libname statement





# Save the temporary SAS dataset created by the import to your library using the following sample code.



When you run the code, you will see the dataset in the ouput data window and can verify its success.

SAS® Studio

Server Files and Folders

- APA1.rtf
- APA2.rtf
- APA\_Styl.sas
- APA\_Style\_outputtesting.sas
- APA\_Style\_test.sas
- APA\_tst.rtf
- bank\_customers.sas7bdat
- bank\_customers2.sas7bdat
- bayview.sas7bdat
- Bayview.xlsx**
- buytest.sas7bdat
- buytest2.sas7bdat
- cf\_ANLT5010\_W1\_WaitTimes.csv
- cf\_ANLT5010\_W2\_InvoiceDayAndAmount.csv
- cf\_ANLT5010\_W3\_Grades.csv
- cf\_ANLT5010\_W4\_AirBNB.csv
- cf\_ANLT5010\_W5\_Sales\_Data.csv
- cf\_ANLT5010\_W8\_HeightShoe\_Size.txt.TXT
- cf\_en\_US.dic

Tasks and Utilities

- Snippets
- Libraries
- File Shortcuts

\*Bayview x \*Program 1 x

CODE LOG RESULTS **OUTPUT DATA**

Table: MYDATA.BAYVIEW View: Column names Filter: (none)

Columns

- ☒ Select all
- ☒ Student
- ☒ Copied from Internet
- ☒ Copied on Exam
- ☒ Collaborated on Individual Proje
- ☒ Gender

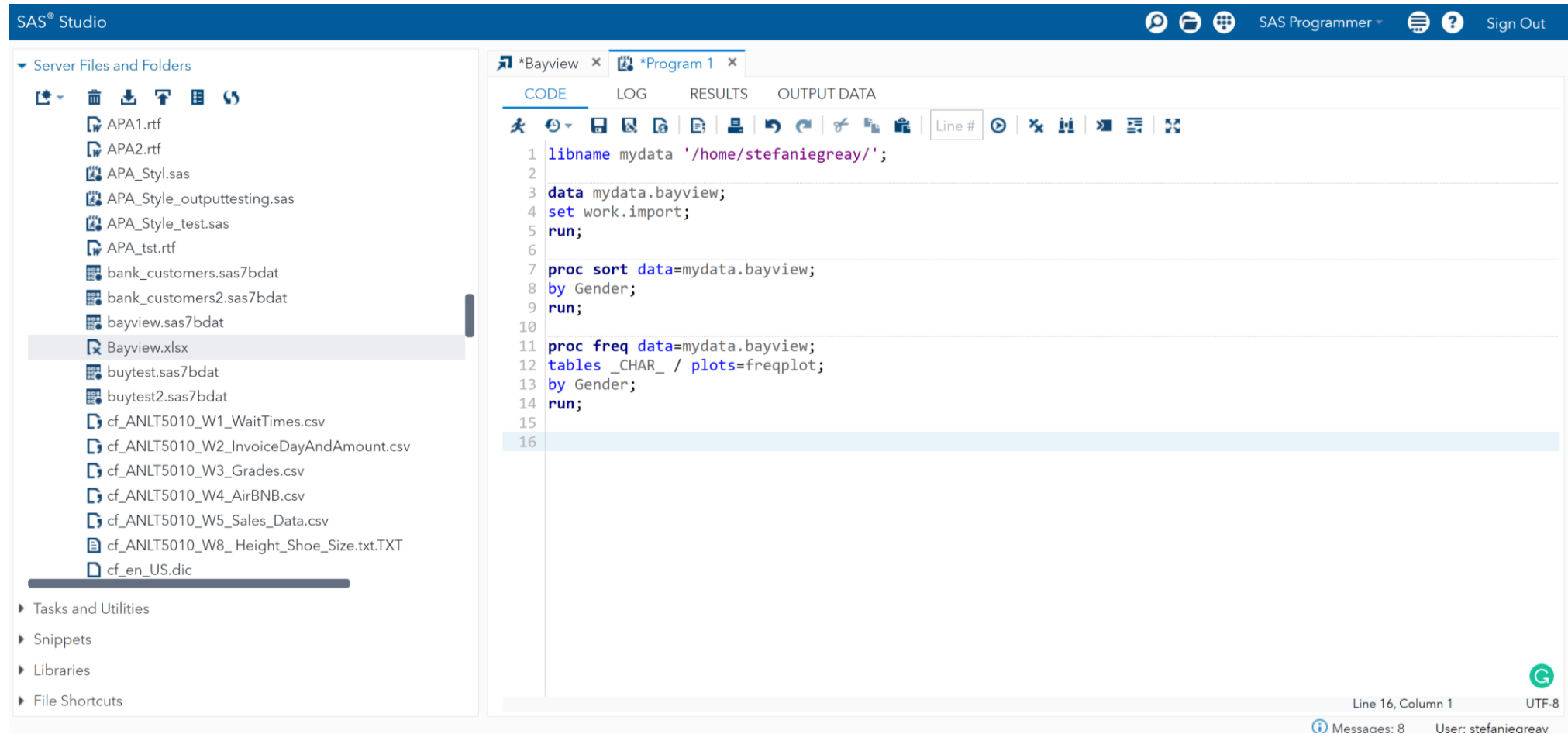
Property Value

	Student	Copied from Internet	Copied on Exam	Collaborated on Individual Proje
1	1	No	No	No
2	6	Yes	No	No
3	7	Yes	Yes	Yes
4	10	Yes	No	No
5	12	No	Yes	Yes
6	18	No	No	No
7	19	No	No	No
8	22	Yes	No	No
9	23	No	No	No
10	25	No	No	Yes
11	26	No	Yes	No
12	27	No	No	No
13	32	Yes	Yes	Yes
14	35	No	No	No
15	36	Yes	No	Yes
16	38	No	No	No
17	40	Yes	Yes	Yes

Messages: 8 User: stefaniecreav



# You can now run any procedures against that dataset via the code window.



# Sample Code for the descriptive portion of this assignment

```
libname mydata '/home/stefaniegreay/';
```

```
data mydata.bayview;  
set work.import;  
run;
```

```
proc sort data=mydata.bayview;  
by Gender;  
run;
```

```
proc freq data=mydata.bayview;  
tables _CHAR_ / plots=freqplot;  
by Gender;  
run;
```



Note that proc freq uses the lowest level for confidence interval and hypothesis test output. This code helps adjust so that it uses the correct level.

```
data mydata.bayview2;  
set mydata.bayview;  
CfI=2;  
CoE=2;  
CoIP=2;  
Cheated=2;  
if 'Copied from Internet'n = 'Yes' then CfI=1;  
if 'Copied on Exam'n = 'Yes' then CoE=1;  
if 'Collaborated on Individual Proje'n = 'Yes' then CoIP=1;  
if ('Copied from Internet'n = 'Yes'  
    or 'Copied on Exam'n = 'Yes'  
    or 'Collaborated on Individual Proje'n = 'Yes' ) then Cheated=1;  
run;
```

Reference:

<https://www.lexjansen.com/phuse/2013/sp/SP05.pdf>



Once you run the code shown below (and on the previous slides), you can review the results to see the confidence intervals and hypothesis test output.

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow binomial;  
exact binomial;  
run;
```

The FREQ Procedure		
Cheated	Frequency	Percent
1	37	41.11
2	53	58.89

Binomial Proportion	
Cheated = 1	
Proportion (P)	0.4111
ASE	0.0519
95% Lower Conf Limit	0.3095
95% Upper Conf Limit	0.5128
Exact Conf Limits	
95% Lower Conf Limit	0.3084
95% Upper Conf Limit	0.5198

Test of H0: Proportion = 0.5	
ASE under H0	0.0527
Z	-1.6865
One-sided Pr < Z	0.0458
Two-sided Pr >  Z	0.0917
Exact Test	
One-sided Pr <= P	0.0567
Two-sided = 2 * One-sided	0.1133

Sample Size = 90

Confidence Interval  
Numeric Results  
(Normal approximation on  
top and then exact Binomial  
below that)

Hypothesis Test Numeric Results  
(Normal approximation on top  
and then exact Binomial below  
that)



# Sample Code for the confidence interval portion of this assignment

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow binomial;  
exact binomial;  
run;
```

```
proc sort data=mydata.bayview2;  
by Gender;  
run;
```

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow binomial;  
by Gender;  
exact binomial;  
run;
```



# Interpreting Confidence Intervals

Example interpretation of a confidence interval for a 95% confidence interval for the average age of students currently enrolled at college ABCD that results in a point estimate of 20 with a margin of error of 2:

- “We are 95% confident that the true average/mean age of students currently enrolled at college ABCD is 20 years old, with a margin of error of 2 years.”

OR

- “We are 95% confident that the true average/mean age of students currently enrolled at college ABCD is between 18 and 22 years old.”





# Sample Code for the hypothesis test portion of this assignment

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow binomial(p=.56);  
exact binomial;  
run;
```

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow binomial(p=.47);  
exact binomial;  
run;
```



Once you run the code shown on the previous slide, you can review the results to see the hypothesis test output.

```
proc freq data=mydata.bayview2;  
tables Cheated / nocum norow  
binomial(p=.47);  
exact binomial;  
run;
```

The FREQ Procedure

Cheated	Frequency	Percent
1	37	41.11
2	53	58.89

Binomial Proportion	
Cheated = 1	
Proportion (P)	0.4111
ASE	0.0519
95% Lower Conf Limit	0.3095
95% Upper Conf Limit	0.5128
Exact Conf Limits	
95% Lower Conf Limit	0.3084
95% Upper Conf Limit	0.5198

Test of H0: Proportion = 0.47	
ASE under H0	0.0526
Z	-1.1194
One-sided Pr < Z	0.1315
Two-sided Pr >  Z	0.2630
Exact Test	
One-sided Pr <= P	0.1553
Two-sided = 2 * One-sided	0.3107

Sample Size = 90

Hypothesis Test Numeric Results  
(Normal approximation on top  
and then exact Binomial below  
that)

Hypothesis Test Numeric Results  
(Normal approximation on top  
and then exact Binomial below  
that)



# Assumptions for Normal Approximation to Binomial Hypothesis tests (for one proportion)

Parameter	Assumption
<b>P</b>	<ul style="list-style-type: none"><li>1) Sample is taken via a SRS or data is from a randomized experiment</li><li>2) <math>n \cdot p \geq 10</math> AND <math>n \cdot (1-p) \geq 10</math></li><li>3) The sampled values are independent of each other</li></ul>



# 5 Steps of a Hypothesis Test

- Hypotheses
  - (null and alternative hypothesis)
- Test Statistic
  - (from software output)
- P-value (or Rejection Region)
  - (from software output)
- Result
  - (reject or fail to reject the null hypothesis)
- Conclusion
  - (result written in terms of claim)



# Step 1: Hypotheses

- $H_o$ : parameter = null value
- $H_a$ : parameter  $>$  null value (*one-tailed test*)  
parameter  $<$  null value (*one-tailed test*)  
parameter  $\neq$  null value (*two-tailed test*)
- Choices for parameters for one sample/population hypothesis test:  
p (proportion) or  $\mu$  (mean)  
(for our case, we are dealing with a proportion, so the parameter will be p)
- The null value comes from the claim we are testing.



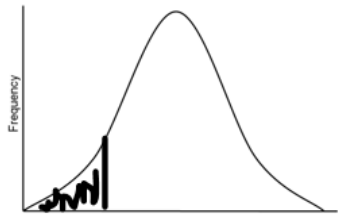
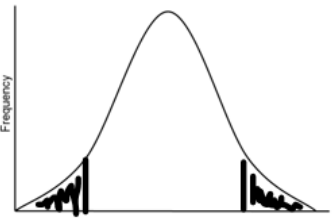
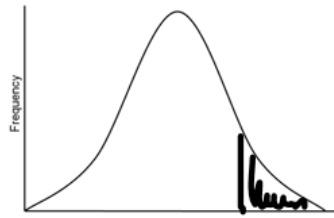
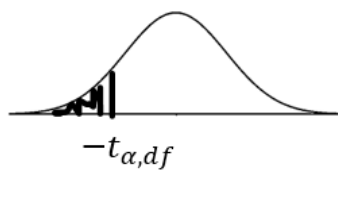
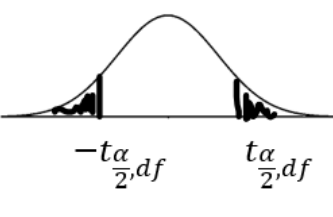
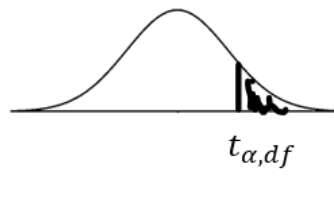
## Step 2: Test Statistic

Parameter	Null Value	Test Statistic
<b>P</b>	$p_0$	$Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 (1 - p_0)}{n}}}$
<b><math>\mu</math></b>	$\mu_0$	$t^* = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$



# Step 3: P-value

p-value and rejection region depend on direction of alternative hypothesis

Alternative Hypothesis Symbol	Ha: Parameter < null value	Ha: Parameter ≠ null value	Ha: Parameter > null value
p-value parameter = p parameter = μ	$P(Z \leq Z^*)$ $P(T \leq t^*)$	$2 * P(Z \geq  Z^* )$ $2 * P(T \geq  t^* )$	$P(Z \geq Z^*)$ $P(T \geq t^*)$
Rejection Region (shaded area)			
parameter = p	$-Z_\alpha$	$-Z_{\alpha/2}$ $Z_{\alpha/2}$	$Z_\alpha$
parameter = μ			
	$-t_{\alpha,df}$ df=n-1	$-t_{\frac{\alpha}{2},df}$ $t_{\frac{\alpha}{2},df}$ df=n-1	$t_{\alpha,df}$ df=n-1



## Step 4: Result

p-value comparison	Result
If $p \leq \alpha$	Reject $H_0$
If $p > \alpha$	Fail to Reject $H_0$





# Step 5: Conclusion

Result	Evidence in favor of $H_a$ /claim?	Statistical Significance
<b>Reject <math>H_0</math></b>	Sufficient Evidence (to suggest claim shown in $H_a$ may be valid)	Statistically Significant
<b>Fail to Reject <math>H_0</math></b>	Insufficient Evidence (or not sufficient evidence) (to suggest claim shown in $H_a$ may be valid)	Statistically Insignificant (or not statistically significant)



# Example (not one of the assignment tests)

If we wanted to test whether or not less than half of the students at Bayview cheated, using an exact binomial test...

- Step 1:  $H_0: p = 0.5$   $H_a: p < 0.5$
- Step 2: test statistic (*no test statistic with exact test*)
- Step 3: p-value: 0.0567
- Step 4: result (tested at  $\alpha=0.05$ )
  - P-value >  $\alpha$
  - $0.0567 > 0.05 \Rightarrow$  Fail to Reject  $H_0$
- Step 5: conclusion
  - There is insufficient evidence to suggest that the proportion of Bayview students who cheated is significantly less than half.

The FREQ Procedure

Cheated	Frequency	Percent
1	37	41.11
2	53	58.89

Binomial Proportion	
Cheated = 1	
Proportion (P)	0.4111
ASE	0.0519
95% Lower Conf Limit	0.3095
95% Upper Conf Limit	0.5128
Exact Conf Limits	
95% Lower Conf Limit	0.3084
95% Upper Conf Limit	0.5198

Test of H0: Proportion = 0.5	
ASE under H0	0.0527
Z	-1.6865
One-sided Pr < Z	0.0458
Two-sided Pr >  Z	0.0917
Exact Test	
One-sided Pr <= P	0.0567
Two-sided = 2 * One-sided	0.1133

Sample Size = 90

Normal approx.  
test output

Binomial  
exact  
test output

