



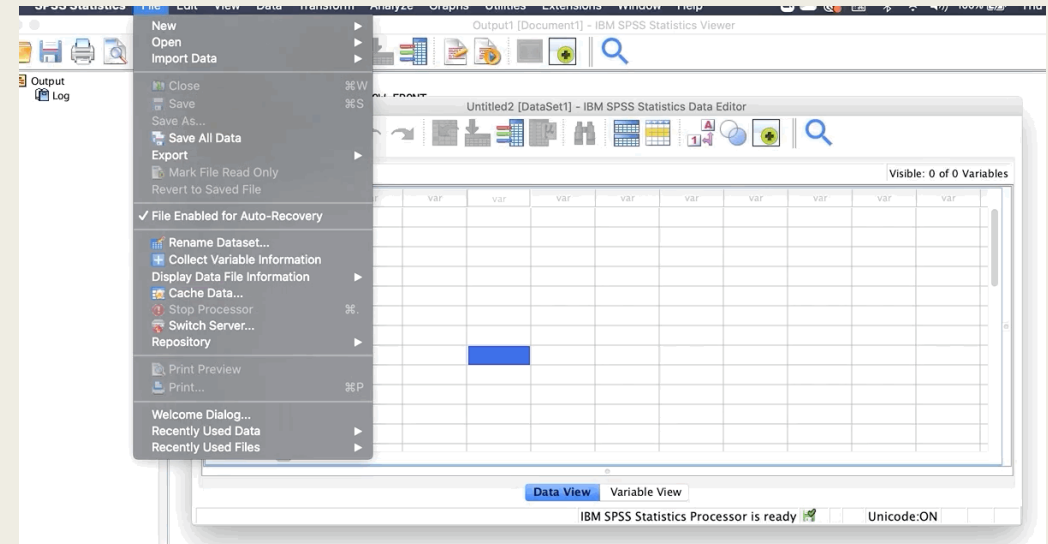
INTRO TO SPSS & DESCRIPTIVE STATISTICS

Marrion Macandog

November 2020

INTRODUCTION

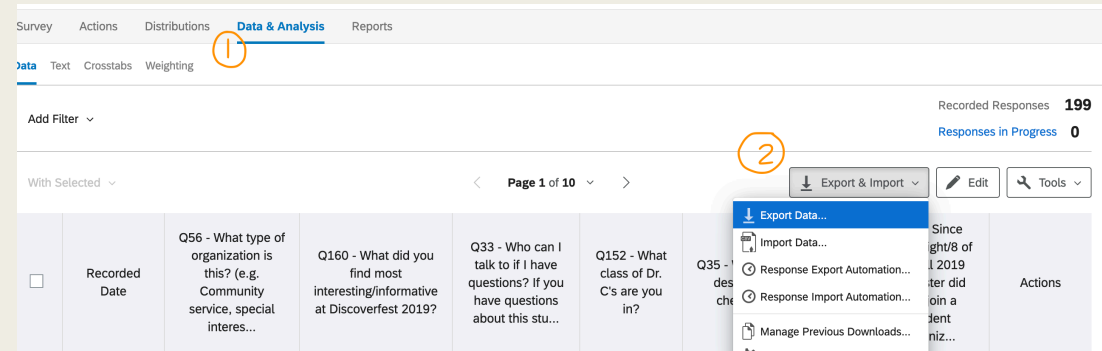
- What is SPSS?
 - *Statistical software for analysis commonly used by Social Scientists*
- Why SPSS?
 - *One of the easier statistical tools to maneuver through*
 - Programming languages don't have drag & drop or a navigation bar!
 - *Valuable technical skill for institutions / companies*
- Key Takeaways on SPSS:
 - *Importing data*
 - *Prep data*
 - *Running basic statistics*



Importing data from Qualtrics into SPSS

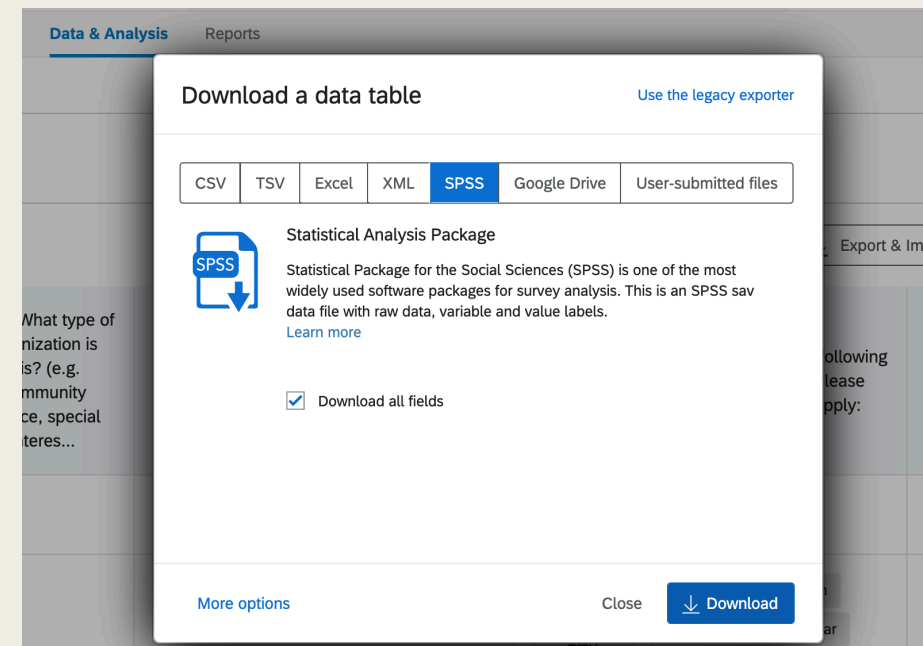
Go to your project on Qualtrics

- Data & Analysis
- Export Data

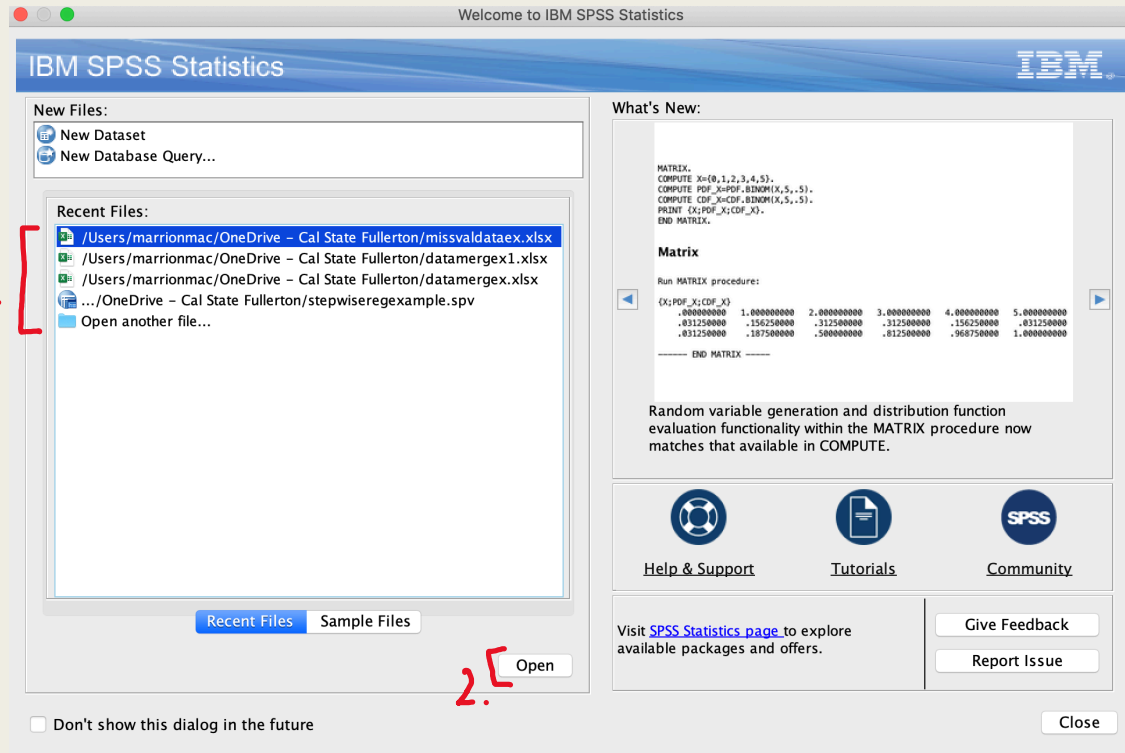


Downloading the data

- Common to use CSV, but we will opt for SPSS since that's the software we'll be using
- SPSS > Download



Opening Datasets in SPSS



1. Choose file under
“Recent Files” or “Open
another file...”

2. Open

Qualtrics Data into SPSS

Data View

Visible: 232 of 232 Variables

	Q33	Q152	Q35_1	Q35_2	Q35_3	Q35_4	Q35_5	Q35_6	Q35_7	Q35_8	Q35_9
1	4	1
2	4	2	.	1	1
3	4	1	1	.	1
4	5	1	.	1
5	4	1	.	1	1
6	4	1	1	.	1	1
7	4	1	.	1
8	4	1	.	1	1
9	4	2
10	4	2	.	1
11	4	1	1
12	4	1	1	.	1	1

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Q33	Numeric	40	0	Who can I talk t...	{4, agree}...	None	6	Right	Scale
2	Q152	Numeric	40	0	What class of D...	{1, T/TH me...	None	8	Right	Scale
3	Q35_1	Numeric	40	0	Which of the fol...	{1, First gen...	None	8	Right	Scale
4	Q35_2	Numeric	40	0	Which of the fol...	{1, Second g...	None	7	Right	Scale
5	Q35_3	Numeric	40	0	Which of the fol...	{1, Commuti...	None	7	Right	Scale
6	Q35_4	Numeric	40	0	Which of the fol...	{1, Commuti...	None	7	Right	Scale
7	Q35_5	Numeric	40	0	Which of the fol...	{1, Commuti...	None	7	Right	Scale
8	Q35_6	Numeric	40	0	Which of the fol...	{1, Commuti...	None	7	Right	Scale
9	Q35_7	Numeric	40	0	Which of the fol...	{1, Recrui...	None	7	Right	Scale
10	Q35_8	Numeric	40	0	Which of the fol...	{1, Internati...	None	7	Right	Scale
11	Q35_9	Numeric	40	0	Which of the fol...	{1, Receivin...	None	9	Right	Scale
12	Q35_10	Numeric	40	0	Which of the fol...	{1, Coming ...	None	8	Right	Scale
13	Q35_11	Numeric	40	0	Which of the fol...	{1, Transfer ...	None	8	Right	Scale
14	Q35_12	Numeric	40	0	Which of the fol...	{1, Transfer ...	None	5	Right	Scale

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

Wow so easy....

Prepping data in SPSS

- Recode missing values
- Specifying “Measure”
- Merging data

Prepping data in SPSS: Missing Values for Binary Questions

Need to account for missing values so our analysis is accurate

3. If you checked "yes", please state what other language(s) you can speak: (Mark all that applies.)

Q3_1 ☐ Spanish
Q3_2 ☐ Vietnamese 1 = marked
Q3_3 ☐ Chinese
Q3_4 ☐ Korean
Q3_5 ☐ Other:
Q3_5_TEXT string

Example dataset -

The screenshot shows the IBM SPSS Statistics Data Editor window. The dataset is named 'exampledata.sav [DataSet1]'. The variable list at the top includes Q1, Q2, Q3_1, Q3_2, Q3_3, Q3_4, Q3_5_TEXT, Q4, Q5_1_TEXT, Q5_1, and Q6. The data grid shows 12 rows. Rows 1-5 have data for Q1, Q2, Q3_1, Q3_2, Q3_3, Q3_4, Q3_5_TEXT, Q4, Q5_1_TEXT, Q5_1, and Q6. Rows 6-12 are empty. A yellow highlight is placed over the Q3_1, Q3_2, Q3_3, and Q3_4 columns for rows 1-5, with the text 'Missing data for Q3' overlaid. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'.

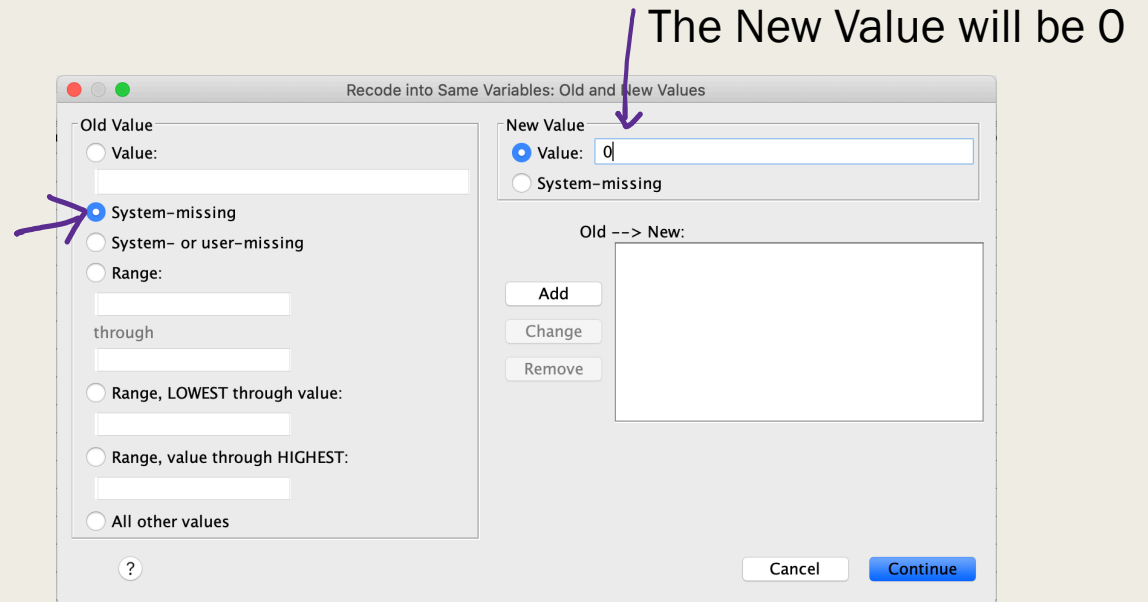
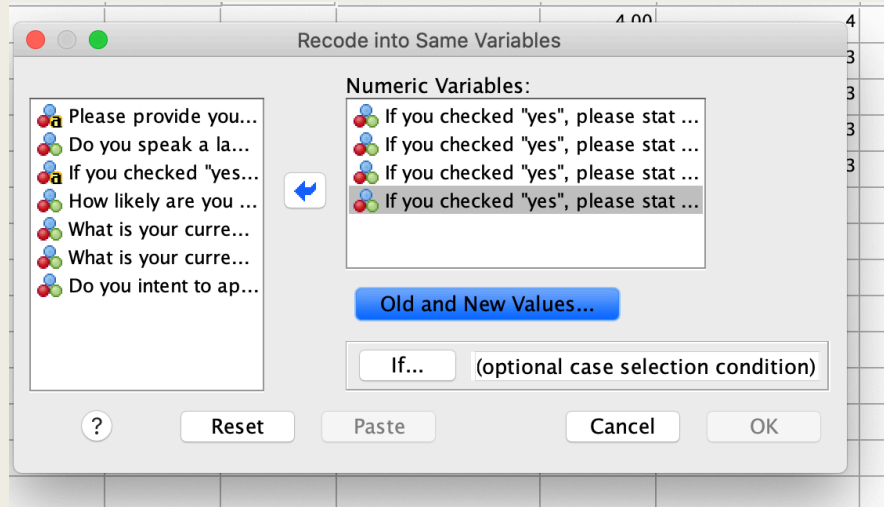
	Q1	Q2	Q3_1	Q3_2	Q3_3	Q3_4	Q3_5_TEXT	Q4	Q5_1_TEXT	Q5_1	Q6
1	007	1.00	1.00					4.00	3.61		1.00
2	008	1.00		1.00				5.00	3.21		1.00
3	009	.00						4.00	2.78		1.00
4	010	1.00	1.00					5.00	3.40		1.00
5	011	1.00	1.00					4.00	3.10		1.00
6											
7											
8											
9											
10											
11											
12											

SPSS assumes that the participant did not answer the question (hence, missing):

Statistics					
	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Spanish	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Vietnamese	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Chinese	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Korean	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Other - Text
N	Valid 3	1	0	0	5
	Missing 2	4	5	5	0
Mean	1.0000	1.0000			
Std. Deviation	.00000				

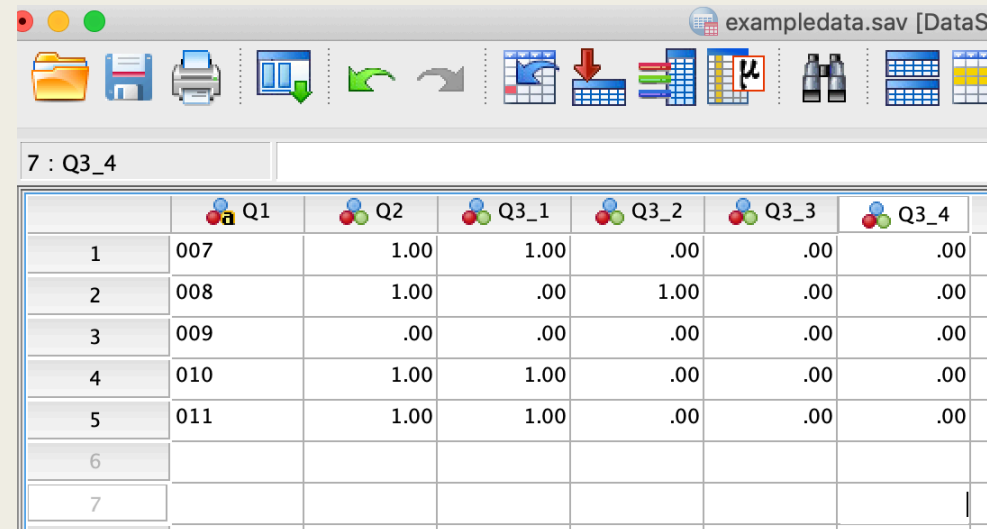
Prepping data in SPSS: Missing Values for Binary Questions

Transform > Recode into Same Variables



Data set with recoded missing variables:

Either 0 or 1, they either speak the language (1) or they don't (0)



The SPSS Data Editor window for 'exampledata.sav'. The title bar shows the file name and '[DataS]'. Below the title bar is a toolbar with various icons. The main area shows a data grid. The first column is labeled '7 : Q3_4'. The subsequent columns are labeled with variable icons and names: Q1, Q2, Q3_1, Q3_2, Q3_3, and Q3_4. The data rows are numbered 1 through 7. The values in the columns are binary (0 or 1) or missing (.00).

	Q1	Q2	Q3_1	Q3_2	Q3_3	Q3_4
1	007	1.00	1.00	.00	.00	.00
2	008	1.00	.00	1.00	.00	.00
3	009	.00	.00	.00	.00	.00
4	010	1.00	1.00	.00	.00	.00
5	011	1.00	1.00	.00	.00	.00
6						
7						

Prepping data in SPSS: Missing Values for Text Responses

Example: Q3_5_TEXT

	Name	Type	Width	Decimals	Label	Values	Missing
7	Q3_5_TEXT	String	8	0	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Other – Text	None	None ...

Missing Values

☐ No missing values

☒ Discrete missing values

☐ Range plus one optional discrete missing value

Low: High:

Discrete value:

Click the “...” in under the Missing column

For missing text responses, we traditionally use -9 at C-REAL

Prepping data: Adjusting our “Measure” column

The 3 options are Scale, Ordinal, and Nominal:

Scale: values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate

Example: score of a student in SAT exam

Ordinal: values represent categories with ranking

Example: 1=Highly satisfied, 2=satisfied, 3= neutral, 4= dissatisfied, 5= highly dissatisfied

Nominal: values represent categories with no ranking

Example: zip code or gender

The image shows a survey form with several questions. Handwritten purple annotations identify the measurement scale for each question:

- Q1:** "Please provide your student I.D.: string" is labeled as **Nominal** with an arrow.
- Q2:** "Do you speak a language other than English?" with options "Yes 1" and "No 0" is labeled as **Nominal** with a bracket.
- Q3:** "If you checked 'yes', please state what other language(s) you can speak: (Mark all that applies.)" with options "Q3_1 Spanish", "Q3_2 Vietnamese 1 = marked", "Q3_3 Chinese", "Q3_4 Korean", and "Q3_5 Other: string" is labeled as **Nominal** with a bracket.
- Q4:** "How likely are you to go to college? (Mark one.)" with options "Extremely unlikely 1", "Unlikely 2", "Neutral 3", "Likely 4", and "Extremely likely 5" is labeled as **Ordinal** with a bracket.
- Q5:** "What is your current GPA for this semester?" with options "Q5_1 TEXT numeric" and "Q5_1 I do not know 0" is labeled as **Scale** with an arrow.
- Q6:** "Do you intend to apply to college?" with options "Yes 1" and "No 0" is labeled as **Nominal** with a bracket.

At the bottom, a table shows the "Measure" column for various items, with a dropdown menu open showing the options: Scale, Ordinal, and Nominal.

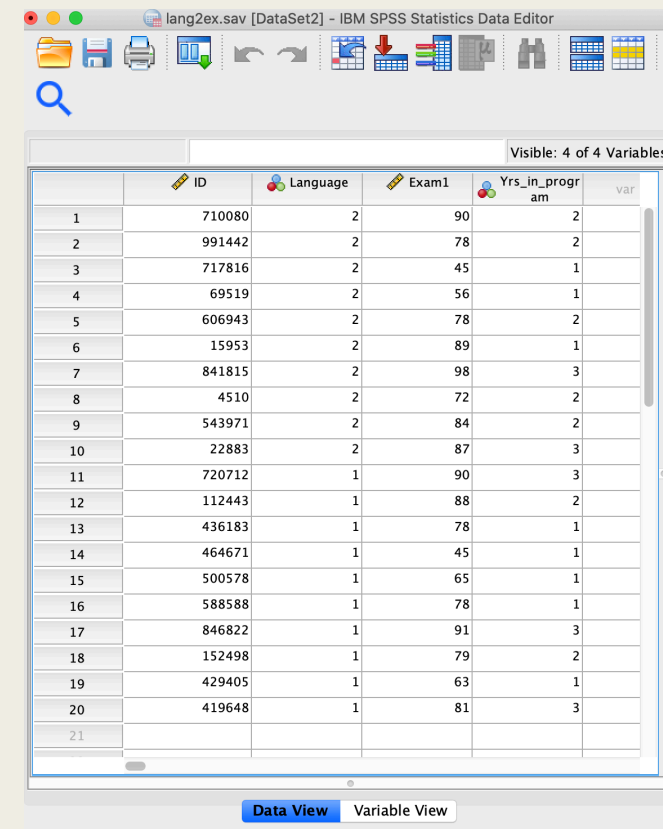
Items	Align	Measure	Role
	Left	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Left	Scale	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input
	Right	Nominal	Input

Prepping data: Merging

- Combining pre/post test data

	ID	Attendance	Score_posttest	Score_pretest	var
1	9975309	0	53	48	
2	9948278	1	80	75	
3	9169427	4	95	90	
4	8643086	3	95	90	
5	7988957	1	79	72	
6	7509550	1	67	61	
7	6133087	2	91	86	
8	6115282	2	89	84	
9	5396215	2	86	81	
10	4365151	3	88	83	
11	4303020	1	83	78	
12	3670778	1	63	56	
13	3316521	0	57	54	
14	3168177	0	67	66	
15	2331702	1	61	56	
16	1767276	0	51	46	
17	1328727	1	71	65	
18	1323956	3	93	88	
19	926172	0	50	48	
20	661007	2	85	79	

- Want to combine datasets



lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor

Visible: 4 of 4 Variables

	ID	Language	Exam1	Yrs_in_program	var
1	710080	2	90	2	
2	991442	2	78	2	
3	717816	2	45	1	
4	69519	2	56	1	
5	606943	2	78	2	
6	15953	2	89	1	
7	841815	2	98	3	
8	4510	2	72	2	
9	543971	2	84	2	
10	22883	2	87	3	
11	720712	1	90	3	
12	112443	1	88	2	
13	436183	1	78	1	
14	464671	1	45	1	
15	500578	1	65	1	
16	588588	1	78	1	
17	846822	1	91	3	
18	152498	1	79	2	
19	429405	1	63	1	
20	419648	1	81	3	
21					

Data View Variable View

Prepping data: Merging different datasets

1. Have both datasets open
2. Make sure matching variables have the same settings under “Variable View”

	Name	Type	Width	Decimals
1	ID	Numeric	18	0
2	Language	Numeric	1	0
3	Exam1	Numeric	2	0
4	Yrs_in_prog...	Numeric	1	0

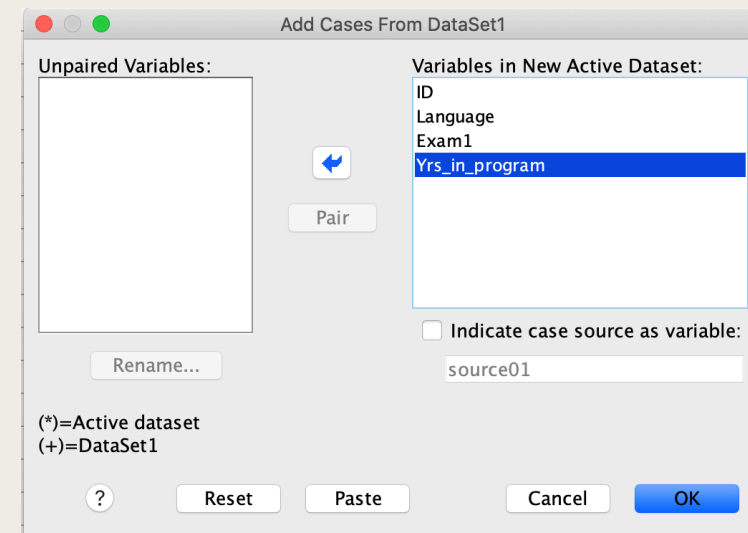
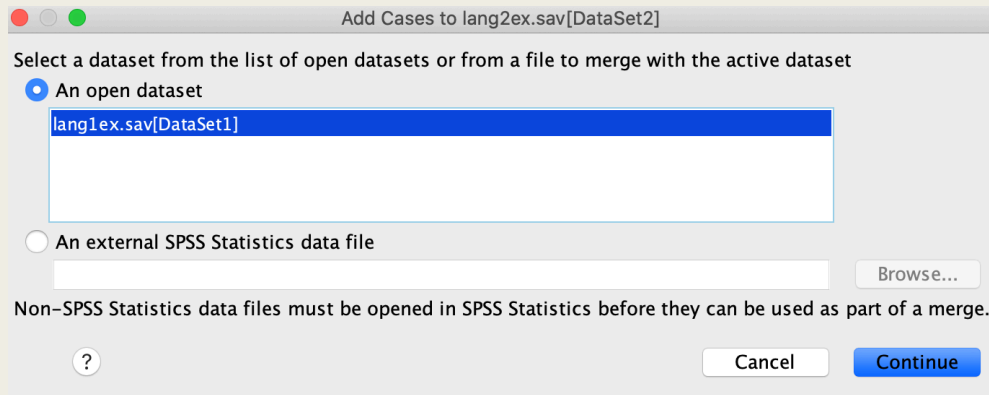
	Name	Type	Width	Decimals	Label
1	ID	Numeric	18	0	
2	Language	Numeric	1	0	
3	Exam1	Numeric	2	0	
4	Yrs_in_prog...	Numeric	1	0	

lang1ex.sav [DataSet1] - IBM SPSS Statistics Data Editor					
Visible: 4 of 4 Variables					
	ID	Language	Exam1	Yrs_in_prog am	var
1	720712	1	90	3	
2	112443	1	88	2	
3	436183	1	78	1	
4	464671	1	45	1	
5	500578	1	65	1	
6	588588	1	78	1	
7	846822	1	91	3	
8	152498	1	79	2	
9	429405	1	63	1	
10	419648	1	81	3	
11					
12					
13					
14					
15					
16					
17					

lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor					
Visible: 4 of 4 Variables					
	ID	Language	Exam1	Yrs_in_prog am	var
1	710080	2	90	2	
2	991442	2	78	2	
3	717816	2	45	1	
4	69519	2	56	1	
5	606943	2	78	2	
6	15953	2	89	1	
7	841815	2	98	3	
8	4510	2	72	2	
9	543971	2	84	2	
10	22883	2	87	3	
11					
12					
13					
14					
15					
16					
17					

Prepping data: Merging different datasets

Data > Merge > Add Cases



Prepping data: Merging different datasets

Before:

	ID	Language	Exam1	Yrs_in_program	var
1	720712	1	90	3	
2	112443	1	88	2	
3	436183	1	78	1	
4	464671	1	45	1	
5	500578	1	65	1	
6	588588	1	78	1	
7	846822	1	91	3	
8	152498	1	79	2	
9	429405	1	63	1	
10	419648	1	81	3	

	ID	Language	Exam1	Yrs_in_program	var
1	710080	2	90	2	
2	991442	2	78	2	
3	717816	2	45	1	
4	69519	2	56	1	
5	606943	2	78	2	
6	15953	2	89	1	
7	841815	2	98	3	
8	4510	2	72	2	
9	543971	2	84	2	
10	22883	2	87	3	

After:

	ID	Language	Exam1	Yrs_in_program	var
1	710080	2	90	2	
2	991442	2	78	2	
3	717816	2	45	1	
4	69519	2	56	1	
5	606943	2	78	2	
6	15953	2	89	1	
7	841815	2	98	3	
8	4510	2	72	2	
9	543971	2	84	2	
10	22883	2	87	3	
11	720712	1	90	3	
12	112443	1	88	2	
13	436183	1	78	1	
14	464671	1	45	1	
15	500578	1	65	1	
16	588588	1	78	1	
17	846822	1	91	3	
18	152498	1	79	2	
19	429405	1	63	1	
20	419648	1	81	3	
21					

Prepping data: Merging pre/post test data

1. Have both datasets open
2. Make sure there's an identification variable; variable we will use to match the two datasets together
 - In this example, we have ID as the matching variable

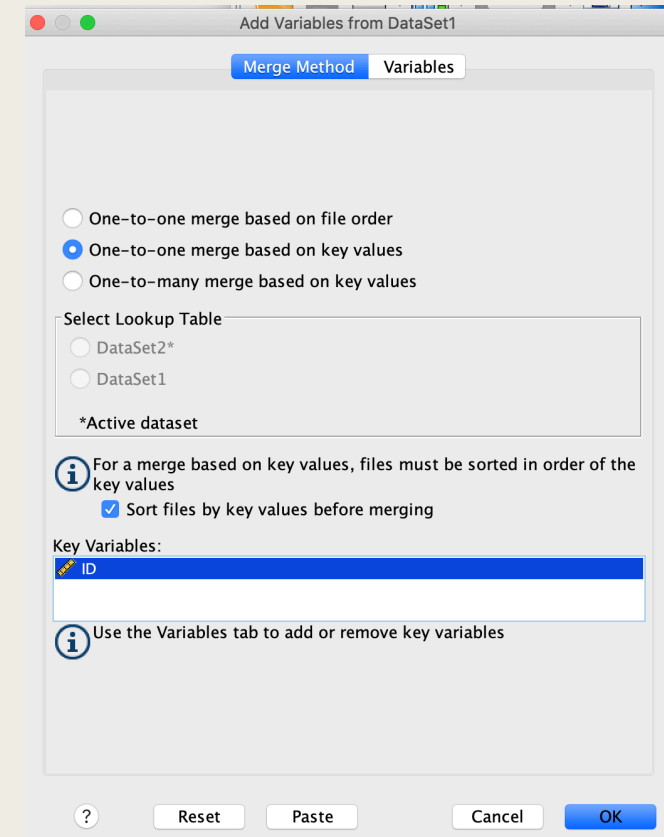
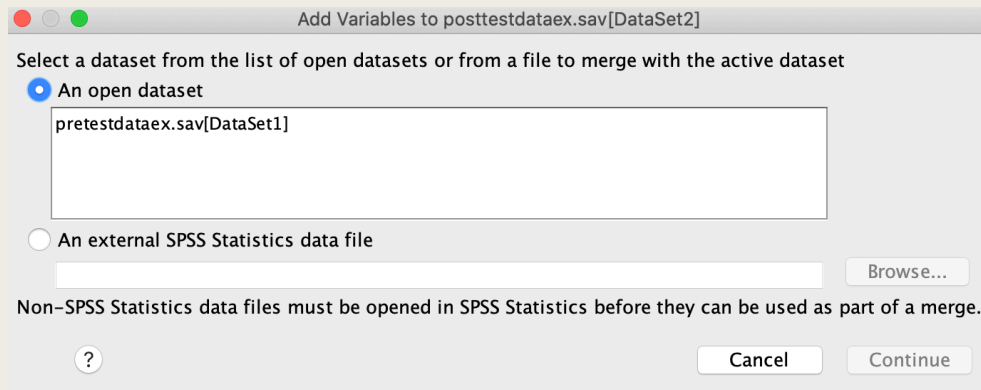
The image displays two side-by-side screenshots of the IBM SPSS Statistics Data View. The left window shows a dataset with three columns: ID, Attendance, and Score_pretest. The right window shows a dataset with three columns: ID, Score_posttest, and var. Both windows show 22 rows of data.

	ID	Attendance	Score_pretest
1	9975309	0	48
2	9948278	1	75
3	9169427	4	90
4	8643086	3	90
5	7988957	1	72
6	7509550	1	61
7	6133087	2	86
8	6115282	2	84
9	5396215	2	81
10	4365151	3	83
11	4303020	1	78
12	3670778	1	56
13	3316521	0	54
14	3168177	0	66
15	2331702	1	56
16	1767276	0	46
17	1328727	1	65
18	1323956	3	88
19	926172	0	48
20	661007	2	79
21			
22			

	ID	Score_posttest	var
1	9975309	53	
2	9948278	80	
3	9169427	95	
4	8643086	95	
5	7988957	79	
6	7509550	67	
7	6133087	91	
8	6115282	89	
9	5396215	86	
10	4365151	88	
11	4303020	83	
12	3670778	63	
13	3316521	57	
14	3168177	67	
15	2331702	61	
16	1767276	51	
17	1328727	71	
18	1323956	93	
19	926172	50	
20	661007	85	
21			
22			

Prepping data: Merging pre/post test data

Data > Merge Files > Add Variables



Prepping data: Merging pre/post test data

Before:

Visible: 2 of 2				
	ID	Attendance	Score_pretest	var
1	9975309	0	48	
2	9948278	1	75	
3	9169427	4	90	
4	8643086	3	90	
5	7988957	1	72	
6	7509550	1	61	
7	6133087	2	86	
8	6115282	2	84	
9	5396215	2	81	
10	4365151	3	83	
11	4303020	1	78	
12	3670778	1	56	
13	3316521	0	54	
14	3168177	0	66	
15	2331702	1	56	
16	1767276	0	46	
17	1328727	1	65	
18	1323956	3	88	
19	926172	0	48	
20	661007	2	79	
21				
22				

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

After:

Visible: 4 of 4 V						
	ID	Score_posttest	Attendance	Score_pretest	var	var
1	661007	85	2	79		
2	926172	50	0	48		
3	1323956	93	3	88		
4	1328727	71	1	65		
5	1767276	51	0	46		
6	2331702	61	1	56		
7	3168177	67	0	66		
8	3316521	57	0	54		
9	3670778	63	1	56		
10	4303020	83	1	78		
11	4365151	88	3	83		
12	5396215	86	2	81		
13	6115282	89	2	84		
14	6133087	91	2	86		
15	7509550	67	1	61		
16	7988957	79	1	72		
17	8643086	95	3	90		
18	9169427	95	4	90		
19	9948278	80	1	75		
20	9975309	53	0	48		
21						
22						
23						

IV: Descriptive Statistics

- A descriptive statistic is a summary statistic that quantitatively describes or summarizes features from a collection of information
 - *I.e. Mean (commonly used as average), standard deviation, frequencies, etc..*
- Some questions we can ask:
 - *What is the average GPA of our sample?*
 - *How much of the sample is Extremely likely to go to college?*

Descriptive Statistics

Frequencies

- How many participants speak a language other than English?
- How much of the sample is Extremely likely to go to college?

Descriptives

- What is the average GPA of our sample?

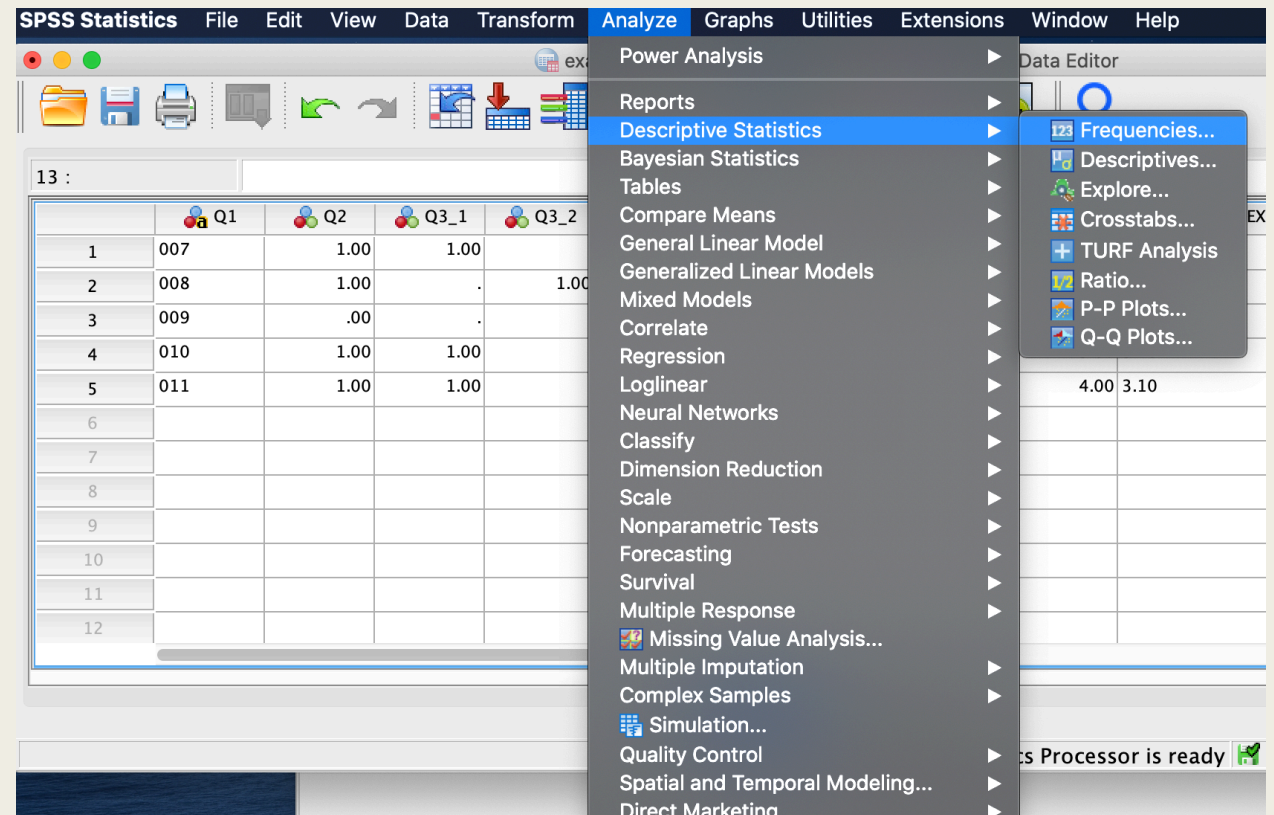
Example dataset:

[illegible]

Answering our questions: Frequencies

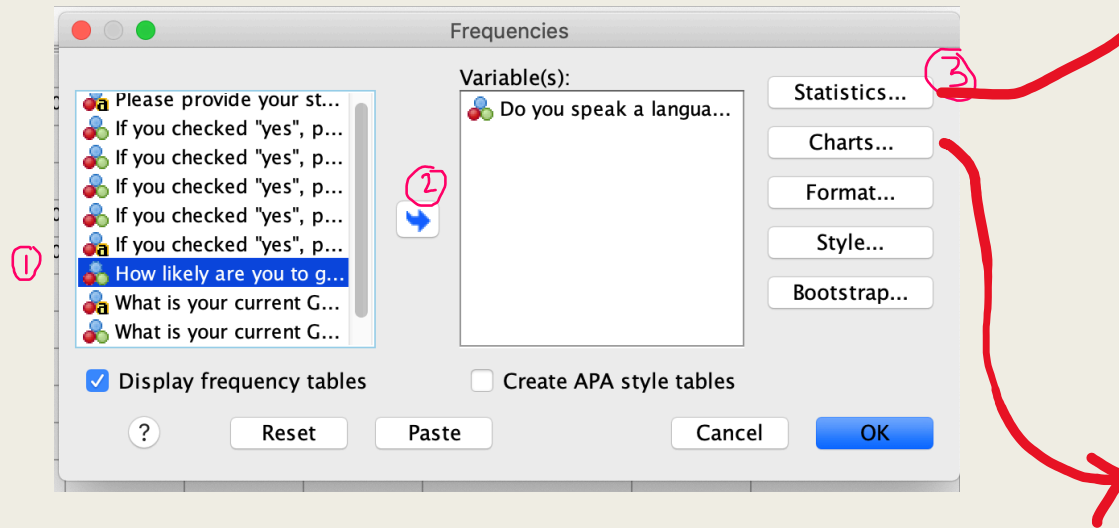
Analyze > Descriptive Statistics > Frequencies

- How many participants speak a language other than English?
- How much of the sample is Extremely likely to go to college?

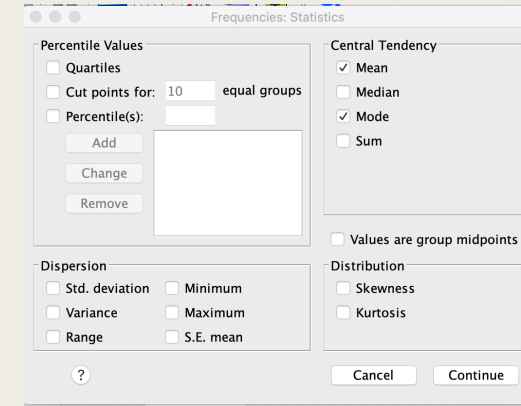


Answering our questions: Frequencies (cont.)

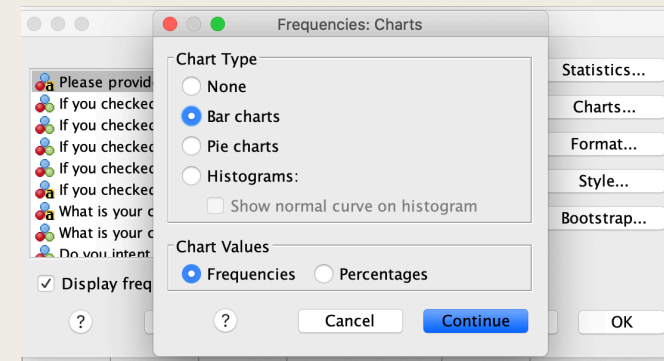
- How many participants speak a language other than English?
- How much of the sample is Extremely likely to go to college?



Statistics



Charts



Our results:

Statistics

→ Frequencies

Statistics

		Do you speak a language other than English?	How likely are you to go to college? (Mark one.)
N	Valid	5	5
	Missing	0	0
Mean		.8000	4.4000
Mode		1.00	4.00

Frequency Table

Do you speak a language other than English?

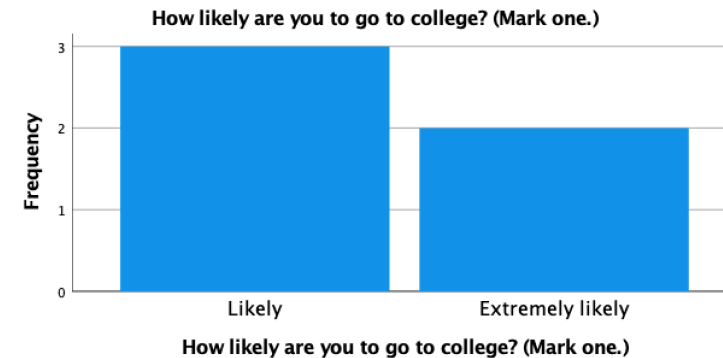
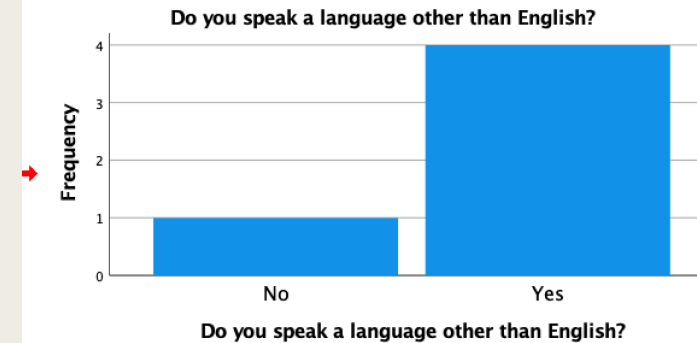
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	1	20.0	20.0	20.0
	Yes	4	80.0	80.0	100.0
Total		5	100.0	100.0	

How likely are you to go to college? (Mark one.)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Likely	3	60.0	60.0	60.0
	Extremely likely	2	40.0	40.0	100.0
Total		5	100.0	100.0	

Charts

Bar Chart



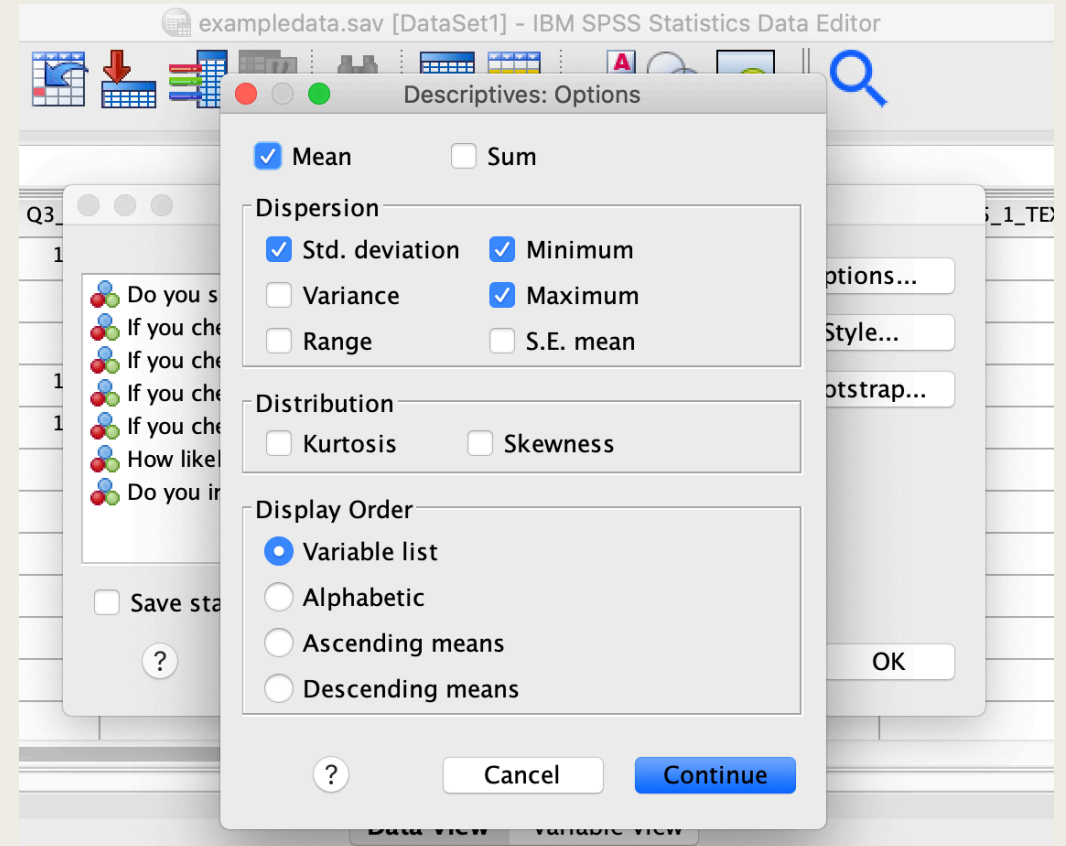
- How many participants speak a language other than English?
 - 80% of the participants in our sample speak a language other than English (N=5)
- How much of the sample is Extremely likely to go to college?
 - 40% of the participants in our sample are Extremely likely to go to college (N=5)

Answering our questions: Descriptives

- What is the average GPA of our sample?

Same path as Frequencies

Analyze > Descriptive Statistics > Descriptives



Our results:

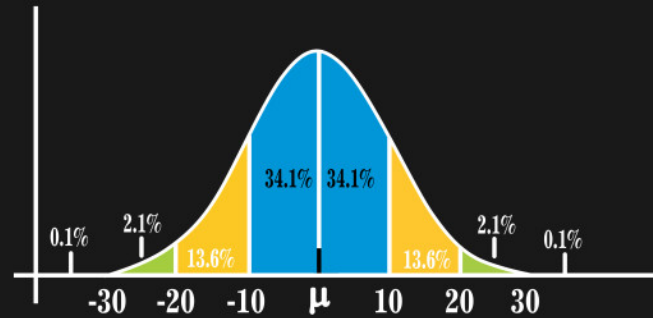
→ Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
What is your current GPA for this semester? – Text	5	3	4	3.22	.313
Valid N (listwise)	5				

- What is the average GPA of our sample?
 - The average GPA of our sample was 3.22 (N=5, SD=.313)

THE END 😊

DON'T BE MEAN



BE ABOVE AVERAGE