SPSS Review

11/09/2020

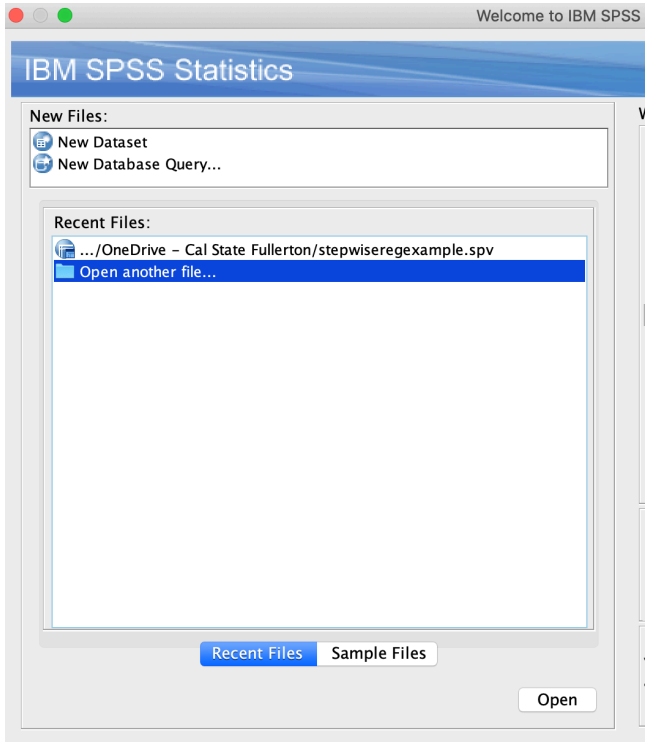
By Marrion Macandog

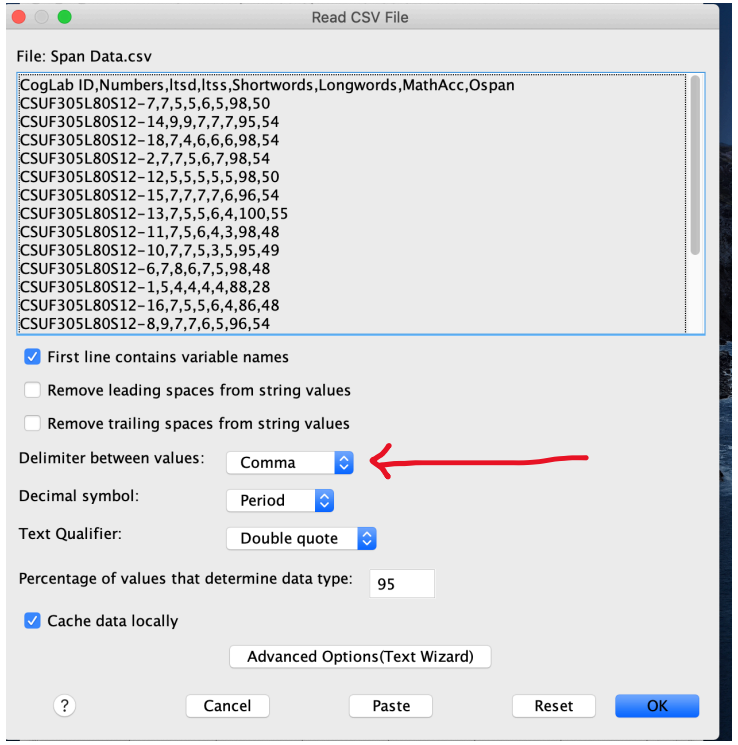
Agenda:

1. Data entry and organization
2. Conducting t-test analyses
3. Multiple linear regression

Importing files into SPSS

Can upload .csv or .xlsx

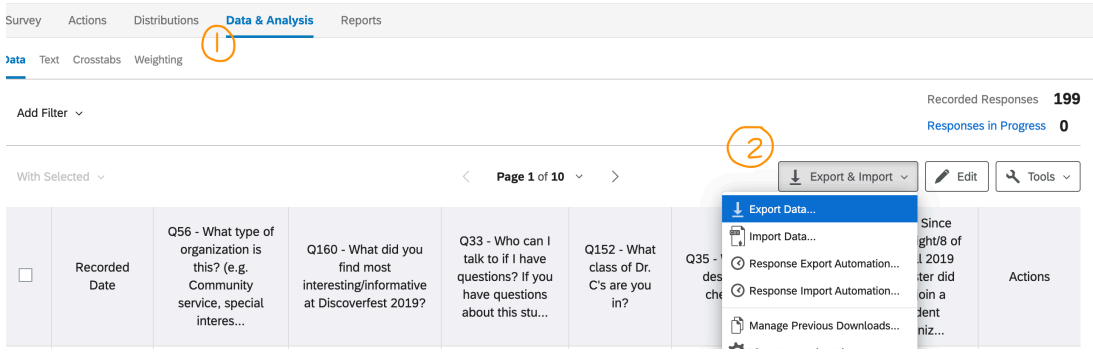




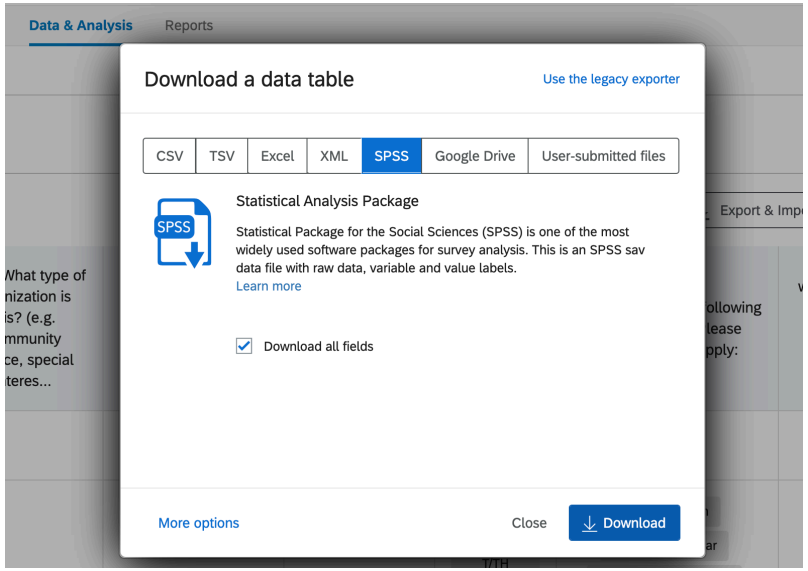
Delimiter: comma

Importing data from Qualtrics into SPSS

Go to your project on Qualtrics
- Data & Analysis
- Export Data

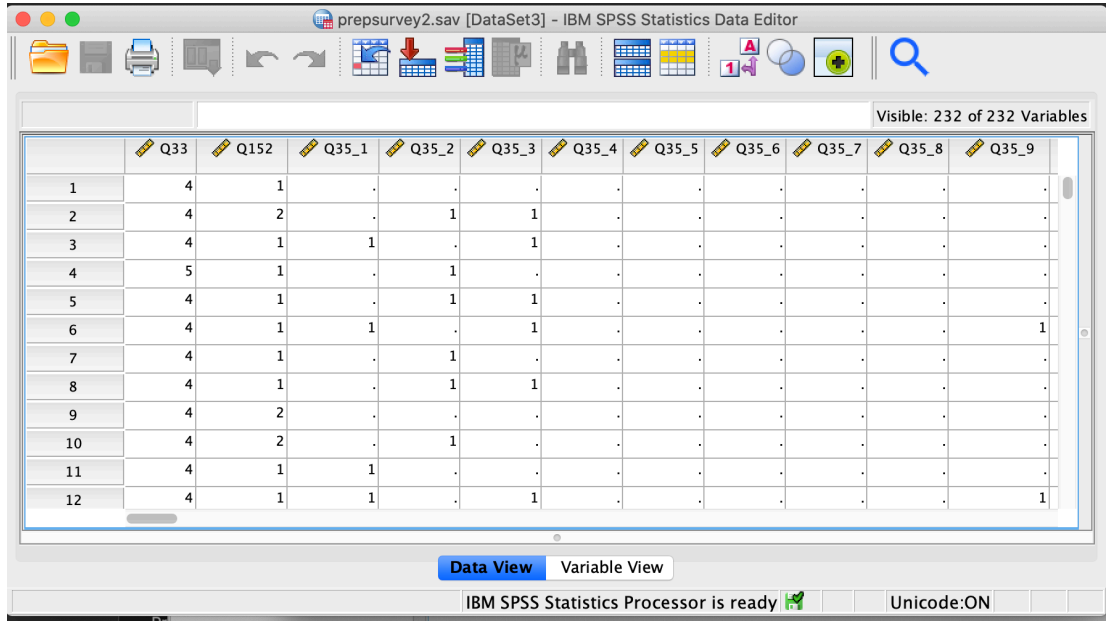


Downloading the data
- Common to use CSV, but we will opt for SPSS since that's the software we'll be using
- SPSS > Download

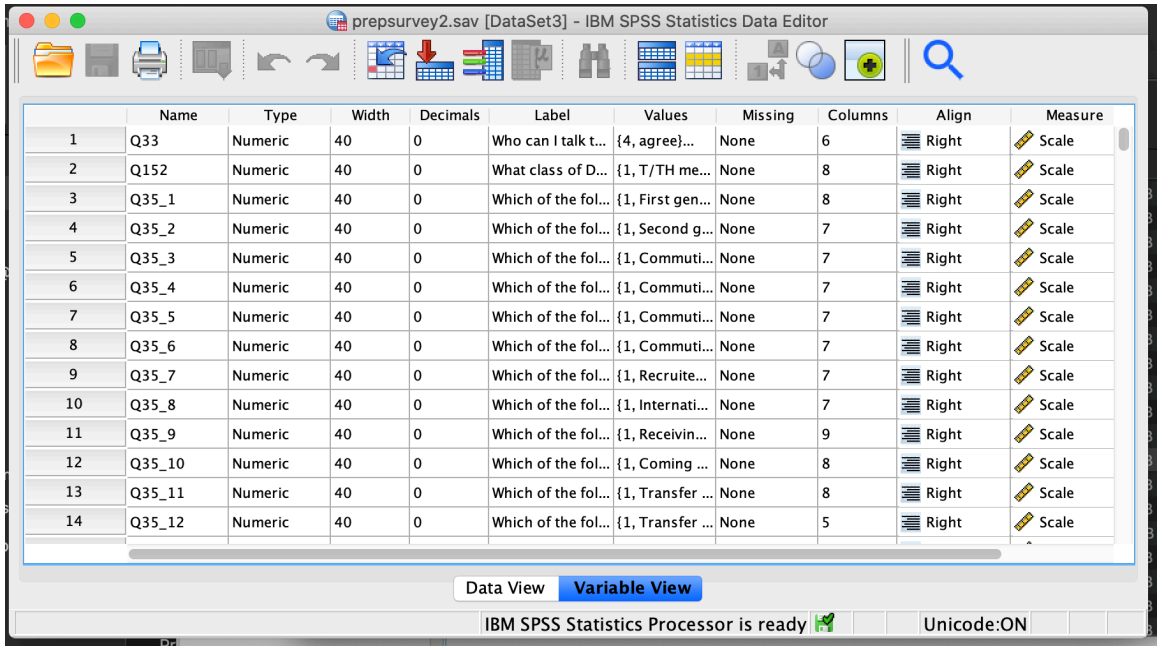


Imported Data into SPSS

Data View



Variable View



Wow so easy….

Prepping data in SPSS

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

Prepping data in SPSS: Missing Values for Multiple Choice Questions

Need to account for missing values so our analysis is accurate

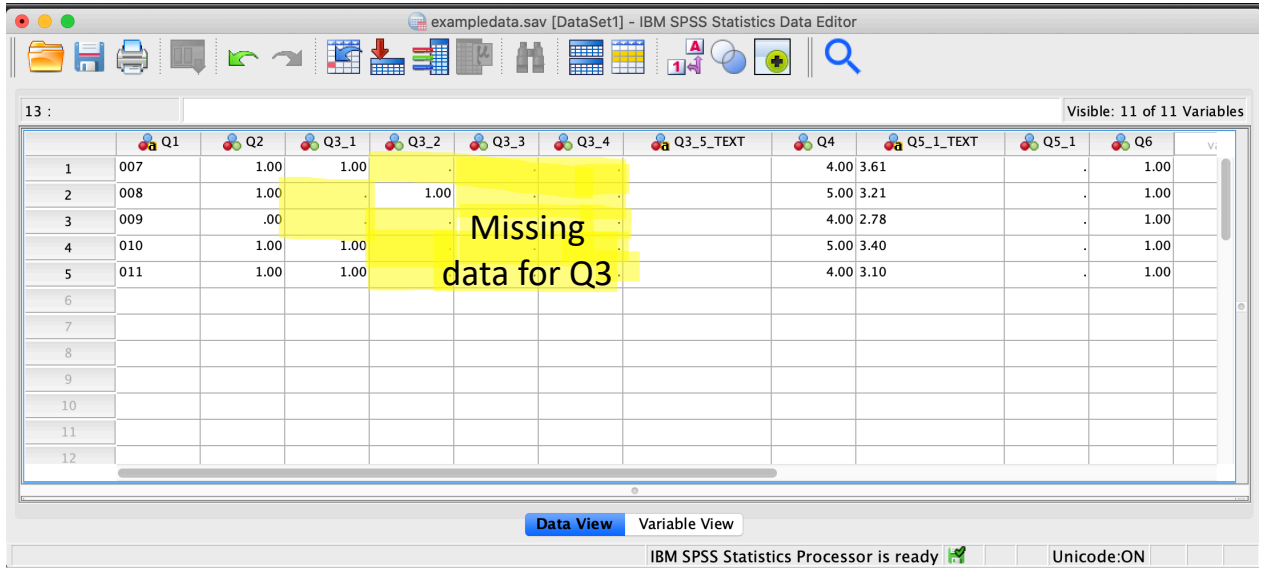
3. If you checked "yes", please state what other language(s) you can speak: (Mark all that applies.)

Q3_1 ☐ Spanish
Q3_2 ☐ Vietnamese
Q3_3 ☐ Chinese
Q3_4 ☐ Korean
Q3_5 ☐ Other:
Q3_5_TEXT _____

1 = marked

string

Example dataset -



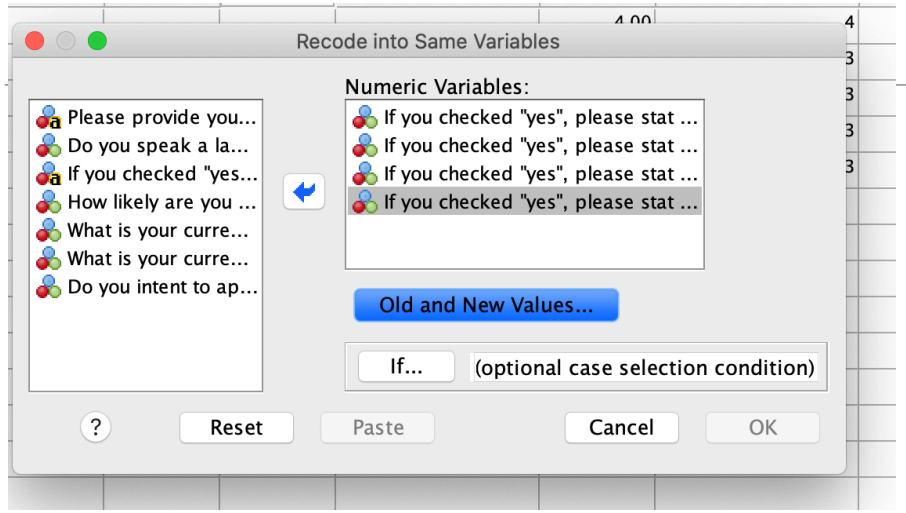
SPSS assumes that the participant did not answer the question (hence, missing):

Statistics

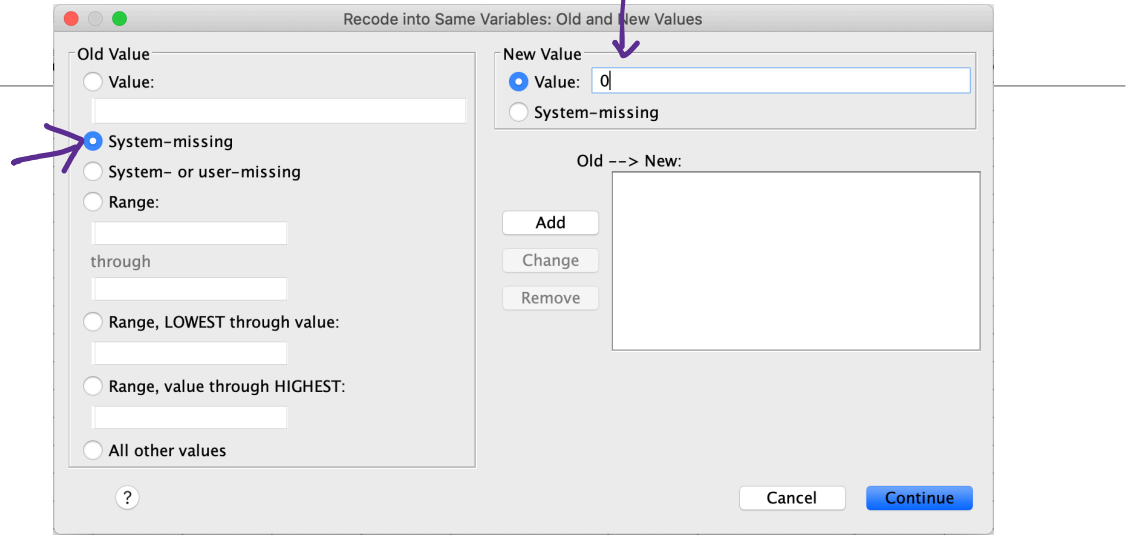
		If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Spanish	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Vietnamese	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): 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 Multiple Choice Questions

Transform > Recode into Same Variables

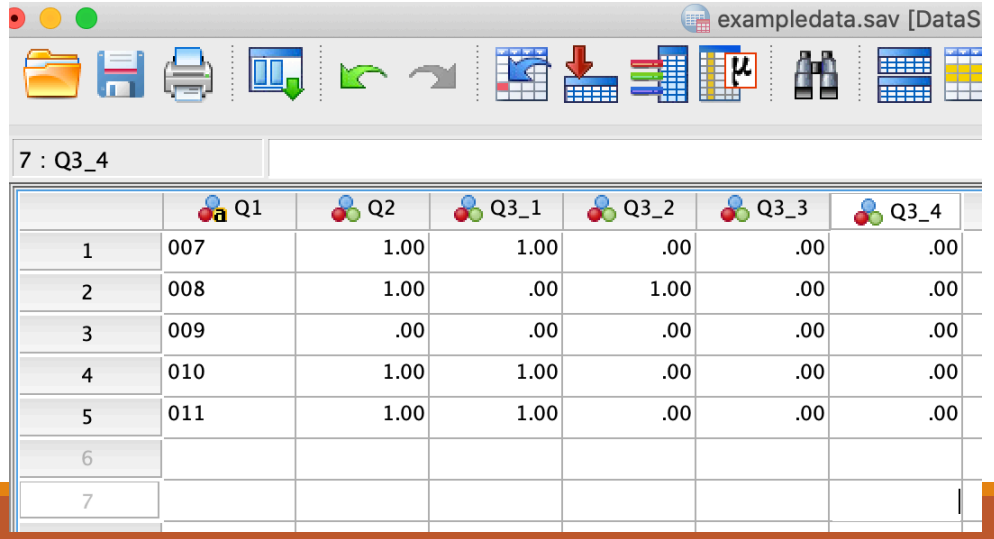


The New Value will be 0



Data set with recoded missing variables:

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

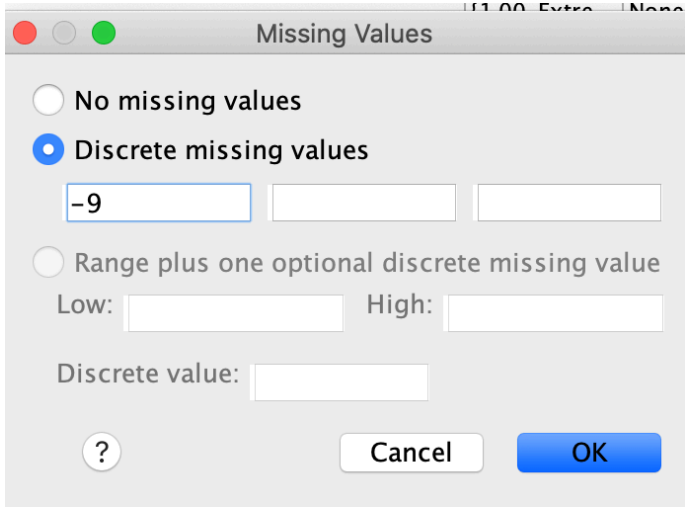


	Q1	Q2	Q3_1	Q3_2	Q3_3	Q3_4
1	007	1.00	1.00	.00	.00	.00
2	008	1.00	.00	1.00	.00	.00
3	009	.00	.00	.00	.00	.00
4	010	1.00	1.00	.00	.00	.00
5	011	1.00	1.00	.00	.00	.00
6						
7						

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 ...	15



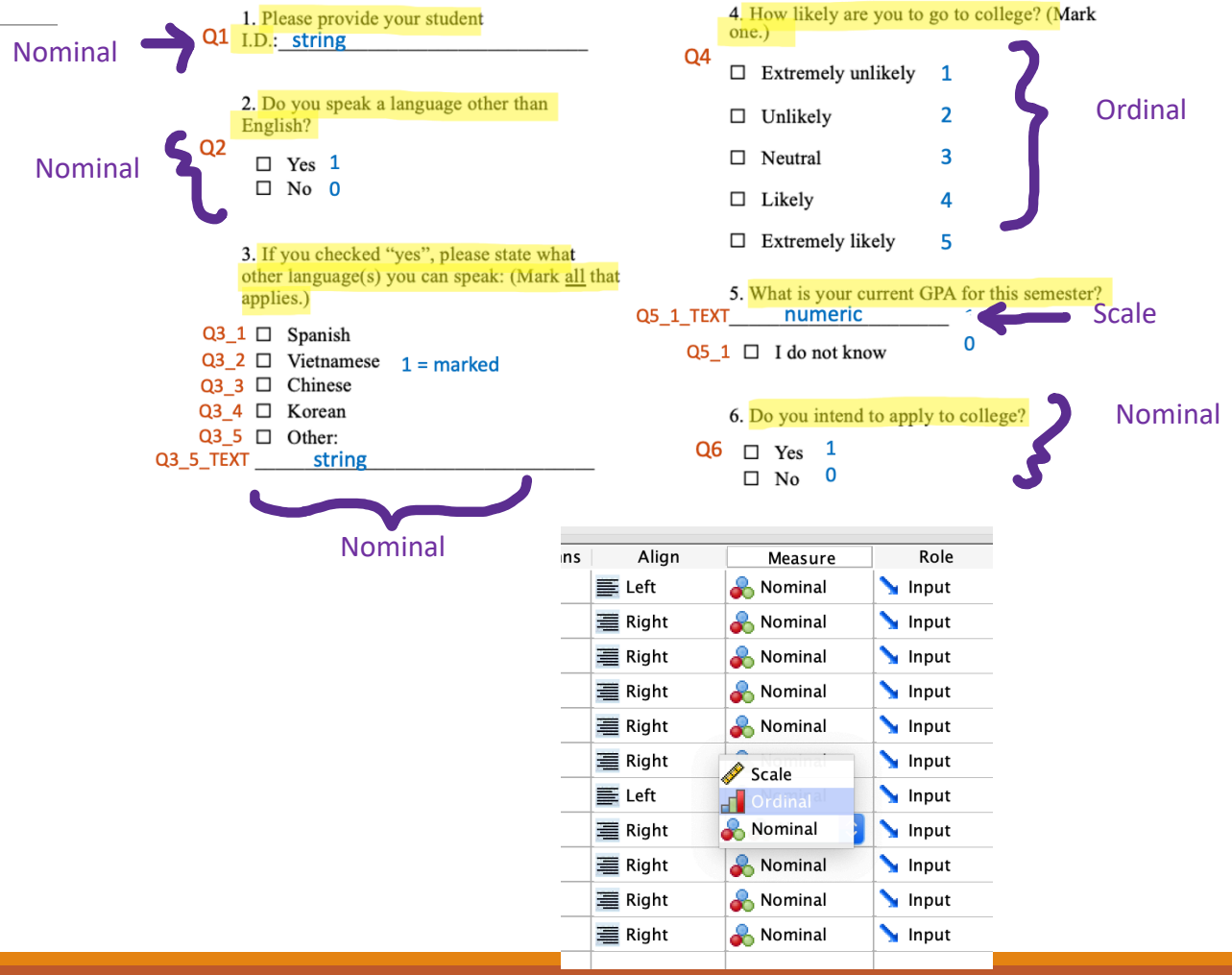
Click the "…" in under the Missing column

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

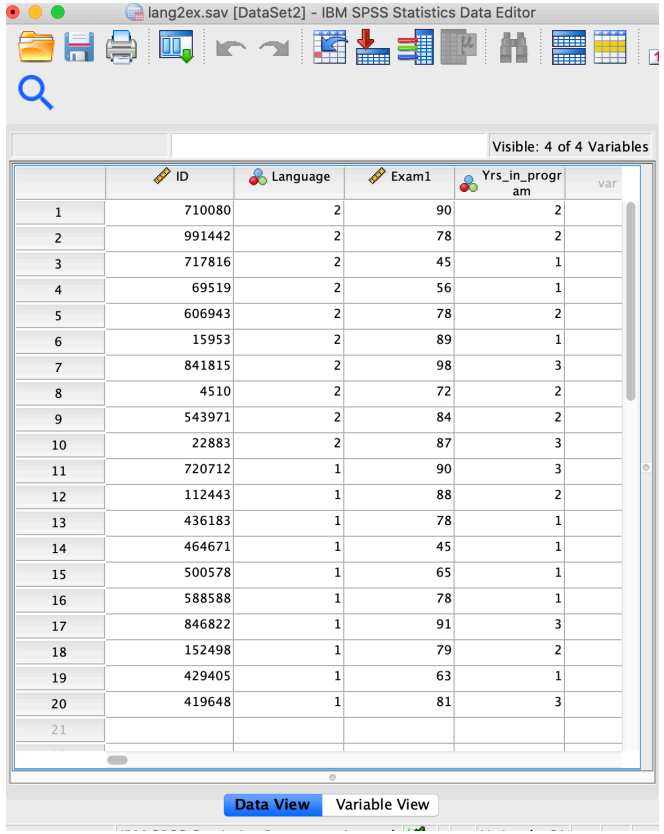


Prepping data: Merging

- 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	

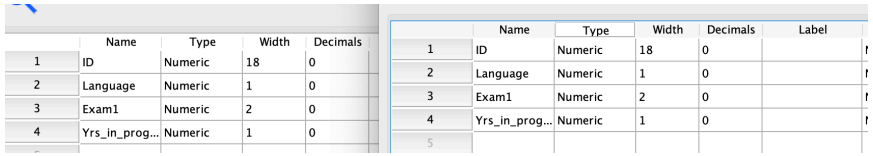
- Want to combine datasets



	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					

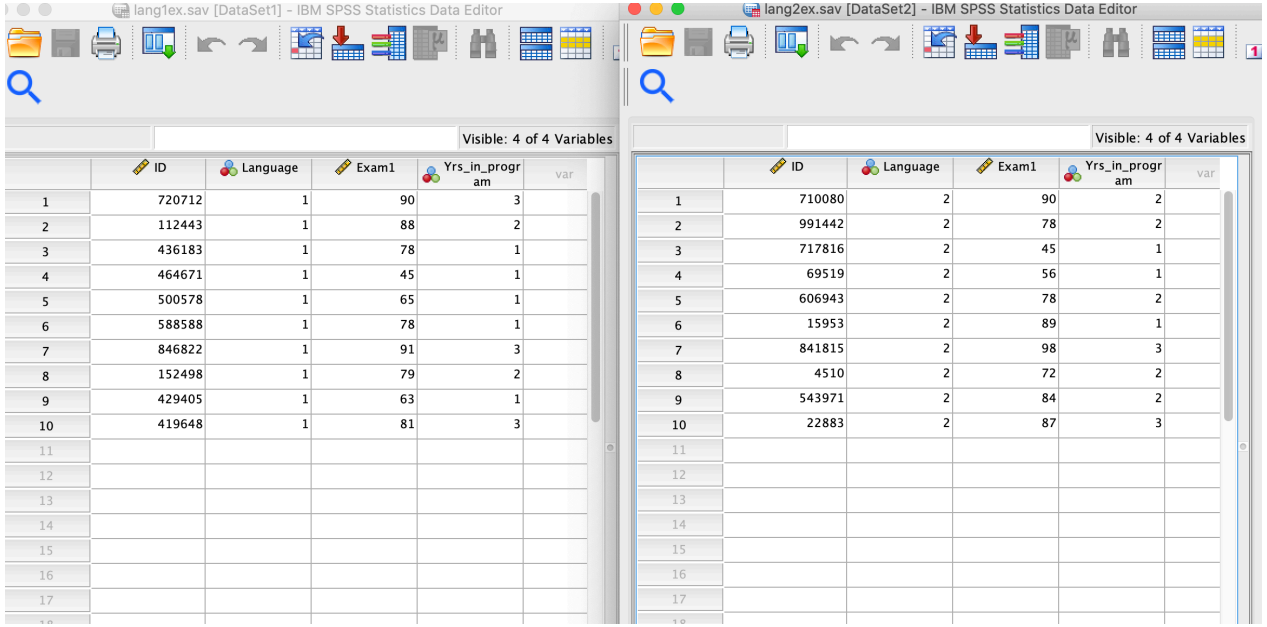
Prepping data: Merging different datasets

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_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	
11					
12					
13					
14					
15					
16					
17					

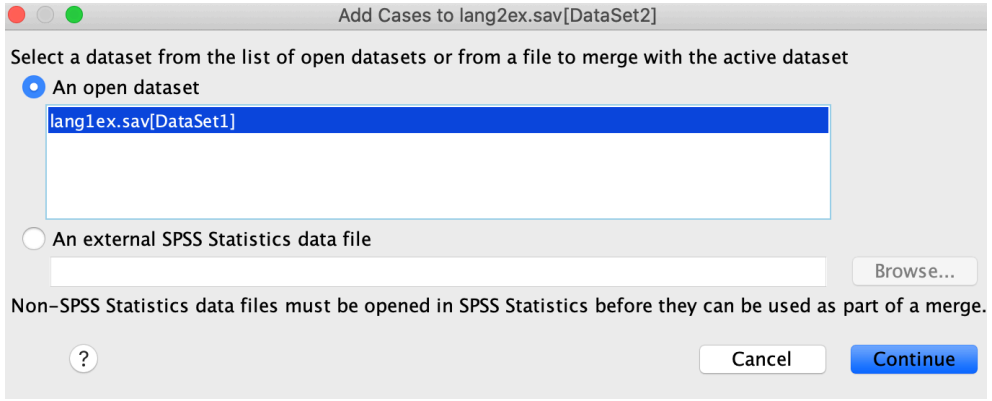
lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor

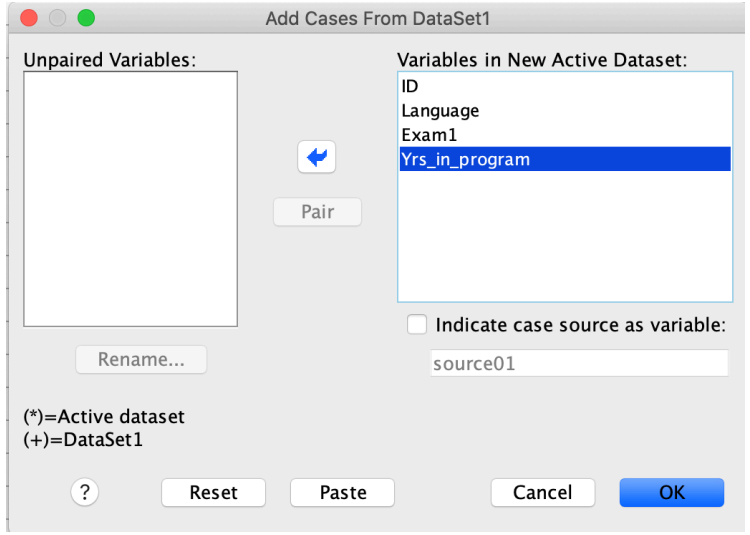
Visible: 4 of 4 Variables

	ID	Language	Exam1	Yrs_in_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					
12					
13					
14					
15					
16					
17					

Prepping data: Merging different datasets

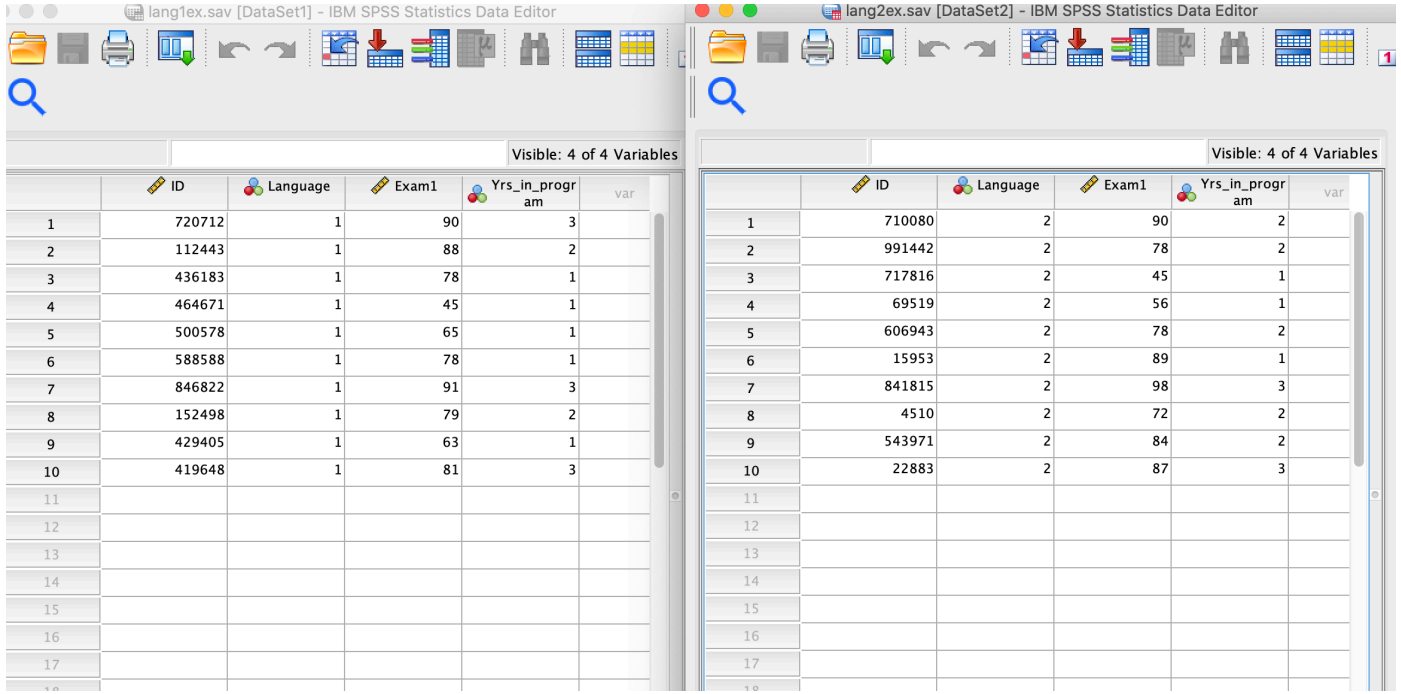
Data > Merge > Add Cases



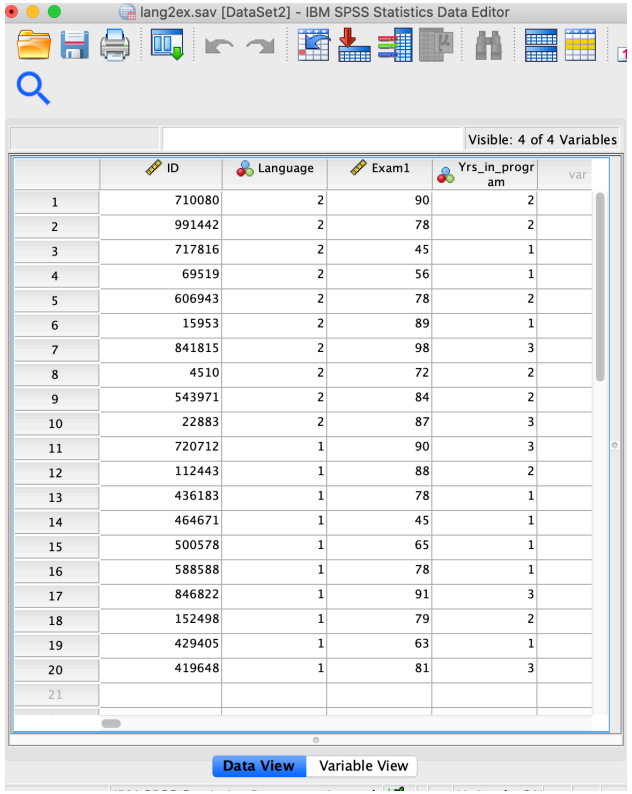


Prepping data: Merging different datasets

Before:

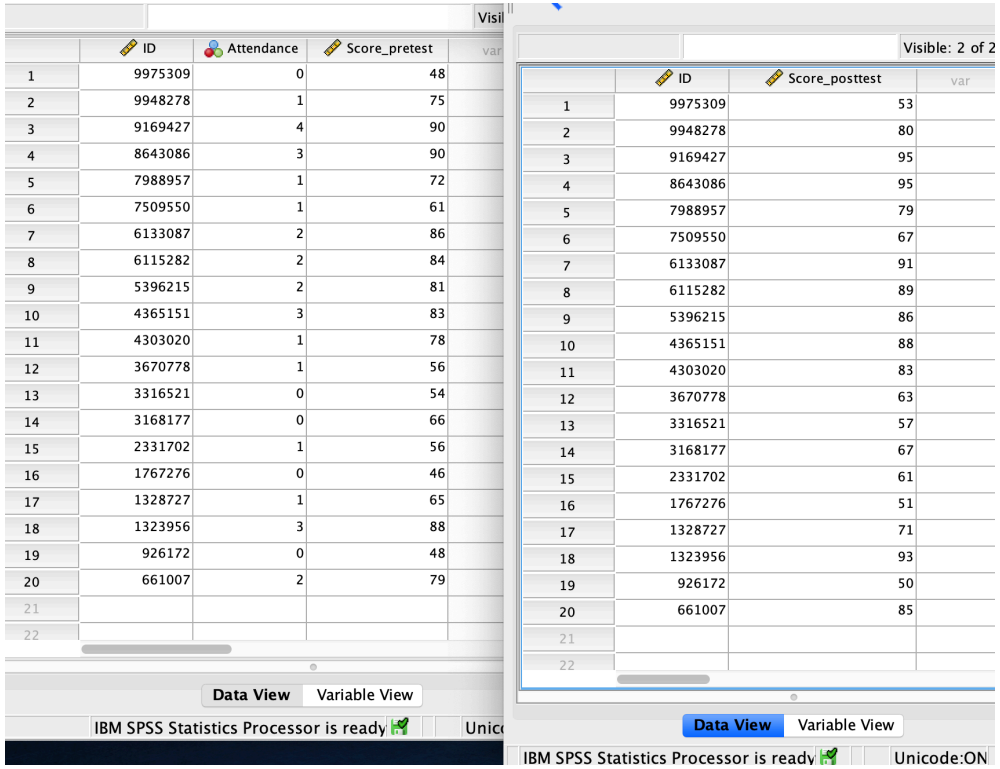


After:



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



	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			

Data View | Variable View

IBM SPSS Statistics Processor is ready

Visible: 2 of 2

	ID	Score_posttest
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		

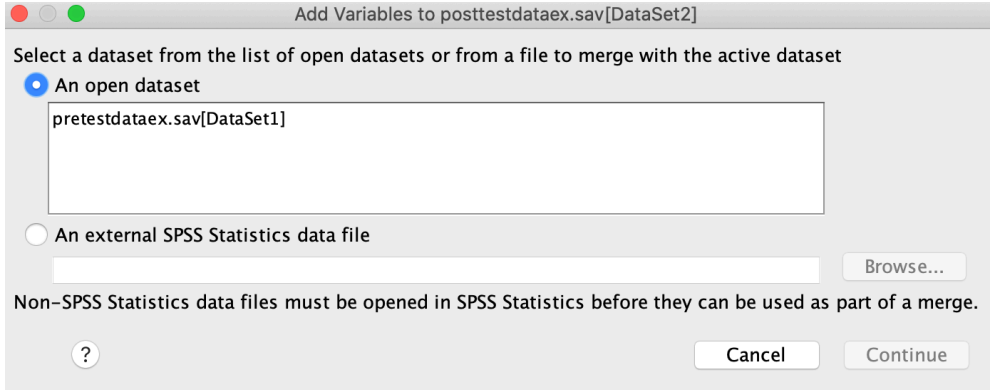
Data View | Variable View

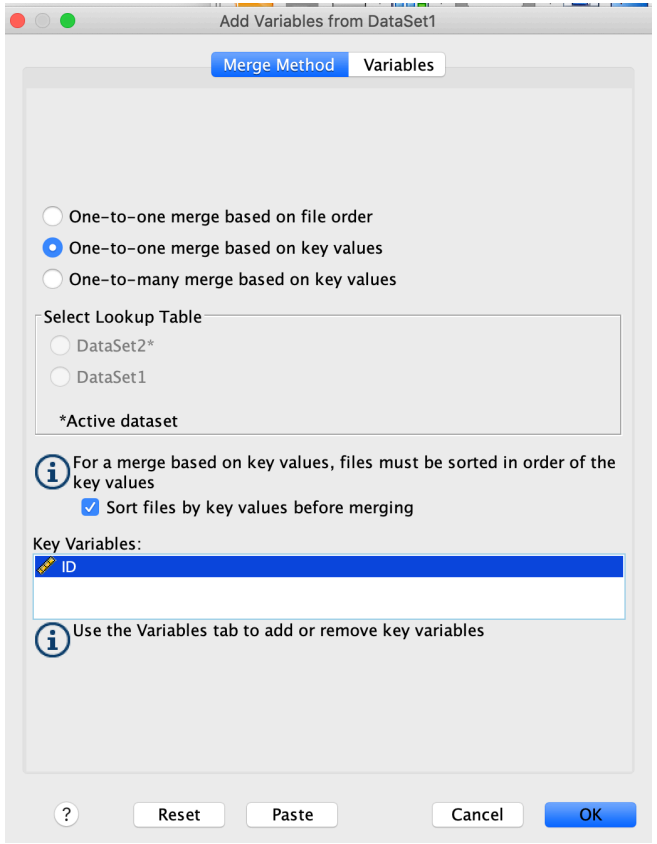
IBM SPSS Statistics Processor is ready

Unicode:ON

Prepping data: Merging pre/post test data

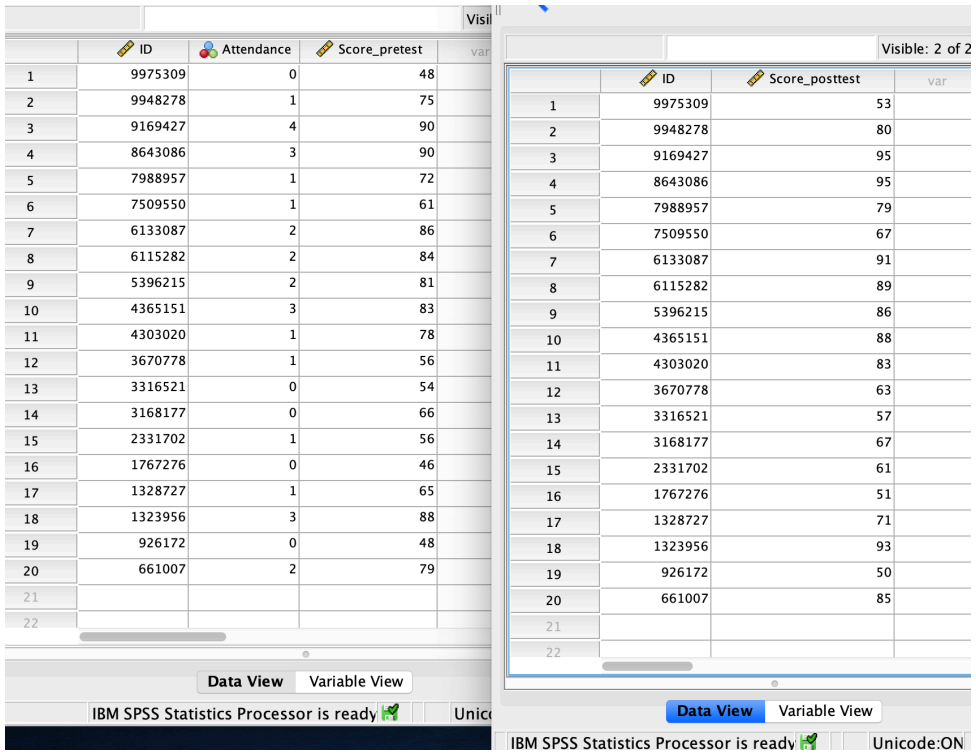
Data > Merge Files > Add Variables





Prepping data: Merging pre/post test data

Before:



	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

Data View | Variable View

IBM SPSS Statistics Processor is ready

Visible: 2 of 2

	ID	Score_posttest
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

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

After:

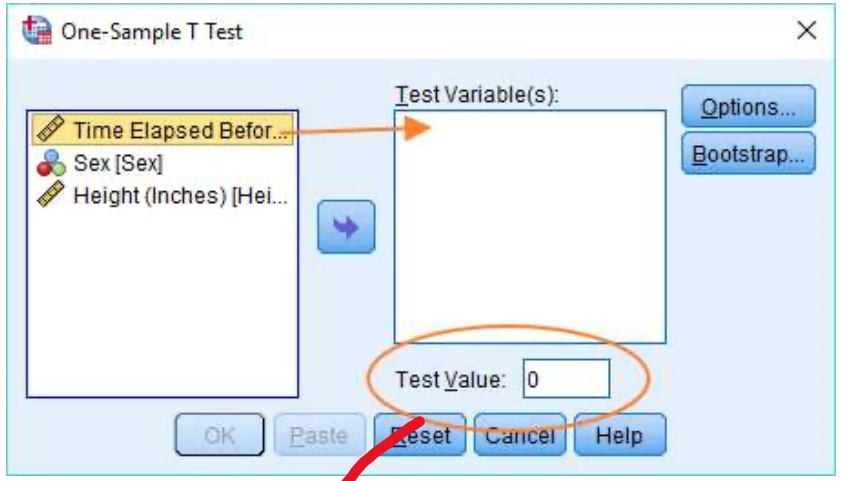
Visible: 4 of 4 V

	ID	Score_posttest	Attendance	Score_pretest
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

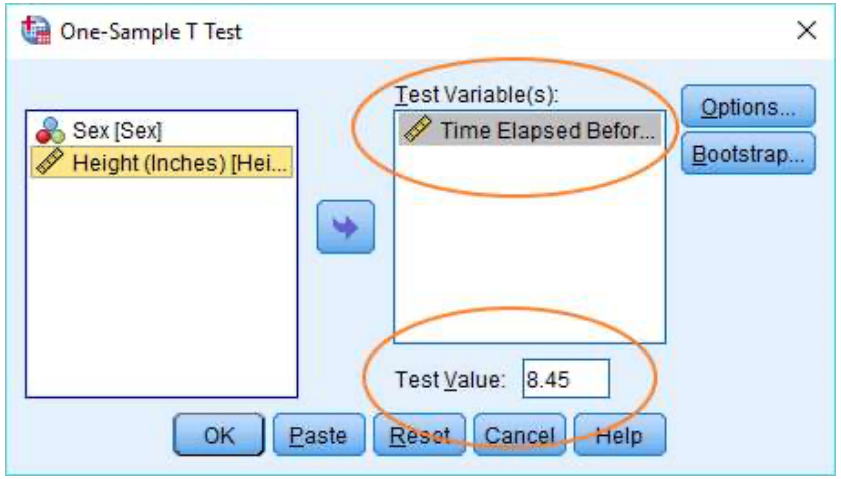
Any questions/comments so far?

SPSS How to: One Sample T-test

- Analyze -> Compare Means -> One-Sample T Test



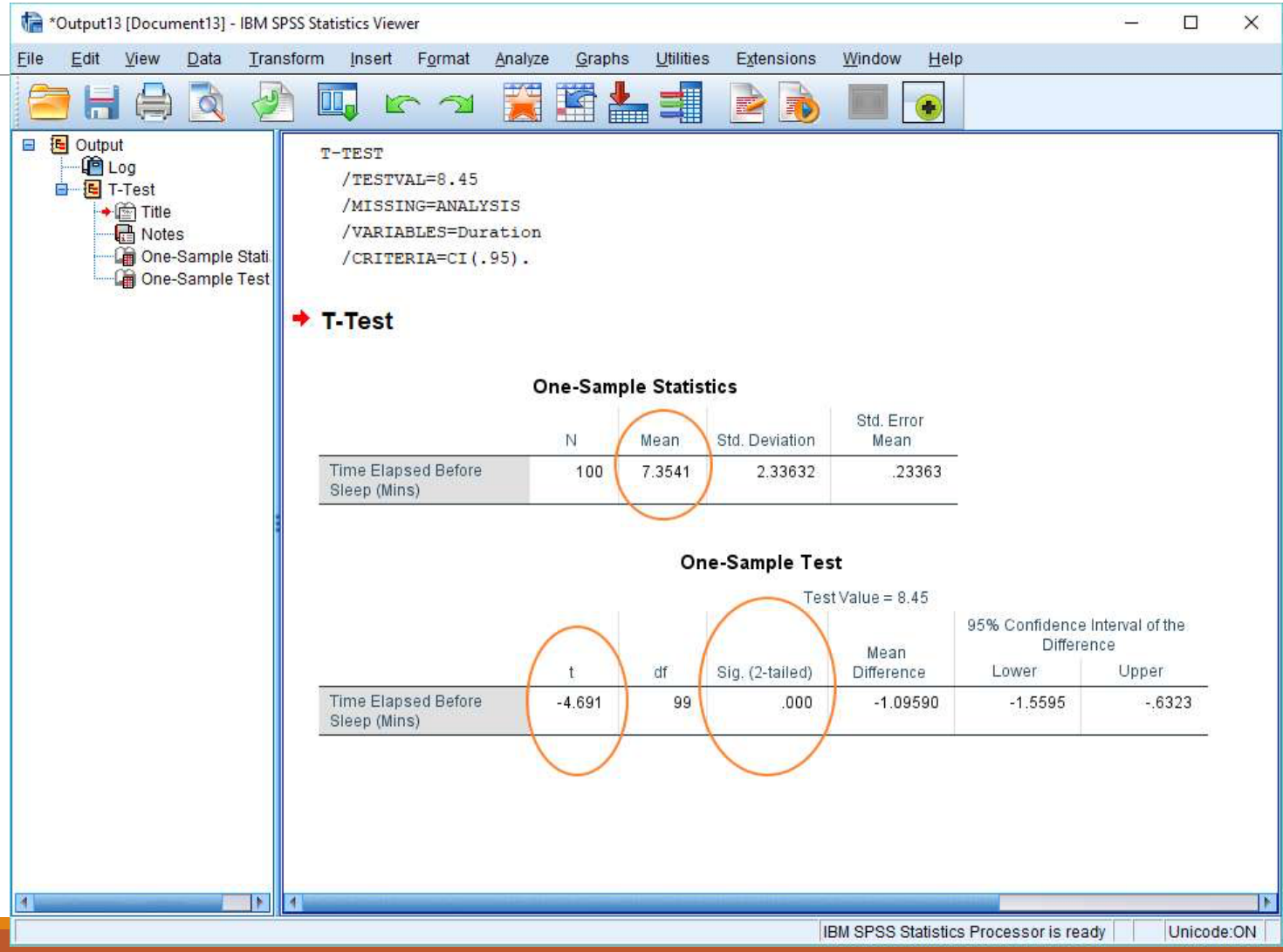
Test Value: Input the population mean here



SPSS How to: One Sample T-test

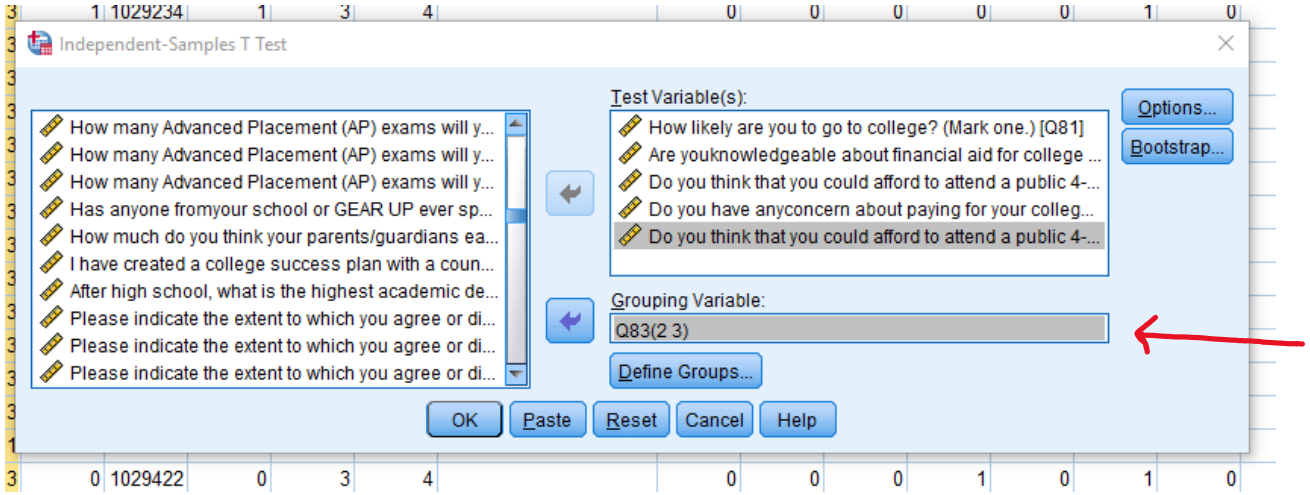
Interpretation:

Our results indicate that the sample's time elapsed before sleep was significantly higher(M=7.354, SD=2.33) than the population average, t(99)=-4.691, p <.001.



SPSS How to: Independent Sample T-test

Analyze -> Compare Means -> Independent-Samples T Test



Grouping variable:
This variable should hold the groups we're comparing, in this care Q83 asked the respondents of their high school and the values represent what the high schools are.
2 = Katella High School
3 = Loara High School

SPSS How to: Independent Sample T-test

Group Statistics

	School you attend:	N	Mean	Std. Deviation	Std. Error Mean
How likely are you to go to college? (Mark one.)	Katella High School	579	4.24	.856	.036
	Loara High School	398	4.17	.908	.045
Are you knowledgeable about financial aid for college and the cost and benefits to you of going to college	Katella High School	586	.67	.472	.019
	Loara High School	403	.78	.417	.021

Descriptive statistics

- Mean, standard deviation, *N*

SPSS How to: Independent Sample T-test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
How likely are you to go to college? (Mark one.)	Equal variances assumed	.004	.948	1.138	975	.256	.065	.057	-.047	.177
	Equal variances not assumed			1.125	819.985	.261	.065	.058	-.048	.178
Are you knowledgeable about financial aid for college and the cost and benefits to you of going to college	Equal variances assumed	61.875	.000	-3.757	987	.000	-.109	.029	-.167	-.052
	Equal variances not assumed			-3.843	927.253	.000	-.109	.028	-.165	-.054

Which results do we report?

First we check the **Levene's Test** sig. value.

If it is BELOW .05 (significant), we look at the equal variances NOT assumed.

If it is ABOVE .05 (not significant), we look at the equal variance assumed.

Interpretation:
There is no significant differences between the high schools in their likeliness to go to college, $t(975) = 1.138$, $p > .05$.

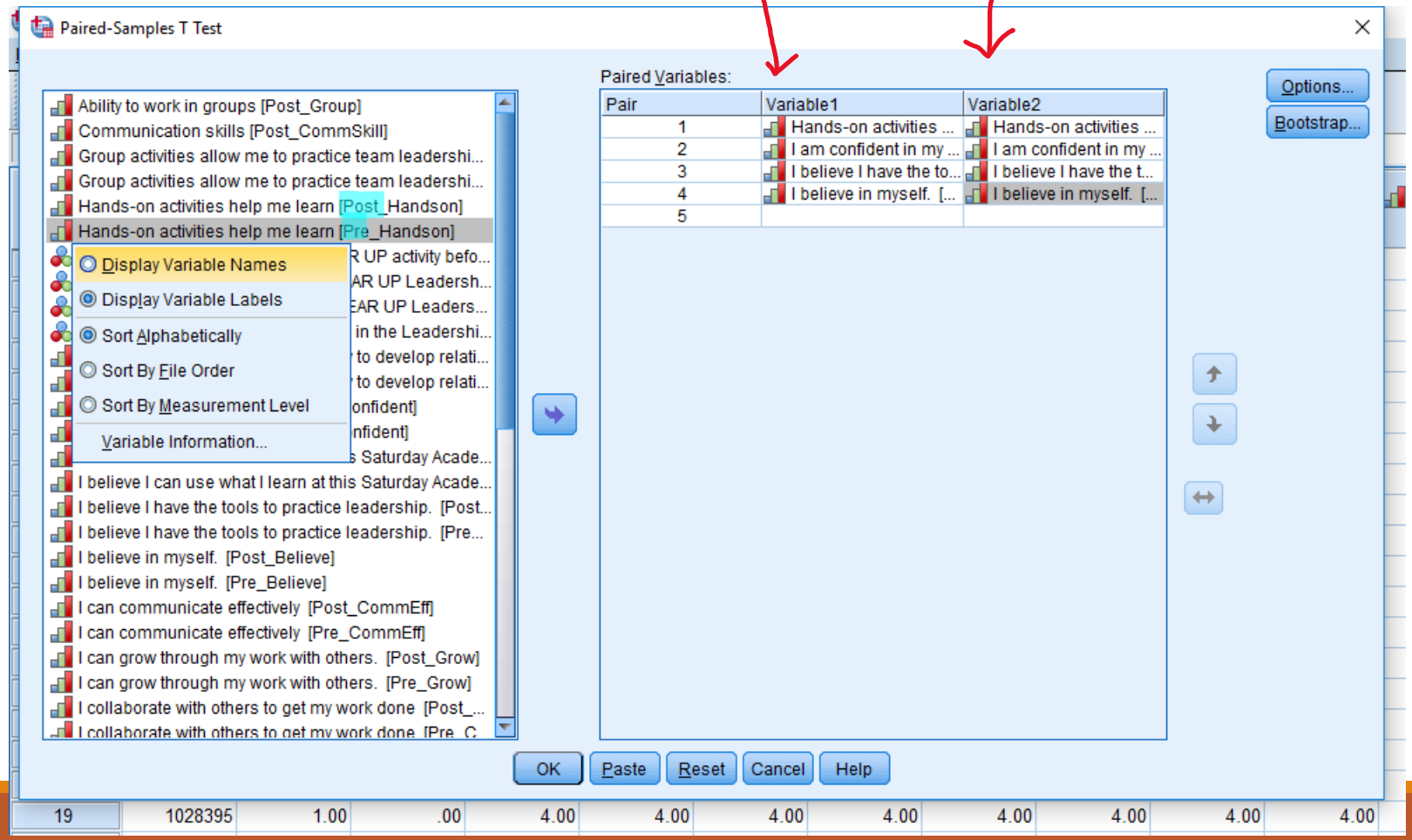
There is a significant difference between Katella High School (M=.67, SD=.472) and Loara High School (M=.78, SD=.417) in their knowledge about college costs, $t(927.25) = -3.843$, $p < .001$.

SPSS How to: Paired Samples T-test

Analyze > Compare Means > Paired-Samples T Test

Variable1:
Pre-test data

Variable2:
Post-test data



*make sure to put them in the right order

Output - Paired Samples t-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Hands-on activities help me learn	3.0909	22	.52636	.11222
	Hands-on activities help me learn	3.3636	22	.49237	.10497
Pair 2	I am confident in my ideas.	3.2174	23	.51843	.10810
	I am confident in my ideas.	3.3478	23	.57277	.11943
Pair 3	I believe I have the tools to practice leadership.	3.1739	23	.57621	.12015
	I believe I have the tools to practice leadership.	3.3913	23	.49901	.10405
Pair 4	I believe in myself.	3.0870	23	.59643	.12436
	I believe in myself.	3.4348	23	.50687	.10569

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Hands-on activities help me learn - Hands-on activities help me learn	-.27273	.55048	.11736	-.51680	-.02866	-2.324	21	.030
Pair 2	I am confident in my ideas. - I am confident in my ideas.	-.13043	.54808	.11428	-.36744	.10657	-1.141	22	.266
Pair 3	I believe I have the tools to practice leadership. - I believe I have the tools to practice leadership.	-.21739	.59974	.12505	-.47674	.04195	-1.738	22	.096
Pair 4	I believe in myself. - I believe in myself.	-.34783	.71406	.14889	-.65661	-.03904	-2.336	22	.029

Interpretation:

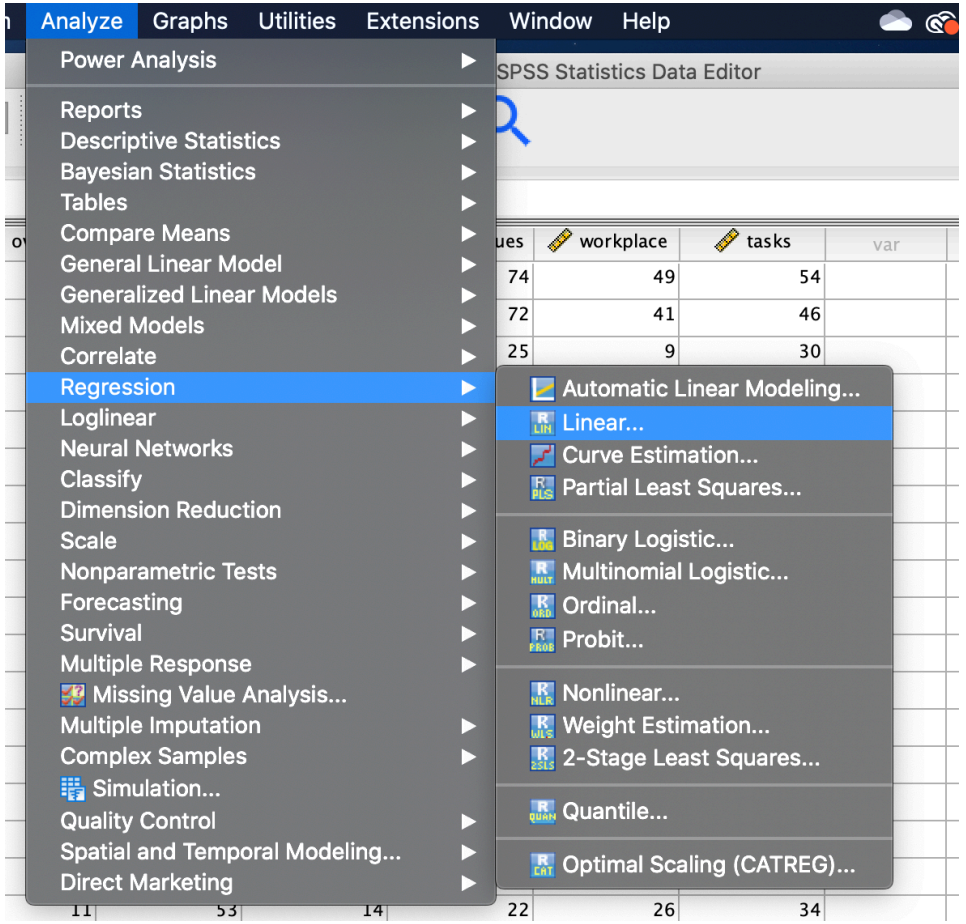
Our sample of students indicated significantly more confidence after the professional development workshop (*M*=3.344, *SD*=.507) in comparison to before (*M=3.087, SD*=.596), $t(33) = -2.336$, $p < .05$.

Any questions/comments so far?

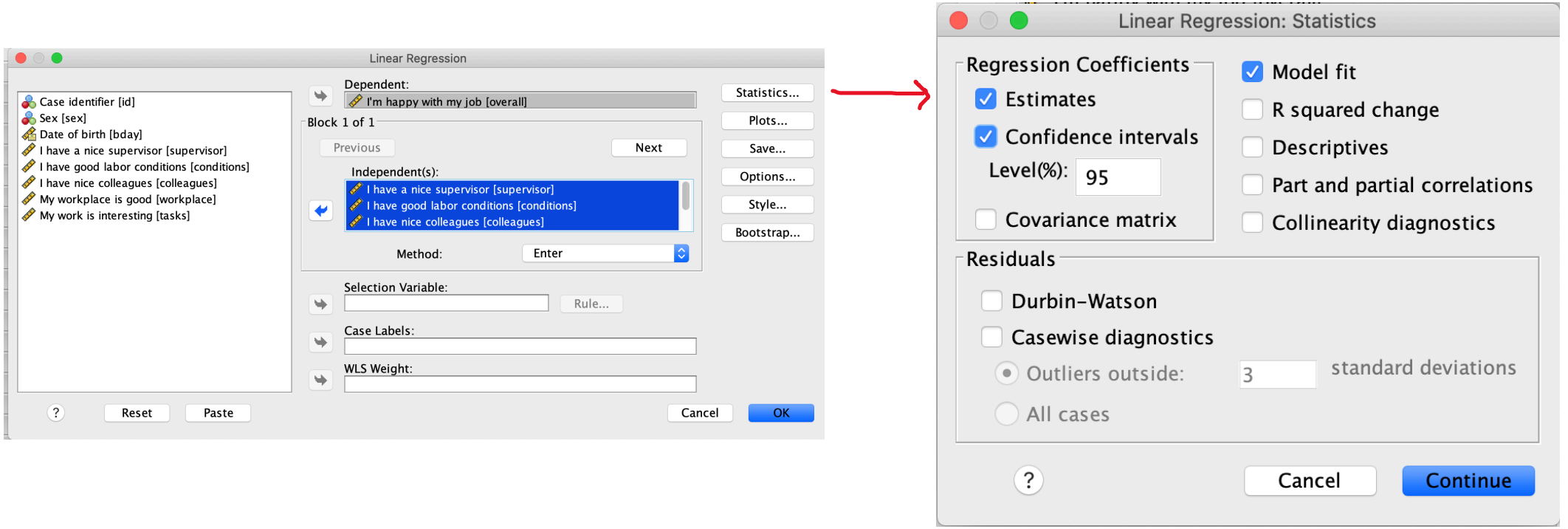
SPSS How to: Multiple Linear Regression

Analyze > Regression > Linear..

Question: What factors predict for job satisfaction? And to what extent?



SPSS How to: Multiple Linear Regression



Output: MLR

R = correlation coefficient
Interpretation: There was 69.5% correlation between X, Y, Z (predictor) variables and A (criterion/dependent) variable..

R-Square = coefficient of determination; proportion of variance explained by the independent variables
Interpretation: Our independent variables account for 48.3% of the variability in our dependent variable..

Adjusted R-Square = only accounts for *significant* variables in the model, which is why it's always lower than R-square

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.695[a]	.483	.424	17.631

a. Predictors: (Constant), My work is interesting, I have good labor conditions, My workplace is good, I have nice colleagues, I have a nice supervisor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12761.123	5	2552.225	8.210	.000[b]
	Residual	13677.757	44	310.858		
	Total	26438.880	49			

a. Dependent Variable: I'm happy with my job

b. Predictors: (Constant), My work is interesting, I have good labor conditions, My workplace is good, I have nice colleagues, I have a nice supervisor

ANOVA results indicate if the model is a good fit.
Interpretation: The table shows that the independent variables statistically significantly predict the dependent variable, $F(5, 55) = 8.210$, $p < .001$

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	5.854	10.120		.578	.566	-14.541	26.250
	I have a nice supervisor	.117	.176	.097	.664	.510	-.238	.472
	I have good labor conditions	.363	.114	.369	3.182	.003	.133	.593
	I have nice colleagues	.103	.145	.098	.707	.483	-.190	.396
	My workplace is good	.256	.139	.225	1.836	.073	-.025	.537
	My work is interesting	.334	.126	.299	2.660	.011	.081	.587

a. Dependent Variable: I'm happy with my job

Output: MLR

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.695[a]	.483	.424	17.631

a. Predictors: (Constant), My work is interesting, I have good labor conditions, My workplace is good, I have nice colleagues, I have a nice supervisor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12761.123	5	2552.225	8.210	.000[b]
	Residual	13677.757	44	310.858		
	Total	26438.880	49			

a. Dependent Variable: I'm happy with my job

b. Predictors: (Constant), My work is interesting, I have good labor conditions, My workplace is good, I have nice colleagues, I have a nice supervisor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	5.854	10.120		.578	.566	-14.541	26.250
	I have a nice supervisor	.117	.176	.097	.664	.510	-.238	.472
	I have good labor conditions	.363	.114	.369	3.182	.003	.133	.593
	I have nice colleagues	.103	.145	.098	.707	.483	-.190	.396
	My workplace is good	.256	.139	.225	1.836	.073	-.025	.537
	My work is interesting	.334	.126	.299	2.660	.011	.081	.587

a. Dependent Variable: I'm happy with my job

Our regression equation:

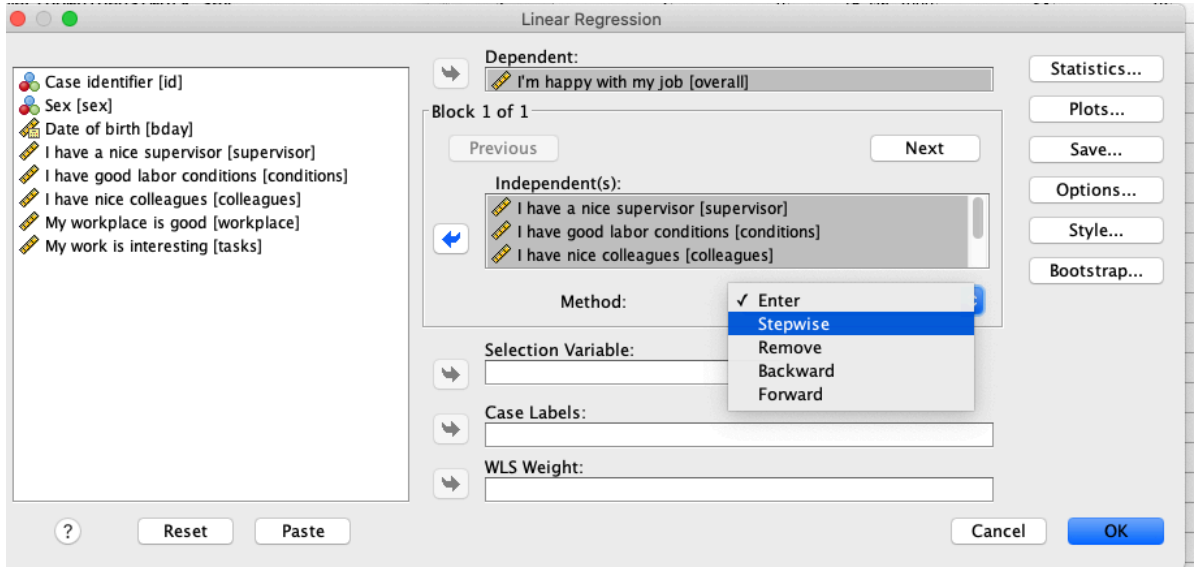
Job satisfaction = 5.854 + .117*supervisor + .363*conditions + .103*colleagues + .256*workplace + .334*interest

MLR: Finding the best model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	5.854	10.120		.578	.566	-14.541	26.250
	I have a nice supervisor	.117	.176	.097	.664	.510	-.238	.472
	I have good labor conditions	.363	.114	.369	3.182	.003	.133	.593
	I have nice colleagues	.103	.145	.098	.707	.483	-.190	.396
	My workplace is good	.256	.139	.225	1.836	.073	-.025	.537
	My work is interesting	.334	.126	.299	2.660	.011	.081	.587

a. Dependent Variable: I'm happy with my job



Which factors contribute the most for predicting job satisfaction?

Not all our predictors are significant; we can further simplify our model to be better fit

- Stepwise Regression: Remove or add predictor that would result in model with the best fit
- Backward Regression: Adds all predictors then sequentially removed
- Forward: Each predictor is added sequentially

Output: Stepwise Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.497[a]	.247	.232	20.362
2	.614[b]	.377	.351	18.719
3	.680[c]	.462	.427	17.587

a. Predictors: (Constant), I have good labor conditions

b. Predictors: (Constant), I have good labor conditions, My work is interesting

c. Predictors: (Constant), I have good labor conditions, My work is interesting, My workplace is good

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6537.379	1	6537.379	15.767	.000[b]
	Residual	19901.501	48	414.615		
	Total	26438.880	49			
2	Regression	9969.814	2	4984.907	14.226	.000[c]
	Residual	16469.066	47	350.406		
	Total	26438.880	49			
3	Regression	12211.599	3	4070.533	13.161	.000[d]
	Residual	14227.281	46	309.289		
	Total	26438.880	49			

a. Dependent Variable: I'm happy with my job

b. Predictors: (Constant), I have good labor conditions

c. Predictors: (Constant), I have good labor conditions, My work is interesting

d. Predictors: (Constant), I have good labor conditions, My work is interesting, My workplace is good

Our regression equation:
Job satisfaction = 10.959 + .408*conditions + .364*workplace + .337*interest

Predictors: Labor conditions, interesting work, and good workplace

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	40.913	7.099		5.763	.000
	I have good labor conditions	.489	.123	.497	3.971	.000
2	(Constant)	21.113	9.089		2.323	.025
	I have good labor conditions	.444	.114	.451	3.890	.000
	My work is interesting	.406	.130	.363	3.130	.003
3	(Constant)	10.959	9.335		1.174	.246
	I have good labor conditions	.408	.108	.415	3.778	.000
	My work is interesting	.364	.123	.326	2.964	.005
	My workplace is good	.337	.125	.296	2.692	.010

a. Dependent Variable: I'm happy with my job

THE END ☺
Any questions/comments?