

SPSS Review

11/09/2020

By Marrion Macandog

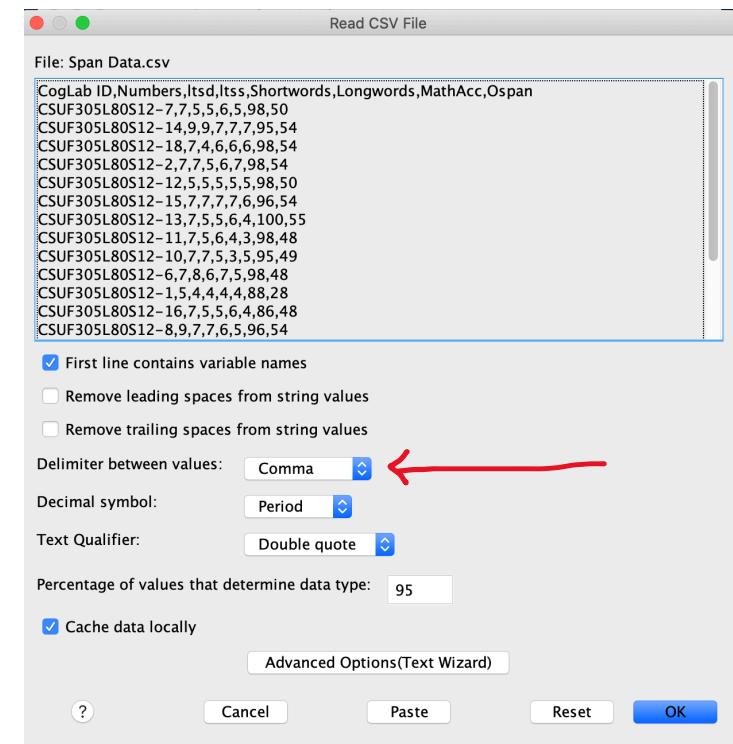
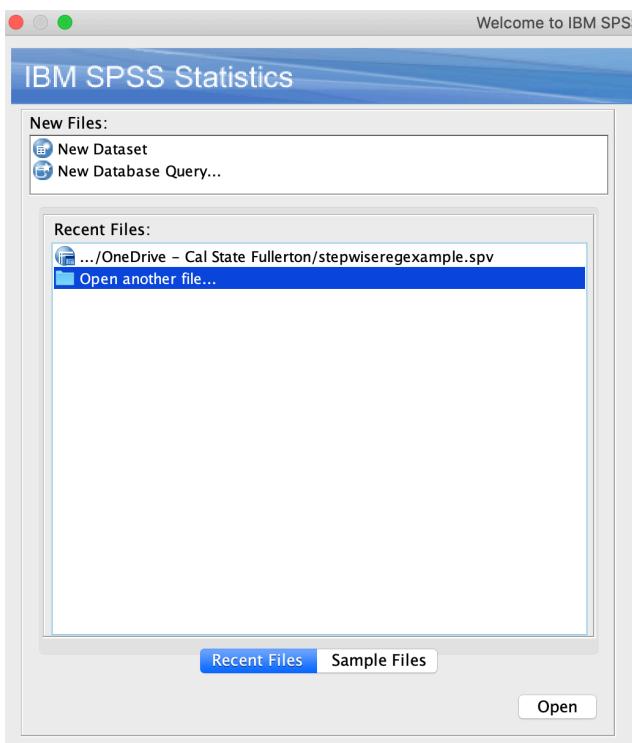


Agenda:

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

The screenshot shows the Qualtrics interface with the 'Data & Analysis' tab selected (circled in orange). On the right, there are statistics: Recorded Responses (199) and Responses in Progress (0). Below the tabs, there's a table with columns for questions and a checkbox for 'Recorded Date'. At the bottom right, a dropdown menu is open under 'Export & Import' (circled in orange), showing options like 'Export Data...', 'Import Data...', and 'Response Export Automation...'.

Downloading the data

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

A modal dialog box titled 'Download a data table' is shown. It has a 'SPSS' tab selected in the file type list (CSV, TSV, Excel, XML, SPSS, Google Drive, User-submitted files). A description of the SPSS format is provided: 'Statistical Package for the Social Sciences (SPSS) is one of the most widely used software packages for survey analysis. This is an SPSS sav data file with raw data, variable and value labels.' There is a checked checkbox for 'Download all fields'. At the bottom are 'More options', 'Close', and a large blue 'Download' button.

Imported Data into SPSS

Data View

The Data View window displays a grid of survey data. The columns are labeled Q33, Q152, Q35_1 through Q35_12. The rows show responses for 12 different participants. The data includes various numerical values and missing data represented by dots.

	Q33	Q152	Q35_1	Q35_2	Q35_3	Q35_4	Q35_5	Q35_6	Q35_7	Q35_8	Q35_9	Q35_10	Q35_11	Q35_12
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	1
7	4	1	.	1
8	4	1	.	1	1
9	4	2
10	4	2	.	1
11	4	1	1	.	.	1
12	4	1	1	.	1	1

Variable View

The Variable View window shows the structure of the imported data. It lists 14 variables: Q33, Q152, Q35_1 through Q35_12. Each variable is defined with a name, type (Numeric), width (40), and decimal places (0). The labels provide context for each question, such as "Who can I talk to..." for Q33 and "Which of the fol..." for Q35_1 through Q35_12. The values column indicates the number of categories for each question, ranging from 2 to 9. The missing column specifies that missing values are represented by dots. The align and measure columns indicate that values are right-aligned and measured on a scale.

	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, Communiti...	None	7	Right	Scale
6	Q35_4	Numeric	40	0	Which of the fol...	{1, Communiti...	None	7	Right	Scale
7	Q35_5	Numeric	40	0	Which of the fol...	{1, Communiti...	None	7	Right	Scale
8	Q35_6	Numeric	40	0	Which of the fol...	{1, Communiti...	None	7	Right	Scale
9	Q35_7	Numeric	40	0	Which of the fol...	{1, Recruitre...	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

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 **1 = marked**
 Q3_3 Chinese
 Q3_4 Korean
 Q3_5 Other:
Q3_5_TEXT string

Example dataset -



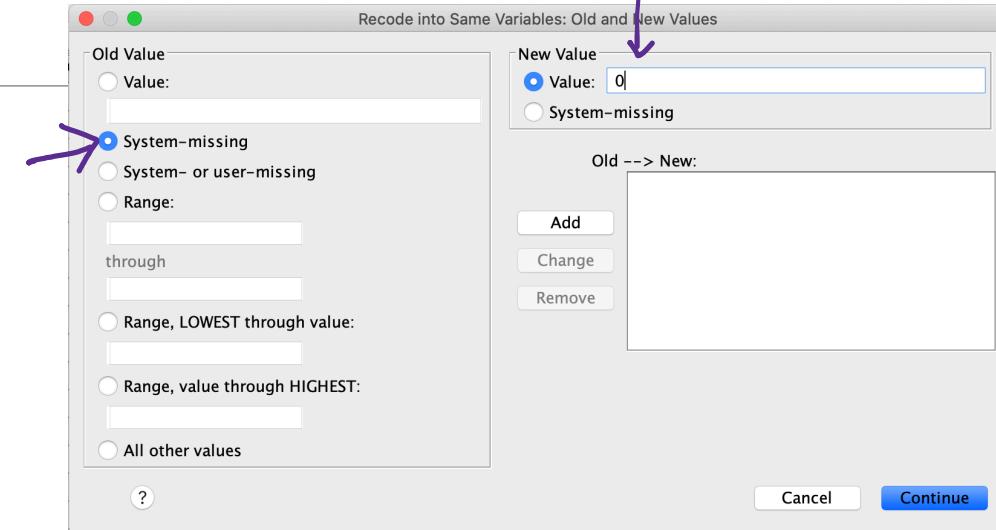
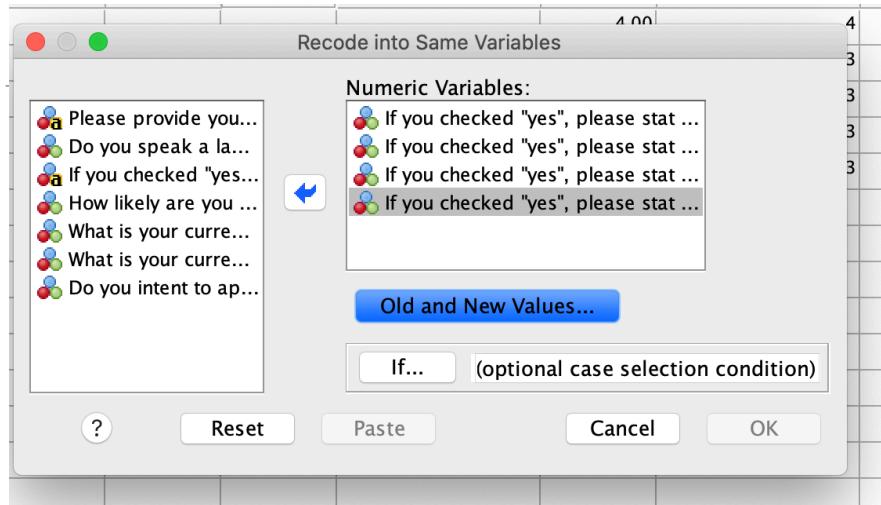
	Q1	Q2	Q3_1	Q3_2	Q3_3	Q3_4	Q3_5_TEXT	Q4	Q5_1_TEXT	Q6	Valid	Missing
1	007	1.00	1.00	4.00	3.61	.	1	0
2	008	1.00	.	1.00	.	.	.	5.00	3.21	.	0	5
3	009	.00	4.00	2.78	.	0	5
4	010	1.00	1.00	5.00	3.40	.	0	5
5	011	1.00	1.00	4.00	3.10	.	0	5
6												
7												
8												
9												
10												
11												
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.):	If you checked "yes", please stat what other language(s) you can speak:(Mark all that applies.): Other – Text
3	1	0	0	5	0
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



Data set with recoded missing variables:

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

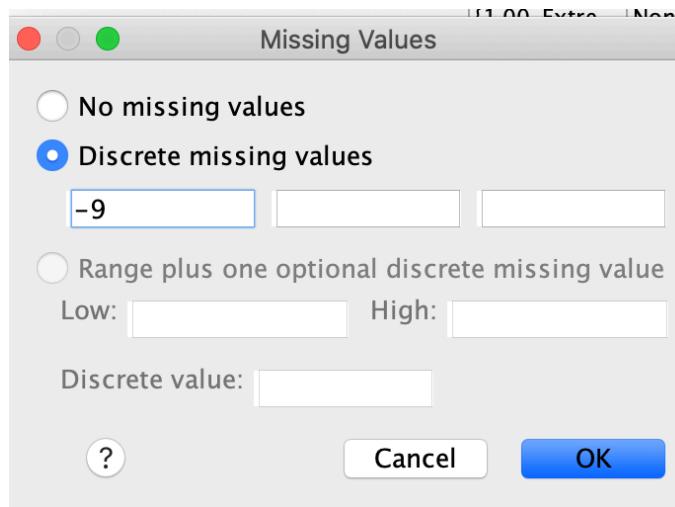
The data editor shows a dataset named 'exampledata.sav' with 7 rows and 7 columns. The columns are labeled Q1, Q2, Q3_1, Q3_2, Q3_3, and Q3_4. The data values are as follows:

	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

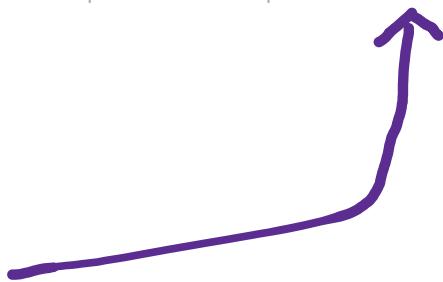
Example: Q3_5_TEXT

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



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

Nominal → Q1 1. Please provide your student I.D.:

Nominal ↳ Q2 2. Do you speak a language other than English?
 Yes 1
 No 0

Q3_1 Spanish
Q3_2 Vietnamese 1 = marked
Q3_3 Chinese
Q3_4 Korean
Q3_5_TEXT

Nominal ↳ Q4 4. How likely are you to go to college? (Mark one.)
 Extremely unlikely 1
 Unlikely 2
 Neutral 3
 Likely 4
 Extremely likely 5

Q5_1_TEXT ← Scale
Q5_1 I do not know 0

Q6 6. Do you intend to apply to college?
 Yes 1
 No 0

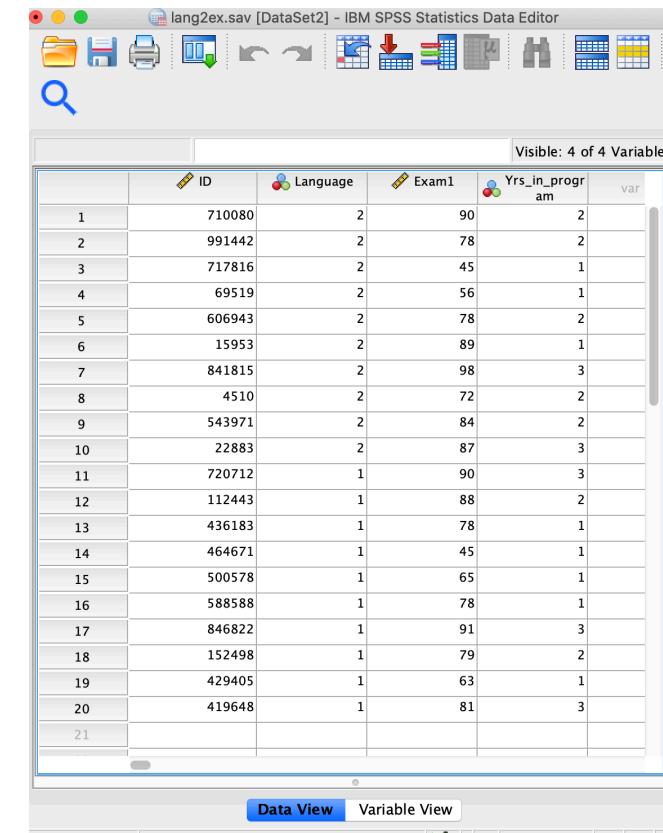
Align	Measure	Role
Left	Nominal	Input
Right	Scale	Input
Left	Ordinal	Input
Right	Nominal	Input

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



The screenshot shows the IBM SPSS Statistics Data Editor window with two datasets displayed in Data View. The left dataset has columns: ID, Attendance, Score_posttest, Score_pretest, and var. The right dataset has columns: ID, Language, Exam1, Yrs_in_program, and var. Both datasets have 21 rows of data.

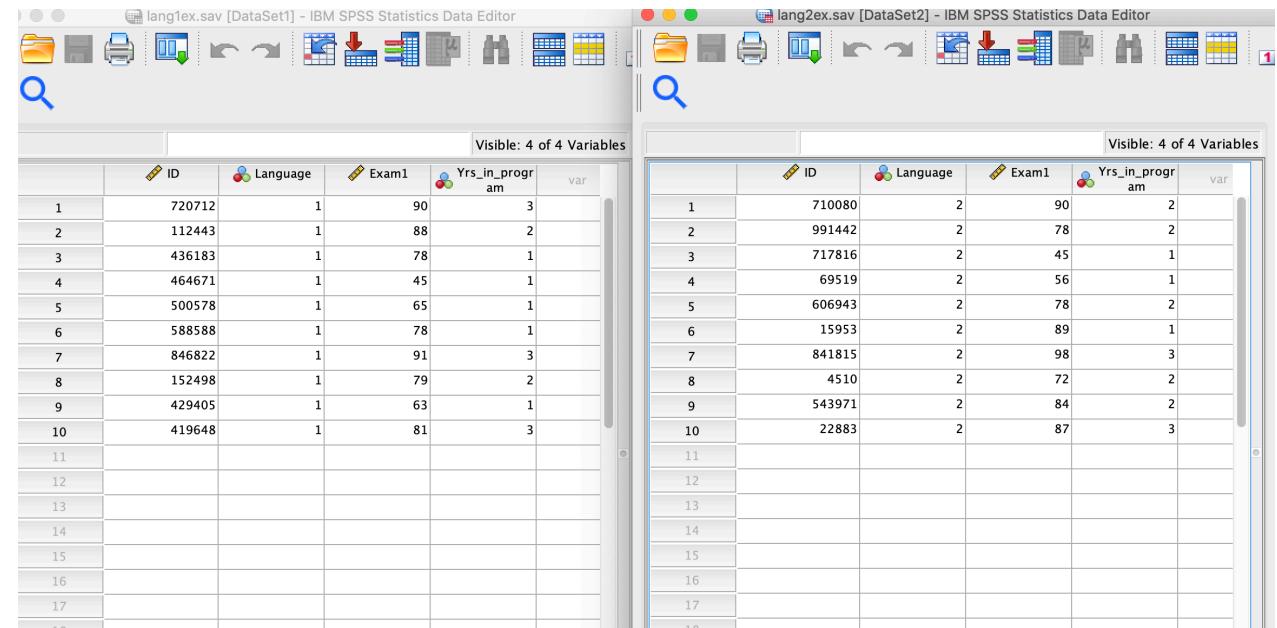
	ID	Attendance	Score_posttest	Score_pretest	var
1	9975309	0	53	48	
2	9948278	1	80	75	
3	9169427	4	95	90	
4	8643086	3	95	90	
5	7988957	1	79	72	
6	7509550	1	67	61	
7	6133087	2	91	86	
8	6115282	2	89	84	
9	5396215	2	86	81	
10	4365151	3	88	83	
11	4303020	1	83	78	
12	3670778	1	63	56	
13	3316521	0	57	54	
14	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	

	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	
-					5



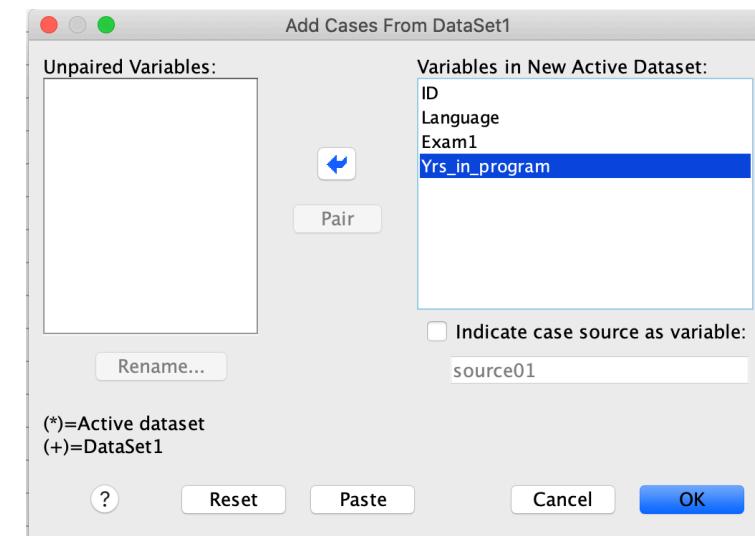
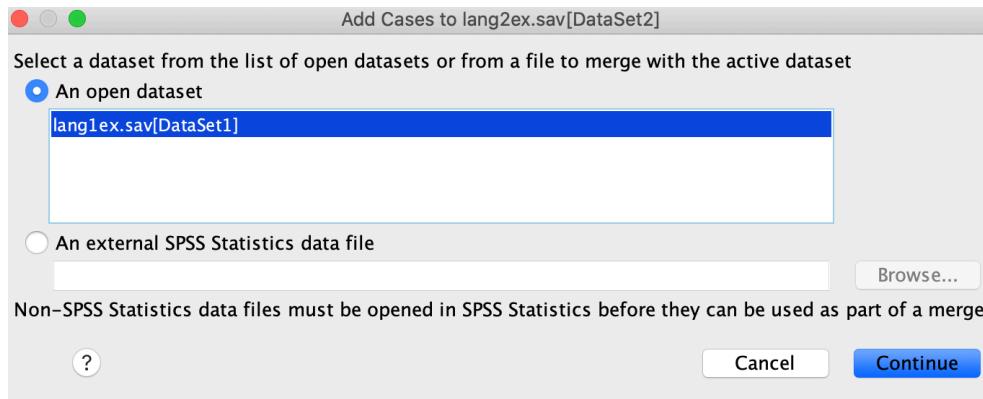
The image shows two side-by-side IBM SPSS Statistics Data Editors. The left window is titled "lang1ex.sav [DataSet1] - IBM SPSS Statistics Data Editor" and the right window is titled "lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor". Both windows show a grid of data with columns labeled ID, Language, Exam1, Yrs_in_program, and var. The data consists of 10 rows of numerical values.

	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					

	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:

The image shows two separate SPSS Data Editors side-by-side. Both windows have the title 'lang1ex.sav [DataSet1] - IBM SPSS Statistics Data Editor' and 'lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor'. Each window displays a data table with four visible variables: ID, Language, Exam1, and Yrs_in_program. The first window (DataSet1) has 18 rows of data, and the second window (DataSet2) has 21 rows of data. The columns are labeled '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					

	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					

After:

The image shows a single SPSS Data Editor window titled 'lang2ex.sav [DataSet2] - IBM SPSS Statistics Data Editor'. The window displays a merged dataset with 21 rows of data across five columns: ID, Language, Exam1, Yrs_in_program, and var. The columns are labeled 'Visible: 4 of 4 Variables'. The 'Data View' tab is selected at the bottom.

	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 shows two separate SPSS Data View windows side-by-side. Both windows have a header row with columns labeled 'ID', 'Attendance', and 'Score_pretest' (in the left window) or 'Score_posttest' (in the right window). The data rows are numbered from 1 to 22. The left window shows 'Attendance' values of 0, 1, 4, 3, 1, 2, 1, 2, 2, 3, 1, 1, 0, 0, 1, 0, 1, 3, 0, 2, 1, 0, 22. The right window shows 'Score_posttest' values of 53, 80, 95, 95, 79, 67, 91, 89, 86, 88, 83, 83, 57, 67, 61, 51, 71, 93, 50, 85, 22. The status bar at the bottom of each window indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'.

	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 Unicode:ON

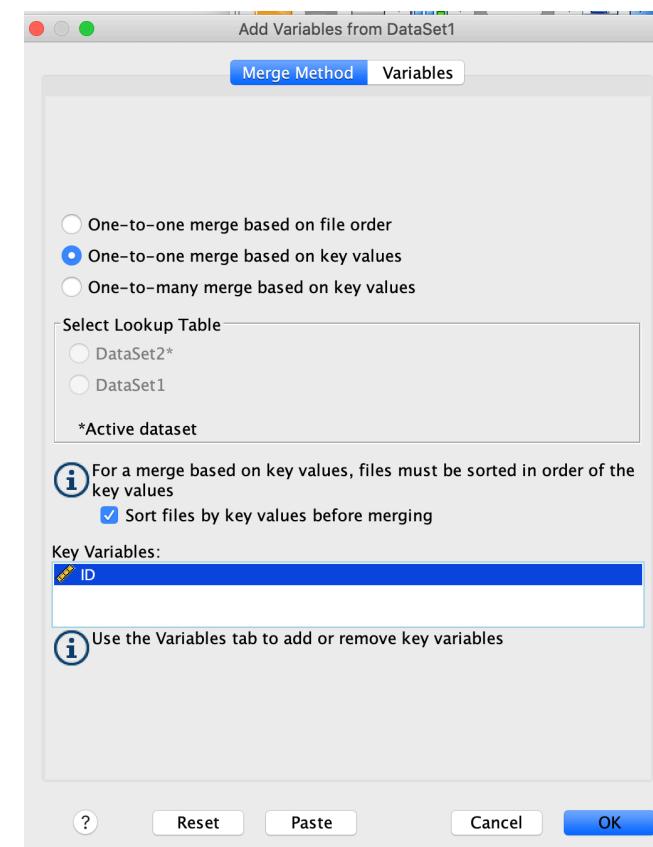
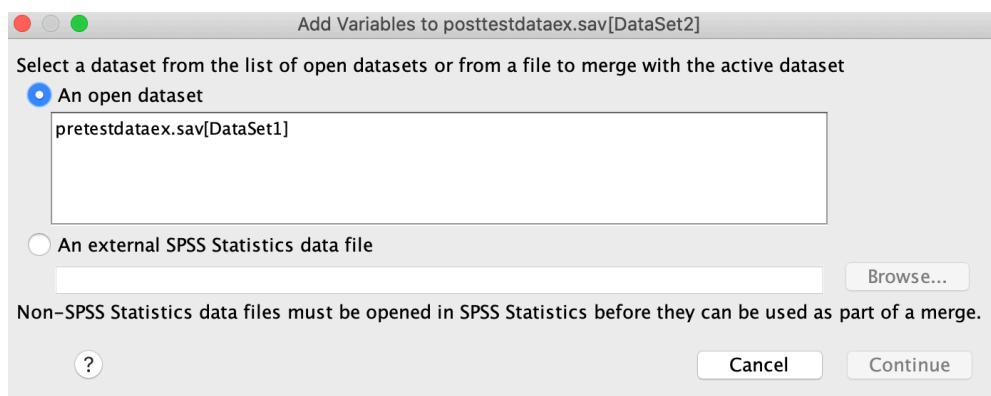
	ID	Score_posttest
1	9975309	53
2	9948278	80
3	9169427	95
4	8643086	95
5	7988957	79
6	7509550	67
7	6133087	91
8	6115282	89
9	5396215	86
10	4365151	88
11	4303020	83
12	3670778	63
13	3316521	57
14	3168177	67
15	2331702	61
16	1767276	51
17	1328727	71
18	1323956	93
19	926172	50
20	661007	85
21		
22		

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
21			
22			

	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		

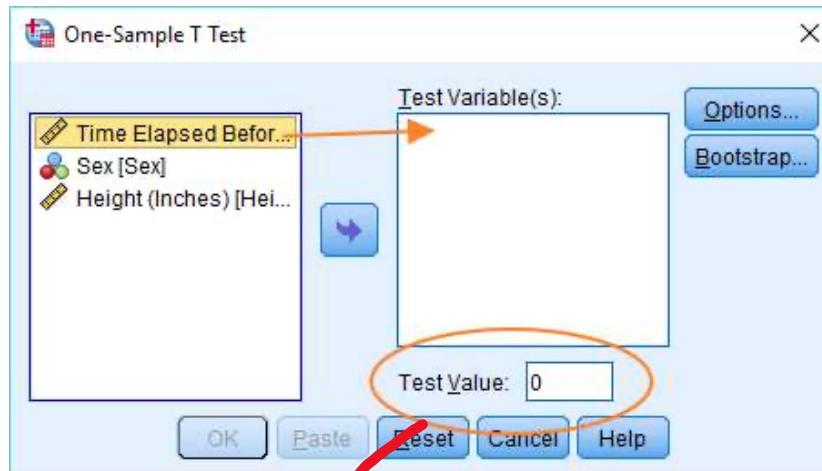
After:

	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						

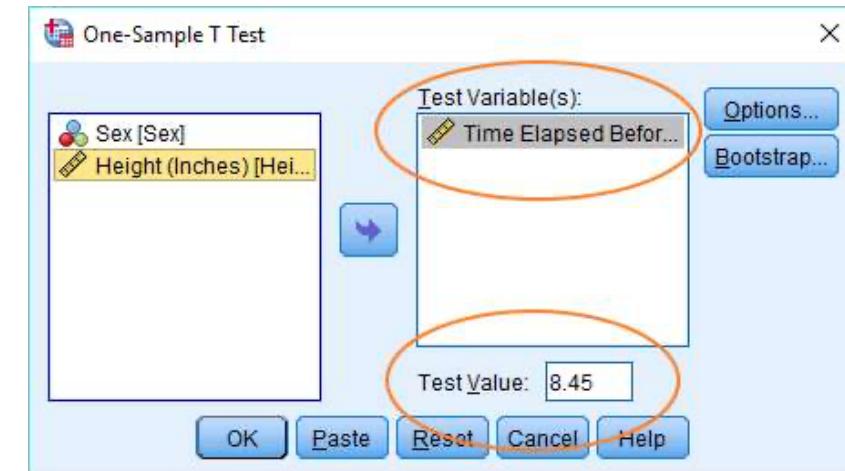
Any questions/comments so far?

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, $t(99)=-4.691$, $p < .001$.

The screenshot shows the SPSS Statistics Viewer window with the following details:

T-TEST
/TESTVAL=8.45
/MISSING=ANALYSIS
/VARIABLES=Duration
/CRITERIA=CI(.95).

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Time Elapsed Before Sleep (Mins)	100	7.3541	2.33632	.23363

One-Sample Test

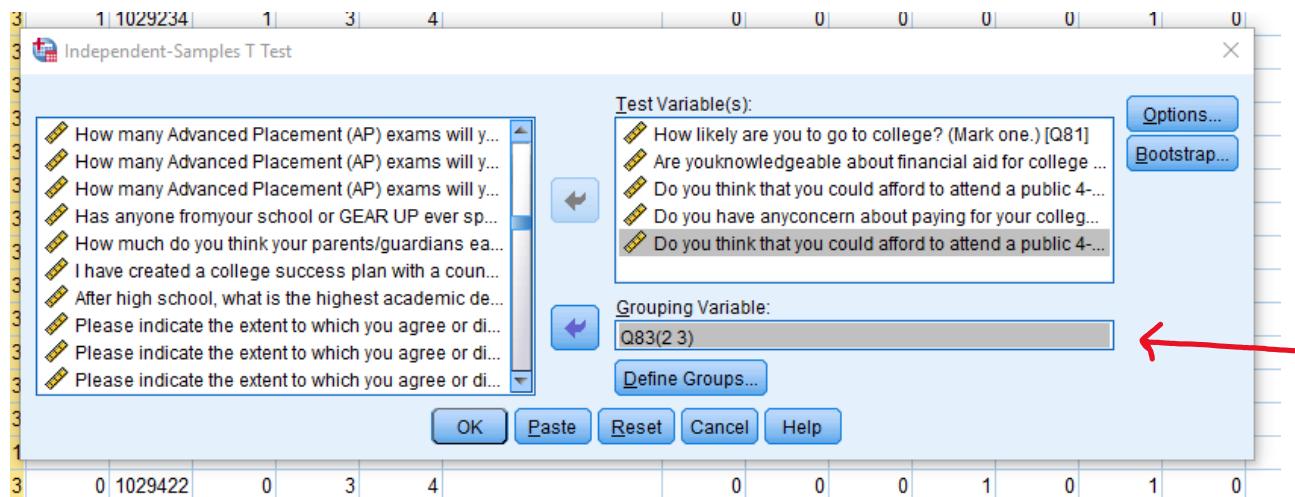
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference
Time Elapsed Before Sleep (Mins)	-4.691	99	.000	-1.09590	Lower: -1.5595, Upper: -.6323

Test Value = 8.45

IBM SPSS Statistics Processor is ready | Unicode:ON

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 case 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				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
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

Pair	Variable1	Variable2
1	Hands-on activities help me learn [Post_Handson]	Hands-on activities help me learn [Pre_Handson]
2	I am confident in myself. [Post_Believe]	I believe in myself. [Pre_Believe]
3	I believe I have the tools to practice leadership. [Post_Leadership]	I believe I can use what I learn at this Saturday Academy. [Pre_Leadership]
4	I can communicate effectively [Post_CommEff]	I can communicate effectively [Pre_CommEff]
5	I can grow through my work with others. [Post_Grow]	I collaborate with others to get my work done. [Post_Collab]

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

Interpretation:

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

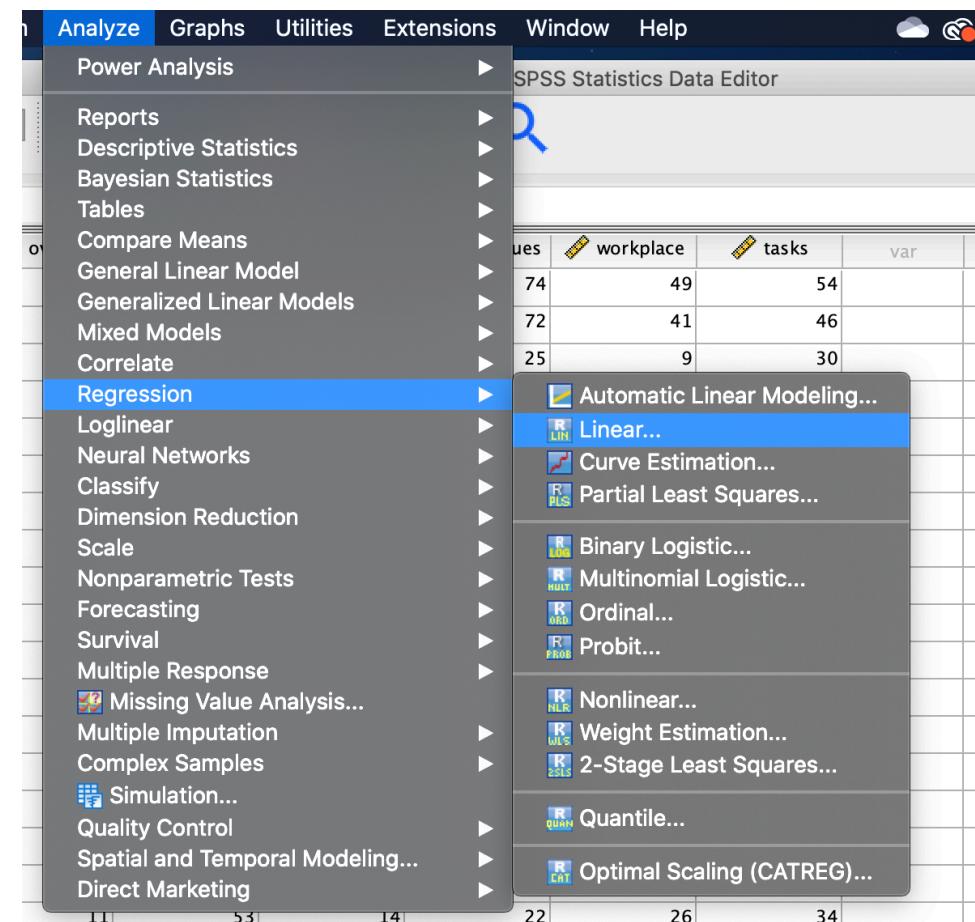
Paired Differences						t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference	Lower	Upper		
Pair 1	Hands-on activities help me learn - Hands-on activities help me learn	-.27273	.55048	.11736	-.51680	-.02866	-2.324	.030
Pair 2	I am confident in my ideas. -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

Any questions/comments so far?

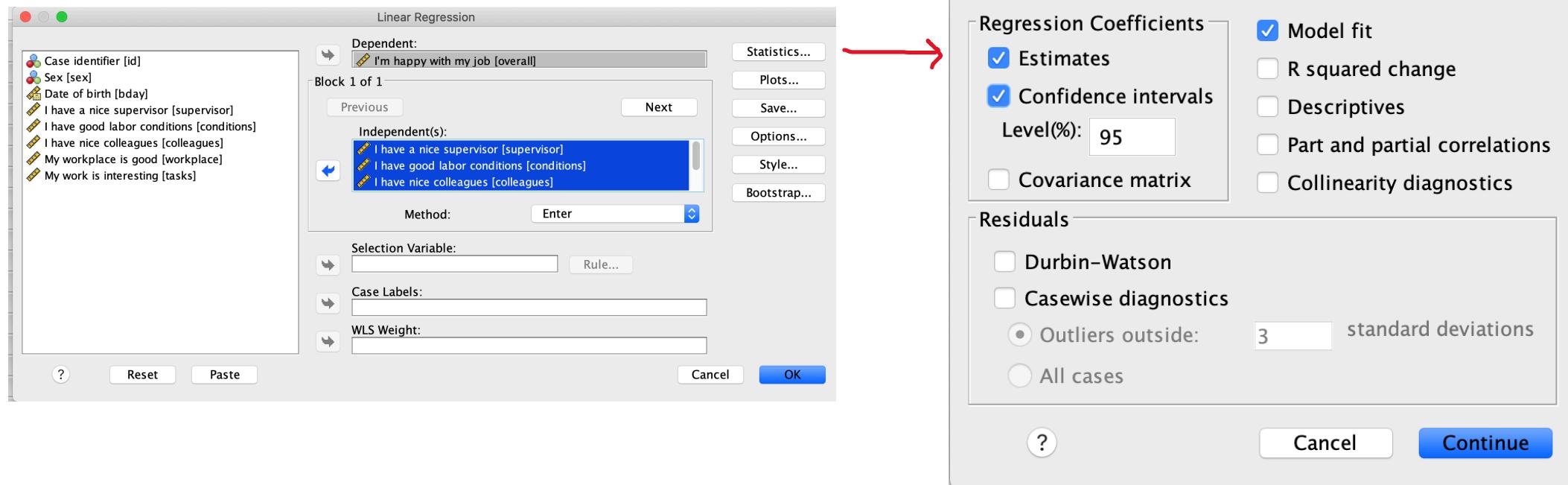
SPSS How to: Multiple Linear Regression

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
	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

Model	Coefficients ^a					95.0% Confidence Interval for B		
	Unstandardized Coefficients		Standardized Coefficients	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

ANOVA results indicate if the model is a good fit.

Interpretation: The table shows that the independent variables statistically significantly predict the dependent variable, $F(5, 55) = 8.210, p < .001$

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

Our regression equation:

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



Coefficients^a

Model		Unstandardized Coefficients		Beta	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error				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

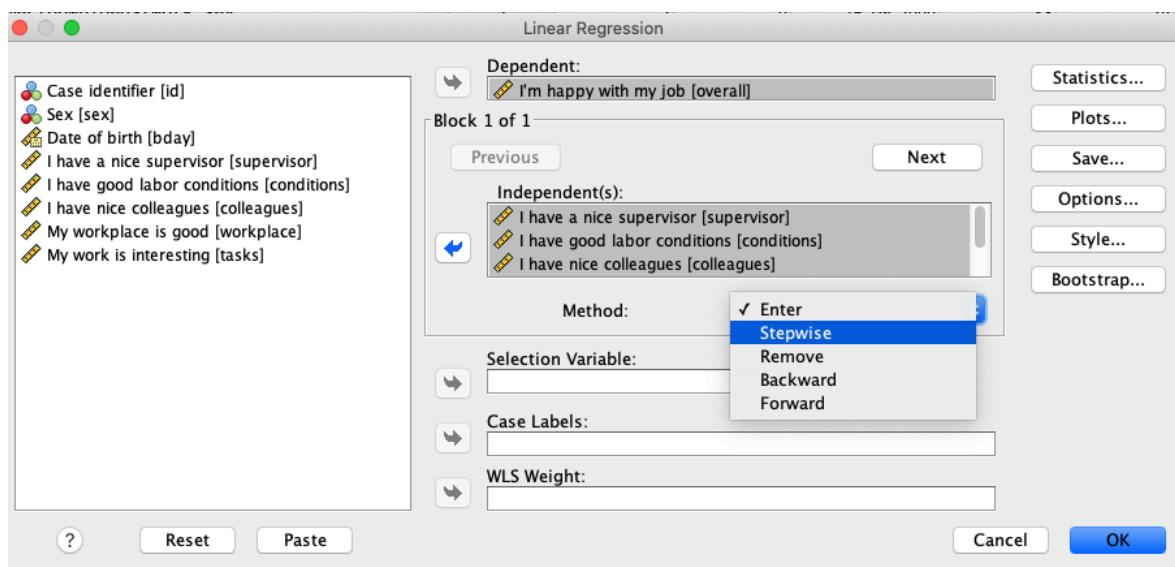
MLR: Finding the best model

Model	Coefficients ^a							
	B	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		
		Beta				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

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
1	Regression	6537.379	1	6537.379	15.767
	Residual	19901.501	48	414.615	
	Total	26438.880	49		
2	Regression	9969.814	2	4984.907	14.226
	Residual	16469.066	47	350.406	
	Total	26438.880	49		
3	Regression	12211.599	3	4070.533	13.161
	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:

$$\text{Job satisfaction} = 10.959 + .408 * \text{conditions} + .364 * \text{workplace} + .337 * \text{interest}$$

Predictors: Labor conditions, interesting work, and good workplace

Model	Coefficients ^a				t	Sig.
		Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta			
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?
