

# A quick introduction to SPSS

To open the SPSS software...

Locate the software under: **All Programs** → **SPSS** → **SPSS 19**.

## SPSS User Windows

Upon opening SPSS, you'll have two windows for which you can toggle back and forth. One looks like a spreadsheet of the actual data (Data View), and the other gives you information on the variables in the data set (Variable View). You can move back and forth between the windows by clicking on the respective buttons at the bottom left of the screen. After you import data or run any options, you will also see an Output Viewer box appear.

- Data View

- Can be used in a manner similar to a spreadsheet.
  - \* Data can be entered.
  - \* A new column can be entered (Highlight column, **Edit** → **Insert Variable**).
  - \* A new row can be entered (Highlight row, **Edit** → **Insert Cases**).
  - \* You can use a formula to create a new variable, or transform a variable.

- Variable View

- Allows user to input characteristics of variables (attributes, coding for missing values, the levels of class variables, etc.).
- Allows user to quickly view overall characteristics of the variables, like how many variables are in the data set, how many are categorical (nominal), or continuous (scale), etc.
- Many SPSS data sets arise from surveys and have 100's of variables, which represent questions on the survey. It can take a long time to input this information, but is quite useful.

- Output Viewer

- Contains the output generated by statistical procedures.
- Output location for graphics and plots.
- Can be saved separately from the .SAV data file as a .SPV file.
- Also serves as a 'log' box.

# SPSS Data Set

UCLA Academic Technology Services has some nice statistical software examples available on-line, and we will utilize one of their SPSS data sets for this introduction. The data file is called binary.SAV, and it is available at the URL below, and also at our class website in the datasets link.

<http://www.ats.ucla.edu/stat/spss/dae/binary.sav>

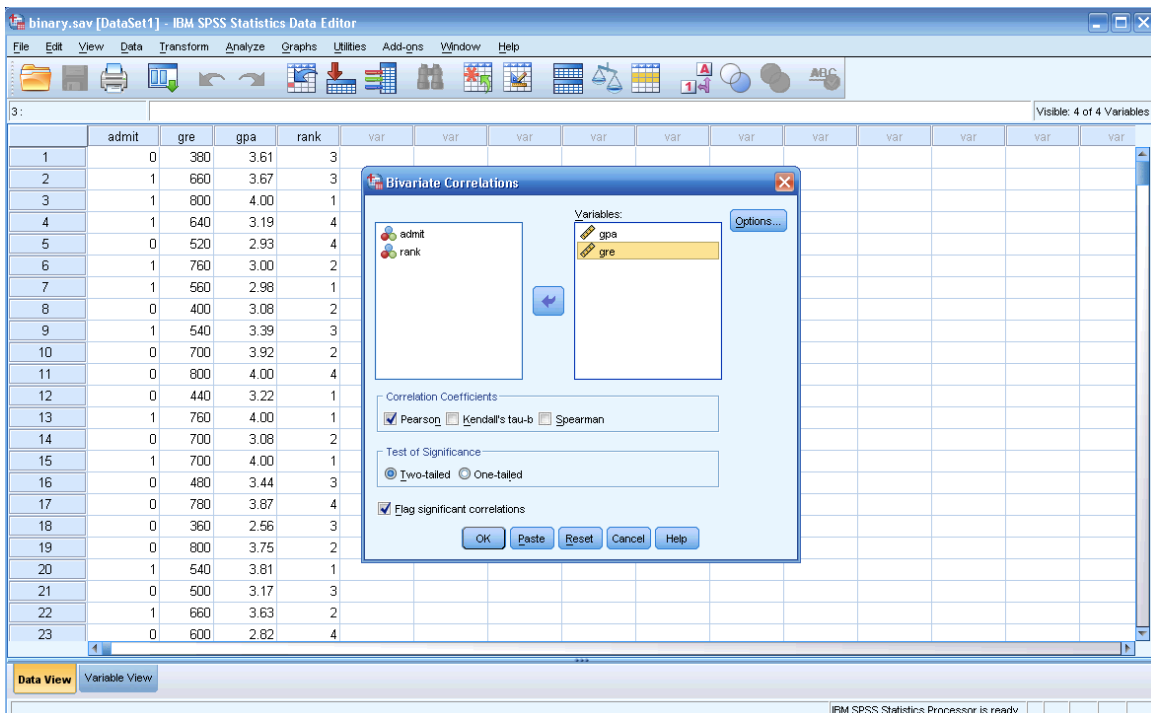
This data set contains variables related to admission to graduate school.

Variables:

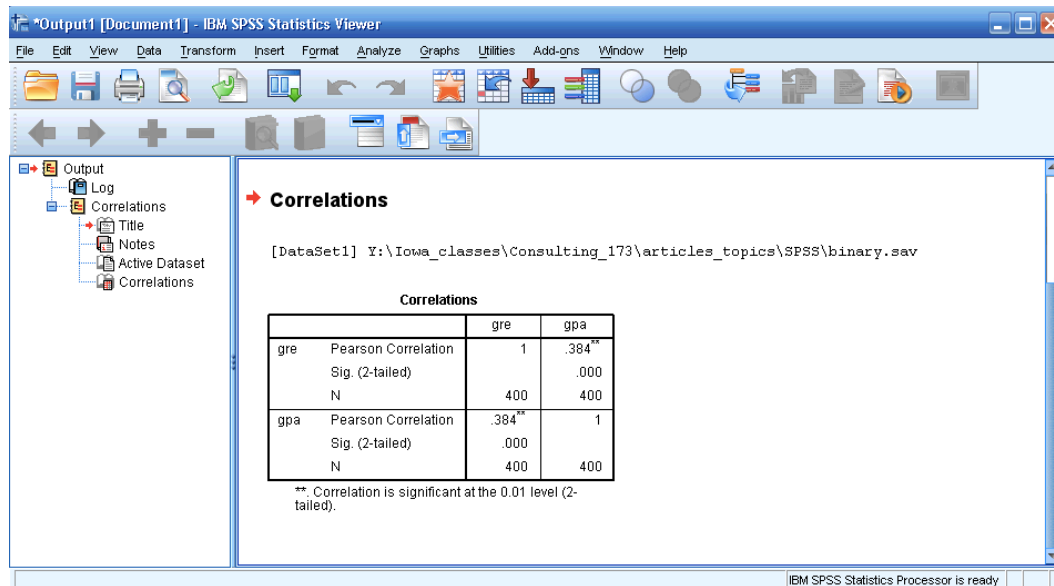
- admit    -- Admission status to graduate school (0=no, 1=yes).
- gre       -- Graduate Record Exam scores. Values range from 200 to 800.
- gpa       -- Grade Point Average.
- rank      -- Prestige of undergraduate school. Values are 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.

## Some Simple Analyses and Plots in SPSS

- Correlation
  - **Analyze** → **Correlate** → **Bivariate**
  - You'll see the "Bivariate Correlations" dialog box appear.
  - Highlight *gre*, then click the arrow to move it to the *Variables* box at the right.
  - Highlight *gpa*, then click the arrow to move it to the *Variables* box at the right.
  - Click OK.

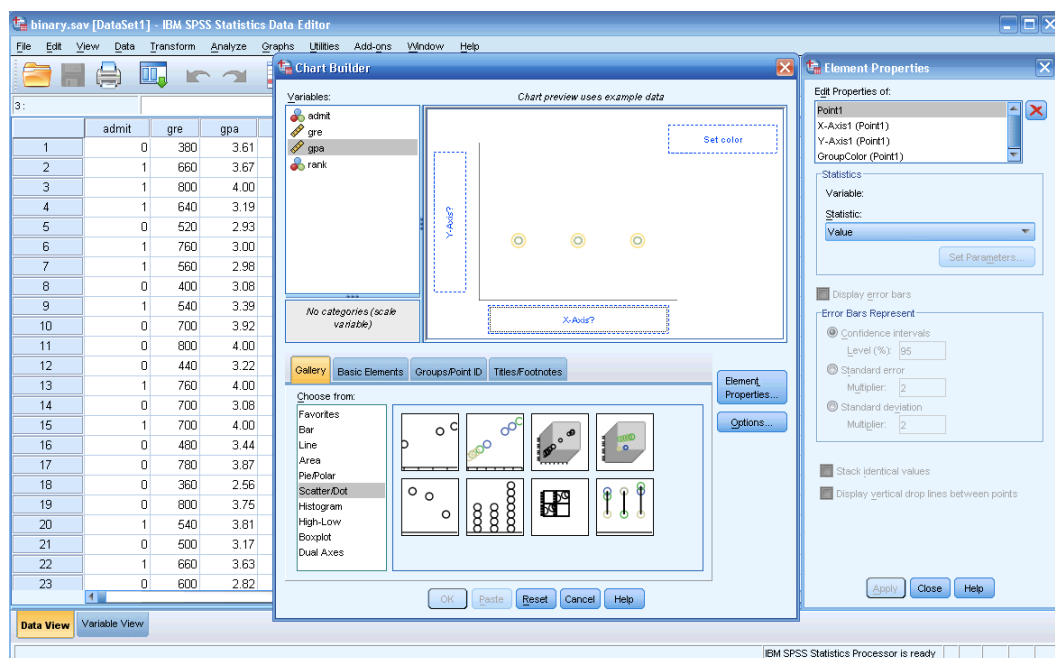


The output will appear in the Output Viewer box.

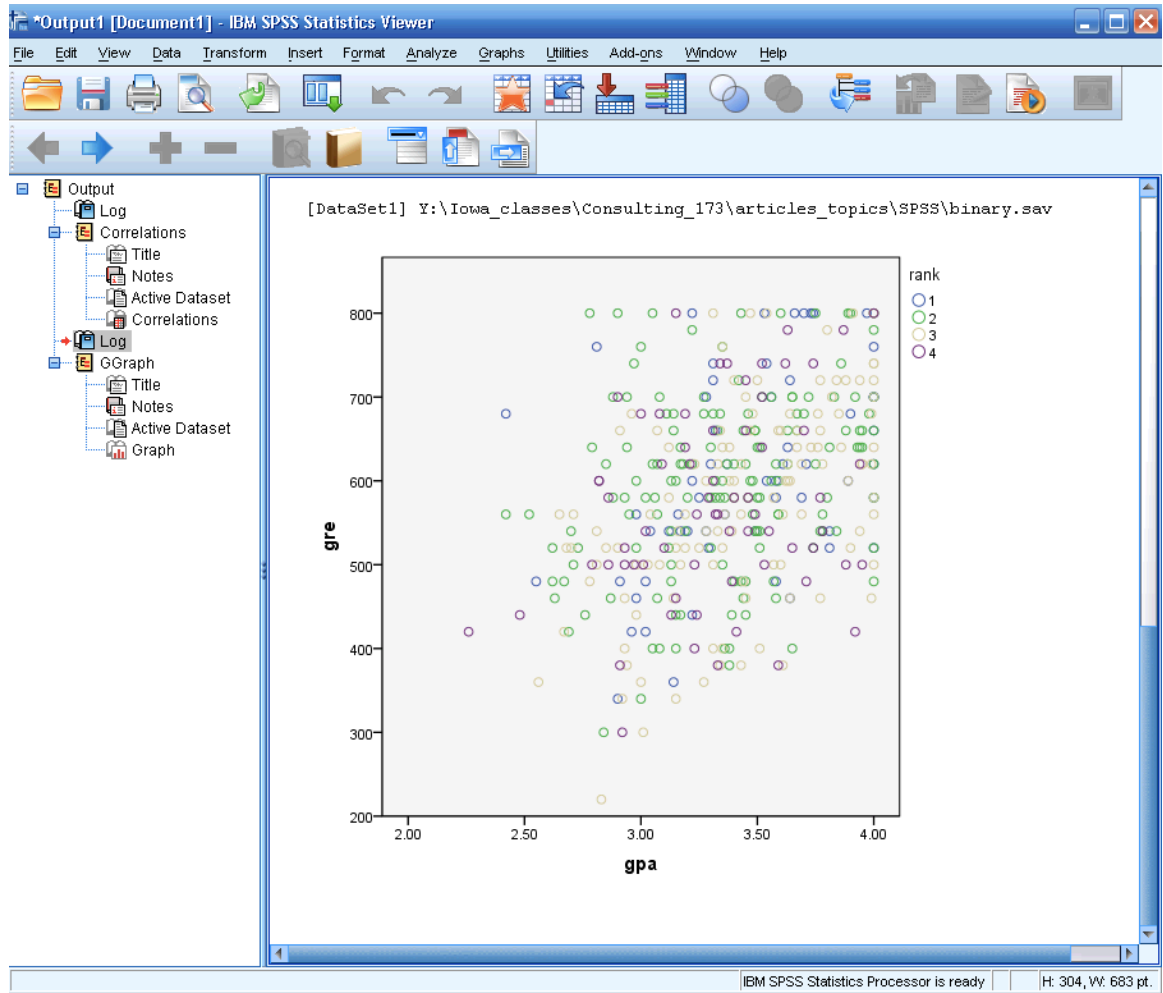


- Scatterplots

- **Graphs** → **Chart Builder**
- You’ll see the “Chart Builder” dialog box appear.
- Highlight *Scatter/dot* in the Gallery box.
- Drag the ‘Grouped Scatter’ visual (2<sup>nd</sup> visual on top) to the plotting area.



- Drag *gre* to the x-axis and *gpa* to the y-axis.
- Drag *rank* to the *set color* box.
- Click OK (output appears in the Output Viewer box).



- Standardizing Variables

- In SPSS, you can quickly standardize a variable and include the standardized ‘z-scores’ as another column in the data set.
- **Analyze** → **Descriptive Statistics** → **Descriptives...**
- Highlight *gre*, then click the arrow to move it to the *Variables* box at the right.
- Highlight *gpa*, then click the arrow to move it to the *Variables* box at the right.
- Check the box: “Save standardized values as variables.”
- Click OK (the new variables will appear in the Data View).

	admit	gre	gpa	rank	Zgre	Zgpa	var	var	var	var	var	var	var
1	0	380	3.61	3	-1.79801	.57835							
2	1	660	3.67	3	.62588	.73601							
3	1	800	4.00	1	1.83783	1.60314							
4	1	640	3.19	4	.45275	-.52527							
5	0	520	2.93	4	-.58606	-1.20846							
6	1	760	3.00	2	1.49156	-1.02452							
7	1	560	2.98	1	-.23979	-1.07708							
8	0	400	3.08	2	-1.62488	-.81431							
9	1	540	3.39	3	-.41293	.00026							
10	0	700	3.92	2	.97216	1.39292							
11	0	800	4.00	4	1.83783	1.60314							
12	0	440	3.22	1	-1.27860	-.44644							
13	1	760	4.00	1	1.49156	1.60314							
14	0	700	3.08	2	.97216	-.81431							
15	1	700	4.00	1	.97216	1.60314							
16	0	480	3.44	3	-.93233	.13165							
17	0	780	3.87	4	1.66470	1.26154							
18	0	360	2.56	3	-1.97115	-2.18070							
19	0	800	3.75	2	1.83783	.94622							
20	1	540	3.81	1	-.41293	1.10388							
21	0	500	3.17	3	-.75920	-.57782							
22	1	660	3.63	2	.62588	.63090							
23	0	600	2.82	4	.10648	-1.49750							

- Transforming or Combining Variables

- Highlight an empty column.
- **Transform** → **Compute Variable...**
- Create a new variable called ‘PreKnowledge’ which is the average of two standardized variables by inputting the formula:  $(Zgre+Zgpa)/2$ , then clicking OK.

The 'Compute Variable' dialog box is shown with the following details:

- Target Variable:** PreKnowledge
- Numeric Expression:** (Zgre + Zgpa)/2
- Function group:** All
- Functions and Special Variables:** (Empty list)
- Optional case selection condition:** If... (optional case selection condition)

IBM SPSS Statistics Data Editor window showing a dataset with 23 rows and 7 variables. The variables are admit, gre, gpa, rank, Zgre, Zgpa, and PreKnowledge. The data is displayed in a grid format.

	admit	gre	gpa	rank	Zgre	Zgpa	PreKnowledge
1	0	380	3.61	3	-1.79801	.57835	-.61
2	1	660	3.67	3	.62588	.73601	.68
3	1	800	4.00	1	1.83783	1.60314	1.72
4	1	640	3.19	4	.45275	-.52527	-.04
5	0	520	2.93	4	-.58606	-1.20846	-.90
6	1	760	3.00	2	1.49156	-1.02452	.23
7	1	560	2.98	1	-.23979	-1.07708	-.66
8	0	400	3.08	2	-1.62488	-.81431	-1.22
9	1	540	3.39	3	-.41293	.00026	-.21
10	0	700	3.92	2	.97216	1.39292	1.18
11	0	800	4.00	4	1.83783	1.60314	1.72
12	0	440	3.22	1	-1.27860	-.44644	-.86
13	1	760	4.00	1	1.49156	1.60314	1.55
14	0	700	3.08	2	.97216	-.81431	.08
15	1	700	4.00	1	.97216	1.60314	1.29
16	0	480	3.44	3	-.93233	.13165	-.40
17	0	780	3.87	4	1.66470	1.26154	1.46
18	0	360	2.56	3	-1.97115	-2.18070	-2.08
19	0	800	3.75	2	1.83783	.94622	1.39
20	1	540	3.81	1	-.41293	1.10388	.35
21	0	500	3.17	3	-.75920	-.57782	-.67
22	1	660	3.63	2	.62588	.63090	.63
23	0	600	2.82	4	.10648	-1.49750	-.70

- Logistic Regression

- Analyze → Generalized Linear Models → Generalized Linear Models...
- Type of model: *binary logistic*.
- Response: *admit*.
- Predictors: enter *rank* as a Factor, enter *gre*, *gpa* as Covariates.
- Model: Include desired terms in the model (main effects and interactions).
- Click OK.

IBM SPSS Statistics Viewer window showing the output of a logistic regression analysis. The dependent variable is 'admit'. The model includes the intercept, rank, gre, and gpa. The output displays Tests of Model Effects and Parameter Estimates.

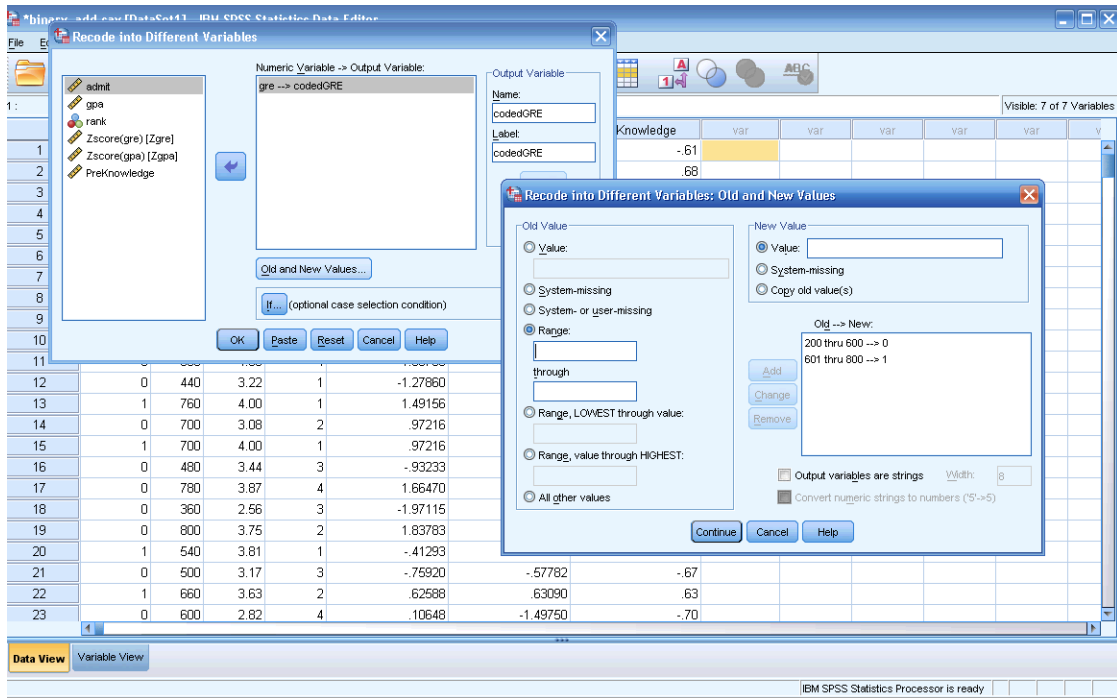
Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	19.234	1	.000
rank	20.895	3	.000
gre	4.284	1	.038
gpa	5.872	1	.015

Dependent Variable: admit  
Model: (Intercept), rank, gre, gpa

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	5.541	1.1381	3.311	7.772	23.709	1	.000
[rank=1]	-1.551	.4178	-2.370	-.733	13.787	1	.000
[rank=2]	-.876	.3667	-1.595	-.157	5.706	1	.017
[rank=3]	-.211	.3929	-.981	.559	.289	1	.591
[rank=4]	0 <sup>a</sup>						
gre	-.002	.0011	-.004	.000	4.284	1	.038
gpa	-.804	.3318	-1.454	-.154	5.872	1	.015

Dependent Variable: admit  
Model: (Intercept), rank, gre, gpa  
a. Set to zero because this parameter is redundant.  
b. Fixed at the displayed value.

- Dichotomizing a continuous variable
  - **Transform** → **Recode into Different Variables...**
  - You'll see the "Recode into Different Variables" dialog box appear.
  - Highlight *gre* and put into the *Variables* box.
  - Provide a new name in the 'Name' box, and press 'Change'.
  - Click on 'Old and New Values'.
  - Old Value Range: 200-600
  - New Value: 0
  - Click 'Add'
  - Old Value Range: 601-800
  - New Value: 1
  - Click 'Add', click 'Continue', click 'OK'.



- Two-way tables (categorical variables)
  - Analyze → Descriptive Statistics → Crosstabs...
  - Highlight *admit* and put into the *Rows* box.
  - Highlight *codedGRE* and put into the *Columns* box.
  - Click on ‘Cells...’ and choose Column Percentages, then Continue.
  - Click on ‘Statistics...’ and choose Chi-Squared, then Continue.

The screenshot shows the IBM SPSS Statistics interface. On the left, the 'Data View' window displays a dataset with 23 rows and 4 columns: admit, gre, gpa, and rank. The 'Output Viewer' window on the right displays the results of a crosstabs analysis.

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
admit * codedGRE	400	100.0%	0	0%	400	100.0%

**admit \* codedGRE Crosstabulation**

admit		codedGRE		Total
		.00	1.00	
0	Count	168	105	273
	% within codedGRE	74.3%	60.3%	68.3%
1	Count	56	69	127
	% within codedGRE	25.7%	39.7%	31.8%
Total	Count	226	174	400
	% within codedGRE	100.0%	100.0%	100.0%

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.881 <sup>a</sup>	1	.003		
Continuity Correction <sup>b</sup>	8.247	1	.004		
Likelihood Ratio	8.845	1	.003		
Fisher's Exact Test				.003	.002
Linear-by-Linear Association	8.859	1	.003		
N of Valid Cases	400				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 55.25.