Outline: Statistical Analysis for Psychological Testing

Sunthud Pornprasertmanit

Entering Data

1) Set up the structure of the data file by 'defining' the variables
2) Enter the data

Defining the Variables

1) In Data Editor Window, changing from Data View to Variable View
2) Defining each of your variables by specifying the required information for each variable
    a. Name
    b. Type (Numeric/String)
    c. Width
    d. Decimals
    e. Label
    f. Values
        i. Value
        ii. Value Label
        iii. Add/Remove/Change
    g. Missing
        i. No missing values
        ii. Discrete missing values
        iii. Range plus one optional discrete missing value
    h. Columns
    i. Align (Right/Left/Center)
    j. Measure (Nominal/Ordinal/Scale)
3) Shortcut in defining variables
    a. Copy and paste
    b. Paste Variables…

Entering Data

1) Go to Data View
2) Type in the data
3) Press Right arrow or Tab on the keyboard
4) Make sure that the values are entered in the correct columns
5) Move back to the first column by press Tab or Home or Ctrl + left arrow

Move an existing variable

1) Right click where you would like the moved variable to appear
2) Choose Insert Variable
3) An empty column, named "Var00001", will appear in which you can enter the new variable.
4) Right click on the variable name of the existing variable you wish to move.
5) Choose Cut
6) Right click on the empty variable, the destination, and choose Paste.

Modifying the data file

Sort cases

Method 1

1) Right click on the variable you want to sort its values.
2) Choose Sort Ascending (from least to most) or Sort Descending (from most to least)

Method 2

1) Data → Sort Cases…
2) Choose Ascending or Descending in Sort Order depend on the direction you want to sort in the first variable
3) Choose the first variable in Sort by blank.
4) If you want to sort second variable within the equal values in first variable, choose Sort Order and then the second variable.
5) For example, select sex (Ascending) and then age ascending

Sex	Age
1	30
1	25
2	29
2	21

Split File

To split your file

1) Data → Split File…
2) Click on "Compare Groups" or "Organize output by groups" (depending on table format you want to see) and specify the grouping variable
3) Click on OK
4) You will see "Split File on" on the right bottom on the window.

To turn the Split File off

1) Data → Split File…
2) Click on the first dot (Analyze all cases, do not create groups)
3) Click on OK
4) You will see "Split File on" disappear from the screen.

Select Cases

To Select Cases

1) Data → Select Cases…
2) Click on If conditions is satisfied
3) Click on the button labeled IF…
4) Choose the variable that defines the group that you are interested in
5) Click on the arrow button to move the variable name into the box
6) Type in the value that corresponds to the group you are interested in
7) Click on Continue, you will see your selected group on the right of IF… box
8) Click on OK
9) You will see "Filter on" on the right bottom of the window and the new variable, named filter_$

To turn the select cases option off

1) Data → Select Cases…
2) Click on the first dot (All cases)
3) Click on OK
4) You will see "Filter on" disappeared from the screen
5) Delete filter_$ from variable lists

Cleaning the Data

Checking for errors

There are many ways to check for errors. Each method is suitable for different kinds of variables. This method I show you is suitable for categorical and numerical variables, not for multiple response sets, set of variables that show respondent's alternative.

1) Analyze → Descriptive Statistics → Frequencies
2) Choose the variables that you wish to check (You can use Ctrl or Shift to help in selecting multiple variables)
3) Click on the arrow button to move these into Variables box
4) Ensure to check on Display frequency tables
5) Click on OK
6) SPSS will show you the Statistics Table and Frequency tables of each selected variables in the SPSS Output Viewer

a. Statistics Table will summarize each of the variables you requested
b. Frequency tables of each variable will range your responses

7) In each variable, check on impossible value and system missing
8) Go on next step, finding the error in the data file

Finding and Correcting the Error in the Data File

Method 1

1) Highlight the variable name of the variable in which the error has occurred
2) Edit → Find or Ctrl + F
3) In the Search for box, type in the incorrect value that you are looking for
4) Click on Search Forward box
5) Take note of the ID number of this case and then find out the correct value in the questionnaire.
6) Click on Search Forward to continue searching for other cases with the same incorrect value and find out the correct value to replace

Method 2

1) Sort cases with Method 1 in variable in which the error has occurred
2) If you sort ascending, the system missing will appear on the first rows.
3) Take note of the ID number of this case and then find out the correct value in the questionnaire.
4) Find out incorrect value that you are looking for, if they find out easily by sort ascending or sort descending.
5) Take note of the ID number of this case and then find out the correct value in the questionnaire.

Rechecking for Errors

After you have corrected your errors, it is a good idea to repeat Frequencies… to double-check. Sometimes in correcting one error, you may have accidentally caused another error.

Descriptive Statistics

Graphical Illustration

There are many ways to build graph. This procedure is the most sophisticated way to build graph.

Building Graph

1) Graph → Chart Builder…
2) There are two adjacent windows: Chart Builder and Element Properties (If the Element Properties window does not appear, click the Element Properties button on the bottom left of Chart Builder window)
3) Choose the graphical illustration that you want to present your data in the bottom of Chart Builder Window.
4) Drag variables you want to represent each component of the graph from Variable box to Destination
5) Edit each element in the graph by click on element in Edit Properties of box in Element Properties window
6) Change this element properties under Edit Properties Box (such as each axis label)
7) Click on OK

Edit Graph

1) Right click on graph that you want to edit
2) SPSS Chart Object → Open

Numerical Illustration

Numerical Illustration for one variable

Percentage for each category in categorical variable

1) Analyze → Descriptive Statistics → Frequencies…
2) Choose the variables that you wish to find out percentage of each category
3) Click on the arrow button to move these into Variables box
4) Click on OK

Mean, Median, Mode, SD, Minimum, Maximum, Sum

1) Analyze → Descriptive Statistics → Descriptive…
2) Choose the variables that you wish to find out these statistics
3) Click on the arrow button to move these into Variables box
4) Click on Options button
5) Choose statistics that you want to find out from these variables
6) Click on Continue
7) Click on OK

Quartile, Decide, Percentile

1) Analyze → Descriptive Statistics → Frequencies…
2) Choose the variables that you wish to find out quartiles, decides or percentiles
3) Click on the arrow button to move these into Variables box
4) Click on Statistics button
5) In Percentiles Values, you can choose quartiles, decides (cut point for 10 equal groups) or specific percentiles (fill the percentiles you want and then click on Add)

6) Click on Continue
7) Click on OK
8) This statistics will appear on Statistics table in SPSS Output.

Numerical Illustration for two variables

Percentage in each group

1) Analyze → Descriptive Statistics → Crosstabs…
2) Suppose that you want to find percentage of drinking alcohol in each sex. You can find it out by two alternatives

Method 1

3) Choose Drinking Alcohol in variables list and click on arrow to Row box
4) Choose Sex in variables list and click on arrow to Column box
5) Click on Cells… button
6) In Percentage, click on Column
7) Click on Continue
8) Click on OK

Method 2

3) Choose Sex in variables list and click on arrow to Row box
4) Choose Drinking Alcohol in variables list and click on arrow to Column box
5) Click on Cells… button
6) In Percentage, click on Row
7) Click on Continue
8) Click on OK

Mean, Median, Mode, SD, Min, Max, and Sum in each group

1) Analyze → Compare Means → Means…
2) Choose group variable in variables list and click on arrow to Independent List (If your grouping variables are combined in factorial design, you can choose another group variable to Independent List)
3) Choose numerical variables that you wish to find out these statistics in groups that you defined in Independent List and click on arrow to Dependent List
4) Click on Options…
5) Choose Statistics from list and click on arrow to Cell Statistics
6) Click on Continue
7) Click on OK

Screening the Data

Normality

In screening data, there are many ways to detect deviation from normality. However, SPSS provide the command for screening data.

1) Analyze → Descriptive Statistics → Explore
2) Choose the variable that you wish to examine the distribution
3) Click on the arrow button to move into Dependent List box
4) Click on Plots…
5) Check Normality plots with tests
6) Check Stem-and-Leaf and Histogram in Descriptive box
7) Click on Continue
8) Click on OK
9) SPSS output show stem-and-leaf display, histogram, Normal q-q plot, Detrended normal q-q plot, Skewness and Kurtosis in Descriptive Table, and Test of Normality Table

*Normal q-q plot is better than normal p-p plot because it is more sensitive in deviating from normality in the tail of distribution.

*The Shapiro-Wilk Test has superior power to other statistics in detecting that the data comes from a wide range of other distributions.

Outlier and Extreme Score

In screening data, there are many ways to detect deviation from normality.

Method 1

1) Analyze → Descriptive Statistics → Explore
2) Choose the variable that you wish to examine outlier and extreme score
3) Click on the arrow button to move into Dependent List box
4) Choose the variable that represent ID, if available
5) Click on the arrow button to move into Label Cases by
6) Click on Plots…
7) Choose Dependents together in Boxplots box
8) Click on Continue
9) Click on Statistics…
10) Check Outliers
11) Click on Continue
12) Click on OK
13) SPSS output show extreme values (both highest and lowest identified by ID number)
14) SPSS output also show boxplot of the variable
    a. The circles represent outlier (outside $Mdn \pm 2(Q_3 - Q_1)$)
    b. The asterisks represent extreme value (outside $Mdn \pm 3.5(Q_3 - Q_1)$)

Method 2

1) Analyze → Descriptive Statistics → Descriptive…
2) Choose the variable that you wish to examine outlier and extreme score
3) Click on the arrow button to move into Variable box
4) Check “Save standardized values as variables”

5) Click on OK
6) Change window to Data Editor. You will see new variable that represent standard score of selected variable.
7) The outlier is outside -2.5 to 2.5

Manipulating the data

Rank Cases

1) Transform → Rank Cases…
2) Choose the variable that you wish to rank
3) Click on the arrow button to move into Variable box
4) Choose Smallest value or Largest value in "Assign Rank 1 to" box
5) Click on Rank Types… button
6) Check Rank
7) Click on Continue
8) Click on Ties… button
9) Choose Mean, Low, High or Sequential ranks to unique values to customize your ranking
10) Click on Continue
11) Click on OK
12) You will see new variable that represent rank of selected variable in Data Editor Window.

Linear z-score

1) Analyze → Descriptive Statistics → Descriptive…
2) Choose the variable that you wish to examine outlier and extreme score
3) Click on the arrow button to move into Variable box
4) Check on "Save standardized values as variables"
5) Click on OK
6) Change window to Data Editor. You will see new variable that represent standard score of selected variable.

Linearly Transforming Variable to Percentile Rank

1) Transform → Rank Cases…
2) Choose the variable that you wish to transform
3) Click on the arrow button to move into Variable box
4) Click on Rank Types… button
5) Check "Fractional rank as %"
6) Click on Continue
7) Click on OK
8) You will see new variable that represent percentiles of selected variable of each case in Data Editor Window.

Normalized z-score

1) Transform → Rank Cases…
2) Choose the variable that you wish to transform
3) Click on the arrow button to move into Variable box
4) Click on Rank Types… button
5) Check “Normal scores”
6) In “Proportion Estimation Formula” box, choose the formula you want to use in normalizing. The default formula is Blom’s formula.
7) Click on Continue
8) Click on OK
9) You will see new variable that represent normalized standard score in Data Editor Window.

Normalized Percentiles

1) Transform → Rank Cases…
2) Choose the variable that you wish to transform
3) Click on the arrow button to move into Variable box
4) Click on Rank Types… button
5) Check “Proportion Estimation Formula”
6) In “Proportion Estimation Formula” box, choose the formula you want to use in normalizing. The default formula is Blom’s formula.
7) Click on Continue
8) Click on OK
9) You will see new variable that represent normalized proportion in Data Editor Window.
10) Transform → Compute Variable…
11) Name the new variable, representing normalized percentile, in Target Variable.
12) In Numeric Expression box, type the name of normalized proportion variable and then follow with “* 100*”
13) Click on OK
14) You will see new variable that represent normalized percentile.

Dividing Groups from Smallest to Largest

1) Transform → Rank Cases…
2) Choose the variable that you wish to divide
3) Click on the arrow button to move into Variable box
4) Click on Rank Types… button
5) Check “Ntiles” and type in the number of groups you want to divide
6) Click on Continue
7) Click on OK
8) You will see new variable that represent grouping in Data Editor Window.

Recoding Value in each Variable

You can use Recode command to Utilize

1) Transform → Recode into Different Variables…
2) Choose the variable that you wish to recode
3) Click on the arrow button
4) Name and label the new variable in Output Variable Box
5) Click on Change
6) Click on "Old and New Values…"
7) Define the old value you want to change in Old Value box
8) Define the new value you want to replace the old value in New Value box.
9) Click Add, then this changing will appear on "Old → New:" box.
10) Define another old value, new value, and then click Add.
11) You can Change or Remove your command, until finishing.
12) Click on Continue
13) Click on OK
14) You will see new variable appear in Data Editor

Calculating data from variables in each case

1) Transform → Compute Variable…
2) Name the new variable in Target Variable.
3) In Numeric Expression box, type the command that you want to calculate from variable in your list.
4) Click on OK
5) You will see new variable appear in Data Editor.

T-test

One Sample T-test

1) Analyze → Compare Means → One-sample T test…
2) Click on the tested variable
3) Type test value
4) Click on OK

Paired-Samples T-test

1) Analyze → Compare Means → Paired-Samples T test…
2) Click on the first variable you want to compare, such as pretest. Then, the clicked variable will appear after Variable 1: in Current Selections box.
3) Click on the second variable you want to compare, such as posttest. Then, the clicked variable will appear after Variable 2: in Current Selections box.
4) Click on the arrow button
5) Click on OK

Independent-Samples T-test

1) Analyze → Compare Means → Independent-Samples T test…
2) Click on the dependent variable
3) Click on the arrow button to move into Test Variables(s): box
4) Click on the independent variable
5) Click on the arrow button to move into Grouping Variables: box
6) Click on Define Groups… button
7) Type in the value of both groups you want to compare in Group 1 and Group 2 boxes.
8) Click on continue
9) Click on OK

Correlation

Pearson Product Moment Correlation and Point-biserial Correlation

1) Analyze → Correlate → Bivariate…
2) Select variables to compute correlation (if you select more than two variables, the SPSS will provide you a correlation matrix of each pair of variables.)
3) Select two-tailed or one-tailed in Test of Significance box
4) Click on OK

Phi Coefficient

1) Analyze → Descriptive Statistics → Crosstabs…
2) Select the first dichotomous variable into Rows box.
3) Select the second dichotomous variable into Column box.
4) Click on Statistics… button
5) Check “Phi and Cramer’s V” in Nominal box
6) Click on Continue
7) Click on OK

Chi-square

1) Analyze → Descriptive Statistics → Crosstabs…
2) Select the first dichotomous variable into Rows box.
3) Select the second dichotomous variable into Column box.
4) Click on Statistics… button
5) Check Chi-square in the top left of window
6) Click on Continue
7) Click on OK

Simple Regression

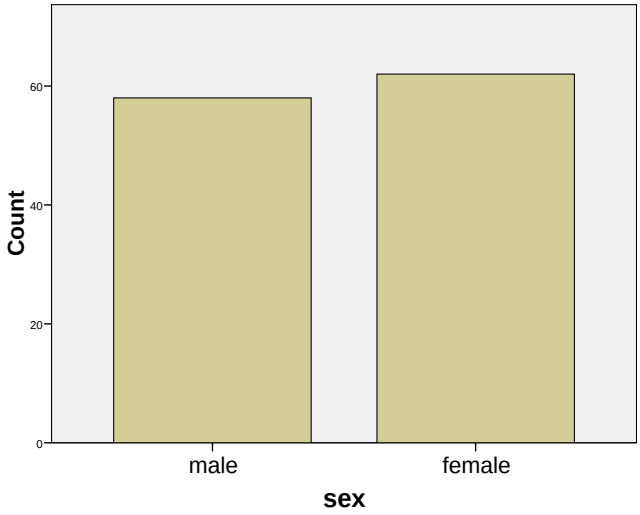
1) Analyze → Regression → Linear…
2) Click on the dependent variable
3) Click on the arrow button to move into Dependent box
4) Click on the independent variable
5) Click on the arrow button to move into Independent box
6) Click on OK

SPSS Output

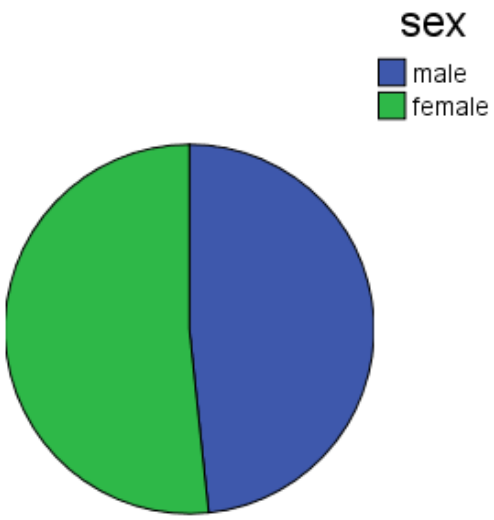
In these printouts, the variables are sex (male, female), education (high school, undergraduate), family support , couple support, and friend support (possible minimum = 5 and maximum = 20).

Descriptive Statistics: Graphical Illustration

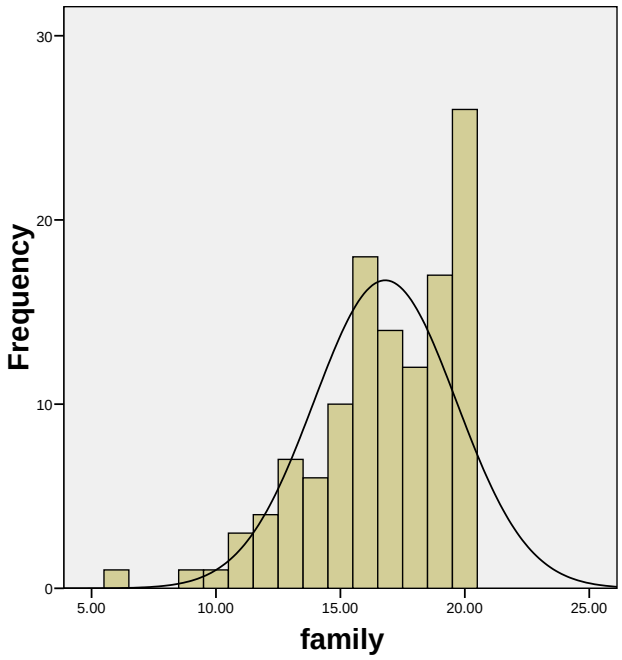
Bar Graph



Pie Chart



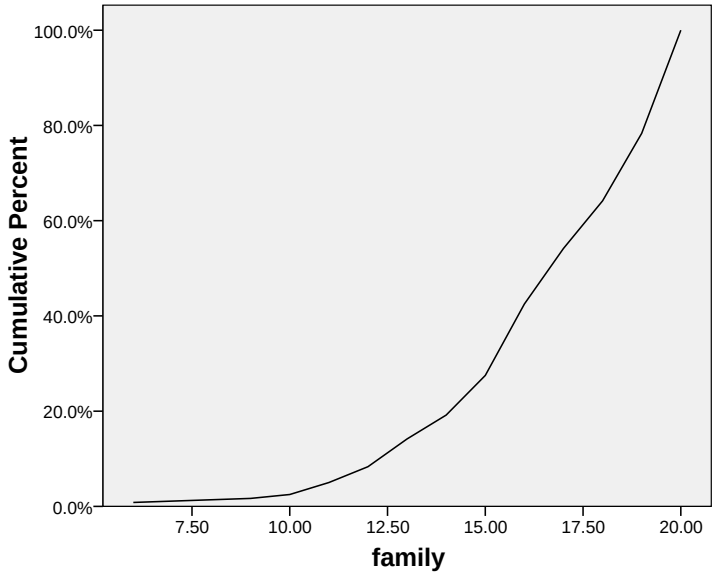
Histogram



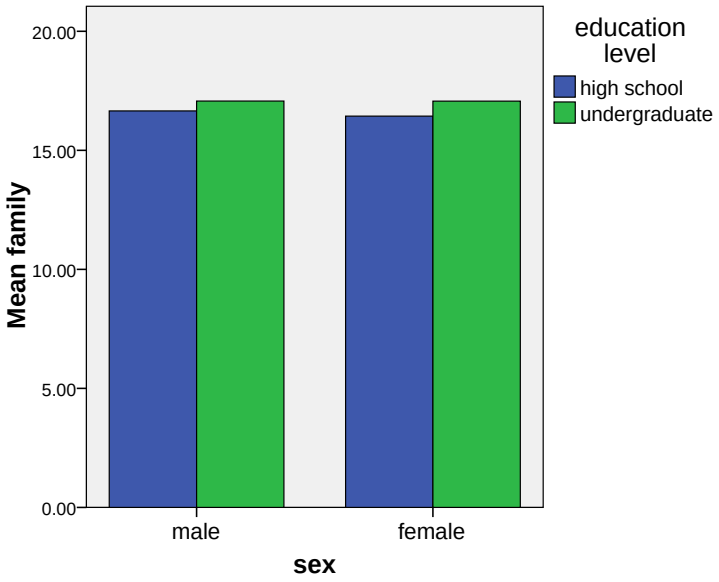
Stem-and-Leaf Display

```
family Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

     1.00 Extremes    (=<6.0)
     1.00         9 .  0
     1.00        10 .  0
     3.00        11 .  000
     4.00        12 .  0000
     7.00        13 .  0000000
     6.00        14 .  000000
    10.00        15 .  0000000000
    18.00        16 .  000000000000000000
    14.00        17 .  00000000000000
    12.00        18 .  000000000000
    17.00        19 .  00000000000000000
    26.00        20 .  00000000000000000000000000

 Stem width:      1.00
 Each leaf:       1 case(s)
```

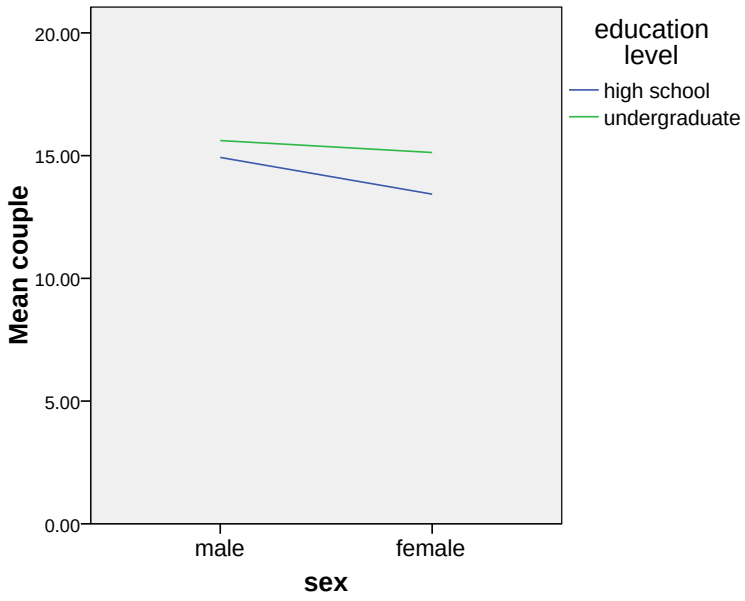
Cumulative Frequencies Distribution



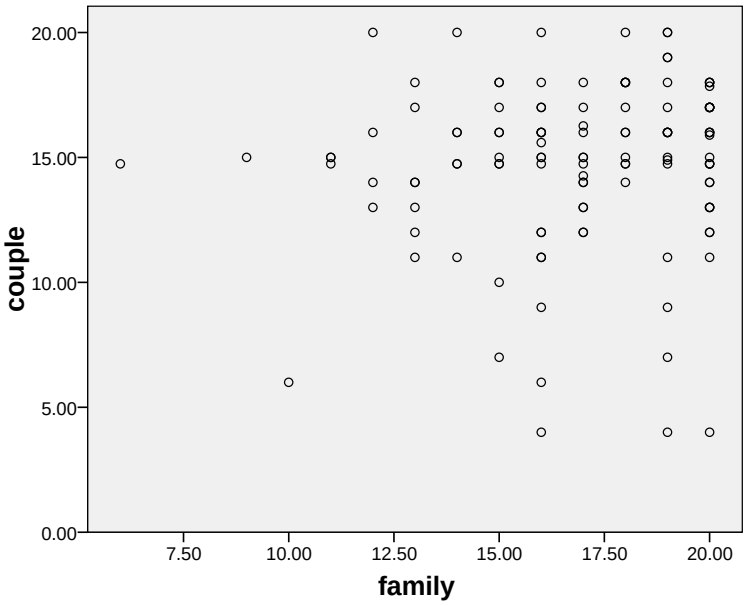
Bar Graph with combination



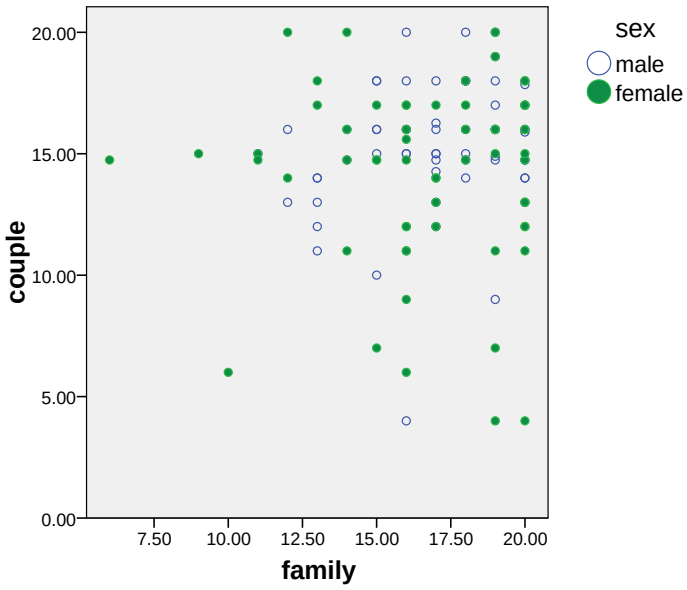
Line Graph representing Interaction



Scatterplot



Scatterplot with multiple type of dots



Descriptive Statistics: Numerical Illustration

Frequencies Distribution

Statistics

sex

N	Valid	119
	Missing	1

sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	57	47.5	47.9	47.9
	female	62	51.7	52.1	100.0
	Total	119	99.2	100.0	
Missing	99	1	.8		
Total		120	100.0		

Descriptive

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
family	120	6.00	20.00	2016.00	16.8000	2.86268	8.195
couple	120	4.00	20.00	1769.16	14.7430	3.32347	11.045
Valid N (listwise)	120						

Quartile, Decide, Percentile

Statistics

family

N	Valid	120
	Missing	0
Percentiles	10	13.0000
	20	15.0000
	25	15.0000
	30	16.0000
	40	16.0000
	50	17.0000
	60	18.0000
	65	19.0000
	70	19.0000
	75	19.0000
	80	20.0000
	90	20.0000

Percentage in each group

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sex * education level	120	100.0%	0	.0%	120	100.0%

sex * education level Crosstabulation

			education level		Total
			high school	undergraduate	high school
sex	male	Count	29	29	58
		% within education level	47.5%	49.2%	48.3%
	female	Count	32	30	62
		% within education level	52.5%	50.8%	51.7%
Total		Count	61	59	120
		% within education level	100.0%	100.0%	100.0%

Descriptive Statistics in each group

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
family * sex	120	100.0%	0	.0%	120	100.0%

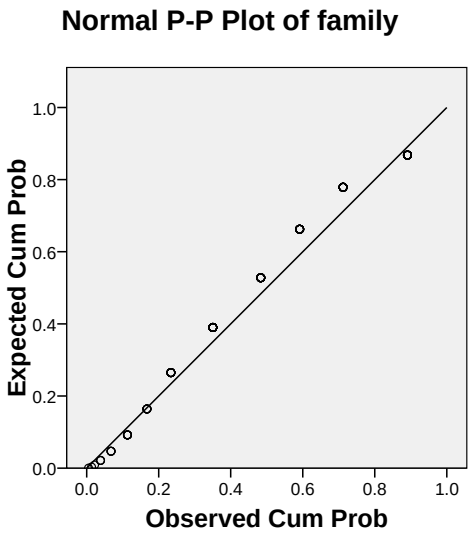
Report

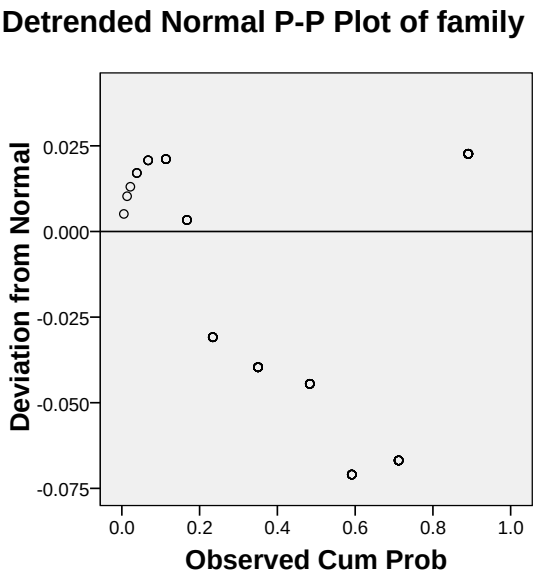
family

sex	Mean	N	Std. Deviation	Minimum	Maximum
male	16.8621	58	2.43835	11.00	20.00
female	16.7419	62	3.22864	6.00	20.00
Total	16.8000	120	2.86268	6.00	20.00

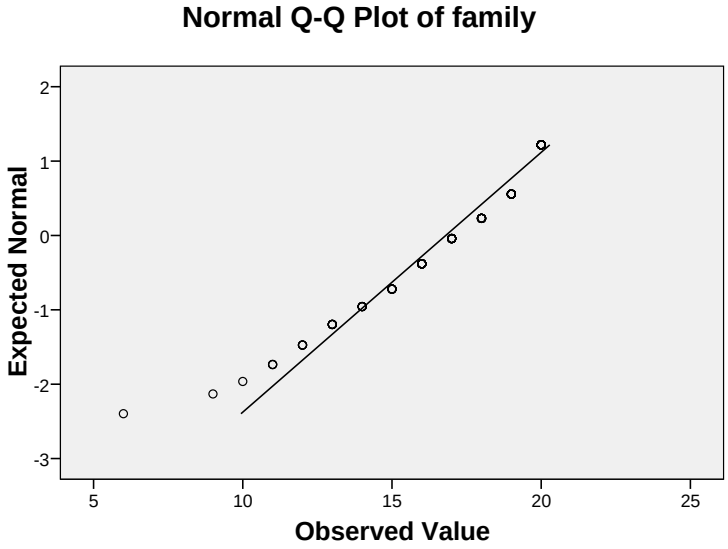
Checking Normality

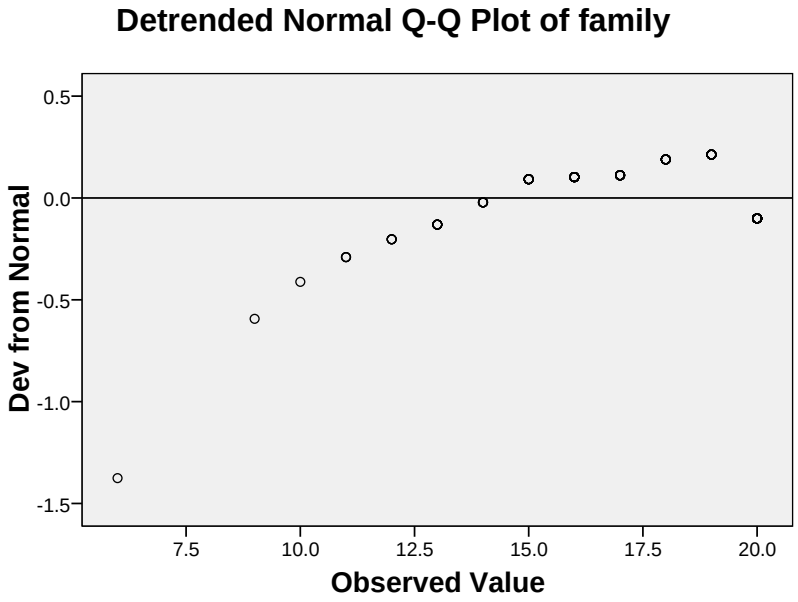
Normal P-P Plot





Normal Q-Q Plot





Skewness and Kurtosis

Descriptives

			Statistic	Std. Error
family	Mean		16.8000	.26133
	95% Confidence Interval for Mean	Lower Bound	16.2825	
		Upper Bound	17.3175	
	5% Trimmed Mean		17.0185	
	Median		17.0000	
	Variance		8.195	
	Std. Deviation		2.86268	
	Minimum		6.00	
	Maximum		20.00	
	Range		14.00	
	Interquartile Range		4.00	
	Skewness		-.949	.221
	Kurtosis		.878	.438

Kolmogorov-Smirnov Test and Shapiro-Wilk Test

Tests of Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
family	.137	120	.000	.906	120	.000

a Lilliefors Significance Correction

Checking Outliers and Extreme Scores

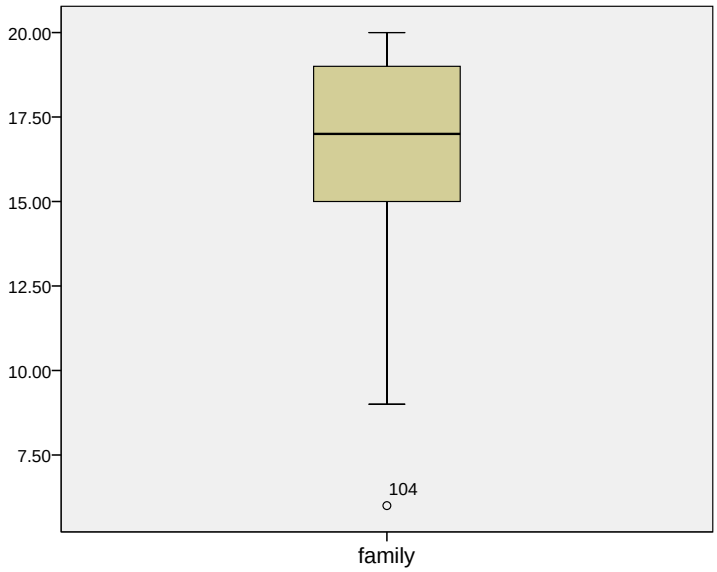
Extreme Value

Extreme Values

			Case Number	number	Value
family	Highest	1	11	59	20.00
		2	12	99	20.00
		3	20	10	20.00
		4	22	115	20.00
		5	23	116	20.00(a)
	Lowest	1	104	52	6.00
		2	57	114	9.00
		3	50	30	10.00
		4	93	76	11.00
		5	74	39	11.00(b)

a Only a partial list of cases with the value 20.00 are shown in the table of upper extremes.
b Only a partial list of cases with the value 11.00 are shown in the table of lower extremes.

Boxplot



Inferential Statistics

One-sample T-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
family	120	16.8000	2.86268	.26133

One-Sample Test

	Test Value = 15					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
family	6.888	119	.000	1.80000	1.2825	2.3175

Paired-sample T-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	family	16.8000	120	2.86268	.26133
	couple	14.7430	120	3.32347	.30339

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	family	16.8000	120	2.86268	.26133
	couple	14.7430	120	3.32347	.30339

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	family - couple	2.05704	4.19622	.38306	1.29854	2.81554	5.370	119	.000

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
family	male	58	16.8621	2.43835	.32017
	female	62	16.7419	3.22864	.41004

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
family	Equal variances assumed	3.558	.062	.229	118	.819	.12013	.52504	-.91958	1.15985
	Equal variances not assumed			.231	113.075	.818	.12013	.52023	-.91053	1.15080

Correlation Matrix

Correlations

		family	friend	couple
family	Pearson Correlation	1	.285(**)	.086
	Sig. (2-tailed)		.002	.352
	N	120	120	120
friend	Pearson Correlation	.285(**)	1	.396(**)
	Sig. (2-tailed)	.002		.000
	N	120	120	120
couple	Pearson Correlation	.086	.396(**)	1
	Sig. (2-tailed)	.352	.000	
	N	120	120	120

** Correlation is significant at the 0.01 level (2-tailed).

Chi-Square and Phi Coefficient

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sex * education level	120	100.0%	0	.0%	120	100.0%

sex * education level Crosstabulation

			education level		Total
			high school	undergraduate	high school
sex	male	Count	29	29	58
		Expected Count	29.5	28.5	58.0
	female	Count	32	30	62
		Expected Count	31.5	30.5	62.0
Total		Count	61	59	120
		Expected Count	61.0	59.0	120.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.031(b)	1	.860		
Continuity Correction(a)	.000	1	1.000		
Likelihood Ratio	.031	1	.860		
Fisher's Exact Test				1.000	.502
Linear-by-Linear Association	.031	1	.860		
N of Valid Cases	120				

a Computed only for a 2x2 table
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 28.52.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.016	.860
	Cramer's V	.016	.860
N of Valid Cases		120	

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Simple Regression

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	couple(a)	.	Enter

a All requested variables entered.
b Dependent Variable: family

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.086(a)	.007	-.001	2.86419

a Predictors: (Constant), couple

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.175	1	7.175	.875	.352(a)
	Residual	968.025	118	8.204		
	Total	975.200	119			

a Predictors: (Constant), couple
b Dependent Variable: family

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	15.711	1.194		13.161	.000
	couple	.074	.079	.086	.935	.352

a Dependent Variable: family