DATA ANALYSIS WITH SPSS

- An SPSS Lab Book -

Hanh Nguyen
THE CITY COLLEGE OF NEW YORK

Contents

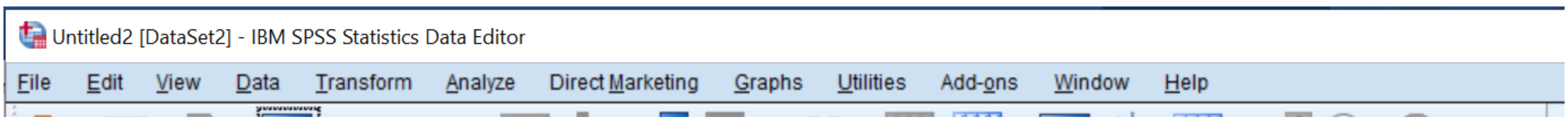
LAB 1: INTRODUCTION TO SPSS

What is SPSS?

- A computer-based program that is commonly used in social science and research for data entry and data analysis.
- Capable of handling large amount of data.

Opening SPSS – The Menu Bar

Menu Bar



- File: includes all of the crucial options, such as open, save, exit or export.
- Edit: includes the typical cut, copy, and paste commands, as well as various options for displaying data and output.
- View: giving options of which toolbars to show, font size, add or remove or how to display raw data and labels.
- Data: giving option to display data sorted by a specific variable to selecting certain cases for subsequent analyses.
- Transform: includes several options to change current variables, such as from continuous variables to categorical variables, scores into rank scores, or adding a constant to variables.
- Analyze: includes all of the commands to carry out statistical analyses and to calculate descriptive statistics. Much of this book will focus on using commands located in this menu.
- Graphs: includes the commands to create various types of graphs including box plots, histograms, line graphs, and bar charts.
- Utilities: list file information which is a list of all variables, there labels, values, locations in the data file, and type.
- Add-ons: programs that can be added to the base SPSS package.
- Window: can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax).
- Help: useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide.

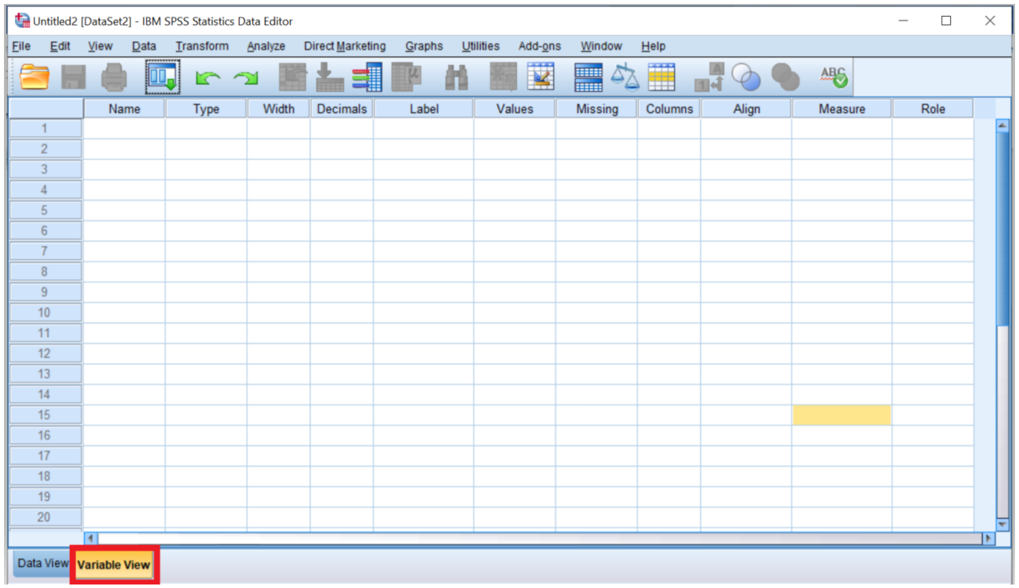
Icon Bar

- Directly under the Menu bar provide shortcuts to many common commands that are available in menus.

SPSS Lay-out

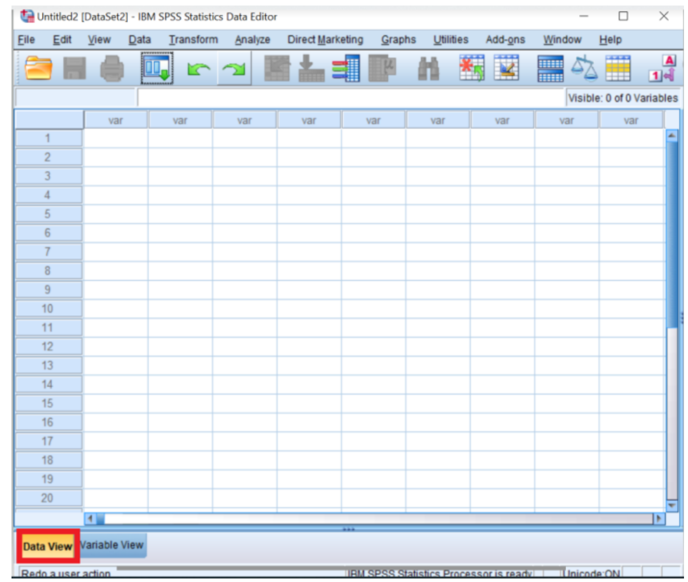
Data Editor

Variable View



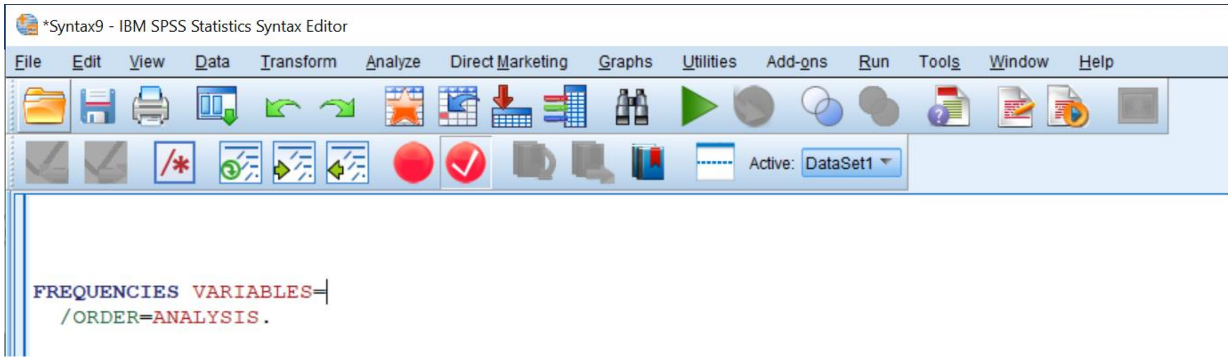
- An overview of each variable in the dataset.
- Each row contains information of a variable, including the variable name, type, various properties of the way in which the data are stored, any label(s) for the variable itself and variable values.
  - Name: Name of the variable.
  - Type: Type of the variable (numeric/string/date/…)
  - Decimals: how many decimals the data of the variable should have.
  - Labels: Detail description of the variable.
  - Values: specific notations needed for the variable.
  - Missing: notation for missing value.
  - Measure: scale of measurement for the variable (Nominal/Ordinal/Scale).

Data View



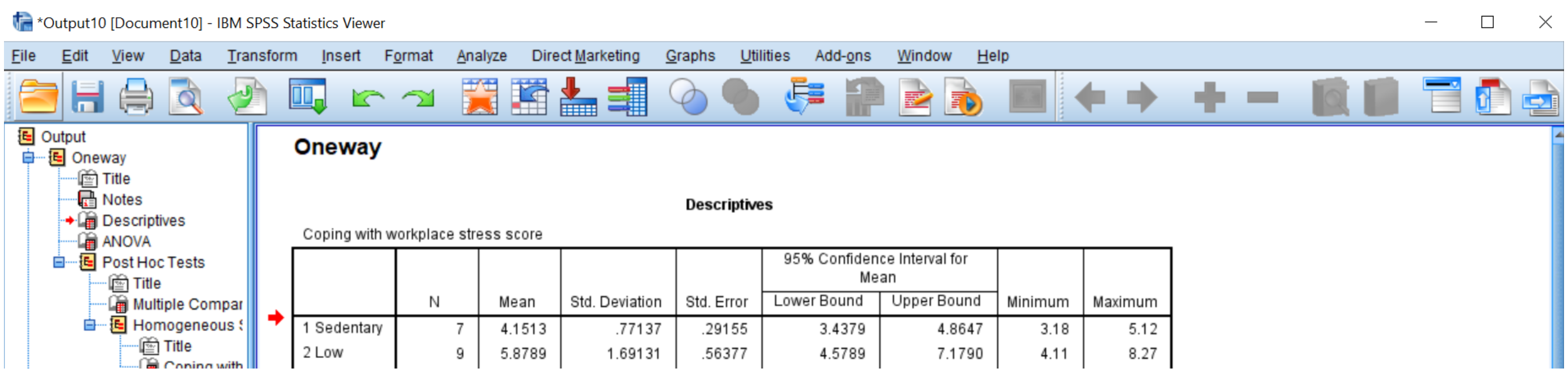
- The actual data in your dataset for each record and each variable.
- Each column represents a variable.
  - The column name will appear as the variable name from Variable View
- Each cell contains data for the column it belongs to.

Syntax Editor



- Where commands in syntax will be printed and run, such as opening files, editing and managing data, undertaking statistical procedures and tests, and saving files.
- Syntax Editor is useful and convenient to directly type in command code to let SPSS know what we want it to do with our data.

Output Viewer



- Keep a record of any commands and results run in SPSS
- Allowing us to view the results of any data/statistical procedures undertaken.

LAB 2: SCALE OF MEASUREMENT & DATA ENTRY IN SPSS

What is Scale of Measurement?

- The way that variables are defined and categorized.
- There are four (4) scales of measurement, each with its own properties which are used to determine the type of statistical analyses.

Qualitative Measurement		Quantitative/Scales Measurement	
Nominal	Ordinal	Interval	Ratio
• Categorical data and numbers that are simply used as identifiers or names Ex: • Gender • Blood type	• Data in ordered series of relationships or rank order Ex: • Level of pain • Educational level	• Data represents quantity and has equal units • Zero represents simply an additional point of measurement is an interval scale Ex: • Temperature in Celsius • pH level	• Similar to the interval scale in that it also represents quantity and has equality of units. • However, this scale also has an absolute zero (no numbers exist below the zero) Ex: • Weight • Length

Data Entry in SPSS Step by Step

Based on what you have learned so far, create a dataset in SPSS as following:

Awards data set

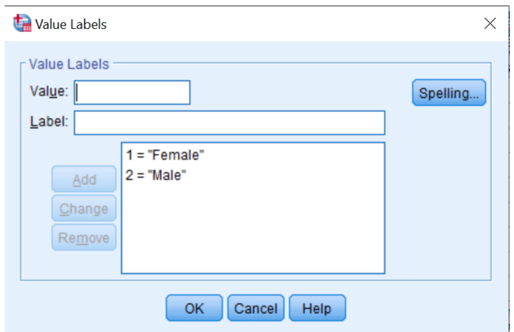
Seven (7) students are randomly selected to take a pre-designed math test for the study. After taking the test, the students were asked to rank the difficulty of the test. The Math scores and the ranking survey are used to predict the number of awards that they can earn.

- Gender: 1 – female, 2 – male
- Difficulty ranking: 1 – comfortable, 2 – moderate, 3 - hard

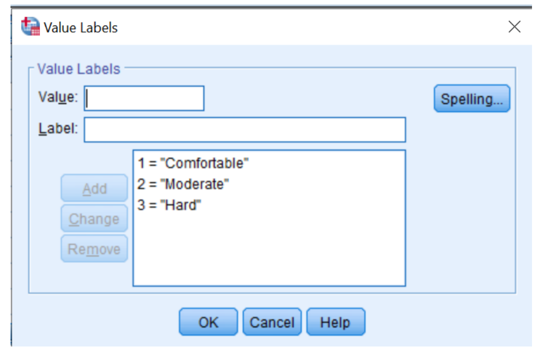
Participant gender	Number of awards	Math Scores	Difficulty ranking
1	1	450	2
2	0	300	3
2	1	400	3
2	2	480	2
1	0	380	3
1	1	460	1
2	3	500	1

1. Open SPSS
2. Variable View: enter Variables and their information.
    - Enter variable name for all four (4) variables
    - Since all data is in whole number, the type of all variable is numeric with zero decimals.
    - Value:

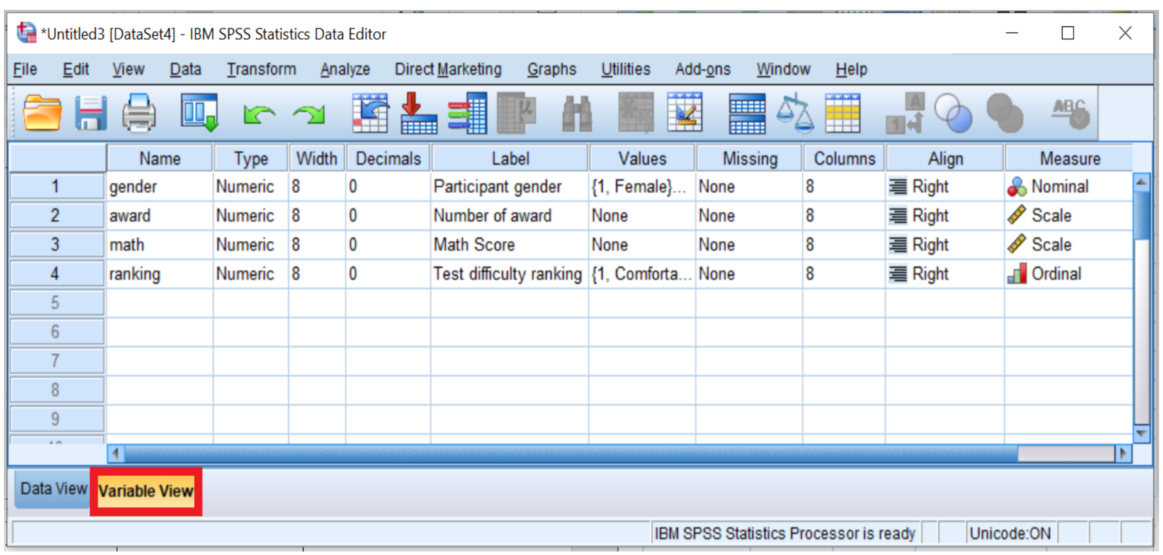
Gender: 1 for female and 2 for male

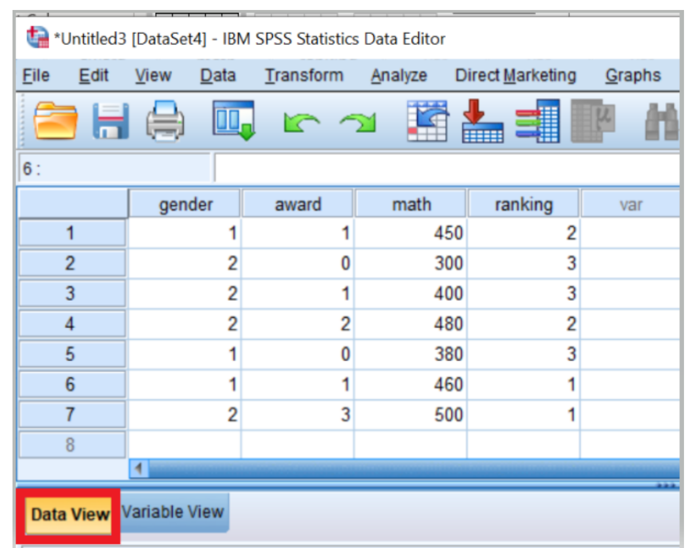


Ranking: 1 for comfortable, 2 for moderate, and 3 for hard



    - Measurement: enter the scale of measurement of each variable based on what you have learned
3. Data View: enter data for each variable

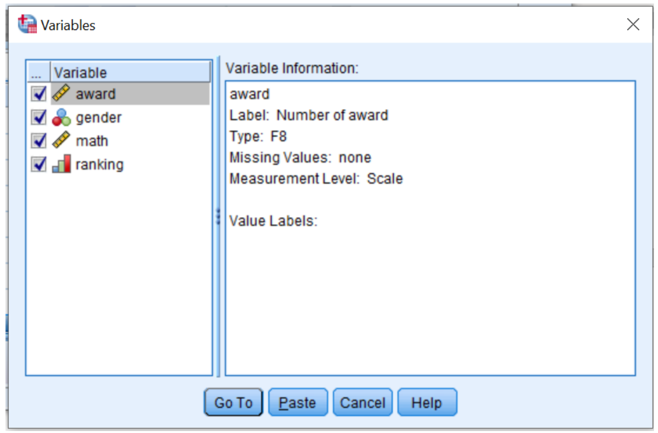




Viewing Variable in SPSS

During your data analysis, it is possible to view the variable information in detail without going back to Variable View to avoid any accidental alteration in the original data by the following pathway:

Utilities Tab – Variables – Select the <u>desired variable</u>



Practice

Now, let spend some time to review what we just learned and apply them to the exercise below by entering the dataset in SPSS.

The dataset includes information of 9 different football players, with:

- Age: the player's age
- College: The College that the player comes from
- YrsInPros: the number of years the player plays as a professional
- Height: the player's height in meter
- Weight: the player's weight in kilogram

*with 99 indicates missing values

Age	College	YrsInPros	Height (m)	Weight (kg)
24	Arizona	1	1.8796	89.81221
23	Louisville	1	1.9304	120.2032
27	Michigan	3	1.9558	122.9248
24	Iowa	2	1.9812	127.9144
25	Texas Tech	3	1.8542	86.63703
23	Georgia	0	1.8542	97.52336
29	Colorado	6	1.8542	105.6881
25	Pittsburgh	99*	1.9304	125.6464
25	Tennessee State	2	1.9558	120.2032

LAB 3: FREQUENCY DISTRIBUTION & BASIC DATA MANIPULATION IN SPSS

What is Frequency Distribution?

- Definition: A frequency distribution is an organized tabulation of the number of individuals located in each category on the scale of measurement.
- Elements:
  1. The set of categories that make up the original measurement scale.
  2. A record of frequency or number of individuals in each category.
- Purpose: help researchers read data in an organized way.

Frequency distribution table						
Score X	Frequency f	fX	$Proportion\ p = \frac{f}{N}$	$\text{Percentage}\ \% = 100\text{p} = 100\frac{f}{N}$	Cumulative frequency cf	Cumulative percentage c% $c\% = \frac{cf}{N}100\%$
Organize score so that the larger scores are at the top.	Frequencies of each score $\sum f = N$	Product of score X and its frequency. $\sum X = \sum fX$	The fraction of the total group that is associated with each score, also known as relative frequency.		The sum of the frequency in and below that category.	

Creating Table of Frequency Distribution in SPSP

Analyze – Descriptive Statistics – Frequency
- ***Select the desired Variable***

OR

```
DATASET ACTIVATE <DataSetName>.
FREQUENCIES VARIABLES = <VariableName>
 /ORDER=ANALYSIS.
```

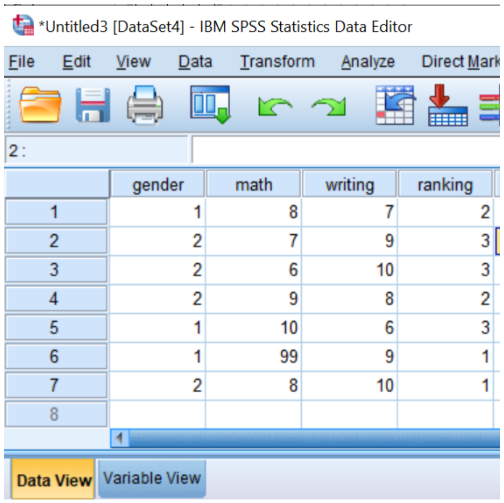
Data Manipulation in SPSS

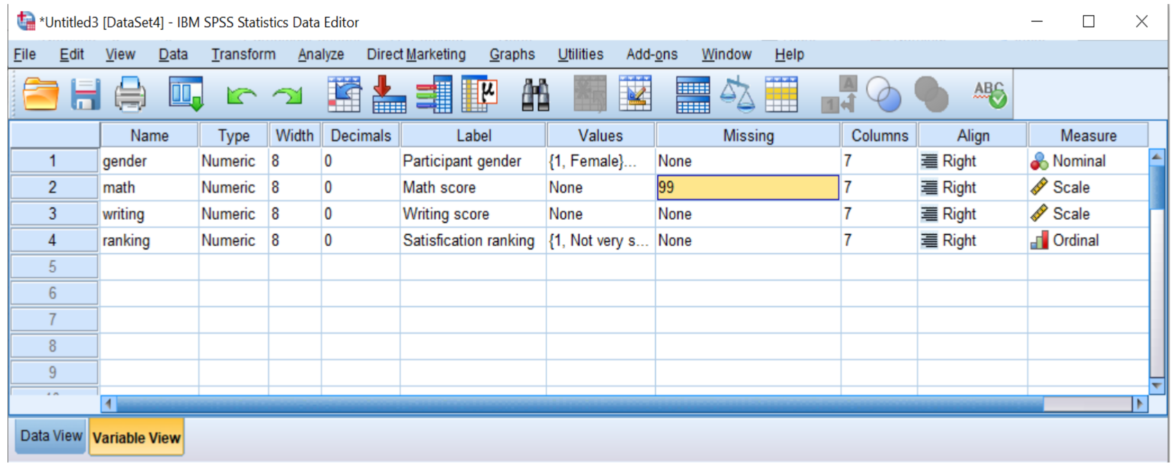
- Data manipulation is the process of changing data to make it easier to read or be more organized.
- In SPSS, it is useful for us to know basic techniques such as filtering data and computing new variable.

Filtering Data	Computing New Variable
Data – Select Cases • Select “If condition is satisfied” ○ Choose Variable – Enter condition – OK	Transform – Compute Variable • Target: New Variable Name • Numeric Expression: the formula for calculation COMPUTE <NewVariableName> = <NumericExpression>. EXECUTE.

Example

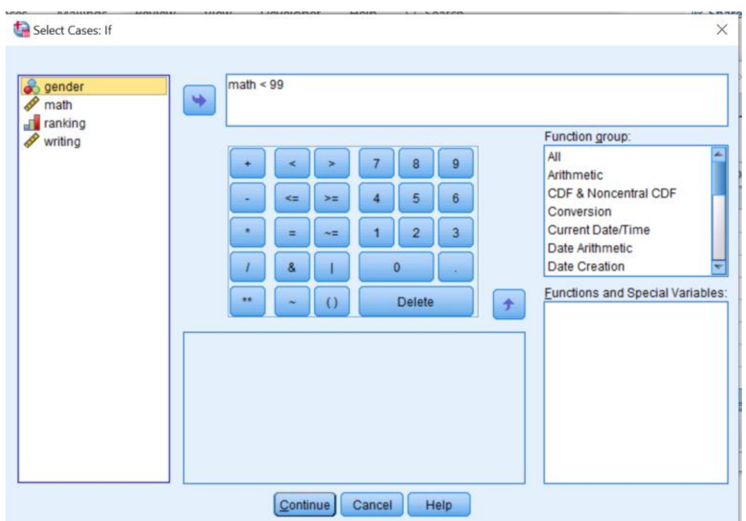
A group of seven (7) students in tenth grade are selected randomly to participate in a new method of testing, containing two parts: mathematics and writing. Based on the total result of the whole test, students are asked to rate their satisfaction. The data of the study are recorded in SPSS as below:

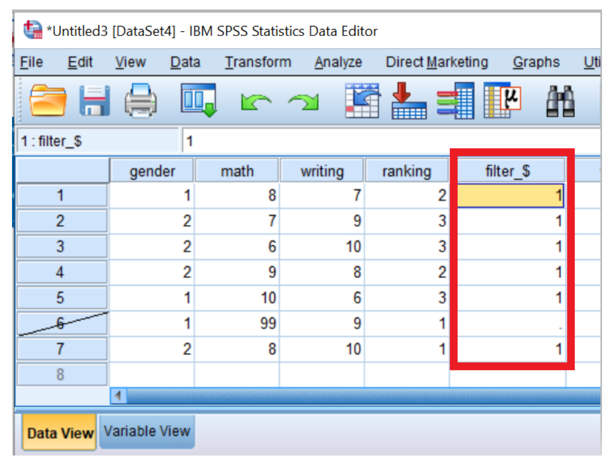




- Gender: 1 for “female,” 2 for “male”
- Ranking: 1 for “Not very satisfied,” 2 for “Satisfied,” 3 for “Very satisfied”
- 99 indicates missing value

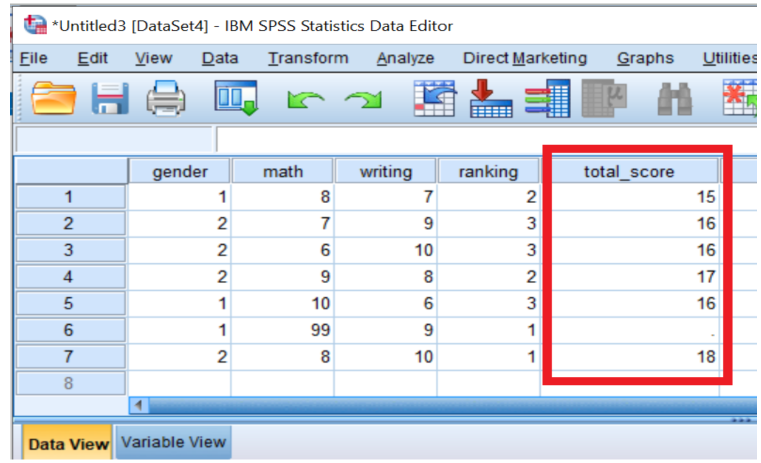
- **Filtering missing data**





- **Computing Total Score as the sum of math score and writing score**

```
COMPUTE total_score=math + writing.
EXECUTE.
```



	gender	math	writing	ranking	total_score
1	1	8	7	2	15
2	2	7	9	3	16
3	2	6	10	3	16
4	2	9	8	2	17
5	1	10	6	3	16
6	1	99	9	1	.
7	2	8	10	1	18
8					

Practice

Now, let spend some time to review what we just learned and apply them to the exercise below:

The dataset includes information of 9 different football players, with:

- Age: the player's age
- College: The College that the player comes from
- YrsInPros: the number of years the player plays as a professional
- HtFt: the player's height in feet
- HtIn: the player's height in inch
- Wt: the player's weight in pound

*with 99 indicates missing values

Age	Wt	HtFt	HtIn	College	YrsInPros
24	198	6	2	Arizona	1
23	265	6	4	Louisville	1
27	271	6	5	Michigan	3
24	282	6	6	Iowa	2
25	191	6	1	Texas Tech	3
23	215	6	1	Georgia	0
29	233	6	1	Colorado	6
25	277	6	4	Pittsburgh	99*
25	265	6	5	Tennessee State	2

a. Compute variable Height as the player's height in meter
b. Compute variable Weight as the player's weight in kilogram
c. Filter missing value
d. Creating the frequency distribution table for variable YrsInPros

LAB 4: CENTRAL TENDENCY & VARIABILITY

What is Central Tendency?

- a statistical measure to determine a single score that defines the center of the distribution.
- Goal: find a single score that is most typical or most representative of the entire a group.

<table>
<tr><th colspan="2">Mean (M)</th><th>Median (Mdn)</th><th>Mode</th></tr>
<tr><td colspan="2">The mean for a distribution is the sum of the scores divided by the number of scores.</td><td rowspan="3">The median is the midpoint of the list that locates at 50 percentiles of the distribution.</td><td>The mode is the score or category that has the greatest frequency.</td></tr>
<tr><td>Population Mean
$\mu = \frac{\sum X}{N}$</td><td>Sample Mean
$\text{M or } \bar{X} = \frac{\sum X}{n}$</td><td>2 modes = bimodal
≥2 modes = multimodal</td></tr>
<tr><td colspan="2">Weighted Mean = Overall Mean
$\mathrm{M} = \frac{\sum X_1 + \sum X_2}{n_1 + n_2}$ or $M = \left(\frac{n_1}{n_1 + n_2}\right) M_1 + \left(\frac{n_2}{n_1 + n_2}\right) M_2$</td><td>Several equally high point -> NO mode

Taller peak: major mode
Shorter peak: minor mode</td></tr>
<tr><td colspan="2">- The most appropriate measure in Interval or Ratio Scale</td><td>- The most appropriate measure in Ordinal Scale <u>or</u> with Quantitative variable in skewed distribution <u>or</u> in Open ended distributions.</td><td>- The most appropriate measure in Nominal Scale or Discrete variables (Whole numbers)
- Describing shapes, locating peaks.</td></tr>
</table>

Central Tendency and Shape of the Distribution

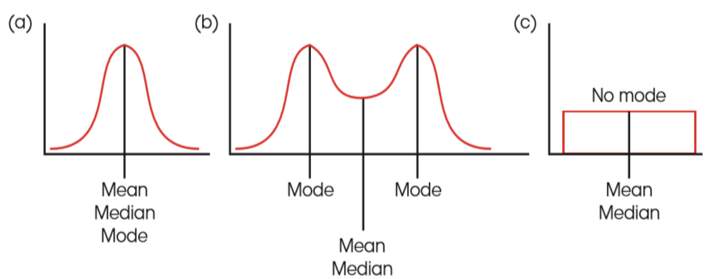


FIGURE 3.11
Measures of central tendency for three symmetrical distributions: normal, bimodal, and rectangular.

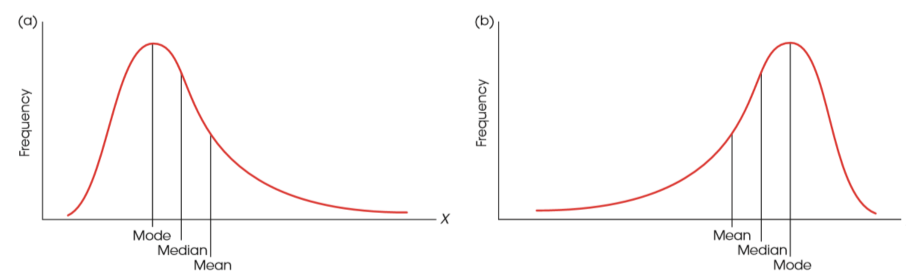


FIGURE 3.12
Measures of central tendency for skewed distributions.

(Source: Gravetter, Frederick. Statistics for the Behavioral Science 10th Ed)

What is Variability?

- A quantitative measure of the differences between scores in the distribution and describes the degree to which the scores are spread out or clustered together.

Range	Standard Deviation	Variance
The range is the distance covered by the scores in distribution, from the smallest scores to the largest score.	• Standard Deviation is the most common and most important measurement. It is used as the reference point to measure variability by considering the distance between each score and the mean.	Variance is equals to the mean of squared deviations.
$Range = X_{max} - X_{min}$	1. Find deviation of each score: $deviation\ score = X - \mu$	
The Range for Continuous Variables: $Range = URL_{for\ X_{max}} - LRL_{for\ X_{min}}$	2. Find sum of square deviation (SS) of each score: $SS = \sum(X-\mu)^2 = \sum X^2 - \frac{(\sum X)^2}{N^*}$ (* n: for sample)	
The Range for Whole Numbers: $Range = X_{max} - X_{min} + 1$	3. Find Variance, aka mean squared deviation (σ^2): $Population\ Variance\ \sigma^2 = \frac{\sum(X-\mu)^2}{N} = \frac{SS}{N}$ $Sample\ Variance\ s^2 = \frac{SS}{n-1} = \frac{SS}{df}$	
The Range is unreliable measure of variability because it does not consider all the scores in the distribution, thus, does not give an accurate description of the variability for the entire distribution.	4. Find Standard Deviation: $Population\ Standard\ Deviation\ \sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}} = \sqrt{\frac{SS}{N}}$ $Sample\ Standard\ Deviation\ s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}}$	

Best Measure of Central Tendency and Distribution Spread for Different Scales of Measurement

	Qualitative Measurement		Quantitative/Scales Measurement	
	Nominal	Ordinal	Interval	Ratio
Best measure of Central Tendency and Spread	Mode	Median (primary) – Interquartile Range (IQR) Mode (secondary)	Mean (primary) – Standard Deviation Median (secondary) – Interquartile Range (IQR) Mode (tertiary) (*Use Median as the primary measurement if the distribution is skewed)	

SPSS Techniques

Access Central Tendency and Variability from FREQUENCIES COMMAND in SPSS

Frequencies Command is considered as one of the most common use commands in SPSS to access various data information such as Frequency Distribution, Central Tendency, Variability, Skewness as well as Kurtosis or Standard Error.

```
FREQUENCIES VARIABLES= <VariableName>
  /NTILES=4
  /STATISTICS= <MeasurementName>
  /<GraphName>
  /ORDER=ANALYSIS.
```

(*we will learn more about graph in the next lab)

Analyze - Descriptive Statistics – Frequencies – Choose ***the desired variable in the list***

- Click Statistics tab – check Quartiles – check ***the desired measurements of Central Tendency/Dispersion*** – Continue.
- Click Charts tab – select the desired chart types – Continue.

OK.

***Notations of Central Tendency Statistics Measurement in codes.**

STDDEV
VARIANCE
RANGE
MINIMUM
MAXIMUM
SEMEAN
MEAN
MEDIAN
MODE
SUM
SKEWNESS
SESKEW
KURTOSIS
SEKURT

Output View Window:

- Statistics table of the variable
- Distribution table of the variable.
- Graph.

Access Variability from DESCRIPTIVE COMMAND in SPSS

Analyze - Descriptive Statistics – Descriptive – Choose ***the desired variable in the list***

- Click Option tab – check ***the desired measurements of the Variability Measurement*** – Continue.

Paste.

```
DESCRIPTIVES VARIABLES= VariableName
/STATISTICS= <MeasurementName>.
```

***Notations of Variability Measurements in codes.**

STDDEV
VARIANCE
RANGE
MIN
MAX

Output View Window:

- Descriptive table of the variable.

Report Central Tendency and Variability in APA Format

When writing a research paper, it is crucial for one to report Central Tendency and Variability of the variable in a correct format so that other scientists can understand how the conclusion is drawn. APA format is the universal scientific format to report such information. Therefore, whenever a variable is mentioned in the conclusion, it is necessary to be report as the formula:

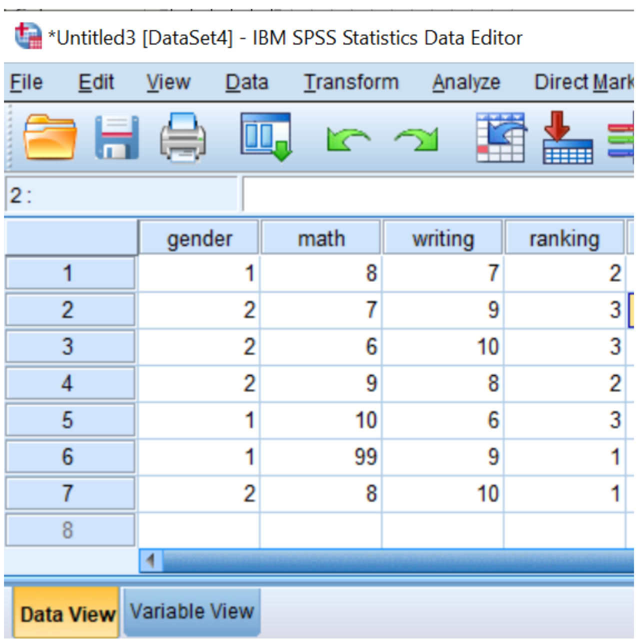
<VariableName> (M = , SD =)

- M: Mean
- SD: Standard Deviation

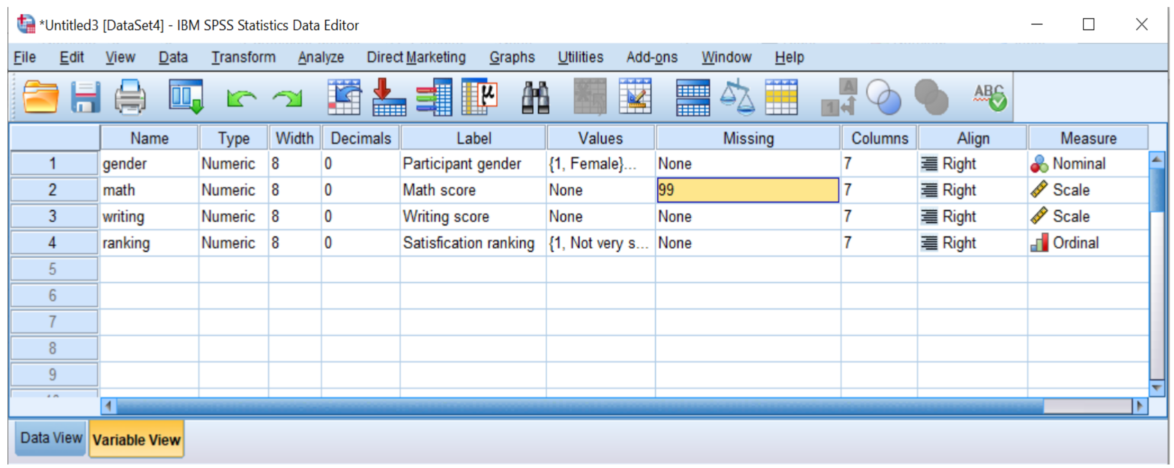
<u>Ex:</u> math (M = 8, SD = 1.414)

Example

A group of seven (7) students in tenth grade are selected randomly to participate in a new method of testing, containing two parts: mathematics and writing. Based on the total result of the whole test, students are asked to rate their satisfaction. The data of the study are recorded in SPSS as below:



	gender	math	writing	ranking
1	1	8	7	2
2	2	7	9	3
3	2	6	10	3
4	2	9	8	2
5	1	10	6	3
6	1	99	9	1
7	2	8	10	1
8				



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	gender	Numeric	8	0	Participant gender	{1, Female}...	None	7	Right	Nominal
2	math	Numeric	8	0	Math score	None	99	7	Right	Scale
3	writing	Numeric	8	0	Writing score	None	None	7	Right	Scale
4	ranking	Numeric	8	0	Satisfaction ranking	{1, Not very s...	None	7	Right	Ordinal

- Gender: 1 for “female,” 2 for “male”
- Ranking: 1 for “Not very satisfied,” 2 for “Satisfied,” 3 for “Very satisfied”
- 99 indicates missing value

Central Tendency and Variability from FREQUENCIES COMMAND for math variable

```
FREQUENCIES VARIABLES=math
 /NTILES=4
 /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
 /ORDER=ANALYSIS.
```

Output View Window

Statistics

Math score

N	Valid	6
	Missing	1
Mean		8.00
Median		8.00
Mode		8
Std. Deviation		1.414
Variance		2.000
Range		4
Minimum		6
Maximum		10
Percentiles	25	6.75
	50	8.00
	75	9.25

Math score

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	6	1	14.3	16.7	16.7
	7	1	14.3	16.7	33.3
	8	2	28.6	33.3	66.7
	9	1	14.3	16.7	83.3
	10	1	14.3	16.7	100.0
	Total	6	85.7	100.0	
Missing	99	1	14.3		
Total		7	100.0		

Variability from DESCRIPTIVE COMMAND for math variable

```
DESCRIPTIVES VARIABLES=math
/STATISTICS=MEAN STDDEV VARIANCE RANGE MIN MAX.
```

Output View Window

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Math score	6	4	6	10	8.00	1.414	2.000
Valid N (listwise)	6						

Practice

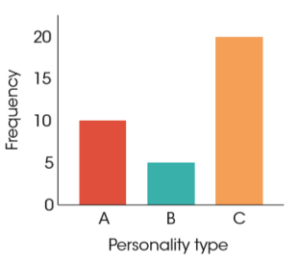
Let's apply what we learned to the other variable of the dataset in the Example section such as: writing, ranking and gender to get familiar with the codes and their outputs.

LAB 5: GRAPHS

Graphs for Different Scales of Measurement

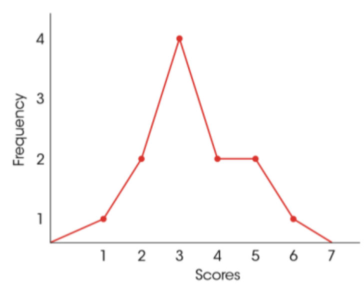
- Graphing data is an important method to visualize data, making it easier to understand or to recognize trends and patterns before interpretation or applying testing procedures.
- In this lab, we will be focusing on different types of graph and how they can be used for different Scales of Measurement.

	Qualitative Measurement		Quantitative/Scales Measurement	
	Nominal	Ordinal	Interval	Ratio
Type of Graph	Bar	Bar	• Histogram • Line • Boxplots – if a Quantitative Continuous Variable pair with a Nominal Variable. • Scatterplot – If two (2) Quantitative Continuous Variables pair with each other.	

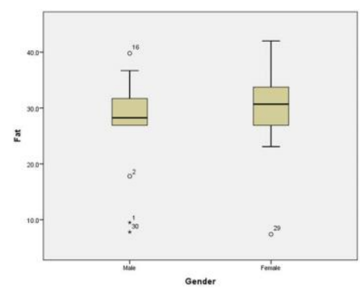


Bar Chart

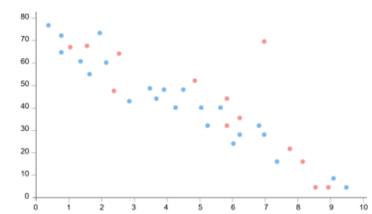
Histogram



Line Graph



Boxplot



Scatterplot

Building Graphs in SPSS

Building Bar Chart and Histogram Using FREQUENCIES COMMAND

Analyze - Descriptive Statistics – Frequencies – Choose ***the desired variable in the list***

- Click Statistics tab – check Quartiles – check ***the desired measurements of <u>Central Tendency/Dispersion</u>*** – Continue.
- Click Charts tab – select the desired chart types – Continue.

OK.

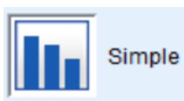
```
FREQUENCIES VARIABLES= <VariableName>
 /NTILES=4
 /STATISTICS= <MeasurementName>
 /<GraphName>
 /ORDER=ANALYSIS.
```

***Notations of <u>Chart Types</u> in codes.** **BARCHART PIECHART HISTOGRAM**

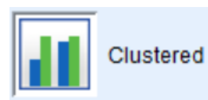
Building Charts Using GRAPH COMMAND

<u>Bar Chart</u>

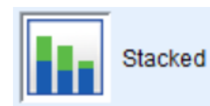
```
GRAPH
 /BAR(SIMPLE)=COUNT BY <VariableName>.
```




```
GRAPH
 /BAR(GROUPED)=COUNT BY <Variable1> BY <Variable2>.
```




```
GRAPH
 /BAR(STACK)=COUNT BY <Variable1> BY <Variable2>.
```



<u>Histogram</u>

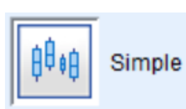
```
GRAPH
 /HISTOGRAM = <QuantitativeVariable> .
```

<u>Scatterplot</u>

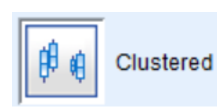
```
GRAPH
 /SCATTERPLOT(BIVAR) = <QuantitativeVariable1> WITH <QuantitativeVariable2>
 /MISSING=LISTWISE.
```

<u>Boxplot</u>

```
EXAMINE VARIABLES= <QuantitativeVariable> BY
<QualitativeVariable>
 /PLOT=BOXPLOT
 /STATISTICS=NONE
 /NOTOTAL.
```

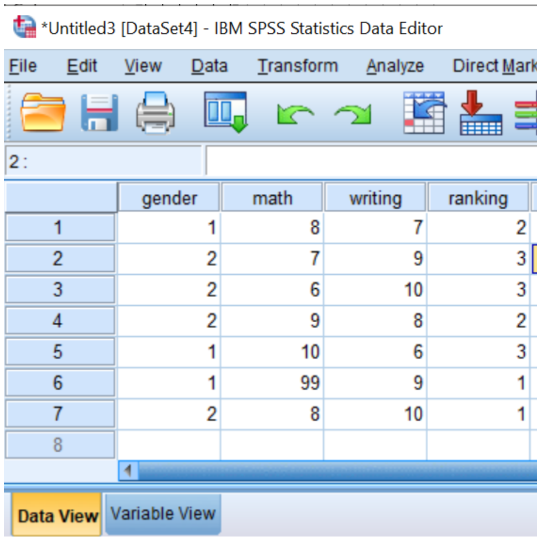



```
EXAMINE VARIABLES = <QuantitativeVariable> BY <QualitativeVariable> BY <QualitativeVariable>
 /PLOT=BOXPLOT
 /STATISTICS=NONE
 /NOTOTAL.
```



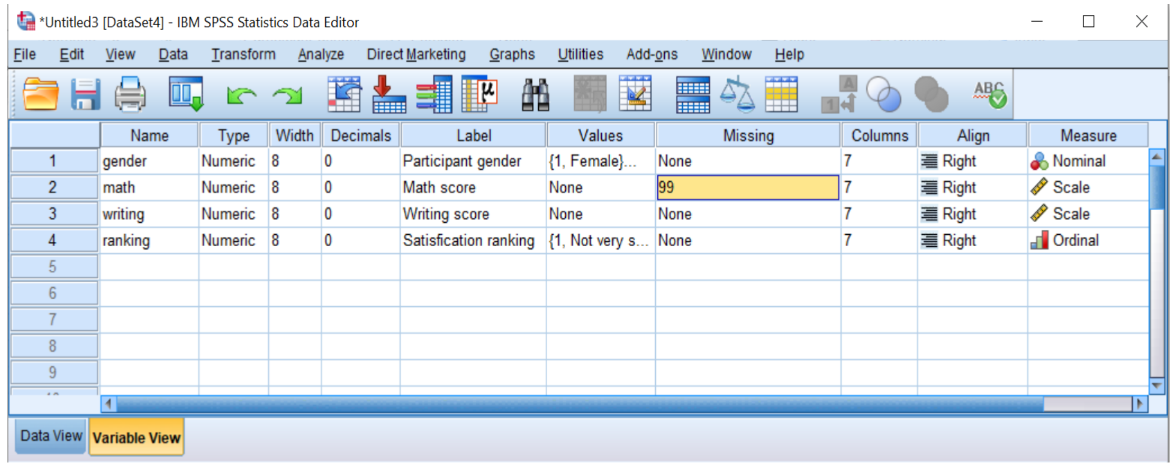
Example

A group of seven (7) students in tenth grade are selected randomly to participate in a new method of testing, containing two parts: mathematics and writing. Based on the total result of the whole test, students are asked to rate their satisfaction. The data of the study are recorded in SPSS as below:



*Untitled3 [DataSet4] - IBM SPSS Statistics Data Editor

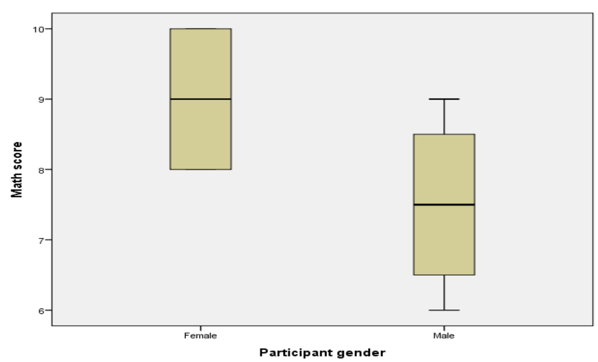
	gender	math	writing	ranking
1	1	8	7	2
2	2	7	9	3
3	2	6	10	3
4	2	9	8	2
5	1	10	6	3
6	1	99	9	1
7	2	8	10	1



*Untitled3 [DataSet4] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	gender	Numeric	8	0	Participant gender	{1, Female}...	None	7	Right	Nominal
2	math	Numeric	8	0	Math score	None	99	7	Right	Scale
3	writing	Numeric	8	0	Writing score	None	None	7	Right	Scale
4	ranking	Numeric	8	0	Satisfication ranking	{1, Not very s...	None	7	Right	Ordinal

- Gender: 1 for “female,” 2 for “male”
- Ranking: 1 for “Not very satisfied,” 2 for “Satisfied,” 3 for “Very satisfied”
- 99 indicates missing value



<u>Boxplot for math score clustered by gender</u>

```
EXAMINE VARIABLES=math BY gender
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL.
```

Practice

The dataset includes information of 9 different football players, with:

- Age: the player’s age
- College: The College that the player comes from
- YrsInPros: the number of years the player plays as a professional
- Height: the player’s height in meter
- Weight: the player’s weight in kilogram

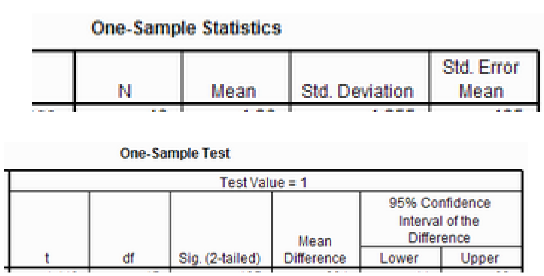
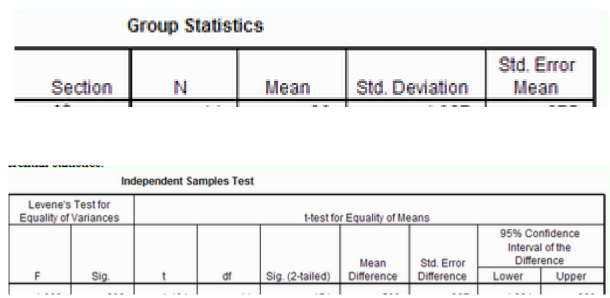
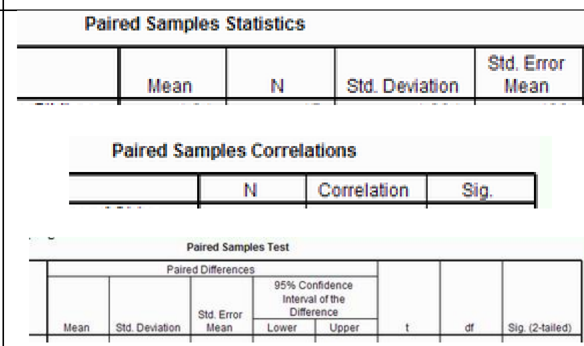
*with 99 indicates missing values

Age	College	YrsInPros	Height (m)	Weight (kg)
24	Arizona	1	1.8796	89.81221
23	Louisville	1	1.9304	120.2032
27	Michigan	3	1.9558	122.9248
24	Iowa	2	1.9812	127.9144
25	Texas Tech	3	1.8542	86.63703
23	Georgia	0	1.8542	97.52336
29	Colorado	6	1.8542	105.6881
25	Pittsburgh	99*	1.9304	125.6464
25	Tennessee State	2	1.9558	120.2032

- Based on what we just learned about visualizing data on SPSS, answer the questions below:

  1. Create a histogram for Age
  2. What Scale of Measurement is Weight? Which type of graph should we use for Weight?
  3. How about Height?

LAB 6: T-TEST

One Sample t-Tests	Independent Sample t-tests	Pair (Dependent) Sample t-tests
Determine if the mean from a sample is different from a particular value	Compare the mean of two different samples	Compare the mean of a variable under different conditions
H_0: no difference between the tested mean and the value H_1: the tested mean is larger/smaller than the value If $p \le \alpha$, then reject H_0; otherwise, fail to reject H_0	H_0: no difference in the tested variable between two groups H_1: there is difference in the tested variable between two groups If $p \le \alpha$, then reject H_0; otherwise, fail to reject H_0	H_0: no difference between two conditions H_1: there is difference between two conditions If $p \le \alpha$, then reject H_0; otherwise, fail to reject H_0
Analyze – Compare Means – One-Sample T Test ➢ Select the desired variable. ➢ Paste	Analyze – Compare Means – Independent-Samples T Test ➢ <u>Test Variable box</u>: select the desired variable. ➢ <u>Group variable box</u>: define the two groups ➢ Continue » Paste	Analyze - Compare Means – Paired-Sample T Test ➢ Select a pair of variables ➢ Paste
t-test **/testval= ?** **variables= Tested-Variable** *? in testval is the particular value that we want to compare our mean with. **<u>For example</u>**, if we want to compare the mean of our tested variable, "grade," with "100," then our test val is 100, thus the code will be: t-test /testval = **100** variables = **grade**	**t-test groups=Grouped-Variable(? ?)** **/variables=Tested-Variable.** *(? ?) is the two independent group. **<u>For example</u>**, if we need to test for a math score, "math," for the "gender" variable, with "0" is male and "1" is female, then the code will be: t-test groups = **gender (0 1)** /variables = **math**	**t-test** **pairs= Variable1 with Variable2 (paired).** **<u>For example</u>**, students need to take an English exam including two sections: reading and writing. Each section gives its own score as 'reading' and 'writing.' We want to compare the test score of students in these two sections, then the code will be: t-test pairs = **reading** with **writing** (paired)
 One-Sample Statistics: N, Mean, Std. Deviation, Std. Error Mean One-Sample Test: Test Value = 1; t, df, Sig. (2-tailed), Mean Difference, 95% Confidence Interval of the Difference (Lower, Upper)	 Group Statistics: Section, N, Mean, Std. Deviation, Std. Error Mean Independent Samples Test: Levene's Test for Equality of Variances (F, Sig.); t-test for Equality of Means (t, df, Sig. (2-tailed), Mean Difference, Std. Error Difference, 95% Confidence Interval of the Difference (Lower, Upper))	 Paired Samples Statistics: Mean, N, Std. Deviation, Std. Error Mean Paired Samples Correlations: N, Correlation, Sig. Paired Samples Test: Paired Differences (Mean, Std. Deviation, Std. Error Mean, 95% Confidence Interval of the Difference (Lower, Upper)), t, df, Sig. (2-tailed)
A one sample *t* test **<failed/succeeded>** to reveal a statistically reliable difference between the mean of **<tested-variable (M, s)>** and **<tested-value>, *t*(<df>), *p***** (* $p \le \alpha$ or $> \alpha$ depending on whether we reject or fail to reject H_0)	An independent sample *t* test **<failed/succeeded>** to reveal a statistically reliable difference between the mean number of **<tested-variable>** in **<group1 (M_1, s_1)>** and **<group 2 (M_2, s_2)>, *t*(<df>), *p***** (* $p \le \alpha$ or $> \alpha$ depending on whether we reject or fail to reject H_0)	A paired samples *t* test **<failed/succeeded>** to reveal a statistically reliable difference between the mean number of **<condition1 (M_1, s_1)>** and **<condition 2 (M_2, s_2), *t*(<df>), *p***** (* $p \le \alpha$ or $> \alpha$ depending on whether we reject or fail to reject H_0)

Example

Cholesterol data set

A study tested whether cholesterol was reduced after using a certain brand of margarine as part of a low fat, low cholesterol diet. This data set contains information on 18 people using margarine to reduce cholesterol over three time points. (Source: www.statstutor.ac.uk)

Before	After
6.42	5.83
6.56	5.83
8.43	7.71
8.05	7.25
5.77	5.31
6.77	6.15
6.44	5.59
6.85	6.4
5.73	5.13

Variable name	Variable	Data type
Before	Cholesterol before the diet (mmol/L)	Scale
After	Cholesterol after 4 weeks on the diet (mmol/L)	Scale

a. Indicate the hypotheses of the research

H_0: $\mu_D = 0$

H_i: $\mu_D \neq 0$

b. What type of t-test should be applied here?

Dependent Sample t-Test

c. Use SPSS to find if the effect on reducing cholesterol is significantly different between before and after the diet.

```
T-TEST PAIRS=Before WITH After4weeks (PAIRED)
 /CRITERIA=CI(.9500)
 /MISSING=ANALYSIS.
```

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	6.7800	9	.91901	.30634
	After 4 weeks	6.1333	9	.86371	.28790

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	Before - After 4 weeks	.64667	.14000	.04667	.53905	.75428	13.857	8	.000

d. Write your conclusion in APA format

A paired samples *t* test indicates that the effect of margarine on reducing cholesterol is significantly different between Before the diet (M = 6.780, SD = 0.919) and Four week after the diet (M = 6.133, SD = 0.864), *t*(8) = 13.857, *p < 0.05, 95%CI [0.539, 0.754]*

Practice

Based on what you have learned so far about t-test both in lecture and in lab, examine the dataset below and answer the questions.

Cholesterol data set

A study tested whether cholesterol was reduced after using a certain brand of margarine as part of a low fat, low cholesterol diet. (Source: www.statstutor.ac.uk)

After8weeks	Margarine
5.75	B
6.13	A
5.71	B
4.15	A
7.67	B
7.05	A
7.1	B
4.67	A
5.33	B
3.66	A
5.96	B
5.64	B
5.51	A
6.96	A
6.82	A
6.29	B
4.45	A
5.17	B

Variable name	Variable	Data type
After8weeks	Cholesterol after 4 weeks on the diet (mmol/L)	Scale
Margarine	Margarine type A or B	Binary

a. Indicate the hypotheses of the research
b. What type of t-test should be applied here?
c. Use SPSS to find if there is any significant difference in effect of different type of margarine.
d. Write you conclusion in APA format

LAB 7: CORRELATION

Bivariate Correlation

- Determine if two numerical continuous variables are linearly related to each other.
- Correlation coefficient is a number between -1 and 1 indicates the strength of the relationship of two variables.

- Sign: direction, positive or negative
  - Positive: higher score on one variable are associated with higher scores on the other variables
  - Negative: higher score on one variable are associated with lower scores on the other variable
- Magnitude: strength
  - A correlation coefficient of 1: near-perfect positive correlation.
  - A correlation coefficient of 0: uncorrelated.
  - A correlation coefficient of -1: near-perfect negative correlation.

Analyze – Correlate – Bivariate

- Select the two desired variables
- Option
  - Select Mean and Standard Deviation
  - Select either Pairwise or Listwise as desired
- Paste

```
correlations
 /variables = Variable1 Variable2
 /print =  twotail-or-onetail nosig.
 /statistics descriptives
 /missing = pairwise-or-listwise
```

Two-tailed test:

H_0: There is no correlations, $\rho = 0$.

H_1: There is correlation, $\rho \neq 0$.

One-tailed test:

H_0: There is no correlation, $\rho = 0$.

H_1: There is a positive correlation, $\rho > 0$.

Or

H_1: There is a negative correlation, $\rho < 0$.

Scatterplot Graph

- A simple two-dimensional plot in which the two coordinates of each dot represent the value of one variable measured on a single observation

- Independent variable on the horizontal axis.
- Dependent variables on the vertical axis.

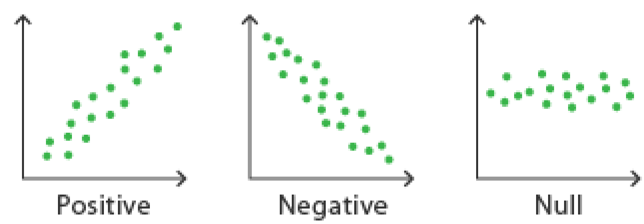
- Underlying the phenomenon based on:

- Form: the overall shape made by the points.
  - Ex: linear, quadratic or nonlinear.
- Direction: positive or negative, whether the two variables tend to move in the same or opposite direction.
- Strength: governed by how much scatter is present, whether the points seem to be clustered to suggest a relationship.
- Outliers: any point that don't fit the overall pattern or lie far away.

Graph – Legacy Dialog – Scatter/Dot

- Select Simple Scatter – Define
- Select the desired independent variable for x-axis
- Select the desired dependent variable for y-axis
- Paste

```
graph
 /scatterplot = Variable1 with Variable2
 /title = "Your-Graph-Name"
```



By default, SPSS has selected:

- Pearson and Two-tailed Test although users have the option to select Kendall's tau-b/Spearman and One-tailed Test if desire.
- A pairwise deletion of missing values.
  - Pairwise: as long as both variables in the correlation have valid values for a case, it will be included in the correlation).
  - Listwise: if a case has missing value for any variable, it will be eliminated from all correlation even though there are valid values for the other variables in the current correlation

Reporting Correlation in APA Format:

A Pearson Correlation test has **<failed/succeeded>** to reveal a statistical correlation between **<Variable1 (M= , SD=)>** and **<Variable2 (M= , SD =)>**, with **<r = >**, thus **<accepting/rejecting>** H_0.

height	weight
58	115
59	117
60	120
61	123
62	126
63	129
64	132
65	135
66	139
67	142
68	146
69	150
70	154
71	159
72	164

*** A full data set can be downloaded from:

https://www.picostat.com/dataset/r-dataset-package-datasets-women.

Example:

A selection of data from the Women dataset is chosen to illustrate the correlation between women's height and weight. The selection includes 15 observations and 2 variables: height (in inches) and weight (in pound)

Hypothesis:

Two-tailed test:
H_0: There is no correlation between women's height and weight, $\rho = 0$.
H_1: There is correlation between women's height and weight, $\rho \neq 0$.

SPSS Code:

```
CORRELATIONS
  /VARIABLES=height weight
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.
```

```
GRAPH
  /SCATTERPLOT(BIVAR)=weight WITH height
  /MISSING=LISTWISE
  /TITLE="Scatterplot of Women's Height Based on Women's Weight".
```

Output

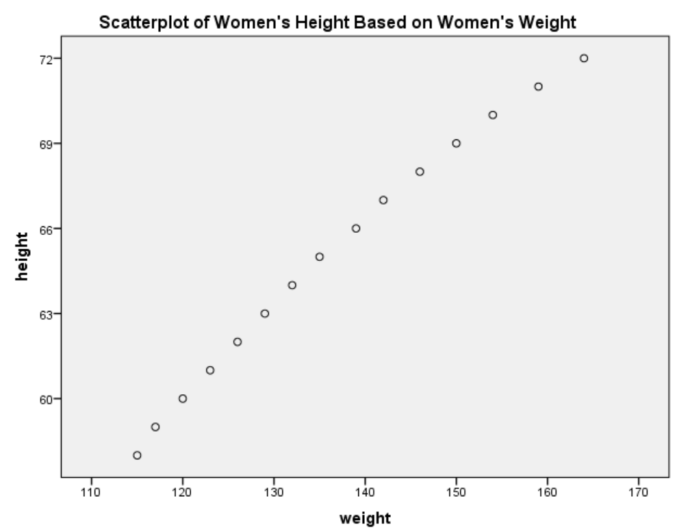
Descriptive Statistics

	Mean	Std. Deviation	N
height	65.00	4.472	15
weight	136.73	15.499	15

Correlations

		height	weight
height	Pearson Correlation	1	.995**
	Sig. (2-tailed)		.000
	N	15	15
weight	Pearson Correlation	.995**	1
	Sig. (2-tailed)	.000	
	N	15	15

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



Conclusion:

A Pearson Correlation test has succeeded to reveal a strong statistical positive correlation between women's height (M = 65, SD = 4.472) and weight (M = 136.73, SD = 15.499), r = 0.995. Therefore, we rejectH_0.

Practice

Birthweight	motherage	fatherage
5.8	24	26
4.2	20	20
6.4	26	25
4.5	41	37
5.8	20	23
6.8	28	39
5.2	20	20
6.1	19	20
7.5	20	22
8	18	20
8.6	29	31
7.1	31	35
6.6	30	38
7	31	32
6.6	27	27
6.3	19	23
7.3	23	32

Birthweight reduced data set
(source: www.statstutor.ac.uk)

This dataset contains information on newborn babies and their parents.
Main dependent variable = Birthweight (lbs)

Name	Variable	Data type
Birthweight	Weight of baby (lbs)	Scale
motherage	Maternal age	Scale
fage	Father's age	Scale

Based on what we have just learned about correlation, use your knowledge and SPSS skill to answer question below:

1. Is child's birthweight correlated with mother's age? How strong is the relationship?
2. How about child's birthweight with father's age? How strong is the relationship?
3. Show your SPSS work and give your conclusion in APA format.

LAB 8: INTRODUCTION TO ANALYSIS OF VARIANCE (ANOVA)

What is ANOVA?

- An extension of t-test in inferential statistics.
- A hypothesis testing procedure that use sample data as a base to make prediction about the population.
- Also known as F-test because the testing statistic is F.

One Way ANOVA

A test that allows one to make comparisons between the means of three or more groups of data.

One Independent Variable

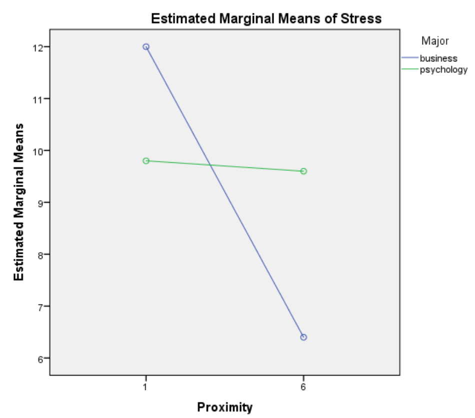
Comparing mean of three of more groups of an independent variable on a dependent variable

Hypotheses

$H_0: \mu_1 = \mu_2 = \mu_3 = ...$

H_1 : at least one mean is different

```
ONEWAY <DependentVariable> BY <IndependentVariable>
  /STATISTICS DESCRIPTIVES
  /MISSING ANALYSIS
  /POSTHOC=TUKEY ALPHA(0.05).
```



Interaction Plot

Two Way ANOVA

A test that allows one to make comparisons between the means of three or more groups of data, where two independent variables are considered.

Two Independent Variables

Comparing the effect of multiple groups of two independent variables on a dependent variable as well as the interaction between the two independent variables

Three (3) null hypothesis (H_0) :

- The population means of the first factor are equal.
- The population means of the second factor are equal.
- There is no interaction between the two factors.

```
UNIANOVA <DependentVariable> BY <IndependentVariable1>  <IndependentVariable2>
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(0.05)
  /DESIGN= <IndependentVariable1>  <IndependentVariable2>  <IndependentVariable1> *  <IndependentVariable2> .
```

Interaction plot – indicates interaction if the two variables cut each other

```
UNIANOVA <DependentVariable> BY <IndependentVariable1>  <IndependentVariable2>
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
 /PLOT=PROFILE( <IndependentVariable1>  *  <IndependentVariable2>)
 /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(0.05)
  /DESIGN= <IndependentVariable1>  <IndependentVariable2>  <IndependentVariable1> *  <IndependentVariable2> .
```

Two Way Between Subject ANOVA	Two Way Within Subject ANOVA	Two Way Mix Mode ANOVA
IV_1 : independent (between subject) measure IV_2 : independent (between subject) measure	IV_1 : repeated (within subject) measure IV_2 : repeated (within subject) measure	IV_1 : independent (between subject) measure IV_2 : repeated (within subject) measure

Reporting ANOVA in APA Format

One Way ANOVA

A one-way ANOVA was conducted to compare the effect of **<IndependentVariable>** on **<DependentVariable>** in **<condition1>, <condition2>, and <condition3>.**

There was **<a significant / not a significant>** effect of **<IndependentVariable>** on **<DependentVariable>** for these conditions:

$F(df_{between}, df_{within})$ = ___ , p = ____

If there is a significant difference, then:

Specifically, Post hoc comparisons indicates that the mean score for **<condition1 (M= , SD=)>** was significant different from **<condition2 (M= , SD=)>**

Two Way ANOVA

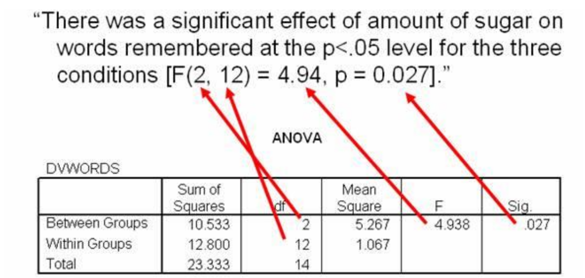
A two-way ANOVA was conducted to compare the main effects of **<IndependentVariable1>** and **<IndependentVariable2>** on **<DependentVariable>.**

The result yielded a main effect for **<IndependentVariable1, $F(df_{IV1}, df_{Error})$ = ____, *p* = ___>,** indicates there **<is / is no>** difference between **<level1 (*M* = , *SD* =)>** and **<level2 (*M* = , *SD* =).**

The result yielded a main effect for **<IndependentVariable2, $F(df_{IV2}, df_{Error})$ = ____, *p* = ___>,** indicates there **<is / is no>** different between **<level1 (*M* = , *SD* =)> and <level2 (*M* = , *SD* =).**

The interaction effect between **<IndependentVariable1>** and **<IndependentVariable2>** was **<significant / not significant>, $F(df_{interaction}, df_{Error})$ = ____, *p* = ____.**

Example (One Way ANOVA)

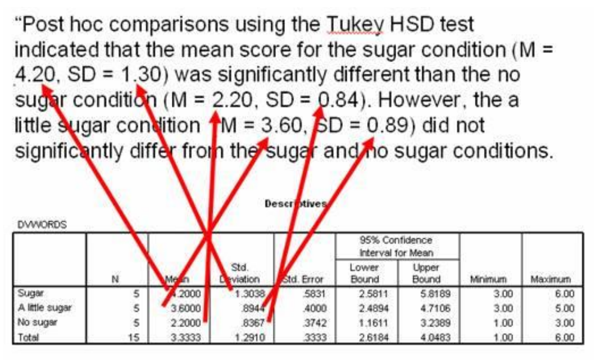


"There was a significant effect of amount of sugar on words remembered at the p<.05 level for the three conditions [F(2, 12) = 4.94, p = 0.027]."

ANOVA

DVWORDS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.533	2	5.267	4.938	.027
Within Groups	12.800	12	1.067		
Total	23.333	14			



"Post hoc comparisons using the Tukey HSD test indicated that the mean score for the sugar condition (M = 4.20, SD = 1.30) was significantly different than the no sugar condition (M = 2.20, SD = 0.84). However, the a little sugar condition (M = 3.60, SD = 0.89) did not significantly differ from the sugar and no sugar conditions.

Descriptives

DVWORDS

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Sugar	5	4.2000	1.3038	.5831	2.5811	5.8189	3.00	6.00
A little sugar	5	3.6000	.8944	.4000	2.4894	4.7106	3.00	5.00
No sugar	5	2.2000	.8367	.3742	1.1611	3.2389	1.00	3.00
Total	15	3.3333	1.2910	.3333	2.6184	4.0483	1.00	6.00

Example (Two Way ANOVA)

Reporting results of major tests in factorial ANOVA; <u>non-significant interaction:</u>

Attitude change scores were subjected to a two-way analysis of variance having two levels of message discrepancy (small, large) and two levels of source expertise (high, low). All effects were statistically significant at the .05 significance level. The main effect of message discrepancy yielded an *F* ratio of *F*(1, 24) = 44.4, *p* < .001, indicating that the mean change score was significantly greater for large-discrepancy messages (*M* = 4.78, *SD* = 1.99) than for small discrepancy messages (*M* = 2.17, *SD* = 1.25). The main effect of source expertise yielded an F ratio of *F*(1, 24)= 25.4, *p* < .01, indicating that the mean change score was significantly higher in the high-expertise message source (*M* = 5.49, *SD* = 2.25) than in the low-expertise message source (*M* = 0.88, *SD* = 1.21). The interaction effect was non-significant, *F*(1, 24) = 1.22, *p* > .05.

Reporting results of major tests in factorial ANOVA; <u>significant interaction:</u>

A two-way analysis of variance yielded a main effect for the diner's gender, *F*(1, 108) = 3.93, *p* < .05, such that the average tip was significantly higher for men (*M* = 15.3%, *SD* = 4.44) than for women (*M* = 12.6%, *SD* =6.18). The main effect of touch was non-significant, *F*(1, 108) = 2.24, *p* > .05. However, the interaction effect was significant, *F*(1, 108) = 5.55, *p* < .05, indicating that the gender effect was greater in the touch condition than in the non-touch condition.

LAB 9: APPLYING ANOVA

Example 1: One Way ANOVA

No alcoh	1 oz. alcohol	3 oz. alcoho
5	8	10
1	9	8
0	3	7
2	12	15
2	10	18
0	8	9
3	8	8
9	0	13
7	4	10
8	7	17
6	11	6
3	13	15

A researcher interested in the effects of alcohol on fine motor control randomly assigned 36 participants to one of three groups. Participants consumed either 0, 1, or 3 ounces of alcohol. Then they traced a star while looking at the image of the star in a mirror. The researcher counted the number of errors each participant made while performing the mirror-image tracing task. Conduct an ANOVA on the data below to determine whether alcohol affected participants' performance on the task. Show your work and state your conclusion about the effect of alcohol on motor control.

- Independent Variable: Alcohol amount
  - No alcohol: Group 1
  - 1oz alcohol: Group 2
  - 3oz alcohol: Group 3
- Dependent Variable: number of errors

Type of test: One Way ANOVA

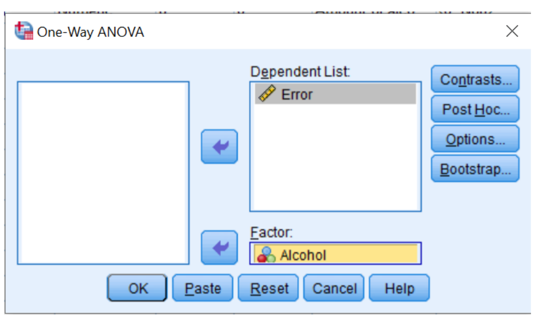
$H_0: \mu_1 = \mu_2 = \mu_3$

H_1 : at least one mean is different

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Alcohol	Numeric	8	0	Amount of alco...	{0, Non}...	None	8	Right	Nominal
2	Error	Numeric	8	0	Number of error	None	None	8	Right	Scale

Non-coding Procedure

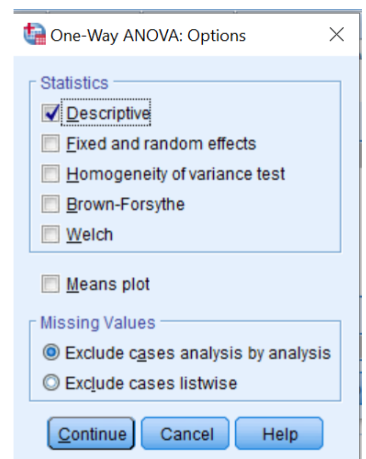
Analyze – Compare Means – One Way ANOVA



Post Hoc Tab

- Select Tukey for equal sample size
- Select Scheffe for unequal sample size
- Enter desired significant level

Option Tab



Coding Procedure

```
ONEWAY Error BY Alcohol
 /STATISTICS DESCRIPTIVES
 /MISSING ANALYSIS
 /POSTHOC=TUKEY ALPHA(0.05).
```

Descriptives

Number of error

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
0 Non	12	3.83	3.099	.895	1.86	5.80	0	9
1 1oz	12	7.75	3.817	1.102	5.32	10.18	0	13
2 3oz	12	11.33	4.097	1.183	8.73	13.94	6	18
Total	36	7.64	4.746	.791	6.03	9.24	0	18

ANOVA

Number of error

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	337.722	2	168.861	12.367	.000
Within Groups	450.583	33	13.654		
Total	788.306	35			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Number of error
Tukey HSD

(I) Amount of alcohol	(J) Amount of alcohol	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
0 Non	1 1oz	-3.917*	1.509	.036	-7.62	-.22
	2 3oz	-7.500*	1.509	.000	-11.20	-3.80
1 1oz	0 Non	3.917*	1.509	.036	.22	7.62
	2 3oz	-3.583	1.509	.059	-7.28	.12
2 3oz	0 Non	7.500*	1.509	.000	3.80	11.20
	1 1oz	3.583	1.509	.059	-.12	7.28

*. The mean difference is significant at the 0.05 level.

Report conclusion in APA format

A one-way ANOVA was conducted to compare the effect of alcohol on fine motor control for the amount of alcohol, no alcohol, 1-ounce alcohol and 3-ounce alcohol. There was a significant effect of amount of alcohol on the error made during tracing image of a star in the mirror, $F(2, 33) = 12.367$, $p < 0.05$. Post hoc comparisons using the Tukey HSD test indicated that the mean difference between no-alcohol condition ($M = 3.83$, $SD = 3.099$) and 1 ounce alcohol condition ($M = 7.75$, $SD = 3.817$) as well as between no-alcohol condition and 3 ounce alcohol condition ($M = 11.33$, $SD = 4.097$). Taken together, these results suggest that larger amount of alcohol that one consumes really do have an effect on fine motor control. Specifically, our results suggest that when humans consume more alcohol, they make more error in tracing image through a mirror.

Example 2: Two Way ANOVA

Stress	Field_of_Study	Proximity
9	psychology	6
13	psychology	6
11	psychology	6
7	psychology	6
8	psychology	6
11	psychology	1
7	psychology	1
11	psychology	1
9	psychology	1
11	psychology	1
5	business	6
7	business	6
6	business	6
10	business	6
4	business	6
10	business	1
15	business	1
8	business	1
17	business	1
10	business	1

- A professor of a statistics course was interested in the effect of proximity to the final exam (5 weeks away, 1 week) on the stress levels of psychology and business students. She measured their level of perceived stress on a standardized questionnaire. In this scenario, stress is the dependent variable while proximity and students' field of study are independent variables.

  (Source: http://www.maths-statistics-tutor.com/two_way_factorial_ANOVA_pasw_spss.php)

- Type of test: Two way ANOVA
  - IV_1: Proximity – 2 levels: one week – six week.
  - IV_2: Field_of_study: psychology – business
  - DV: Stress

- In this example, we have three sets of hypotheses.

  Hypothesis 1
  - Null hypothesis: Proximity to the final exam has no effect on students' stress level.
  - Alternative hypothesis: Proximity to the final exam has an effect on students' stress level.

  Hypothesis 2
  - Null hypothesis: The stress levels of psychology students and business students are the same.
  - Alternative hypothesis: The stress levels of psychology students and business students are not the same.

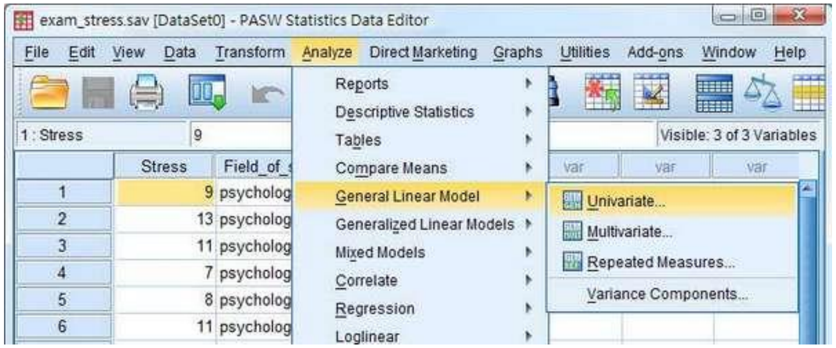
  Hypothesis 3
  - Null hypothesis: There is no interaction between students' field of study and proximity to the final exam. That is, the effect of proximity to the final exam is the same for psychology student and business student.
  - Alternative hypothesis: There is an interaction between students' field of study and proximity to the final exam. That is, the effect of proximity to the final exam is different for psychology student and business student.

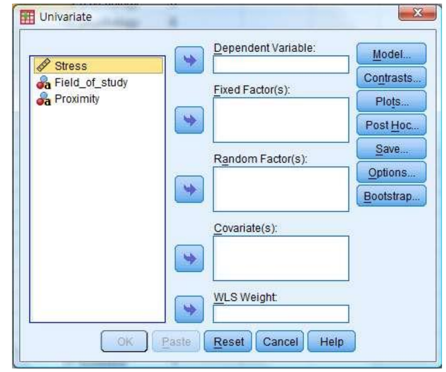
- In the data, the first column is stress score, the second column is field of study and the third is proximity to the final exam.

Non-coding Procedure

Step 1

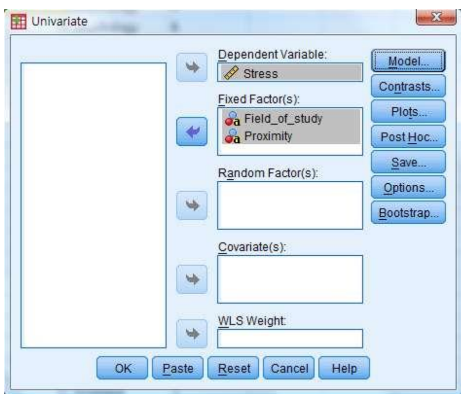
Select "Analyze -> General Linear Model -> Univariate".



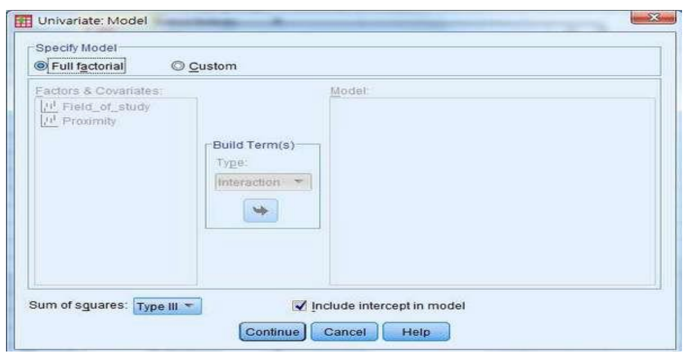


Step 2

Select "Stress" as "Dependent Variable" and "Field of study" and "Proximity" as "Fixed Factor(s)".



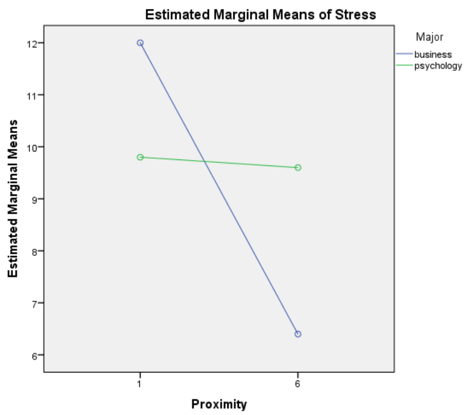
Model Tab:



Plot Tab

Coding Procedure

```
UNIANOVA Stress BY Major Proximity
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PLOT=PROFILE(Proximity*Field_of_study)
 /PRINT=DESCRIPTIVE
 /CRITERIA=ALPHA(0.05)
 /DESIGN= Field_of_study Proximity Field_of_study *Proximity.
```



Between-Subjects Factors

		N
Field_of_study	business	10
	psychology	10
Proximity	1	10
	6	10

Tests of Between-Subjects Effects

Dependent Variable:Stress

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	79.750[a]	3	26.583	3.692	.034
Intercept	1786.050	1	1786.050	248.062	.000
Field_of_study	1.250	1	1.250	.174	A .682
Proximity	42.050	1	42.050	5.840	B .028
Field_of_study * Proximity	36.450	1	36.450	5.063	C .039
Error	115.200	16	7.200		
Total	1981.000	20			
Corrected Total	194.950	19			

a. R Squared = .409 (Adjusted R Squared = .298)

Descriptive Statistics

Dependent Variable: Stress

Major	Proximity	Mean	Std. Deviation	N
business	1	12.00	3.808	5
	6	6.40	2.302	5
	Total	9.20	4.185	10
psychology	1	9.80	1.789	5
	6	9.60	2.408	5
	Total	9.70	2.003	10
Total	1	10.90	3.035	10
	6	8.00	2.789	10
	Total	9.45	3.203	20

Data Interpretation:

- A: p = 0.682 > 0.05 → The effect of field of study is not significant
  → On average, the stress levels of psychology students and business students are the same.
- B: p = 0.028 < 0.05 → The main effect of proximity is significant
- C: p = 0.039 < 0.05 → The interaction is significant
  → The effect of proximity on stress levels for psychology students and business students are not the same.
- The interaction plot suggests that as the final exam approaches, the stress level of business students soars but that of psychology students remains pretty much the same.

Report Conclusion in APA Format:

A two-way analysis of variance was conducted to compare the main effects of type of proximity (one week, six week) and field of study (psychology, business) on students' level of stress. All effect was significantly different except for field of study. The result yielded a main effect for proximity, $F(1, 16) = 5.840$, $p < .05$, indicates the different between one week ($M = 10.9$, $SD = 3.05$) and six week ($M = 8$, $SD = 2.789$). The main effect for field of study was not significant, $F(1, 16) = 0.174$, $p > .05$ between psychology major (M = 9.7, SD = 2.003) and business major (M = 9.2, SD = 4.185). The interaction effect between proximity and field of study was significant, $F(1, 15) = 5.063$, $p < .05$.

LAB 10: APPLYING SPSS

In today lab, we will spend time to apply all of our SPSS skills on real data set.

You will be working on the questions on your own. Feel free to review your notes from our previous labs.

Diet data set

Download the data set for today practice from:
https://www.sheffield.ac.uk/polopoly_fs/1.886048!/file/Diet_SPSS.sav
or
https://www.dropbox.com/s/wo5lnt67v7v7g1l/Diet_SPSS-1.xlsx?dl=0

This data set contains information on 78 people using one of three diets.

Variable name	Variable	Data type
Person	Participant number	
gender	Gender, 1 = male, 0 = female	Binary
Age	Age (years)	Scale
Height	Height (cm)	Scale
preweight	Weight before the diet (kg)	Scale
Diet	Diet	Binary
weight10weeks	Weight after 10 weeks (kg)	Scale
weightLOST	Weight lost after 10 weeks (kg)	Scale

1. Ignore diet and test to see if weight has been lost (Hint: Paired t-test)
2. Remove weight lost and get students to calculate it using before/after weights (Hint: Compute Variable)
3. Summary statistics by diet (Hint: Summary Statistics)
4. Which diet was best for losing weight? (Hint: One Way ANOVA)
5. Are there gender differences for weight lost? (Hint: One Way ANOVA)
6. How is the effect of diet and gender on weight lost? (Hint: Two Way ANOVA)
7. Means plot of weight lost by diet and gender (Hint: Interaction plot)

(source: www.statstutor.ac.uk)

Acknowledgement

My deep gratitude goes to ***Professor Sophia Barret****, who has been guiding me through the forest and lightening up the pathway ahead.*

I have learned a lot during this semester. Without you, I would not be able to go this far.

I cannot thank you enough for giving me all these opportunities from which I find great motivation.

Many thanks and love to ***My Vien****, my best friend and roommate,*

the person who would spend all of her time to drag me and make me move forward.

It is a pleasure to have you as a friend for such a long time and still counting.

I also would like to thank ***my parents*** *and* ***my brother*** *who are always there for me, support and love me without condition.*

I am who I am because of you.

Lastly, designing this Lab Book has been a fun and unexpected journey that I enjoy every moment of it.

"Not all those who wander are lost"

-Gandaff in The Lord of the Rings