



Introduction to Statistical Software Package (PSPP)

Chapter 2

Introduction to Statistical Software Package

- ▶ The statistical software package used – PSPP v1.2.0
- ▶ GNU PSPP is a program for statistical analysis of sampled data. It is a free as in freedom replacement for the proprietary program SPSS, and appears very similar to it with a few exceptions..
- ▶ Compatible with the Windows environment, Mac OS and Linux.

Data Input

- ▶ Defining variables (in Variable view):
 - ▶ Variable names
 - ▶ Labels (Variable and value)
 - ▶ Missing values
 - ▶ Variable types
 - ▶ Column format
 - ▶ Measurement level



Data Input

▶ Variable names

- ▶ Must begin with letter. The remaining can be letter, digit, period or the symbols @, #, _ or \$.
- ▶ Cannot end with period or underscore
- ▶ Blanks and special characters cannot be used (E.g.: !, ?, ' and *)
- ▶ Must be unique; Duplication is not allowed
- ▶ The length in one variable cannot exceed 64 characters
- ▶ Reserved keywords cannot be used as variable names (E.g.: ALL, AND, BY, EQ, GE, LE, LT, NE, NOT, OR, TO and WITH)
- ▶ Are not case sensitive



Data Input

► Labels

- Variable label is a full description of the variable name
- Optional
- It has a value label that can be in alphanumeric and numeric code

Variable name	Label
gender	optional
age	optional
allday	desire for 24-hour shopping facilities
area	choice of shopping area
cost	amount spent on groceries per week

Alphanumeric code

Variable name	Label
id	not applicable
gender	1 = female 2 = male
age	not applicable
allday	1 = would use 24-hour shopping 2 = would not use 24-hour shopping
area	1 = shop in suburb where living 2 = travel to next suburb to shop 3 = travel further to shop
cost	not applicable

Numeric code

Data Input

▶ Missing values

- ▶ How to deal with missing data – Leave the cell blank or assign missing value codes
- ▶ Rules for assigning missing value codes:
 - ▶ Must be the same data type as they represent (E.g.: Missing numeric data must use numeric missing value code)
 - ▶ Missing value code cannot occur as data in the data set
 - ▶ By default, the choice of digit is usually 9

▶ Variable types

- ▶ By default, PSPP assumes all new variables are numeric with two decimal places.
- ▶ Can be changed to other type (date, currency, etc) and have different number of decimal



4.1 Data Input

- ▶ **Column format**

- ▶ It is possible to adjust the width of the Data Editor columns and change the alignment of data in the column (left, centre or right)

- ▶ **Measurement level**

- ▶ Can be specified as nominal, ordinal, interval or ratio



Data Editor

- ▶ Specifically, in editing and viewing data and result in SPSS, it has seven different types of windows:
 - ▶ Data editor (Our focus)
 - ▶ Viewer and draft viewer (Output)
 - ▶ Pivot table editor
 - ▶ Chart editor
 - ▶ Text output editor
 - ▶ Syntax editor
 - ▶ Script editor



Data Editor

► Data view

The screenshot displays the SPSS Data Editor interface for the file *students.sav. The main window shows the Data View tab with a list of cases (1 to 21) and their corresponding data for variables like id, lastname, firstnam, gender, ethnicit, YEAR IN SCHOOL, LOWER OR UPPER DIVISION, section, High School GPA, College GPA, DID EXTRA CREDIT PROJECT?, ATTENDED REVIEW SESSIONS?, quiz3, quiz4, quiz5, final, and total. A Descriptives dialog box is open, showing the list of variables and the statistics to be calculated. The statistics selected are Mean, Standard deviation, Minimum, Maximum, Range, Sum, Standard error, Variance, and Kurtosis. The dialog box also includes options for excluding cases with missing values and saving Z-scores as new variables.

Descriptives Dialog Box:

- Variables: quiz1, quiz2
- Statistics:
 - ☒ Mean
 - ☒ Standard deviation
 - ☒ Minimum
 - ☒ Maximum
 - ☐ Range
 - ☐ Sum
 - ☐ Standard error
 - ☒ Variance
 - ☐ Kurtosis
- Options:
 - ☐ Exclude entire case if any selected variable is missing
 - ☒ Include user-missing data in analysis
 - ☐ Save Z-scores of selected variables as new variables

Data View Table:

Case	id	lastname	firstnam	gender	ethnicit	YEAR IN SCHOOL	LOWER OR UPPER DIVISION	section	High School GPA	College GPA	DID EXTRA CREDIT PROJECT?	ATTENDED REVIEW SESSIONS?	quiz3	quiz4	quiz5	final	total
1	908754	MARQUEZ	CHYRELLE														
2	192627	MISCHKE	ELAINE														
3	822485	VALENZUELA	KATHRYN														
4	106484	VILLARRUIZ	ALFRED														
5	725987	BATILLER	FRED														
6	354601	CARPIO	MARY														
7	307894	TORRENCE	GWEN														
8	762813	DAEL	IVAN														
9	958384	RONCO	SHERRY														
10	985700	CHA	LILY														
11	807963	LEWIS	CARL														
12	378446	SAUNDERS	TAMARA														
13	905109	JENKINS	ERIC														
14	920656	LIAO	MICHELLE														
15	911355	LESKO	LETICIA														
16	154441	LIAN	JENNY														
17	273611	WU	VIDYUTH	1	2	2	1	2	3.70	3.60	1	2	3	5			
18	574170	HURRIA	WAYNE	2	1	2	1	2	3.84	2.98	1	1	4	6			
19	515586	FIALLOS	LAUREL	1	4	2	1	2	3.90	3.15	1	1	7	3			
20	762308	GOUW	BONNIE	1	4	2	1	3	3.90	3.65	1	2	8	4			
21	700978	WEBSTER	DEANNA	1	3	2	1	3	3.90	3.95	1	2	8	2			

4.2 Data Editor

► Variable view

*students.sav [DataSet1] – PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Value Labels

Value: 2

Value Label: ASIAN

+ Add

Apply

Remove

1 = 'AMERICAN INDIAN'

2 = 'ASIAN'

3 = 'AFRO-AMERICAN'

4 = 'CAUCASIAN'

5 = 'HISPANIC'

OK

Cancel

Help

Variable	Name	Type	Width	Decimals	Label	Value Labels	Missing Values	Columns	Align	Measure
1	id	Numeric	6	0		None	None	8	Right	Scale
2	lastname	String	10			None	None	10	Left	Nominal
3	firstname	String	10			None	None	10	Left	Nominal
4	gender	Numeric	1	0		{1, FEMALE}...	None	8	Right	Ordinal
5	ethnicity	Numeric	1	0						Ordinal
6	year	Numeric	1	0	YEAR IN SCHOOL					Ordinal
7	lowup	Numeric	1	0	LOWER OR UPPER DIVISION					Ordinal
8	section	Numeric	1	0						Ordinal
9	hsgpa	Numeric	4	2	High School GPA					Scale
10	colgpa	Numeric	8	2	College GPA					Scale
11	extrcred	Numeric	1	0	DID EXTRA CREDIT PROGRAM					Ordinal
12	review	Numeric	1	0	ATTENDED REVIEW SESSION					Ordinal
13	quiz1	Numeric	2	0						Ordinal
14	quiz2	Numeric	2	0						Ordinal
15	quiz3	Numeric	2	0						Ordinal
16	quiz4	Numeric	2	0		None	None	8	Right	Ordinal
17	quiz5	Numeric	2	0		None	None	8	Right	Ordinal
18	final	Numeric	2	0		None	None	8	Right	Scale
19	total	Numeric	3	0		None	None	8	Right	Scale
20	percent	Numeric	4	1		None	None	8	Right	Scale
21	grade	String	8			None	None	8	Left	Nominal
22	passfail	String	1			None	None	1	Left	Nominal

Data View Variable View

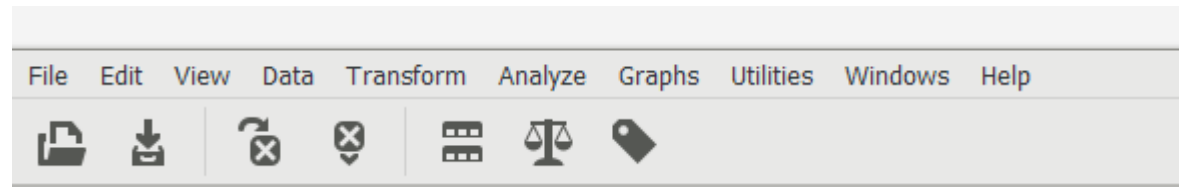
Filter off Weights off Split by year

Data Editor

- ▶ Menu-driven and has variety of pull-down menus.

- ▶ The main menu bar contains 10 menu:

- ▶ File
- ▶ Edit
- ▶ View
- ▶ Data
- ▶ Transform
- ▶ Analyze*
- ▶ Graphs*
- ▶ Utilities
- ▶ Windows
- ▶ Help

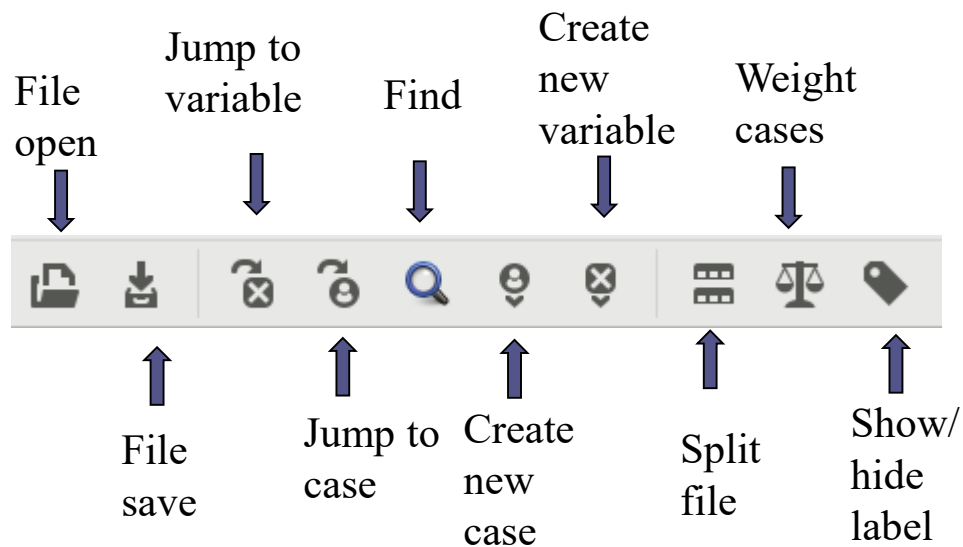


**Available in all windows (Easier to generate new output)*



Data Editor

► Toolbar



Data Editor

- ▶ Exercise
- ▶ Enter the following case:



Exercise

Case	Id	Gender	Hometown	CGPA	Height	Age	Weight	Laptop
1	1	Female	Perlis	3.23	157	21	48.00	Yes
2	2	Female	Selangor	2.93	155	20	40.00	Yes
3	3	Male	Johor	3.40	158	21	44.00	Yes
4	4	Female	Kuala Lumpur	3.09	154	21	47.00	Yes
5	5	Female	Pulau Pinang	3.74	160	21	42.00	Yes
6	6	Female	Pulau Pinang	2.23	160	21	65.00	No
7	7	Female	Melaka	2.53	163	21	50.00	Yes
8	8	Female	Perak	2.41	158	21	56.00	Yes
9	9	Female	Johor	3.00	157	21	48.00	Yes
10	10	Female	Kuala Lumpur	2.90	161	21	86.00	Yes
11	11	Female	Pahang	3.26	157	21	45.00	Yes
12	12	Male	Sarawak	3.13	150	21	60.00	Yes
13	13	Male	Selangor	2.28	152	21	45.00	Yes
14	14	Male	Johor	2.91	163	21	63.00	No
15	15	Female	Negeri Sembilan	2.66	145	21	46.00	Yes

Exercise

► Key in the following case :

16	16	Female	Johor	3.55	155	21	48.00	Yes
17	17	Female	Pulau Pinang	2.50	173	21	90.00	Yes
18	18	Female	Pulau Pinang	2.82	160	21	53.00	Yes
19	19	Female	Pulau Pinang	2.48	165	21	80.00	Yes
20	20	Female	Kedah	2.46	156	21	63.00	Yes
21	21	Female	Pulau Pinang	2.96	160	21	45.00	Yes
22	22	Female	Perak	3.45	152	21	39.00	Yes
23	23	Female	Pulau Pinang	2.94	159	21	50.00	Yes
24	24	Male	Kuala Lumpur	3.63	171	22	83.00	Yes
25	25	Female	Selangor	3.30	156	21	80.00	Yes
26	26	Female	Kedah	2.94	168	21	51.00	Yes
27	27	Female	Perak	3.03	153	21	53.00	Yes
28	28	Female	Pulau Pinang	2.95	167	21	58.00	Yes

Data Computation and Transformation

- ▶ Data screening

- ▶ Objectives:

- ▶ Making sure that data have been entered correctly (Pulau pinang / Pulau Pinang)
 - ▶ Making sure that the distributions are normal

- ▶ The importance of data screening

- ▶ May affect the validity of the results that are produced
 - ▶ May assist in deciding; the need of data transformation, the usage of nonparametric techniques

- ▶ This subtopic will be demonstrated using data of work3.sav



Data Computation and Transformation

► Errors in data entry

- Error in data entry are common and therefore data files must be carefully screened
- Use Analyze > Descriptive Statistics > Frequencies... command
- Output example:

att1

<i>Value Label</i>	<i>Value</i>	<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cum Percent</i>
strongly agree	1	22	22.22	22.22	22.22
agree	2	31	31.31	31.31	53.54
neither agree nor disagree	3	38	38.38	38.38	91.92
disagree	4	3	3.03	3.03	94.95
strongly disagree	5	5	5.05	5.05	100.00
<i>Total</i>		99	100.0	100.0	

att1

<i>N</i>	<i>Valid</i>	99
	<i>Missing</i>	0
<i>Mean</i>		2.37
<i>Std Dev</i>		1.03
<i>Minimum</i>		1.00
<i>Maximum</i>		5.00

Data Computation and Transformation

▶ Assessing normality

- ▶ The assumption of normality is prerequisite of many inferential statistical techniques
- ▶ Normality can be explored using a combined assessment of graphic:
 - ▶ Histogram
 - ▶ Boxplot
 - ▶ Normal probability plot
 - ▶ Detrended normal plot
- ▶ And normality test:
 - ▶ Kolmogorov-Smirnov
 - ▶ Skewness
 - ▶ Kurtosis



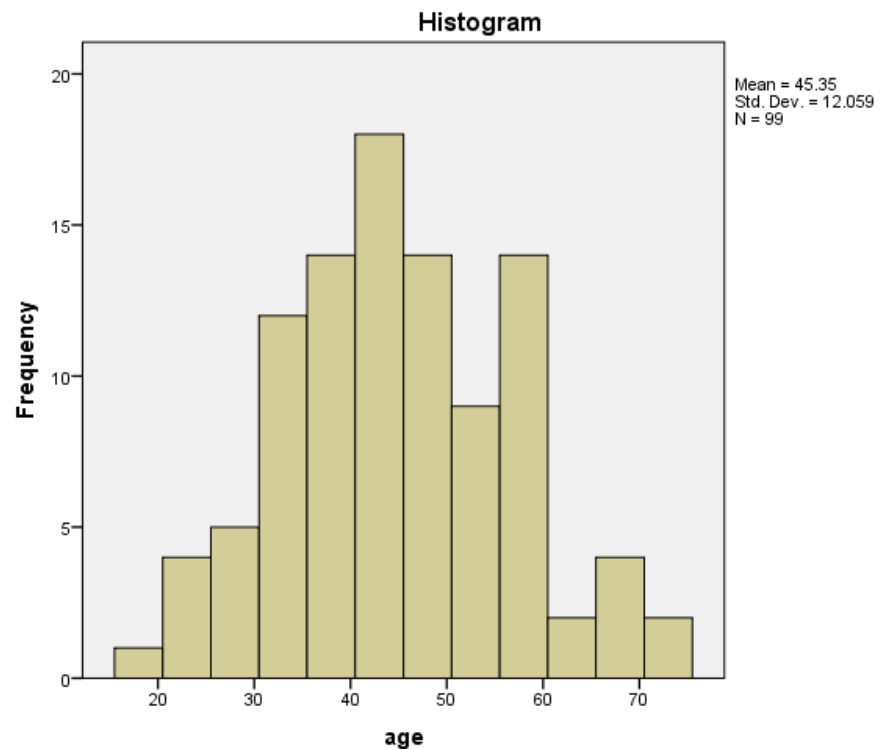
Data Computation and Transformation

► Assessing normality

► Performed using the **Explore** command in **Descriptive Statistics** under **Analyze**

► Histogram

- Need to be Bell shaped /symmetry



Data Computation and Transformation

► Assessing normality

► Boxplot

- Median should be at the center of the body (symmetric)

File > New > Syntax

EXAMINE CGPA

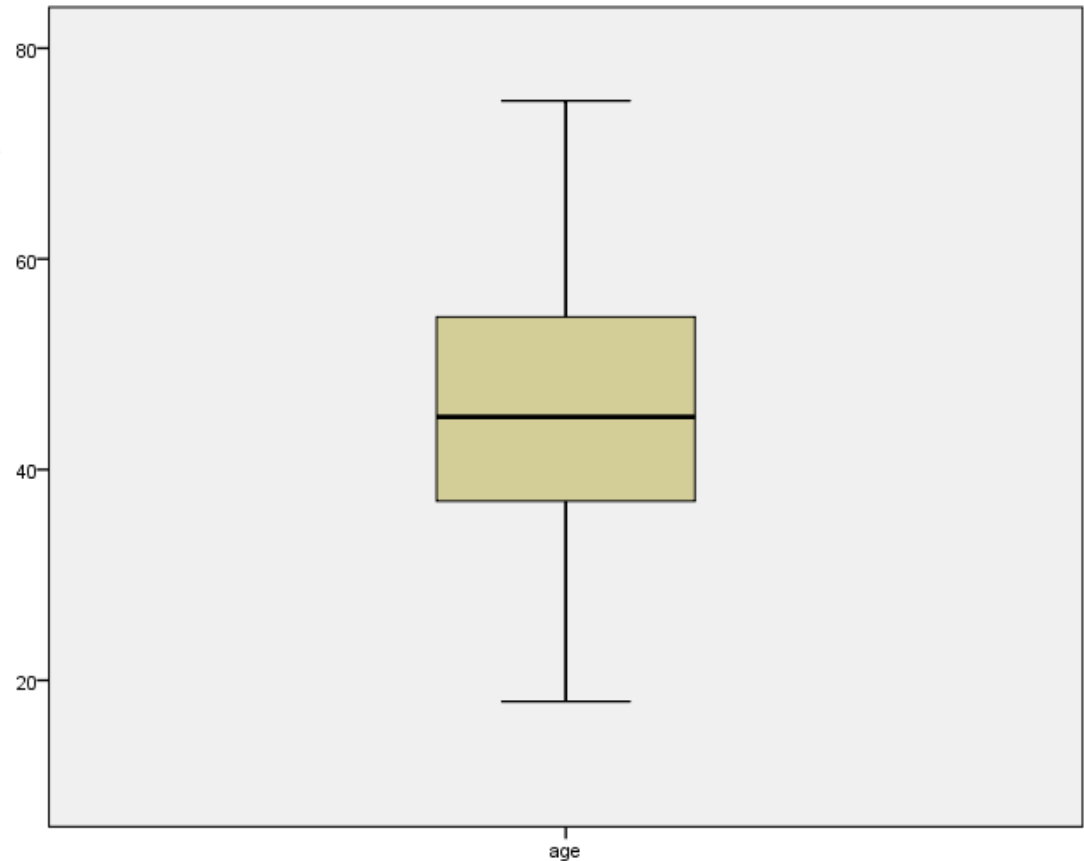
/STATISTICS = DESCRIPTIVES

/PLOT = BOXPLOT

EXAMINE height by Gender

/STATISTICS = DESCRIPTIVES

/PLOT = BOXPLOT



4.3 Data Computation and Transformation

► Assessing normality

► Skewness and Kurtosis

- Their value need to be close to zero to be assumed normal

Descriptives				
			Statistic	Std. Error
Height	Mean		158.74	1.18
	95% Confidence Interval for Mean	Lower Bound	156.31	
		Upper Bound	161.17	
	5% Trimmed Mean		158.66	
	Median		158.00	
	Variance		39.29	
	Std. Deviation		6.27	
	Minimum		145.00	
	Maximum		173.00	
	Range		28.00	
	Interquartile Range		7.50	
	Skewness		.36	.44
	Kurtosis		.41	.86

► Basically, the data of age can be assumed normal

- There is no need for transformation
- Data can be analyzed using parametric test

