

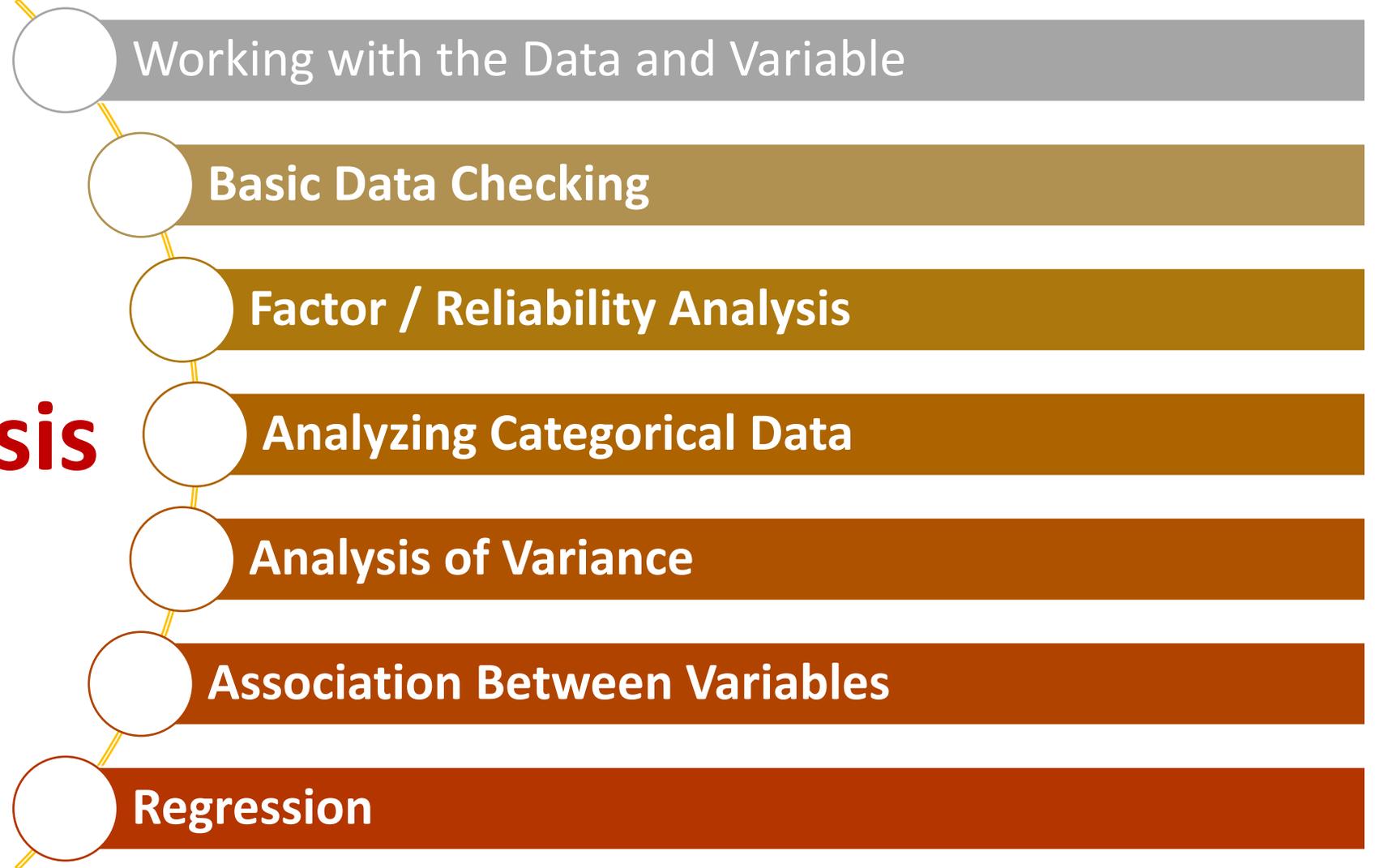


Survey Analysis

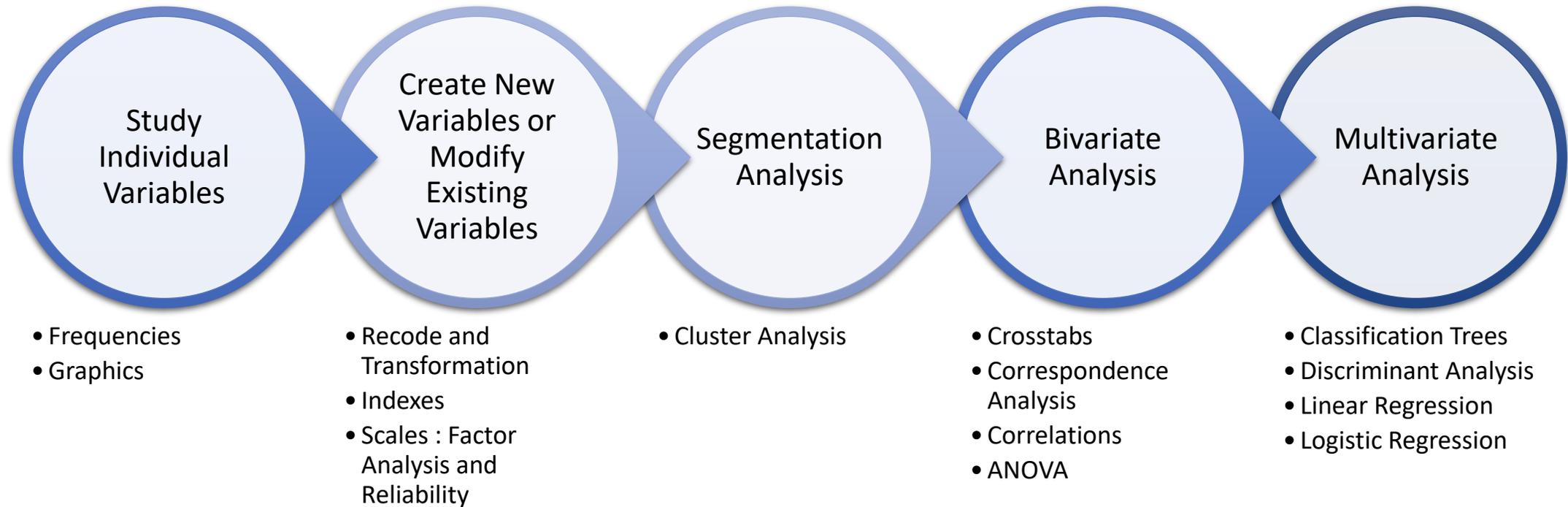
IBM SPSS Statistics

Survey Analysis

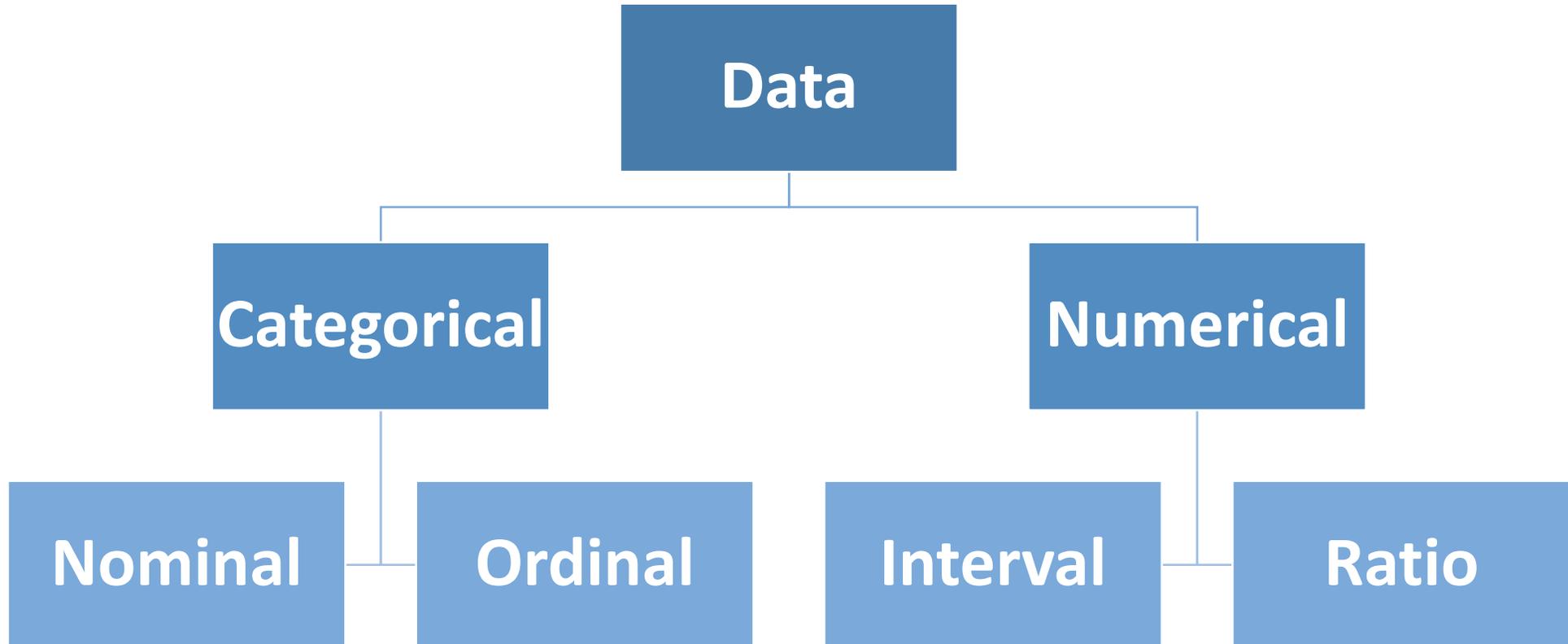
IBM SPSS Statistics



The Process of Survey Analysis



Level of Measurement



Data View

*banks.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

24 : start 07-Mar-1983 Visible: 22 of 22 Variables

	id	completed	first_name	last_name	gender	dob	educ	marit	start	itype	whours	sale
1	0001	20-Jan-2017 11:37:28	Kevin	Garcia	1	03-Oct-1993	.	2	07-May-2016	1	28.25	\$
2	0002	21-Jan-2017 06:30:03	Ayden	Carter	1	31-Oct-1996	4	1	27-Oct-2026	1		\$
3	0003	21-Jan-2017 16:35:48	Madelyn	Williams	0	13-Dec-1985	5	2	07-Nov-2008	1	22.75	\$
4	0004	21-Jan-2017 17:37:33	Madelyn	Baker	0	10-Jun-1988	1	2	21-Feb-2011	.	27.25	\$
5	0005	22-Jan-2017 12:04:23	Tristan	Hernandez	1	23-Dec-1995	3	2	01-Mar-2016	1	.	\$
6	0006	22-Jan-2017 13:05:0	Isaac	Carroll	1	19-Apr-1996	6	2	08-Dec-2016	2	43.75	\$
7	0007	22-Jan-2017 15:44:23	Victoria	Carroll	0	24-Apr-1989	2	2	09-Jun-2016	1	28.50	\$
8	0008	23-Jan-2017 08:58:32	Christopher	Taylor	1	30-Nov-1983	4	2	20-Apr-2005	3	160.00	\$
9	0009	23-Jan-2017 13:37:00	Caroline	Taylor	0	22-Aug-1981	3	1	13-Nov-2005	3	35.25	\$
10	0010	23-Jan-2017 15:11:51	Daniel	Clark	1	30-Mar-1995	3	3	24-Dec-2016	1	28.50	\$
11	0011	23-Jan-2017 15:41:43	Samuel	Perez	1	09-Mar-1980	5	2	11-Mar-2012	4	49.00	\$1
12	0012	23-Jan-2017 16:02:12	Henry	Thomas	1	09-Jan-1992	4	1	10-Feb-2014	3	43.25	\$
13	0013	23-Jan-2017 16:57:29	Brianna	Nelson	0	22-Jan-1992	3	1	01-Mar-2012	3	39.00	\$
14	0014	24-Jan-2017 09:38:08	Ayden	Scott	1	24-Mar-1986	4	1	15-Nov-2009	3	33.00	\$
15	0015	24-Jan-2017 15:01:12	Avery	Moore	0	06-Dec-1988	2	1	14-May-2014	2	33.50	\$
16	0016	24-Jan-2017 21:30:59	Evan	Young	1	11-Sep-1992	4	2	11-Aug-2016	1	33.25	\$
17	0017	25-Jan-2017 11:29:10	Gianna	Green	0	06-Oct-1986	4	1	26-Jul-2016	2	27.75	\$
18	0018	25-Jan-2017 11:36:51	Gabriella	Carter	0	04-Jul-1990	.	1	15-Dec-2014	2	33.00	\$
19	0019	25-Jan-2017 16:07:27	Chloe	Clark	0	14-Sep-1995	4	1	12-Dec-2015	1	180.00	\$
20	0020	25-Jan-2017 16:41:13	Gianna	Jones	0	13-May-1992	4	3	09-Oct-2013	6	26.25	\$
21	0021	25-Jan-2017 17:14:21	Claire	King	0	26-Feb-1988	3	3	22-Jul-2011	1	27.00	\$
22	0022	25-Jan-2017 20:01:05	Hailey	Collins	0	24-Oct-1995	3	1	21-Mar-2045	1	29.50	\$
23	0023	26-Jan-2017 05:36:12	Isaac	Adams	1	16-Mar-1981	5	2	17-Apr-2008	2	31.75	\$
24	0024	26-Jan-2017 08:21:35	Alexa	Campbell	0	25-Feb-1959	6	3	07-Mar-1983	3	24.25	\$
25	0025	26-Jan-2017 08:47:30	Wyatt	Martinez	1	13-Mar-1972	5	4	09-Dec-2012	3	32.00	\$
26	0026	26-Jan-2017 12:43:18	Harper	Roberts	0	05-Feb-1993	3	3	27-Nov-2014	1	21.50	\$
27	0027	26-Jan-2017 13:45:03	Hailey	Edwards	0	08-Dec-1991	4	2	10-Sep-2016	1	30.25	\$
28	0028	26-Jan-2017 14:54:21	Joshua	Smith	1	17-Nov-1980	3	3	09-Nov-2004	1	30.50	\$
29	0029	26-Jan-2017 15:29:28	Ellie	Anderson	0	30-Dec-1988	3	1	17-Jan-2015	1	26.00	\$

Data View Variable View 1

IBM SPSS Statistics Processor is ready Unicode:OFF

1 tabs for switching between Data View and Variable View

2 Columns of cells are called **variables**.

3 Rows of cells are called **observation**

4 **values** refer to cell contents

Variable View

*banks.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Restricted ...	4	0	Questionnaire identifier	None	None	9	Right	Ordinal	Input
2	completed	Date	22	0	Date and time that questionnaire was completed	None	None	17	Right	Scale	Input
3	first_name	String	11	0		None	None	11	Left	Nominal	Input
4	last_name	String	30	0		None	None	26	Left	Nominal	Input
5	gender	Numeric	1	0		{0, female}...	None	10	Right	Nominal	Input
6	dob	Date	11	0	Date of birth	None	None	11	Right	Scale	Input
7	educ	Numeric	1	0	Highest completed education level	{1, Middle school or lower}...	None	10	Right	Nominal	Input
8	marit	Numeric	1	0	Marital status	{1, never married}...	None	10	Right	Nominal	Input
9	start	Date	11	0	Date of enrollment in company	None	None	11	Right	Scale	Input
10	jtype	Numeric	1	0	Current job type	{1, Sales}...	None	10	Right	Nominal	Input
11	whours	Numeric	8	2	On average, how many hours do you work per week?	None	None	15	Right	Scale	Input
12	salary	Dollar	8	0	Gross monthly salary	None	None	10	Right	Scale	Input
13	overall	Numeric	1	0	I'm happy with my job	{1, Totally disagree}...	None	10	Right	Nominal	Input
14	q1	Numeric	1	0	This company takes good care of its employees.	{1, Totally disagree}...	None	10	Right	Nominal	Input
15	q2	Numeric	1	0	This company supports me in my work.	{1, Totally disagree}...	None	10	Right	Nominal	Input
16	q3	Numeric	1	0	My daily tasks are interesting.	{1, Totally disagree}...	None	10	Right	Nominal	Input
17	q4	Numeric	1	0	I like my colleagues.	{1, Totally disagree}...	None	10	Right	Nominal	Input
18	q5	Numeric	1	0	My workspace is good.	{1, Totally disagree}...	None	10	Right	Nominal	Input
19	q6	Numeric	1	0	My salary is good.	{1, Totally disagree}...	None	10	Right	Nominal	Input
20	q7	Numeric	1	0	My secondary labor conditions are good.	{1, Totally disagree}...	None	10	Right	Nominal	Input
21	q8	Numeric	1	0	My work is meaningful.	{1, Totally disagree}...	None	10	Right	Nominal	Input
22	q9	Numeric	1	0	The cooperation with my colleagues is good.	{1, Totally disagree}...	None	10	Right	Nominal	Input
23											
24											
25											
26											
27											
28											
29											
30											
31											

Data View Variable View

Name	Label
id	Questionnaire identifier
completed	Date and time that questionnaire was completed
first_name	
last_name	
gender	
dob	Date of birth
educ	Highest completed education level
marit	Marital status
start	Date of enrollment in company
jtype	Current job type
whours	On average, how many hours do you work per week?
salary	Gross monthly salary
overall	I'm happy with my job
q1	This company takes good care of its employees.
q2	This company supports me in my work.
q3	My daily tasks are interesting.
q4	I like my colleagues.
q5	My workspace is good.
q6	My salary is good.
q7	My secondary labor conditions are good.
q8	My work is meaningful.
q9	The cooperation with my colleagues is good.

Value

5 Gender

0 = "female"
1 = "male"

7 Educ

1 = "Middle school or lower"
2 = "High school"
3 = "Undergraduate"
4 = "Bachelor's"
5 = "Master's"
6 = "PhD or higher"
7 = "No answer"

8 Marit

1 = "never married"
2 = "currently married"
3 = "divorced"
4 = "widowed"
5 = "No answer"

10 Jtype

1 = "Sales"
2 = "Marketing"
3 = "IT"
4 = "Middle Management"
5 = "Upper Management"
6 = "No answer"

13+ Overall – q9

1 = "Totally disagree"
10 = "Totally agree"
11 = "No answer"

5

7

8

10

13
+

Basic Data Checking

The Common Data Problems

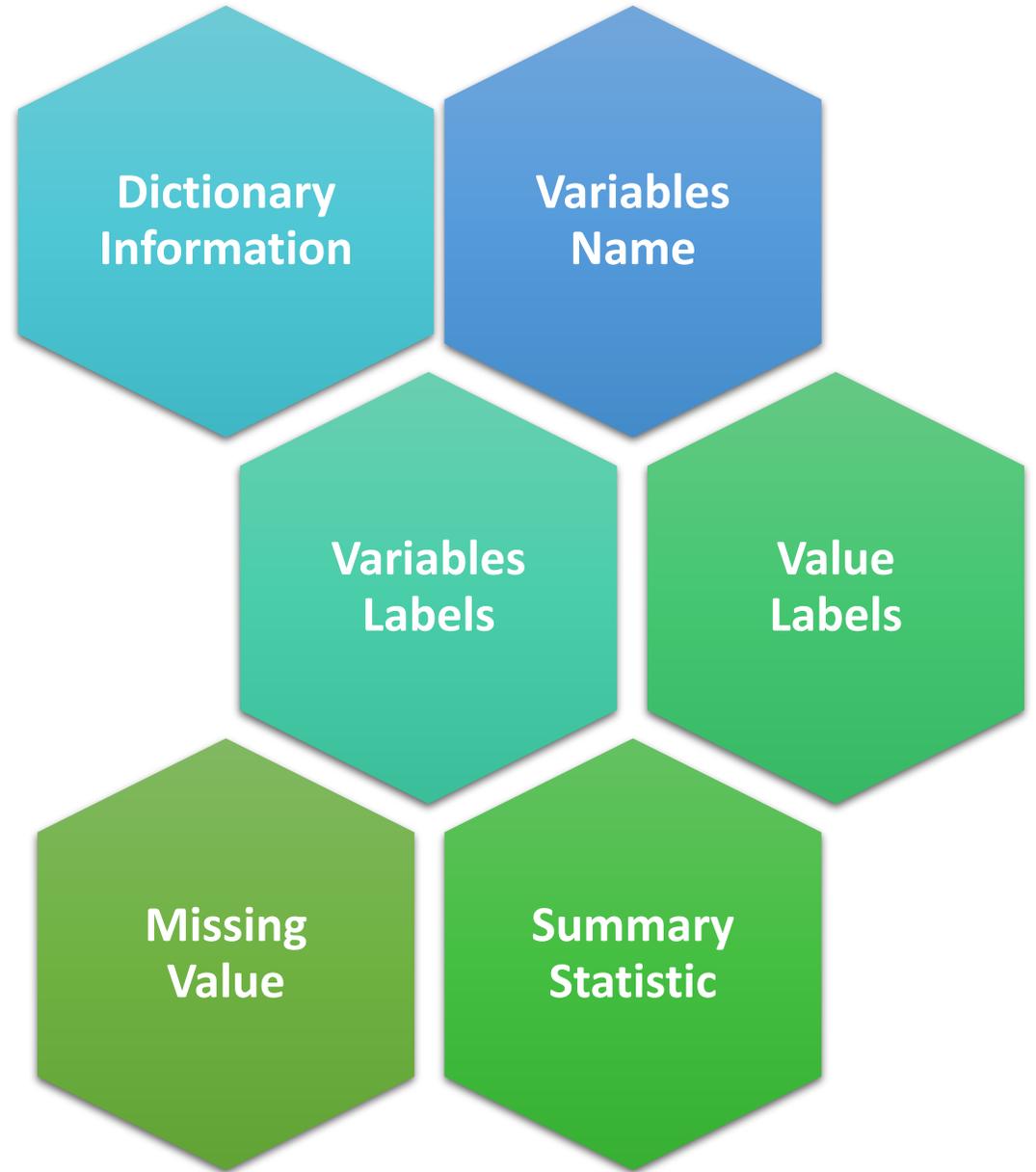
Run frequencies on all variables

Run descriptive on all scale variables

Run crosstabulation to check whether skip were correctly followed

Run additional crosstabulation to look for odd pattern

Create a Variable Codebook



1.1 Codebook

Data : "SPSS_CUST.SAV"

Analyze >> Reports >> CodeBook

Codebook

Variables Output Statistics

Variables:

- Caseid [caseid]

Codebook Variables:

- SPSS prods are an integr prt of wrk...
- SPSS prods/srvcs make job easier ...
- SPSS prods are a gd val [gdvalue]
- Am a satisfied SPSS customer [sat...]
- SPSS offers hi-quality prods [hiqualty]
- SPSS s/w has stat procs needed [s...]
- SPSS prods are easy to learn [easy...]
- SPSS prods are easy to use [easyu...]
- SPSS prods are updated freqntly [u...]
- Am satisfd w/receivd info on new pr...
- SPSS products are priced right [gdp...]
- SPSS'licensing terms are flexible [li...]
- SPSS shipments receivd whn promi...
- SPSS shipments are correct & com...
- Am satisfied w/SPSScommn about ...
- Spoken w/SPSS sales rep w/in pas...
- Sales dept returns my calls promptl...
- Sales rep undrstnds my stat data a...
- Sales rep undrstnds computng env...
- Sales rep undrstnds your business...
- Sales rep relates prods to my need...
- Sales rep informs about all prods & ...
- Sales rep treats customer w/courte...

OK Paste Reset Cancel Help

freqspss

		Value	Count	Percent
Standard Attributes	Position	86		
	Label	How frequently do you use SPSS		
	Type	Numeric		
	Format	F1		
	Measurement	Ordinal		
	Role	Input		
Valid Values	1	Don't use very much at all	85	8.9%
	2	Only for special projects	95	9.9%
	4	Monthly	160	16.8%
	5	Weekly	120	12.6%
	6	Daily/multiple times in a week	250	26.2%
	7	Multiple times in a day	115	12.0%
	Missing Values	3	Quarterly	115
System			15	1.6%

1.2 Using Frequencies to Check Data

Analyze >> Descriptive >> Frequencies

Variable(s):

- SPSS prods are an integr prt of wrk [i...]
- SPSS prods/srvics make job easier [j...]
- SPSS prods are a gd val [gdvalue]
- Am a satisfied SPSS customer [satsc...]
- SPSS offers hi-quality prods [hiquality]
- SPSS s/w has stat procs needed [sta...]
- SPSS prods are easy to learn [easylm]
- SPSS prods are easy to use [easyuse]
- SPSS prods are updated freqntly [upd...]
- Am satisfd w/receivd info on new prod...
- SPSS products are priced right [gdpri...]
- SPSS'licensing terms are flexible [lice...]
- SPSS shipments receivd whn promis...
- SPSS shipments are correct & compl...
- Am satisfied w/SPSScommn about pr...
- Spoken w/SPSS sales rep w/in past ...
- Sales dept returns my calls promptly ...
- Sales rep undrstnds my stat data ana...
- Sales rep undrstnds computng envmt ...
- Sales rep undrstnds your business/o...

Display frequency tables Create APA style tables

OK Paste Reset Cancel Help

Frequencies

Statistics

How frequently do you use SPSS

N	Valid	825
	Missing	130

How frequently do you use SPSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Don't use very much at all	85	8.9	10.3	10.3
	Only for special projects	95	9.9	11.5	21.8
	Monthly	160	16.8	19.4	41.2
	Weekly	120	12.6	14.5	55.8
	Daily/multiple times in a week	250	26.2	30.3	86.1
	Multiple times in a day	115	12.0	13.9	100.0
	Total	825	86.4	100.0	
Missing	Quarterly	115	12.0		
	System	15	1.6		
	Total	130	13.6		
Total		955	100.0		

1.3 Checking Skip Pattern

Analyze >> Descriptive >> Crosstabs

Row(s):

- Satisfied w/tech support [techsat]

Column(s):

- Ever called SPSS for tech supprt? [techcall]

Layer 1 of 1

Previous Next

Display clustered bar charts

Suppress tables

OK Paste Reset Cancel Help

Crosstabs

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Satisfied w/tech support * Ever called SPSS for tech supprt?	571	59.8%	384	40.2%	955	100.0%

Satisfied w/tech support * Ever called SPSS for tech supprt? Crosstabulation

Count		Ever called SPSS for tech supprt?		Total
		Yes	No	
Satisfied w/tech support	Strongly Agree	75	1	76
	Agree	245	0	245
	Neither Agree Nor Disagree	60	0	60
	Disagree	140	0	140
	Strongly Disagree	50	0	50
Total		570	1	571

Exclude Incorrect Value

Data >> Select Cases

Select Cases

Select

All cases

If condition is satisfied

if... ~ (techcall = 2 & MISSIN...

Random sample of cases

Sample...

Based on time or case range

Range...

Use filter variable:

Output

Filter out unselected cases

Copy selected cases to a new dataset

Dataset name:

Delete unselected cases

Current Status: Filter cases by values of filter_\$

OK Paste Reset Cancel Help

Select Cases: If

~ (techcall = 2 & MISSING(techsat) = 0)

Function group:

All

Arithmetic

CDF & Noncentral CDF

Conversion

Current Date/Time

Date Arithmetic

Date Creation

Date Extraction

Inverse DF

Functions and Special Variables:

	caseid	intgresp	jobeasy
1	1	2	2
2	2	1	4
3	3	2	1
4	4	1	1
5	5	2	2

Continue Cancel Help

2. Data Validation

Data : "SPSS_CUST.SAV"

Data >> Validation >> Validate Data

Select Variables

Validate Data

Variables Basic Checks Single-Variable Rules Cross-Variable Rules Output Save

Variables:

- Caseid [caseid]
- Am a satisfied SPSS customer [satcust]
- SPSS offers hi-quality prods [hiqualty]
- SPSS s/w has stat procs needed [statneed]
- SPSS prods are easy to learn [easylrn]
- SPSS prods are easy to use [easyuse]
- SPSS prods are updated freqntly [update]
- Am satisfd w/receivd info on new prods [satinfo]
- SPSS products are priced right [gdprice]
- SPSS'licensing terms are flexible [license]
- SPSS shipments receivd whn promisd [shipget]
- SPSS shipments are correct & complete [shipcomp]
- Am satisfied w/SPSScommn about probs [satcomun]
- Spoken w/SPSS sales rep w/in past year? [salesrep]
- Sales dept returns my calls promptly [repretrn]
- Sales rep undrstnds my stat data anal needs [repanal]
- Sales rep undrstnds computng envmt [repenv]
- Sales rep undrstnds your business/organ [repbu]
- Sales rep relates prods to my needs [repneeds]
- Sales rep informs about all prods & svcs [repinfrm]
- Sales rep treats customer w/courtesy [repcourt]
- Sales rep gives info in right amt of time [reptime]
- Ever called SPSS for tech supprt? [techcall]
- Satisfied w/tech support [techsat]
- Tech supprt answers questns quickly [techtime]
- Tech supprt returns calls promptly [techrtn]
- Specialist gives response promptly [techspec]

Analysis Variables:

- SPSS prods are an integri prt of wrk [intgresp]
- SPSS prods/srvics make job easier [jobeas]
- SPSS prods are a gd val [gdvalue]

Case Identifier Variables:

OK Paste Reset Cancel Help

Identify Basic Checks

Validate Data

Variables Basic Checks Single-Variable Rules Cross-Variable Rules Output Save

Analysis Variables

Flag variables that fail any of the following checks

Maximum percentage of missing values: 70 (Applies to all variables)

Maximum percentage of cases in a single category: 95 (Applies to categorical variables only)

Maximum percentage of categories with count of 1: 90 (Applies to categorical variables only)

Minimum coefficient of variation: 0.001 (Applies to scale variables only)

Minimum standard deviation: 0 (Applies to scale variables only)

Case Identifiers

Flag incomplete IDs

Flag duplicate IDs

Flag empty cases Define Cases By: All variables in dataset except ID variables

A case is considered empty if all relevant variables are missing or blank.

OK Paste Reset Cancel Help

To apply rules to a variable, select the variable then check one or more rules.

The Analysis Variables list shows distributions of nonmissing values based on a scan of the data. The Rules list shows all rules that can be applied to selected variables.

Analysis Variables:

Variable	Distribution	Minimum	Maximum	Rules
SPSS p...		1	5	1
SPSS p...		1	5	1
SPSS p...		1	5	1

Rules:

Apply	Name
<input type="checkbox"/>	Likert
<input type="checkbox"/>	1,2 dichotomy
<input checked="" type="checkbox"/>	1 to 5 integer
<input type="checkbox"/>	1 to 10 integer
<input type="checkbox"/>	Nonnegative number
<input type="checkbox"/>	Nonnegative integer
<input type="checkbox"/>	0 to 100 number
<input type="checkbox"/>	Flag system-missing values
<input type="checkbox"/>	Flag user-missing values
<input type="checkbox"/>	Flag missing values
<input type="checkbox"/>	Flag noninteger values
<input type="checkbox"/>	Flag unlabeled values

Display: All variables Cases Scanned: 955 [Define Rules...](#)

Variable Distributions
 Limit number of cases scanned Cases: 5000 [Rescan](#) Limiting the number of cases scanned does not affect how many cases are validated.

[OK](#) [Paste](#) [Reset](#) [Cancel](#) [Help](#)

All cases passed the requested checks

Validate Data

Warnings

Some or all requested output is not displayed because all cases, variables, or data values passed the requested checks.

Single-Variable Rules

Rules:

Name	Type
Likert	Numeric
1,2 dichotomy	Numeric
1 to 5 integer	Numeric
1 to 10 integer	Numeric
Nonnegative ...	Numeric
Nonnegative i...	Numeric
0 to 100 num...	Numeric
Flag system-...	Numeric
Flag user-mis...	Numeric
Flag missing ...	Numeric
Flag noninteg...	Numeric
Flag unlabele...	Numeric
Sex (1 char.)	String
Sex (full)	String
Day of week (...)	String
Day of week (...)	String
Month (3 char.)	String
Month (full)	String
U.S. states (...)	String

New Duplicate Delete

Rule Definition

Name: Type: Format:

Valid Values:

Values:

Ignore case when checking values

Allow user-missing values

Allow system-missing values

Allow blank values

Continue Cancel Help

Single-Variable Rules

Rule Descriptions	
Rule	Description
1 to 5 integer	Type: Numeric Domain: Range Flag user-missing values: No Flag system-missing values: No Minimum: 1 Maximum: 5 Flag unlabeled values within range: No Flag noninteger values within range: Yes \$VD.SRule[3]: Rule
Rules violated at least once are displayed.	

Variable Summary

	Rule	Number of Violations
SPSS prods are an integr prt of wrk	1 to 5 integer	1
	Total	1

Case Report

Case	Validation Rule Violations
	Single-Variable ^a
1	1 to 5 integer (1)

a. The number of variables that violated the rule follows each rule.

3. Variable Binning

Data : "GSS2008.SAV"

Graphs >> Chart Builder

Chart Builder

Variables: *Chart preview uses example data*

- GENDER OF 7...
- GENDER OF 8...
- GENDER OF 9...
- GENERAL HAP...
- GEOGRAPHIC ...
- GOD CONCER...
- GOVMNT SHO...
- GOVT OR PRIV...
- GSS YEAR FO...

No categories (scale variable)

Simple Histogram of AGE OF RESPONDENT

Histogram

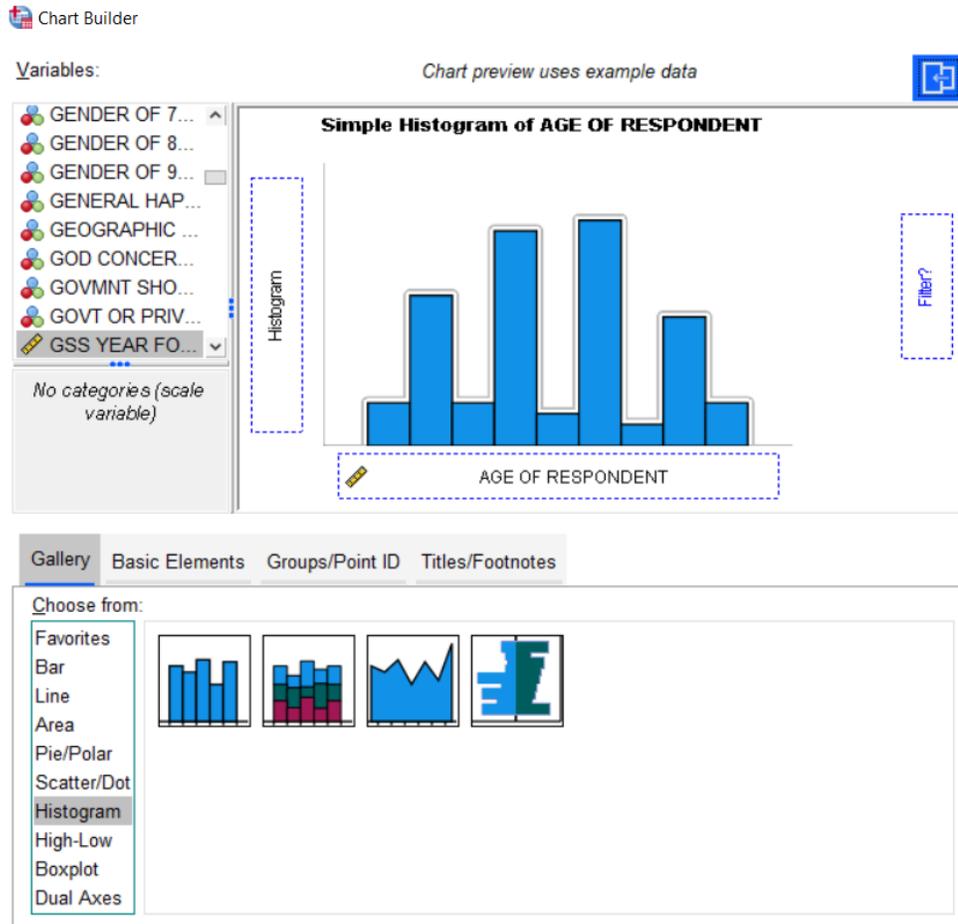
AGE OF RESPONDENT

Filter?

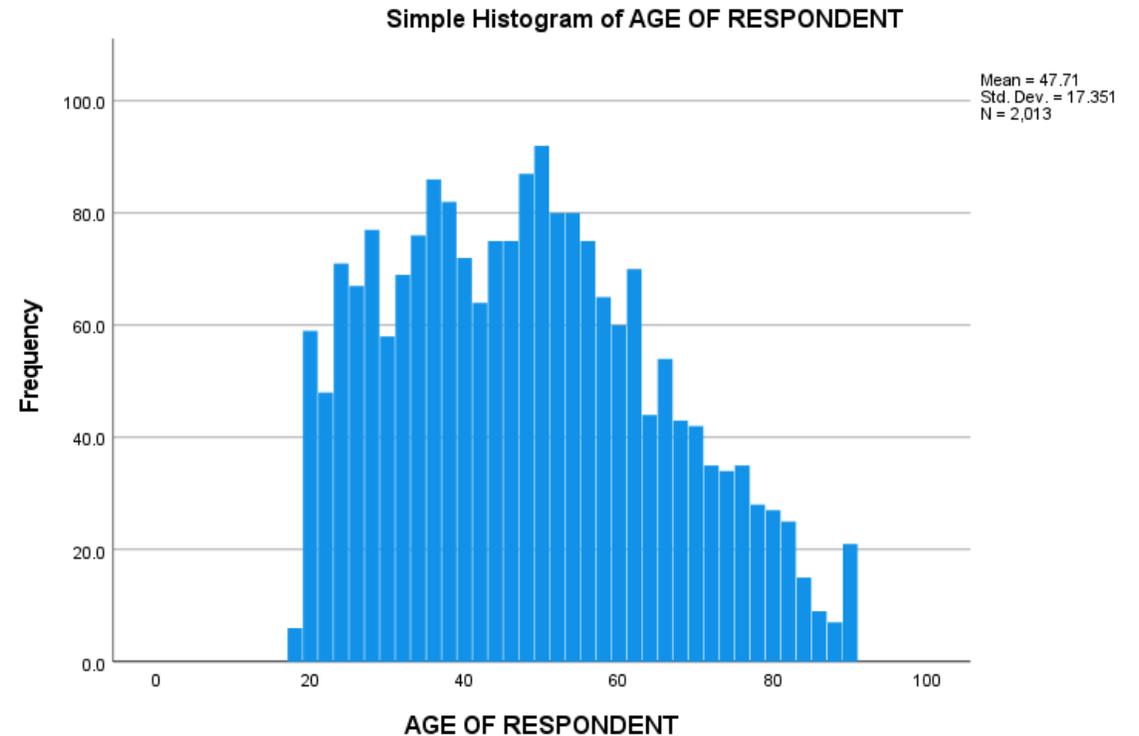
Gallery Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

- Favorites
- Bar
- Line
- Area
- Pie/Polar
- Scatter/Dot
- Histogram**
- High-Low
- Boxplot
- Dual Axes



Histogram of age



Data : "GSS2008.SAV"

Transform >> Visual Binning

Visual Binning

Select the variables whose values will be grouped into bins. Data will be scanned when you click Continue.

The Variables list below contains all numeric ordinal and scale variables.

Variables:

- agecmeus
- agekdbm
- bornsp
- cohort
- colscinm
- curempyr
- dateintv
- denkid
- denom
- denom16
- densp
- emphplan
- eth1
- eth2
- eth3
- ethnic

Variables to Bin:

- age

Limit number of cases scanned to:

Continue Cancel Help

Visual Binning

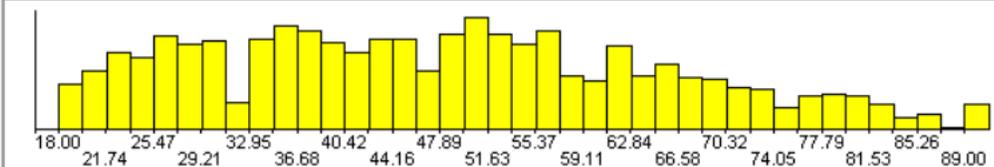
Scanned Variable List:

- age

Name: Label:

Current Variable: Binned Variable:

Minimum: Nonmissing Values Maximum:



Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid:

	Value	Label
1		HIGH
2		

Upper Endpoints

Included (<=)

Excluded (<)

Make Cutpoints...

Make Labels

Reverse scale

OK Paste Reset Cancel Help

Make Cut points

 Make Cutpoints ✕

Equal Width Intervals

Intervals - fill in at least two fields

First Cutpoint Location:

Number of Cutpoints:

Width:

Last Cutpoint Location: 79

Equal Percentiles Based on Scanned Cases

Intervals - fill in either field

Number of Cutpoints:

Width(%):

Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases

+/- 1 Std. Deviation

+/- 2 Std. Deviation

+/- 3 Std. Deviation

 Apply will replace the current cutpoint definitions with this specification.
A final interval will include all remaining values: N cutpoints produce N+1 intervals.

Equal width Interval

Equal percentile intervals

Mean and standard deviation interval

Scanned Variable List:

age

Name:

Label:

Current Variable: age

AGE OF RESPONDENT

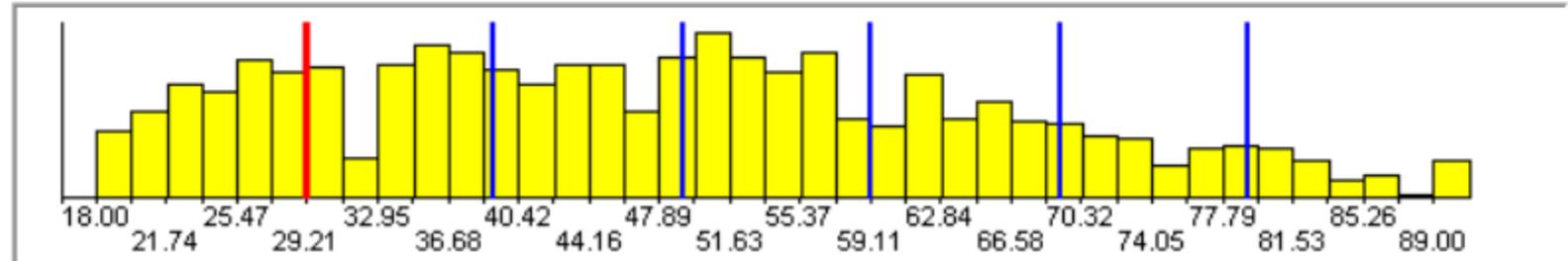
Binned Variable: agebinned

AGE OF RESPONDENT (Binned)

Minimum: 18

Nonmissing Values

Maximum: 89



i Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Cases Scanned: 2023

Missing Values: 10

Copy Bins

From Another Variable...

To Other Variables...

Grid:

	Value	Label
1	29.0	
2	39.0	
3	49.0	
4	59.0	
5	69.0	
6	79.0	
7	HIGH	
8		

Upper Endpoints

Included (\leq)

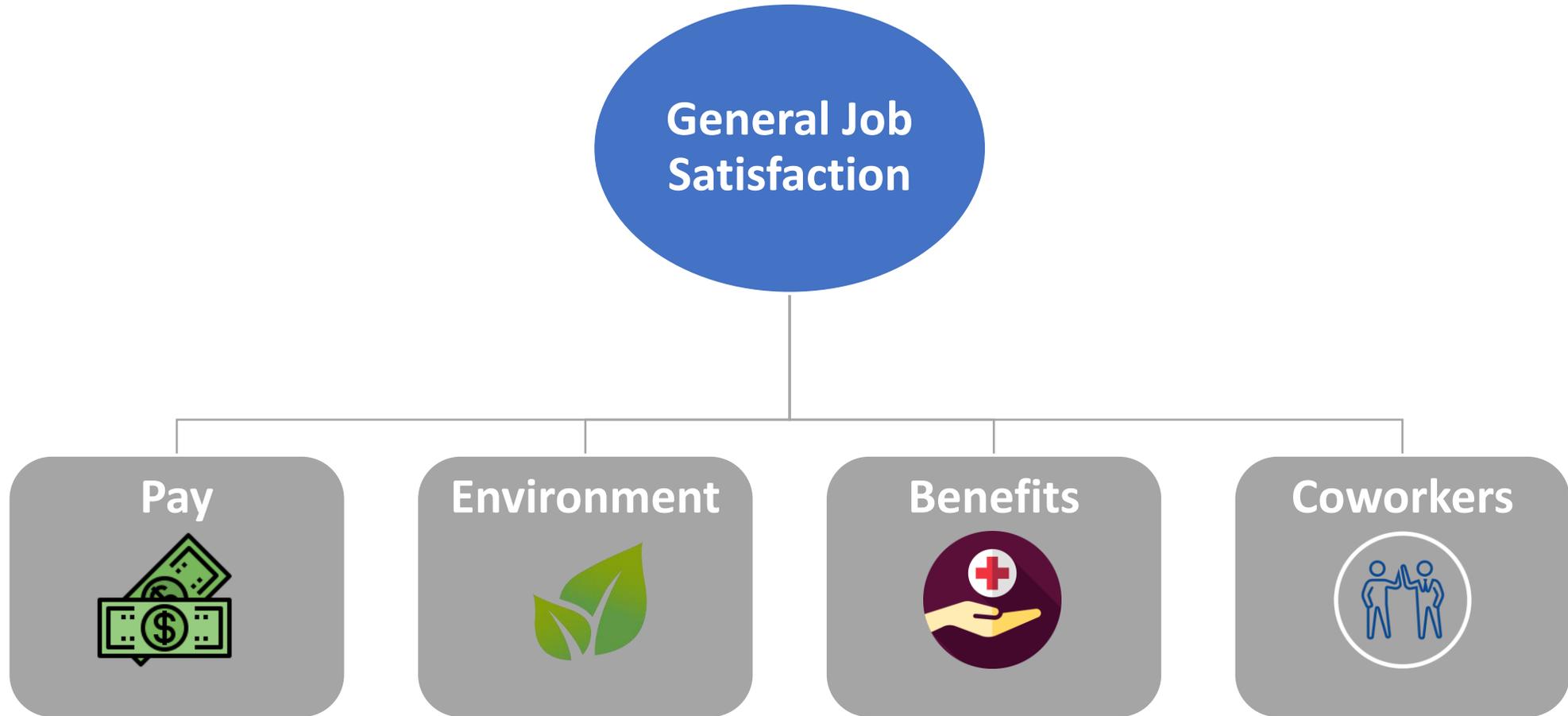
Excluded ($<$)

Make Cutpoints...

Make Labels

Reverse scale

4. Factor Analysis



Reviewing Correlation Among the Sales Rep Items

Analyze >> Correlate >> Bivariate

Data : "SPSS_CUST.SAV"

Bivariate Correlations

Caseid [caseid]

SPSS prods are an integrl prt of wrk [intgresp]

SPSS prods/srvics make job easier [jobeasy]

SPSS prods are a gd val [gdvalue]

Am a satisfied SPSS customer [satcust]

SPSS offers hi-quality prods [hiquality]

SPSS s/w has stat procs needed [statneed]

SPSS prods are easy to learn [easylrn]

SPSS prods are easy to use [easyuse]

SPSS prods are updated freqntly [update]

Am satisfd w/receivd info on new prods [satinfo]

SPSS products are priced right [gdprice]

SPSS'licensing terms are flexible [license]

SPSS shipments receivd whn promisd [shipget]

SPSS shipments are correct & complete [shi...]

Am satisfied w/SPSScommn about probs [sat...]

Seeker w/SPSS sales rep w/in past year? [e...]

Variables:

Sales rep undrstnds my stat data anal needs [re...]

Sales rep undrstnds computng envmt [repenv]

Sales rep undrstnds your business/organ [repb...]

Sales rep relates prods to my needs [repneeds]

Sales rep informs about all prods & srvcs [repinfr...]

Sales rep treats customer w/courtesy [repcourt]

Sales rep gives info in right amt of time [reptime]

Correlation Coefficients

Pearson Kendall's tau-b Spearman

Test of Significance

Two-tailed One-tailed

Flag significant correlations Show only the lower triangle Show diagonal

All Variables are positively related to each other

Correlations

		Sales rep undrstnds my stat data anal needs	Sales rep undrstnds computng envmt	Sales rep undrstnds your business/org an	Sales rep relates prods to my needs	Sales rep informs about all prods & svcs	Sales rep treats customer w/courtesy	Sales rep gives info in right amt of time
Sales rep undrstnds my stat data anal needs	Pearson Correlation	1	.623**	.689**	.693**	.478**	.499**	.480**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	585	555	510	540	575	580	580
Sales rep undrstnds computng envmt	Pearson Correlation	.623**	1	.702**	.593**	.534**	.418**	.523**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	555	575	520	535	555	570	570
Sales rep undrstnds your business/organ	Pearson Correlation	.689**	.702**	1	.719**	.507**	.290**	.398**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	510	520	530	515	520	525	525
Sales rep relates prods to my needs	Pearson Correlation	.693**	.593**	.719**	1	.561**	.353**	.390**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	540	535	515	555	550	545	545
Sales rep informs about all prods & svcs	Pearson Correlation	.478**	.534**	.507**	.561**	1	.576**	.544**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	575	555	520	550	610	600	600
Sales rep treats customer w/courtesy	Pearson Correlation	.499**	.418**	.290**	.353**	.576**	1	.605**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	580	570	525	545	600	625	620
Sales rep gives info in right amt of time	Pearson Correlation	.480**	.523**	.398**	.390**	.544**	.605**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	580	570	525	545	600	620	625

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

Requesting a Factor Analysis

Choose the set of variables to be factors

Choose a method of factor extraction

Choose a method of factor rotation

Request factor display options

Analyze >> Dimension Reduction >> Factor

Factor Analysis

- SPSS products are priced right [gdprice]
- SPSS'licensing terms are flexible [license]
- SPSS shipments received when promised [shi...]
- SPSS shipments are correct & complete [...]
- Am satisfied w/SPSScommn about probs [...]
- Spoken w/SPSS sales rep w/in past year?...
- Sales dept returns my calls promptly [repr...]
- Ever called SPSS for tech supprt? [techcall]
- Satisfied w/tech support [techsat]
- Tech supprt answers questns quickly [tech...]
- Tech supprt returns calls promptly [techrtn]
- Specialist gives response promptly [techsp...]
- Provided needed answers [techans]
- Knowledge of SPSS products [techknow]
- Knowledge of stats [techstat]
- Understands computing envmt [techenv]
- Treated w/courtesy [techcour]
- Do you receive Keywords? [reckeyw]
- Read Keywords? [readkeyw]
- Overall satisfaction w/Keywords [satkeyw]
- Amount/kind of prod info in Keywords [info...]
- Amount/kind of how-to info in Keywords [h...]

Variables:

- Sales rep undrstnds my stat data anal needs...
- Sales rep undrstnds computng envmt [repenv]
- Sales rep understands your business/organ [re...]
- Sales rep relates prods to my needs [repneeds]
- Sales rep informs about all prods & svcs [rep...]
- Sales rep treats customer w/courtesy [repco...]
- Sales rep gives info in right amt of time [repti...]

Selection Variable:

Value...

Extraction

Rotation

Options

Factor Analysis: Extraction

Method: **Principal components**

Analyze

- Correlation matrix
- Covariance matrix

Display

- Unrotated factor solution
- Scree plot

Extract

- Based on Eigenvalue

Eigenvalues greater than:

- Fixed number of factors

Factors to extract:

Maximum Iterations for Convergence:

Factor Analysis: Rotation

Method

- None
- Varimax**
- Direct Oblimin
- Quartimax
- Equamax
- Promax

Delta: Kappa:

Display

- Rotated solution
- Loading plot(s)**

Maximum Iterations for Convergence:

Factor Analysis: Options

Missing Values

- Exclude cases listwise
- Exclude cases pairwise
- Replace with mean

Coefficient Display Format

- Sorted by size**
- Suppress small coefficients

Absolute value below:

7 Possible Factors

The extracted factors will have eigenvalues above 1

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.334	61.914	61.914	4.334	61.914	61.914	2.912	41.598	41.598
2	1.017	14.532	76.446	1.017	14.532	76.446	2.439	34.848	76.446
3	.477	6.811	83.257						
4	.408	5.829	89.087						
5	.335	4.784	93.871						
6	.232	3.311	97.182						
7	.197	2.818	100.000						

Extraction Method: Principal Component Analysis.

We look for a factor solution where variables have a high loading

Rotated Component Matrix^a

	Component	
	1	2
Sales rep understands your business/organ	.890	.188
Sales rep relates prods to my needs	.843	.234
Sales rep understands my stat data anal needs	.822	.330
Sales rep understands computng envmt	.681	.468
Sales rep treats customer w/courtesy	.192	.860
Sales rep gives info in right amt of time	.245	.837
Sales rep informs about all prods & srvc	.416	.762

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Factor Interpretation



Component 1



Component 2

5. Reliability Analysis

Measure of the consistency of a measurement overtime

“with questions about the performance of sales reps for SPSS software, if reliability is high, respondents who strongly agree that the sale rep understand their business are also likely to agree that the sales rep understand their computing environment”

Cronbach's Alpha : varies from 0 to 1

Analyze >> Scale >> Reliability Analysis

Data : "SPSS_CUST.SAV"

Reliability Analysis

- SPSS s/w has stat procs needed [sta...]
- SPSS prods are easy to learn [easylrn]
- SPSS prods are easy to use [easyuse]
- SPSS prods are updated freqntly [upd...]
- Am satisfd w/receivd info on new prod...
- SPSS products are priced right [gdprice]
- SPSS'licensing terms are flexible [lice...]
- SPSS shipments receivd whn promisd...
- SPSS shipments are correct & compl...
- Am satisfied w/SPSScommn about pr...
- Spoken w/SPSS sales rep w/in past y...
- Sales dept returns my calls promptly [...]
- Sales rep undrstnds my stat data anal...
- Sales rep undrstnds computng envmt ...
- Sales rep undrstnds your business/o...
- Sales rep relates prods to my needs [r...
- Sales rep informs about all prods & sr...
- Sales rep treats customer w/courtesy ...
- Sales rep gives info in right amt of tim...
- Ever called SPSS for tech supprt? [te...
- Satisfied w/tech support [techsat]
- Tech supprt answers questns quickly

- Items:
- Sales rep undrstnds my stat data anal...
 - Sales rep undrstnds computng envmt [re...
 - Sales rep undrstnds your business/orga...
 - Sales rep relates prods to my needs [rep...

Statistics...

- Ratings:

Model: Alpha

Scale label:

Reliability Analysis: Statistics

X

- Descriptives for
- Item
 - Scale
 - Scale if item deleted

- Inter-Item
- Correlations
 - Covariances

- Summaries
- Means
 - Variances
 - Covariances
 - Correlations

- ANOVA Table
- None
 - F test
 - Friedman chi-square
 - Cochran chi-square

- Interrater Agreement: Fleiss' Kappa
- Display agreement on individual categories
 - Ignore string cases
 - String category labels are displayed in uppercase
- Asymptotic significance level (%): 95

- Missing
- Exclude both user-missing and system missing values
 - User-missing values are treated as valid

- Hotelling's T-square
- Tukey's test of additivity
- Intraclass correlation coefficient

Model: Two-Way Mixed

Type: Consistency

Confidence interval: 95 %

Test value: 0

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.886	.889	4

Item Statistics

	Mean	Std. Deviation	N
Sales rep undrstnds my stat data anal needs	2.30	.878	500
Sales rep undrstnds computng envmt	2.35	1.082	500
Sales rep undrstnds your business/organ	2.55	1.034	500
Sales rep relates prods to my needs	2.63	1.056	500

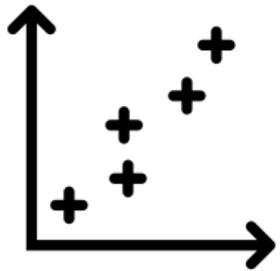
Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Sales rep undrstnds my stat data anal needs	7.53	7.644	.797	.642	.842
Sales rep undrstnds computng envmt	7.48	7.104	.696	.527	.876
Sales rep undrstnds your business/organ	7.28	6.895	.795	.634	.835
Sales rep relates prods to my needs	7.20	7.054	.734	.599	.860

6. Analyzing Categorical Data



- Crosstabs are commonly used to explore how demographic characteristics are related to attitudes and behaviors



- Use to study the relationships between two, or more categorical variables

Business context

We are concerned in surveys with both descriptive and causal relationship we might want to know whether:

- Satisfaction with the instructor in a training workshop was related to satisfaction with the course material
- Eating more often at fast-food restaurant was related to more frequent shopping at convenience store

Data : "GSS2008.SAV"

Analyze >> Descriptive >> Crosstabs

Crosstabs

Row(s):
hapmar

Column(s):
happy

Layer 1 of 1
Previous Next

Display layer variables in table layers

Display clustered bar charts
 Suppress tables

Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...

Cells

Crosstabs: Cell Display

Counts
 Observed
 Expected
 Hide small counts
Less than 5

z-test
 Compare column proportions
 Adjust p-values (Bonferroni method)

Percentages
 Row
 Column
 Total
 Create APA style table

Residuals
 Unstandardized
 Standardized
 Adjusted standardized

Noninteger Weights
 Round cell counts
 Round case weights
 Truncate cell counts
 Truncate case weights
 No adjustments

HAPPINESS OF MARRIAGE * GENERAL HAPPINESS Crosstabulation

		GENERAL HAPPINESS				
			VERY HAPPY	PRETTY HAPPY	NOT TOO HAPPY	Total
HAPPINESS OF MARRIAGE	VERY HAPPY	Count	356	215	24	595
		% within HAPPINESS OF MARRIAGE	59.8%	36.1%	4.0%	100.0%
		% within GENERAL HAPPINESS	89.4%	43.9%	29.6%	61.4%
	PRETTY HAPPY	Count	40	259	44	343
		% within HAPPINESS OF MARRIAGE	11.7%	75.5%	12.8%	100.0%
		% within GENERAL HAPPINESS	10.1%	52.9%	54.3%	35.4%
	NOT TOO HAPPY	Count	2	15	13	30
		% within HAPPINESS OF MARRIAGE	6.7%	50.0%	43.3%	100.0%
		% within GENERAL HAPPINESS	0.5%	3.1%	16.0%	3.1%
	DK	Count	0	1	0	1
		% within HAPPINESS OF MARRIAGE	0.0%	100.0%	0.0%	100.0%
		% within GENERAL HAPPINESS	0.0%	0.2%	0.0%	0.1%
Total	Count	398	490	81	969	
	% within HAPPINESS OF MARRIAGE	41.1%	50.6%	8.4%	100.0%	
	% within GENERAL HAPPINESS	100.0%	100.0%	100.0%	100.0%	

Determining Whether 2 Variables are Related

File Edit View Insert Pivot Format Analyze Graphs Utilities Window Help

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
HAPPINESS OF MARRIAGE * GENERAL HAPPINESS	969	47.9%	1054	52.1%	2023	100.0%

HAPPINESS OF MARRIAGE * GENERAL HAPPINESS Crosstabulation

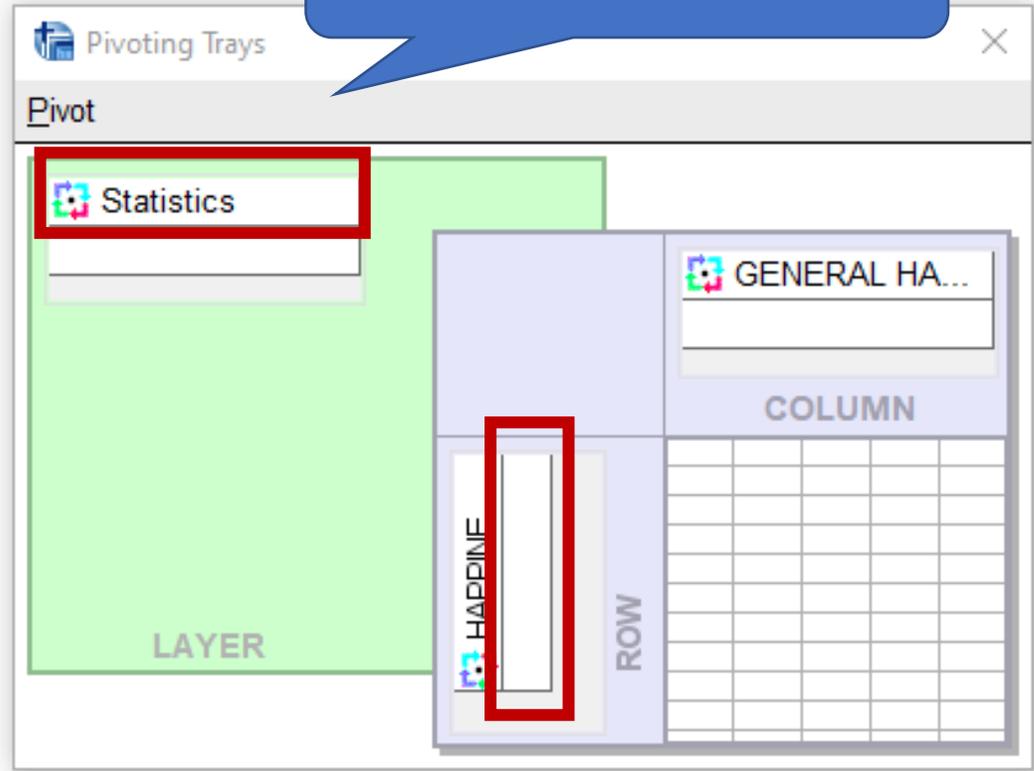
		GENERAL HAPPINESS			Total	
		VERY HAPPY	PRETTY HAPPY	NOT TOO HAPPY		
HAPPINESS OF MARRIAGE	VERY HAPPY	Count	356	215	24	595
	% within HAPPINESS OF MARRIAGE		59.8%	36.1%	4.0%	100.0%
	% within GENERAL HAPPINESS		89.4%	43.9%	29.6%	61.4%
PRETTY HAPPY	Count		40	259	44	343
	% within HAPPINESS OF MARRIAGE		11.7%	75.5%	12.8%	100.0%
	% within GENERAL HAPPINESS		10.1%	52.9%	54.3%	35.4%
NOT TOO HAPPY	Count		2	15	13	30
	% within HAPPINESS OF MARRIAGE		6.7%	50.0%	43.3%	100.0%
	% within GENERAL HAPPINESS		0.5%	3.1%	16.0%	3.1%
DK	Count		0	1	0	1
	% within HAPPINESS OF MARRIAGE		0.0%	100.0%	0.0%	100.0%
	% within GENERAL HAPPINESS		0.0%	0.2%	0.0%	0.1%
Total	Count		398	490	81	969
	% within HAPPINESS OF MARRIAGE		41.1%	50.6%	8.4%	100.0%
	% within GENERAL HAPPINESS		100.0%	100.0%	100.0%	100.0%

Double-Click at table

Pivot >> Pivot Trays

Select % within HAPPINESS OF MARRIAGE Statistics

Move 'Statistics' to Layer



HAPPINESS OF MARRIAGE * GENERAL HAPPINESS Crosstabulation

% within HAPPINESS OF MARRIAGE

		GENERAL HAPPINESS			Total
		VERY HAPPY	PRETTY HAPPY	NOT TOO HAPPY	
HAPPINESS OF MARRIAGE	VERY HAPPY	59.8%	36.1%	4.0%	100.0%
	PRETTY HAPPY	11.7%	75.5%	12.8%	100.0%
	NOT TOO HAPPY	6.7%	50.0%	43.3%	100.0%
	DK	0.0%	100.0%	0.0%	100.0%
Total		41.1%	50.6%	8.4%	100.0%

Count

HAPPINESS OF MARRIAGE * GENERAL HAPPINESS Crosstabulation

Count

		GENERAL HAPPINESS			Total
		VERY HAPPY	PRETTY HAPPY	NOT TOO HAPPY	
HAPPINESS OF MARRIAGE	VERY HAPPY	356	215	24	595
	PRETTY HAPPY	40	259	44	343
	NOT TOO HAPPY	2	15	13	30
	DK	0	1	0	1
Total		398	490	81	969

Testing Relationships for Categorical Data

Analyze >> Descriptive >> Crosstabs

Cells >> Chi-square

Crosstabs

Row(s):
happy

Column(s):
marital

Layer 1 of 1
Previous Next

Display layer variables in table layers

Display clustered bar charts

Suppress tables

Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...

Crosstabs: Statistics

Chi-square

Correlations

Nominal

Contingency coefficient

Phi and Cramer's V

Lambda

Uncertainty coefficient

Ordinal

Gamma

Somers' d

Kendall's tau-b

Kendall's tau-c

Nominal by Interval

Eta

Kappa

Risk

McNemar

Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	159.200 ^a	12	.000
Likelihood Ratio	161.952	12	.000
Linear-by-Linear Association	113.242	1	.000
N of Valid Cases	2012		

a. 5 cells (25.0%) have expected count less than 5. The minimum expected count is .07.

GENERAL HAPPINESS * MARITAL STATUS Crosstabulation

			MARITAL STATUS					
			MARRIED	WIDOWED	DIVORCED	SEPARATED	NEVER MARRIED	Total
GENERAL HAPPINESS	VERY HAPPY	Count	398	31	56	11	101	597
		% within GENERAL HAPPINESS	66.7%	5.2%	9.4%	1.8%	16.9%	100.0%
	PRETTY HAPPY	Count	490	95	169	41	304	1099
		% within GENERAL HAPPINESS	44.6%	8.6%	15.4%	3.7%	27.7%	100.0%
	NOT TOO HAPPY	Count	81	37	55	18	123	314
		% within GENERAL HAPPINESS	25.8%	11.8%	17.5%	5.7%	39.2%	100.0%
	DK	Count	0	1	0	0	1	2
		% within GENERAL HAPPINESS	0.0%	50.0%	0.0%	0.0%	50.0%	100.0%
Total		Count	969	164	280	70	529	2012
		% within GENERAL HAPPINESS	48.2%	8.2%	13.9%	3.5%	26.3%	100.0%

7. Analysis of Variance

Univariate analysis of variance to test for mean difference

Business context

When we examine mean difference between three or more group, we would like to know whether relationship we observe is likely to exist on our target population or instead is caused by random sampling variation.

- Statistical testing tells us whether the mean of an outcome variable is different or statistically the same in several categories of interest, e.g., customer type. Without that, we might make decision based on observed mean difference that are not likely to exist in population of customers,

Analyze >> Compare Mean >> One-Way ANOVA

Data : "SPSS_CUST.SAV"

One-Way ANOVA

Dependent List:

- Am a satisfied SPSS customer [satcust]

Factor:

- How long have you used SPSS products? [usespss]

Estimate effect size for overall tests

Descriptives

Am a satisfied SPSS customer

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LT 1 yr	115	2.83	.920	.086	2.66	3.00	1	5
1 yr-3 yrs	235	2.64	.911	.059	2.52	2.76	1	4
GT 3 yrs-5 yrs	95	2.47	.885	.091	2.29	2.65	1	4
GT 5 yrs-10 yrs	175	2.37	1.047	.079	2.22	2.53	1	5
GT 10 yrs	295	2.53	1.016	.059	2.41	2.64	1	5
Total	915	2.56	.979	.032	2.49	2.62	1	5

ANOVA

Am a satisfied SPSS customer

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16.860	4	4.215	4.466	.001
Within Groups	858.878	910	.944		
Total	875.738	914			

Dependent

Am a satisfied SPSS customer [satcus]

Factor

How long have you used SPSS products? [usespss]

Post Hoc Multiple Comparison

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

LSD S-N-K Waller-Duncan
 Bonferroni Tukey Type I/Type II Error Ratio: 100
 Sidak Tukey's-b Dunnett
 Scheffe Duncan Control Category: Last
 R-E-G-W F Hochberg's GT2 Test
 R-E-G-W Q Gabriel 2-sided < Control > Control

Equal Variances Not Assumed

Tamhane's T2 Dunnett's T3 Games-Howell Dunnett's C

Null Hypothesis test

Use the same significance level [alpha] as the setting in Options
 Specify the significance level [alpha] for the post hoc test
Level: 0.05

Equality of Error Variances

Option >> Homogeneity of variance test

Levene's Test of Equality of Error Variances^{a,b}

		Levene Statistic	df1	df2	Sig.
Am a satisfied SPSS customer	Based on Mean	2.715	4	910	.029
	Based on Median	2.399	4	910	.049
	Based on Median and with adjusted df	2.399	4	894.925	.049
	Based on trimmed mean	2.515	4	910	.040

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Am a satisfied SPSS customer

b. Design: Intercept + usespss

Does the satisfaction vary according to the time of use?

Multiple Comparisons

Dependent Variable: Am a satisfied SPSS customer

Games-Howell

(I) How long have you used SPSS products?	(J) How long have you used SPSS products?	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
LT 1 yr	1 yr-3 yrs	.19	.104	.377	-.10	.47
	GT 3 yrs-5 yrs	.35*	.125	.042	.01	.70
	GT 5 yrs-10 yrs	.45*	.117	.001	.13	.78
	GT 10 yrs	.30*	.104	.035	.01	.59
1 yr-3 yrs	LT 1 yr	-.19	.104	.377	-.47	.10
	GT 3 yrs-5 yrs	.16	.109	.553	-.13	.46
	GT 5 yrs-10 yrs	.27	.099	.057	.00	.54
	GT 10 yrs	.11	.084	.663	-.12	.34
GT 3 yrs-5 yrs	LT 1 yr	-.35*	.125	.042	-.70	-.01
	1 yr-3 yrs	-.16	.109	.553	-.46	.13
	GT 5 yrs-10 yrs	.10	.120	.915	-.23	.43
	GT 10 yrs	-.05	.108	.989	-.35	.25
GT 5 yrs-10 yrs	LT 1 yr	-.45*	.117	.001	-.78	-.13
	1 yr-3 yrs	-.27	.099	.057	-.54	.00
	GT 3 yrs-5 yrs	-.10	.120	.915	-.43	.23
	GT 10 yrs	-.15	.099	.525	-.42	.12
GT 10 yrs	LT 1 yr	-.30*	.104	.035	-.59	-.01
	1 yr-3 yrs	-.11	.084	.663	-.34	.12
	GT 3 yrs-5 yrs	.05	.108	.989	-.25	.35
	GT 5 yrs-10 yrs	.15	.099	.525	-.12	.42

Based on observed means.

The error term is Mean Square(Error) = .944.

*. The mean difference is significant at the .05 level.

8. Associations Between Variables

Business context

Testing for associations between variables is quite common with survey data:

- We might want to learn how a customer's age is related to the number of purchases they have made, or the total revenue from those purchases.
- In a survey of patients, we might want to learn whether satisfaction is higher physicians is correlated with overall satisfaction.

Using Scatterplots to Examine Relationships

Data : "Boston House.SAV"

Graph >> Chart Builder

Select Variables

X axis : median value of owner-occupied homes [medv]

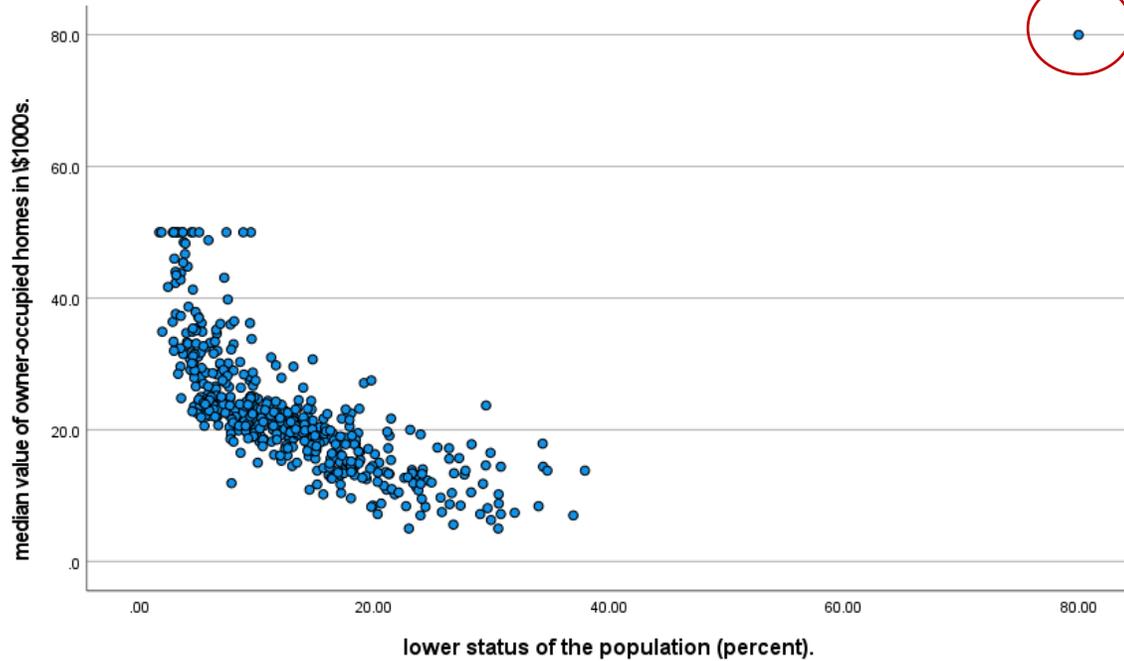
Y axis : %lower status of the population [lstat]

Select Chart Type

The screenshot displays the Minitab Chart Builder interface. On the left, a list of variables includes 'crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis', and 'rad'. The 'Y-axis' is set to 'weighted mean of distance...' and the 'X-axis' is set to 'median value of owner-occupied homes in \$1000s'. The chart preview shows a scatter plot with blue circular markers. Below the chart, there are options to 'Set color?', 'Set size?', and 'Filter?'. At the bottom, a 'Choose from:' menu lists various chart types: Favorites, Bar, Line, Area, Pie/Polar, Scatter/Dot (which is highlighted), Histogram, High-Low, Boxplot, and Dual Axes. The 'Gallery' tab is selected, and the 'Basic Elements' sub-tab is active.

Remove Outlier and create chart again

Extreme Value



Select Cases

per capita crime rate by town. [crim]
proportion of residential land zoned for lots over 2...
proportion of non-retail business acres per town. [l...
Charles River dummy variable (= 1 if tract bounds ...
nitrogen oxides concentration (parts per 10 million...
average number of rooms per dwelling. [rm]
proportion of owner-occupied units built prior to 19...
weighted mean of distances to five Boston emplo...
index of accessibility to radial highways. [rad]
full-value property-tax rate per \$10,000. [tax]
pupil-teacher ratio by town. [ptratio]
1000(Bk - 0.63)^2 where Bk is the proportion of bl...
lower status of the population (percent). [lstat]
median value of owner-occupied homes in \$1000...
medv < 60 (FILTER) [filter_\$]

Select

All cases
 If condition is satisfied
if... medv < 60
 Random sample of cases
Sample...
 Based on time or case range
Range...
 Use filter variable:
[]

Output

Filter out unselected cases
 Copy selected cases to a new dataset
Dataset name: []
 Delete unselected cases

Create Condition

Data >> Select Cases

Select Cases: If

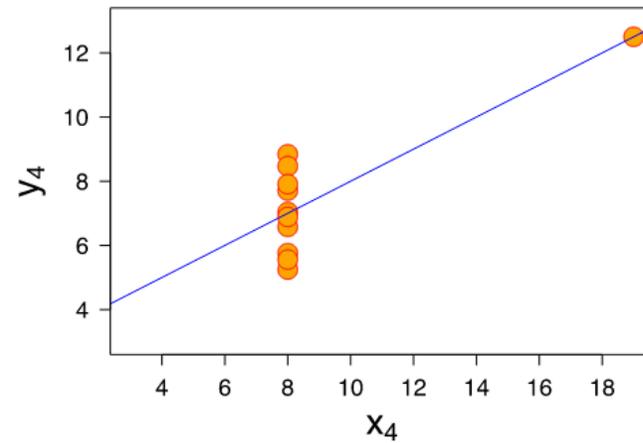
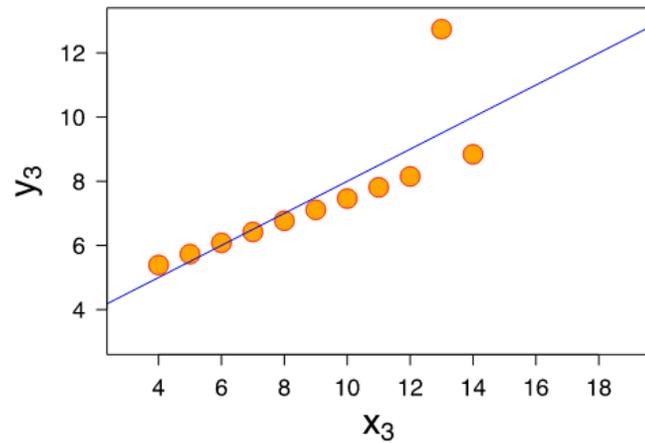
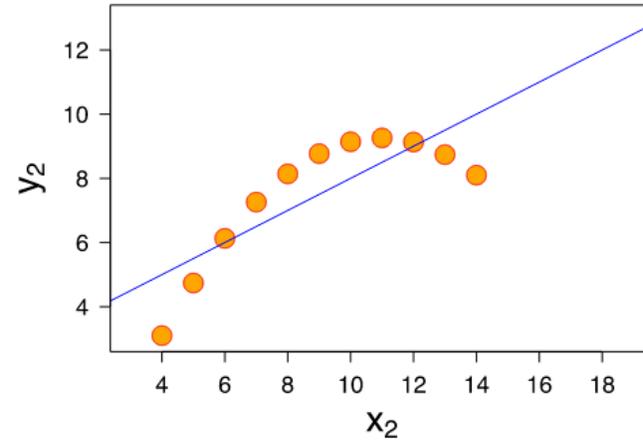
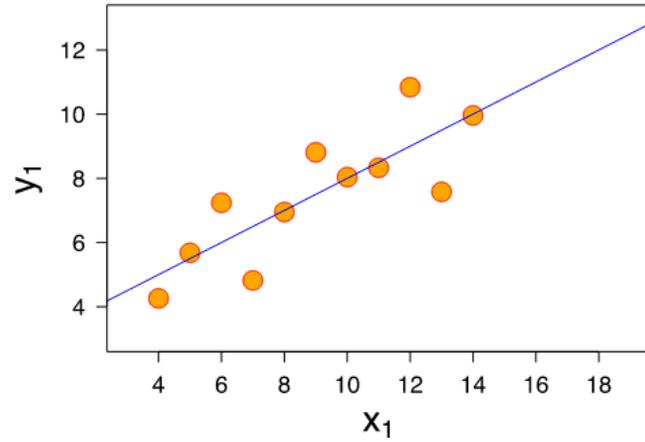
medv < 60

Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation

Functions and Special Variables:

Select if medv < 60

Anscombe's Quartet



The correlation between x and y is **0.816** for each dataset

Correlations Coefficient

Analyze >> Correlate >> Bivariate Correlations

Bivariate Correlations

Variables:

- per capita crime rate by town. [crim]
- proportion of residential land zoned f...
- proportion of non-retail business acr...
- Charles River dummy variable (= 1 if...
- nitrogen oxides concentration (parts...
- average number of rooms per dwelli...
- proportion of owner-occupied units b...
- weighted mean of distances to five ...
- index of accessibility to radial highw...
- full-value property-tax rate per \ \$10,...
- pupil-teacher ratio by town. [ptratio]
- 1000(Bk - 0.63)^2 where Bk is the p...
- medv < 60 (FILTER) [filter \$]

Options...
Style...
Bootstrap...

Correlation Coefficients

Pearson Kendall's tau-b Spearman

Test of Significance

Two-tailed One-tailed

Flag significant correlations Show only the lower triangle Show diagonal

Correlations

		lower status of the population (percent).	median value of owner-occupied homes in \ \$1000s.
lower status of the population (percent).	Pearson Correlation	1	-.738**
	Sig. (2-tailed)		.000
	N	506	506
median value of owner-occupied homes in \ \$1000s.	Pearson Correlation	-.738**	1
	Sig. (2-tailed)	.000	
	N	506	506

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

9. Regression Analysis

Business context

Multivariate regression is the basic technique used to create models to predict an outcome or dependent variable. It is used in almost all industries.

- Models can be developed to predict customer satisfaction based on rating of various aspects of product/services
- Models can be developed to predict customer revenue based on previous revenue and other customer characteristics.

Non linear correlation

Analyze >> Regression >> Linear

Model Summary

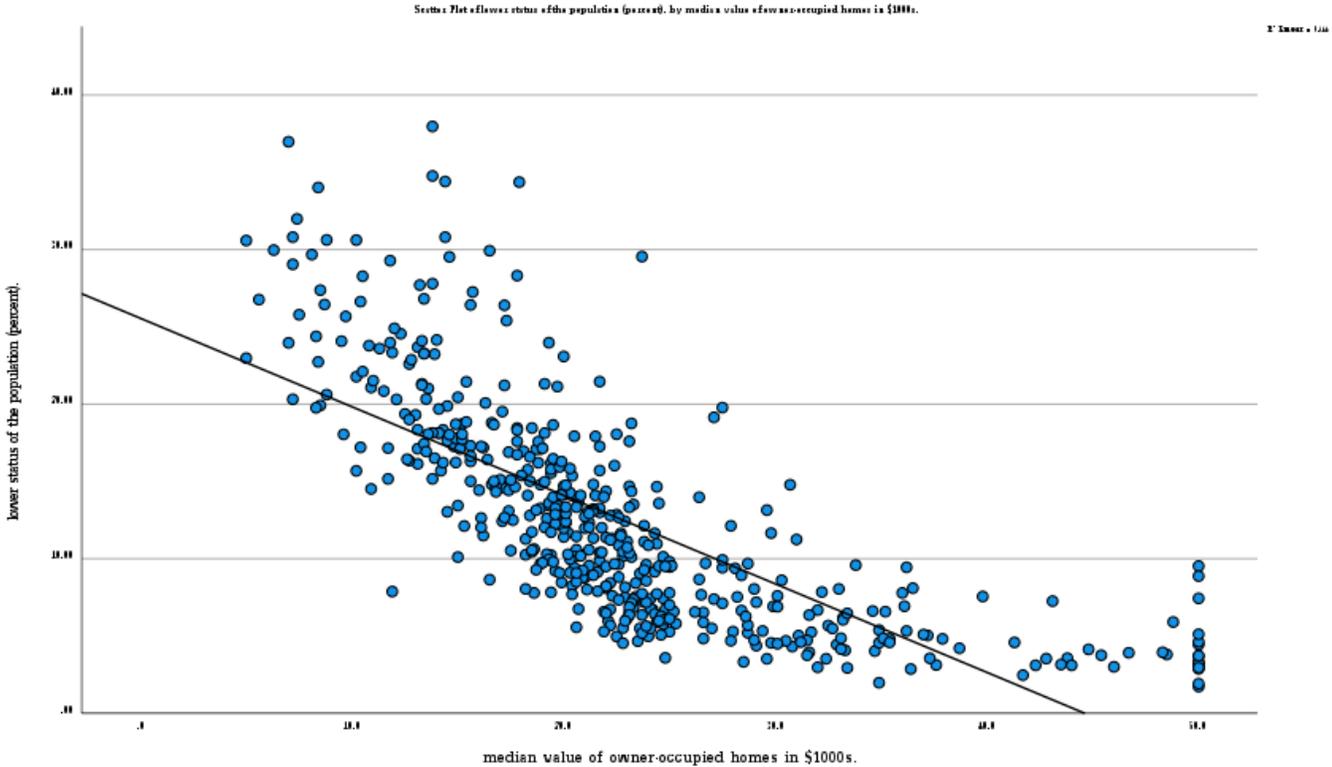
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.738 ^a	.544	.543	6.2158

a. Predictors: (Constant), lower status of the population (percent).

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error			
1	(Constant)	34.554	.563		61.415	.000
	lower status of the population (percent).	-.950	.039	-.738	-24.528	.000

a. Dependent Variable: median value of owner-occupied homes in \$1000s.



Which line is the best fit?

Analyze >> Regression >> Curve Estimate

Curve Estimation

per capita crime rate by town. [crim]
 proportion of residential land zoned for lots ov...
 proportion of non-retail business acres per to...
 Charles River dummy variable (= 1 if tract bou...
 nitrogen oxides concentration (parts per 10 m...
 average number of rooms per dwelling. [rm]
 proportion of owner-occupied units built prior t...
 weighted mean of distances to five Boston e...
 index of accessibility to radial highways. [rad]
 full-value property-tax rate per \$10,000. [tax]
 pupil-teacher ratio by town. [ptratio]
 1000(Bk - 0.63)^2 where Bk is the proportion ...
 medv < 60 (FILTER) [filter_\$]

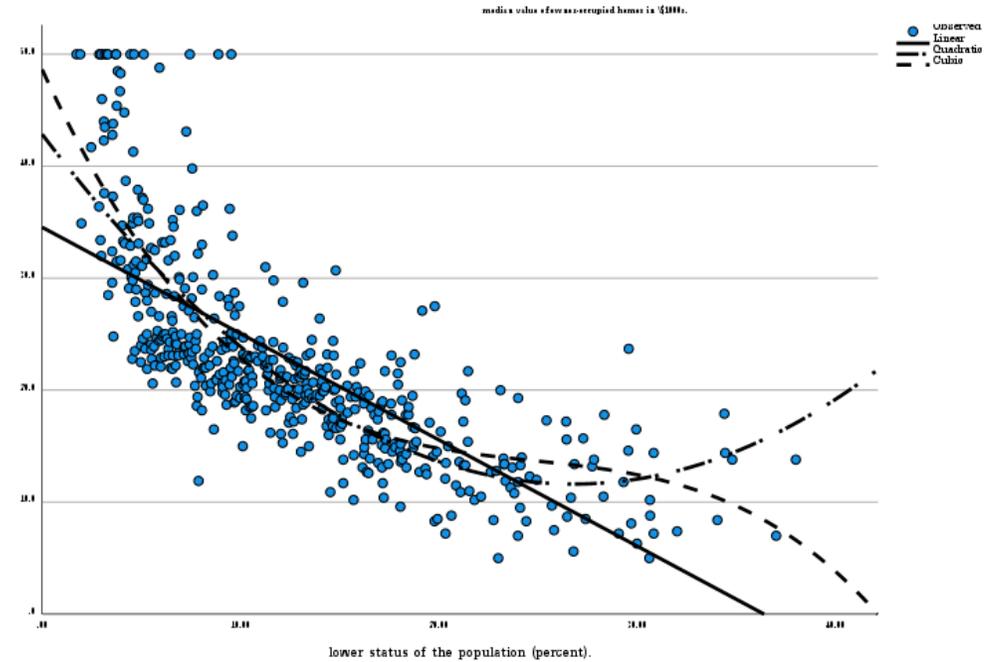
Dependent(s):
 median value of owner-occupied homes in \$1...

Independent Variable:
 lower status of the population (percent). [lstat]

Case Labels:
 Include constant in equation
 Plot models

Models
 Linear Quadratic Compound Growth
 Logarithmic Cubic S Exponential
 Inverse Power: Logistic
 Upper bound:

Display ANOVA table



Model Summary and Parameter Estimates

Dependent Variable: median value of owner-occupied homes in \$1000s.

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.544	601.618	1	504	.000	34.554	-.950		
Quadratic	.641	448.505	2	503	.000	42.862	-2.333	.044	
Cubic	.658	321.728	3	502	.000	48.650	-3.866	.149	-.002

The independent variable is lower status of the population (percent)..

Stepwise Regression

Linear Regression

- per capita crime rate by town. [crim]
- proportion of residential land zoned for lots over 25,0...
- proportion of non-retail business acres per town. [in...
- Charles River dummy variable (= 1 if tract bounds riv...
- nitrogen oxides concentration (parts per 10 million). ...
- average number of rooms per dwelling. [rm]
- proportion of owner-occupied units built prior to 194...
- weighted mean of distances to five Boston employm...
- index of accessibility to radial highways. [rad]
- full-value property-tax rate per \$10,000. [tax]
- pupil-teacher ratio by town. [ptratio]
- $1000(Bk - 0.63)^2$ where Bk is the proportion of blac...
- lower status of the population (percent). [lstat]
- medv < 60 (FILTER) [filter_\$]

Dependent:

median value of owner-occupied homes in \$1000s. [medv]

Block 1 of 1

Previous Next

Independent(s):

- per capita crime rate by town. [crim]
- proportion of residential land zoned for lots over 25,000 sq...
- proportion of non-retail business acres per town. [indus]
- Charles River dummy variable (= 1 if tract bounds river; 0 ...
- nitrogen oxides concentration (parts per 10 million). [nox]
- average number of rooms per dwelling. [rm]
- proportion of owner-occupied units built prior to 1940. [age]
- weighted mean of distances to five Boston employment c...

Method: Stepwise

Selection Variable: Rule...

Case Labels:

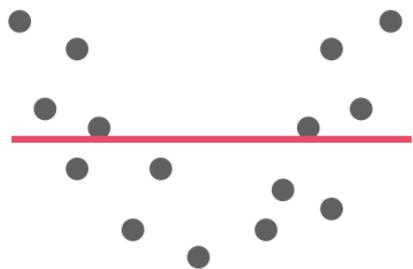
WLS Weight:

Model Summary

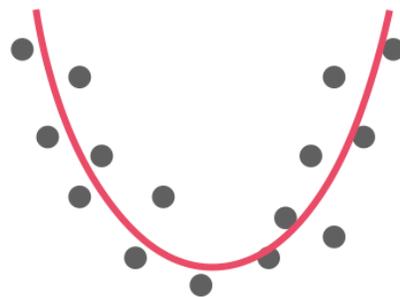
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.738 ^a	.544	.543	6.2158
2	.799 ^b	.639	.637	5.5403
3	.824 ^c	.679	.677	5.2294
4	.831 ^d	.690	.688	5.1386
5	.841 ^e	.708	.705	4.9939
6	.846 ^f	.716	.712	4.9326
7	.850 ^g	.722	.718	4.8818
8	.852 ^h	.727	.722	4.8474
9	.854 ⁱ	.729	.724	4.8326
10	.857 ^j	.734	.729	4.7895
11	.861 ^k	.741	.735	4.7362

Overfit Problem

Regression



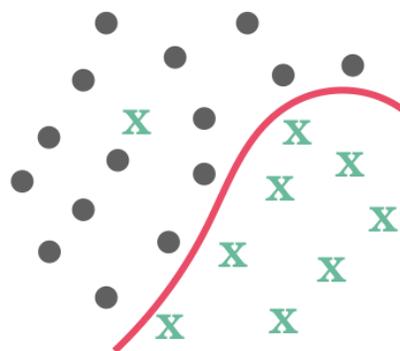
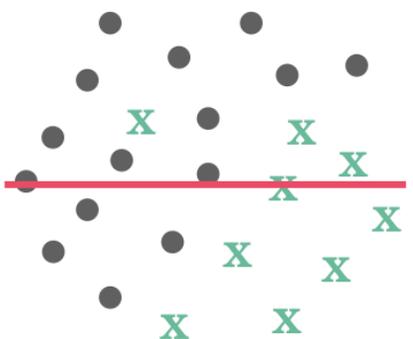
Underfitting



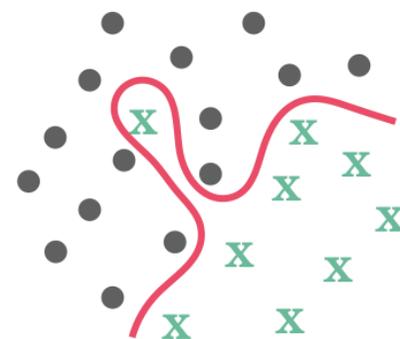
Desired



Overfitting



Classification



Q & A