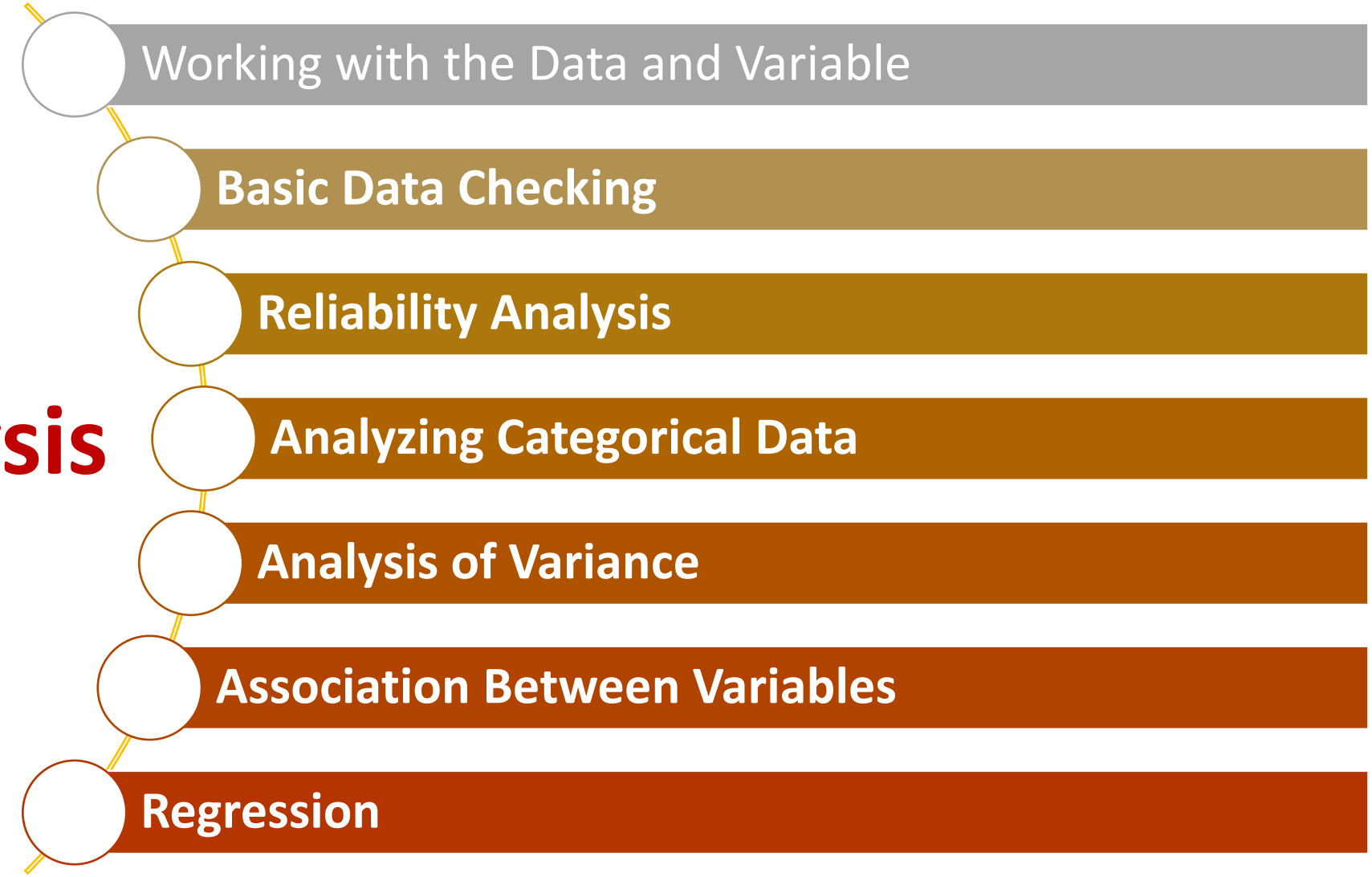
Survey Analysis

IBM SPSS Statistics

SIFT

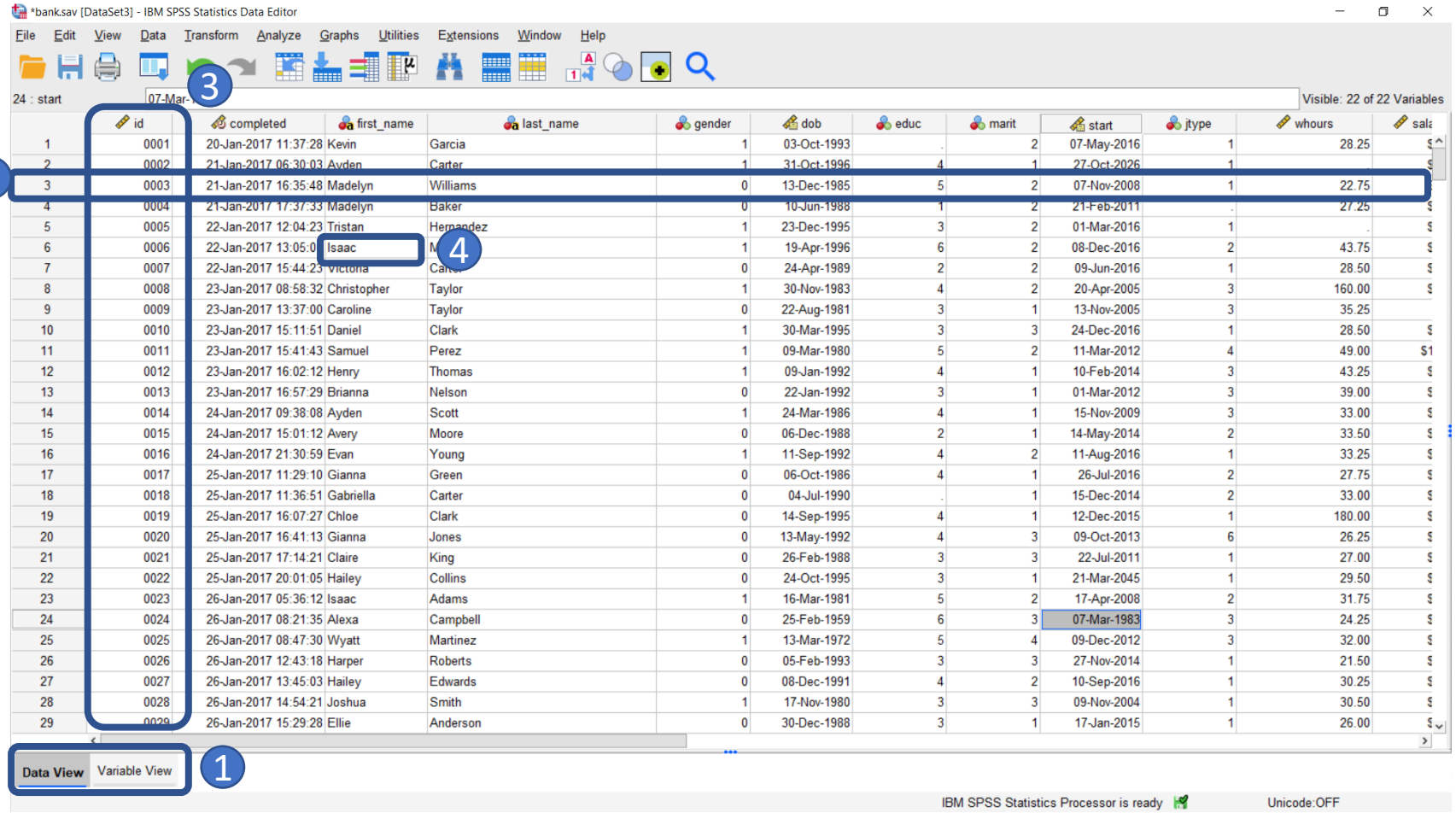
Survey Analysis

IBM SPSS Statistics



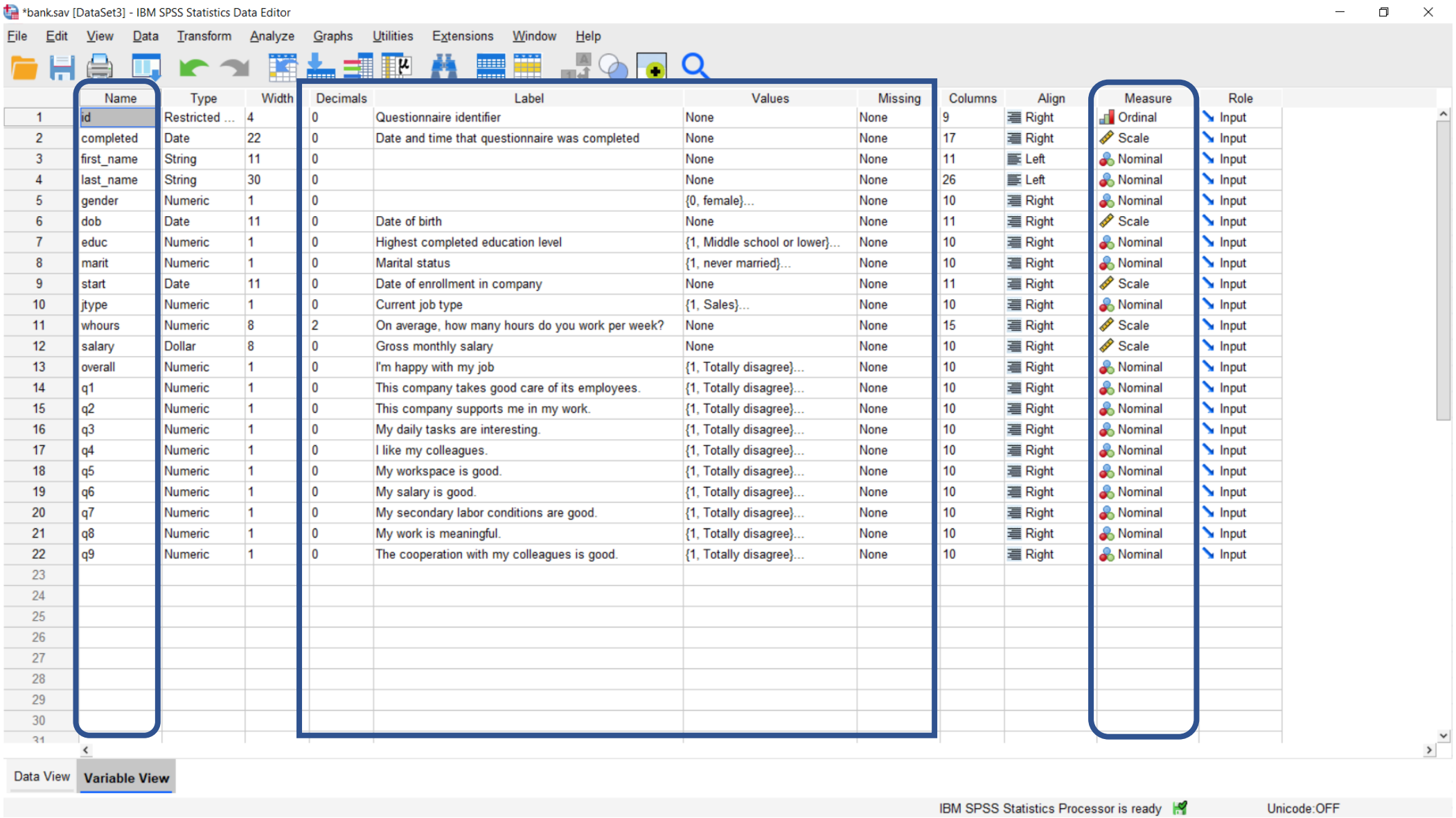
1.Working with the Data and Variable

Data View



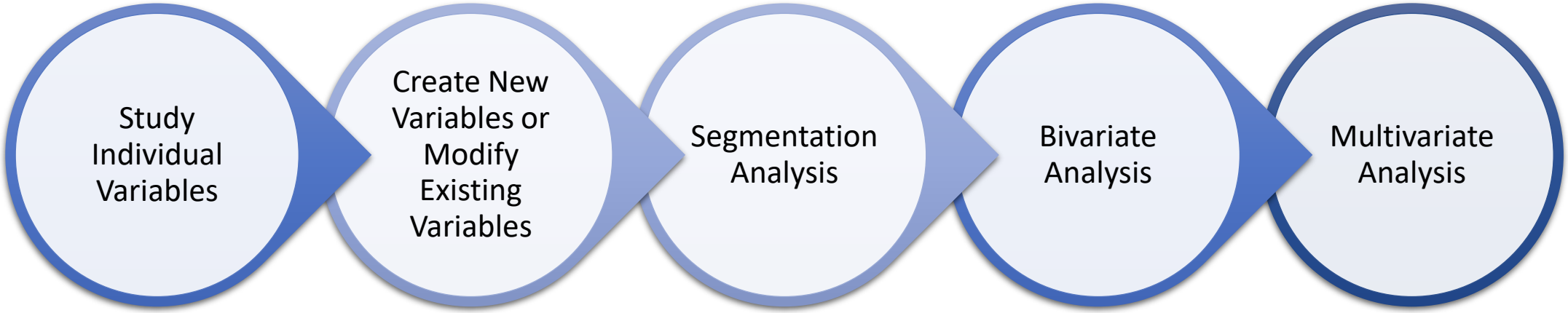
1. **tabs** for switching between Data View and Variable View
2. Columns of cells are called **variables**.
3. Rows of cells are called **observation**
4. **values** refer to cell contents

Variable View



	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											

The Process of Survey Analysis



- Frequencies
- Graphics

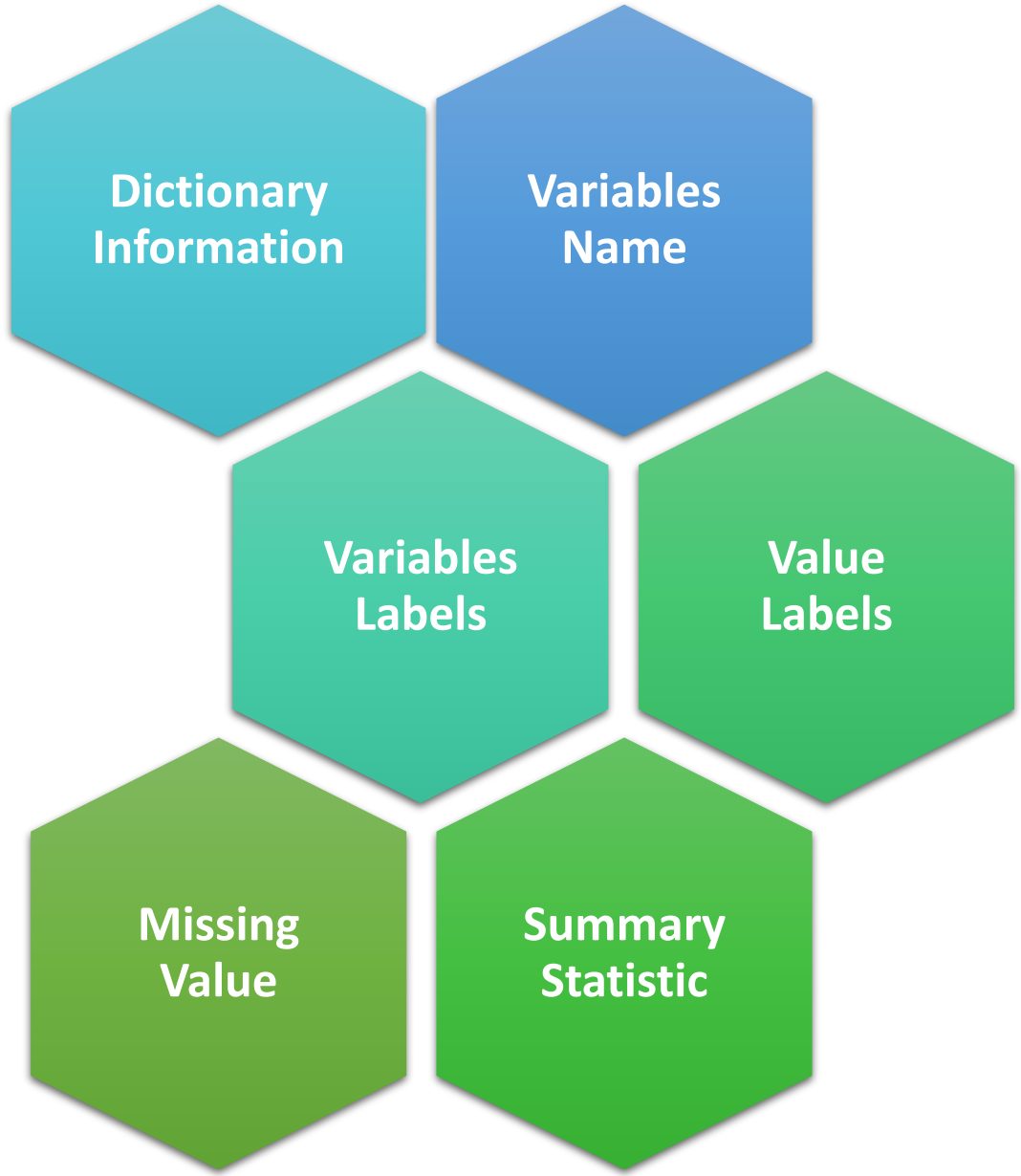
- Recode and Transformation
- Indexes
- Scales : Factor Analysis and Reliability

- Cluster Analysis

- Crosstabs
- Correspondence Analysis
- Correlations
- ANOVA

- Classification Trees
- Discriminant Analysis
- Linear Regression
- Logistic Regression

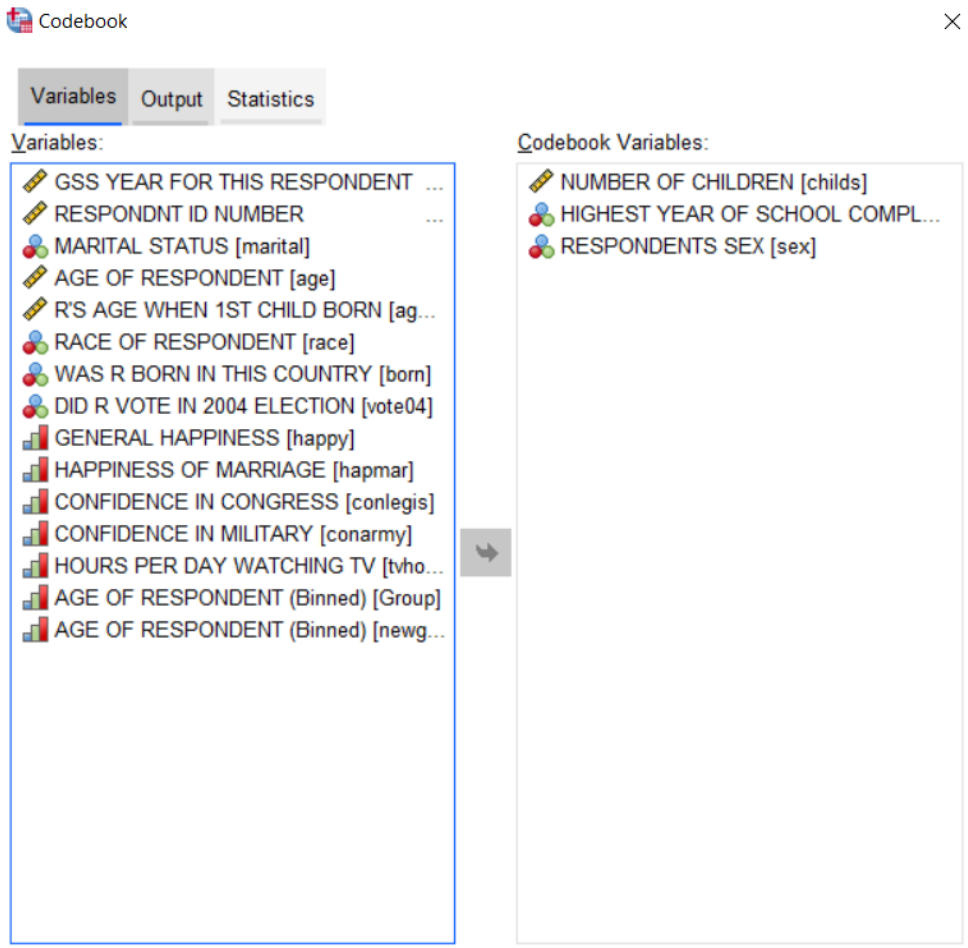
2. Basic Data Checking With Codebook



2.1 Codebook

Data : "GSS2008.SAV"

Analyze >> Reports >> CodeBook



childs

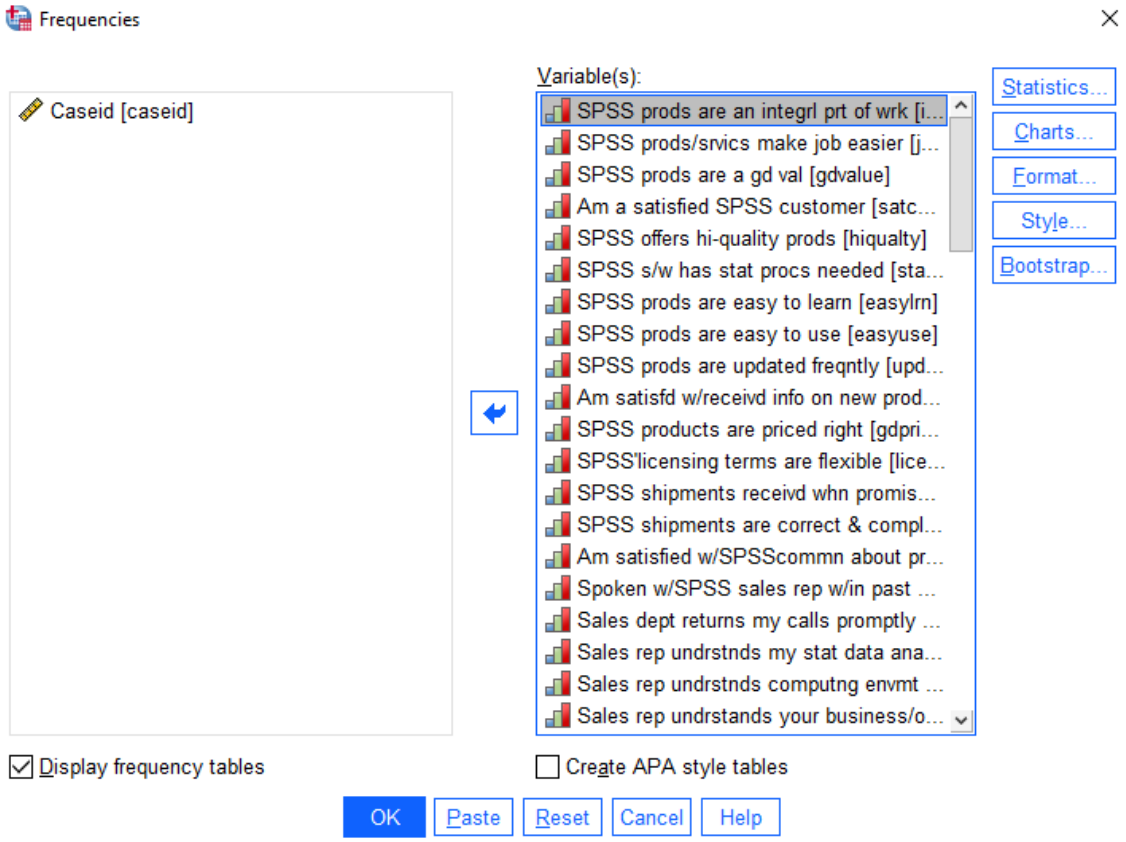
		Value
Standard Attributes	Position	4
	Label	NUMBER OF CHILDREN
	Type	Numeric
	Format	F1
	Measurement	Scale
	Role	Input
N	Valid	2020
	Missing	3
Central Tendency and Dispersion	Mean	1.94
	Standard Deviation	1.698
	Percentile 25	.00
	Percentile 50	2.00
	Percentile 75	3.00

sex

		Value	Count	Percent
Standard Attributes	Position	8		
	Label	RESPONDENTS SEX		
	Type	Numeric		
	Format	F1		
	Measurement	Nominal		
	Role	Input		
Valid Values	1	MALE	929	45.9%
	2	FEMALE	1094	54.1%

2.2 Using Frequencies to Check Data

Analyze >> Descriptive >> Frequencies



Frequencies

Statistics

How frequently do you use SPSS

N	Valid	825
	Missing	130

How frequently do you use SPSS

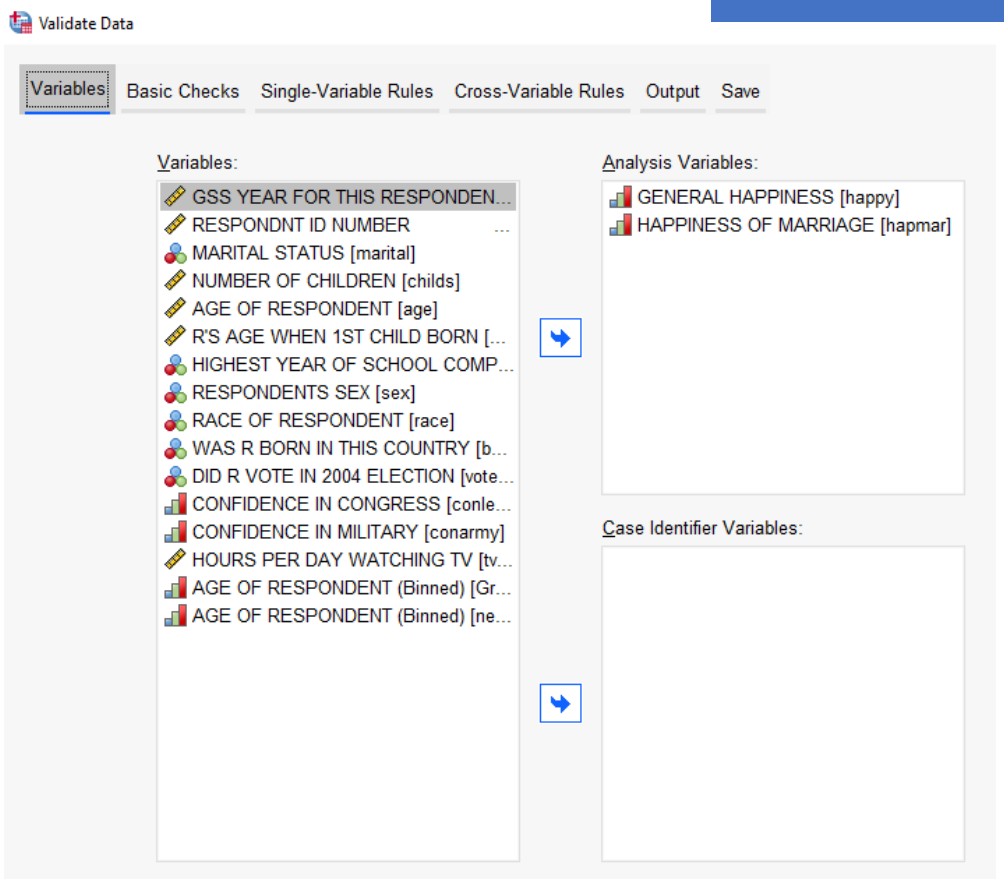
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Don't use very much at all	85	8.9	10.3	10.3
	Only for special projects	95	9.9	11.5	21.8
	Monthly	160	16.8	19.4	41.2
	Weekly	120	12.6	14.5	55.8
	Daily/multiple times in a week	250	26.2	30.3	86.1
	Multiple times in a day	115	12.0	13.9	100.0
	Total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

2.3 Data Validation

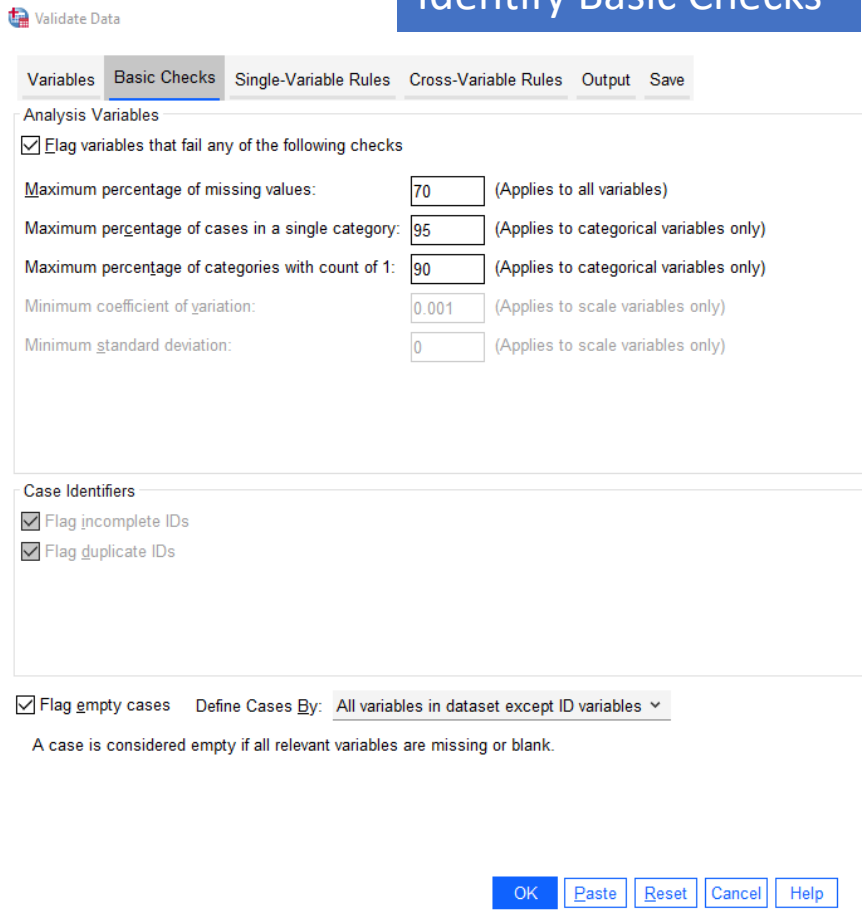
Data : “SPSS_CUST.SAV”

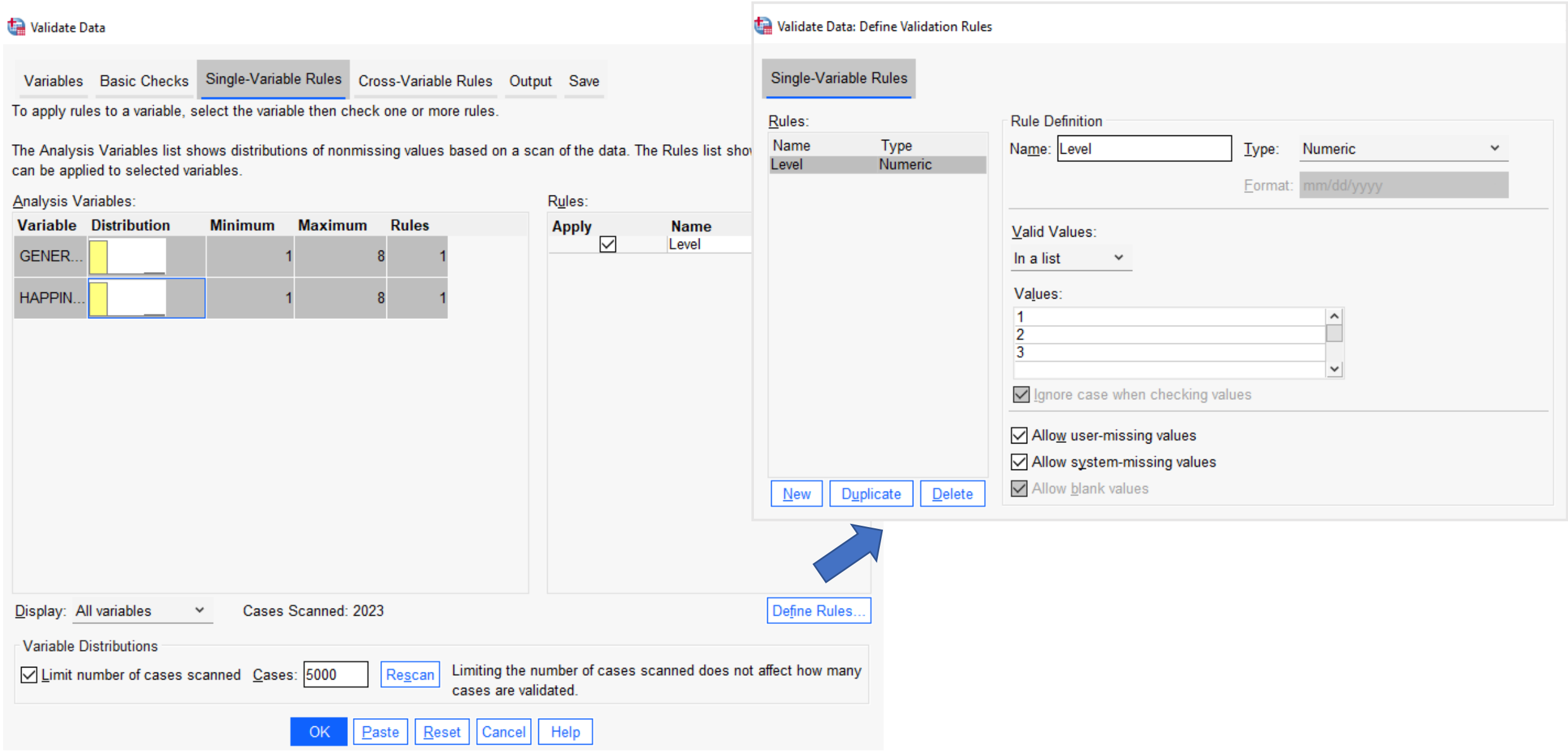
Data >> Validation >> Validate Data

Select Variables



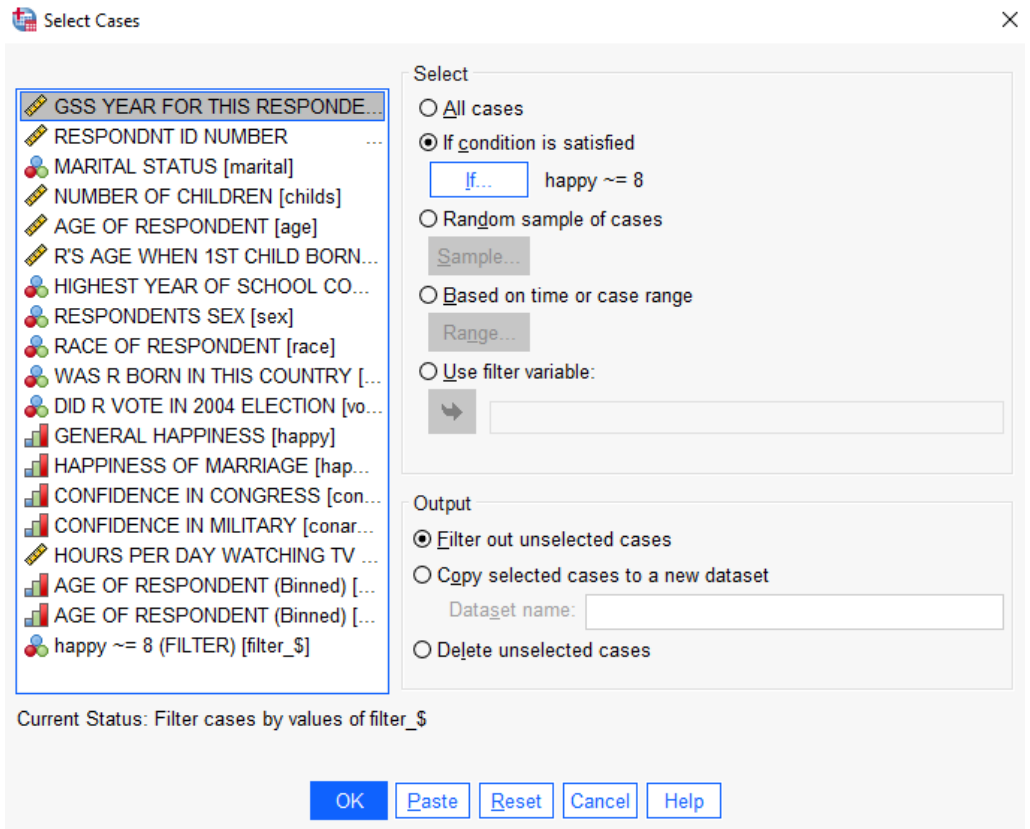
Identify Basic Checks

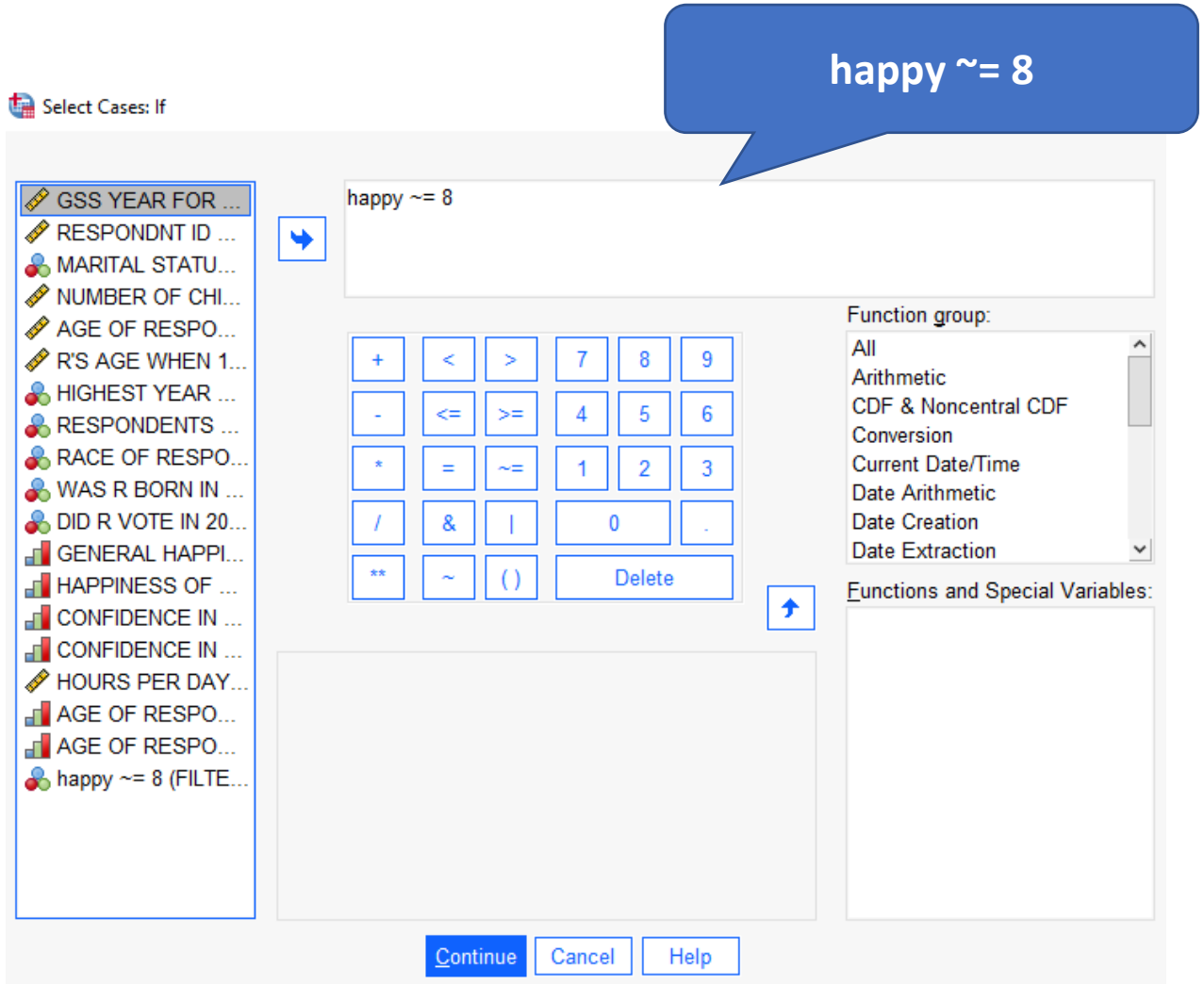



Validate Data
Variables
Basic Checks
Single-Variable Rules
Cross-Variable Rules
Output
Save
To apply rules to a variable, select the variable then check one or more rules.
The Analysis Variables list shows distributions of nonmissing values based on a scan of the data. The Rules list sho
can be applied to selected variables.
Analysis Variables:
Variable
Distribution
Minimum
Maximum
Rules
GENER...
1
8
1
HAPPIN...
1
8
1
Rules:
Apply
Name
Level
Display:
All variables
Cases Scanned: 2023
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
Validate Data: Define Validation Rules
Single-Variable Rules
Rules:
Name
Type
Level
Numeric
New
Duplicate
Delete
Rule Definition
Name:
Level
Type:
Numeric
Format:
mm/dd/yyyy
Valid Values:
In a list
Values:
1
2
3
Ignore case when checking values
Allow user-missing values
Allow system-missing values
Allow blank values

Exclude Incorrect Value

Data >> Select Cases

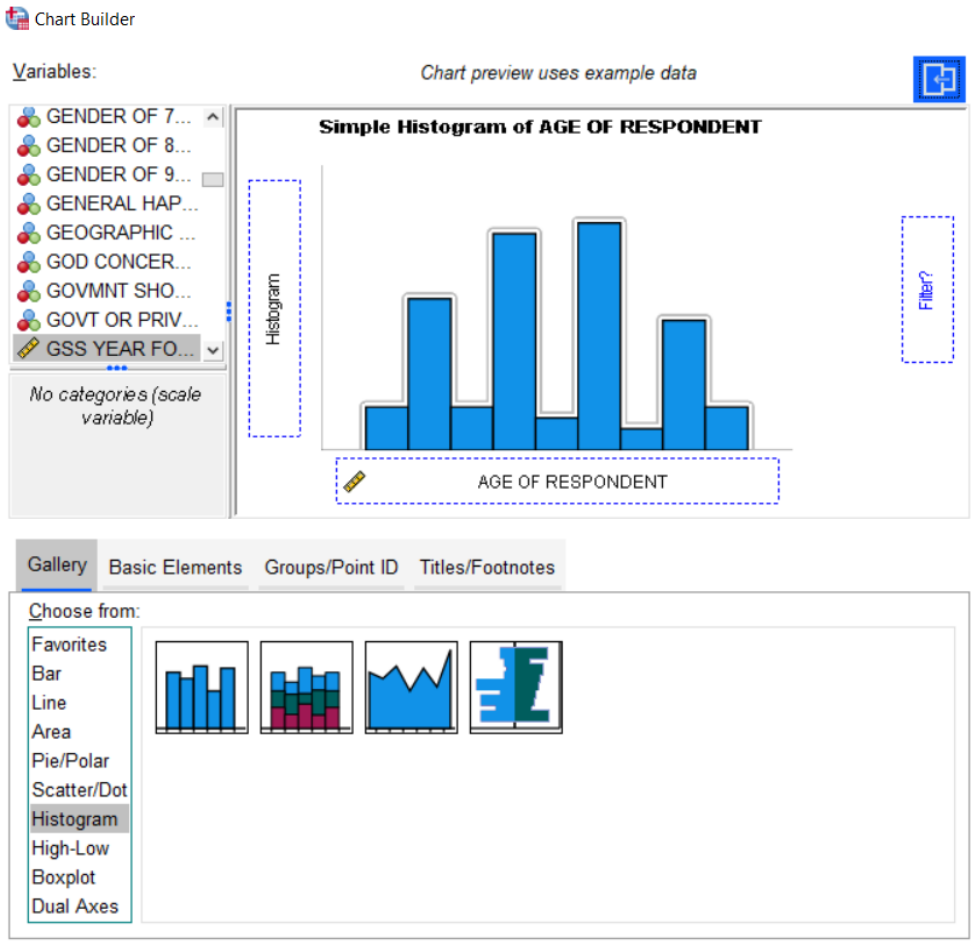




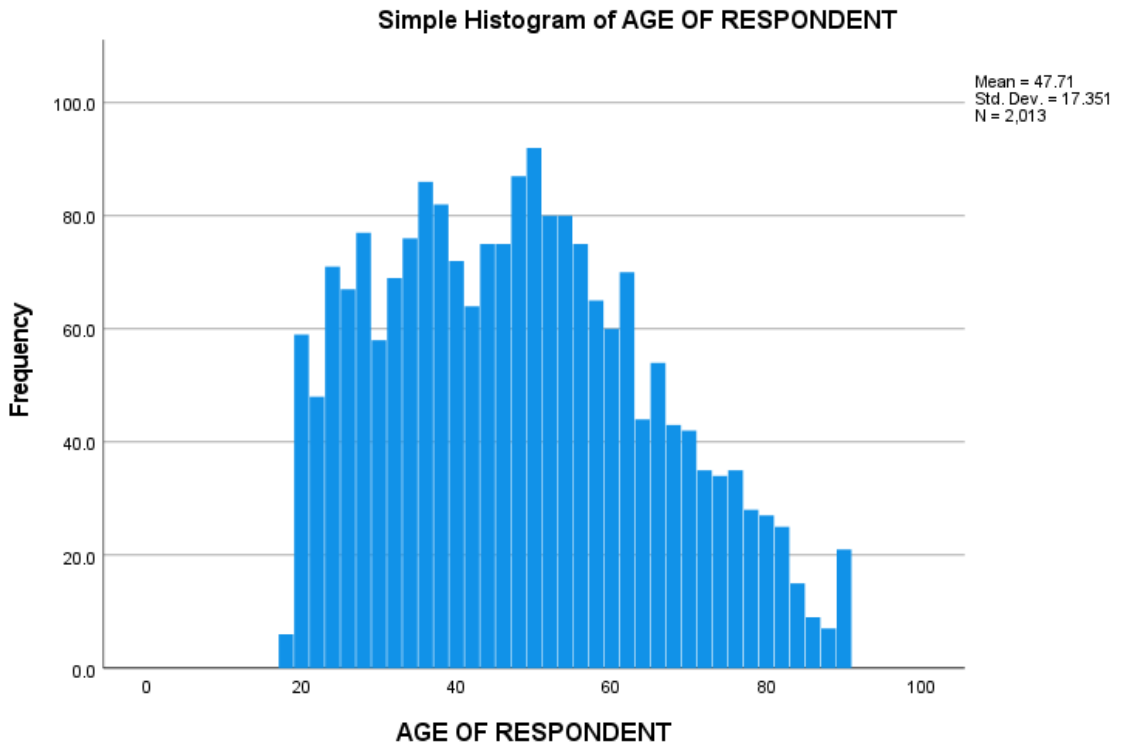
2.4 Variable Binning

Data : “GSS2008.SAV”

Graphs >> Chart Builder

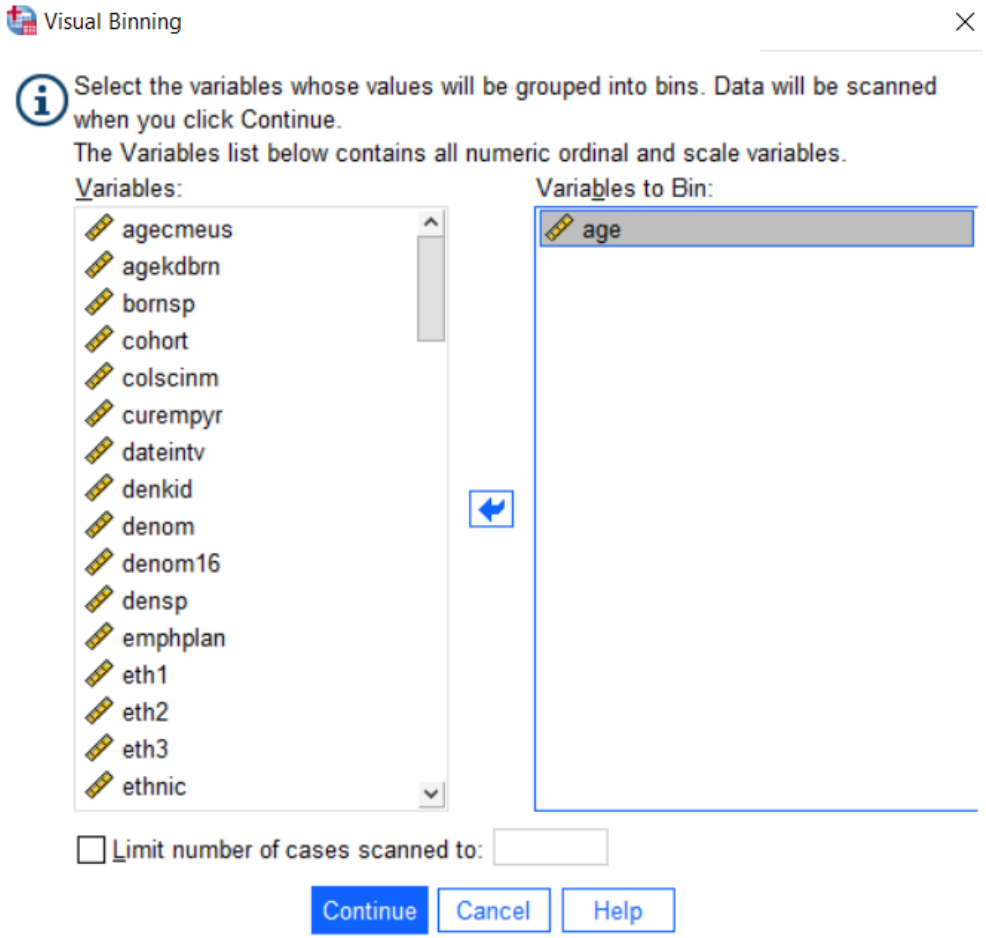


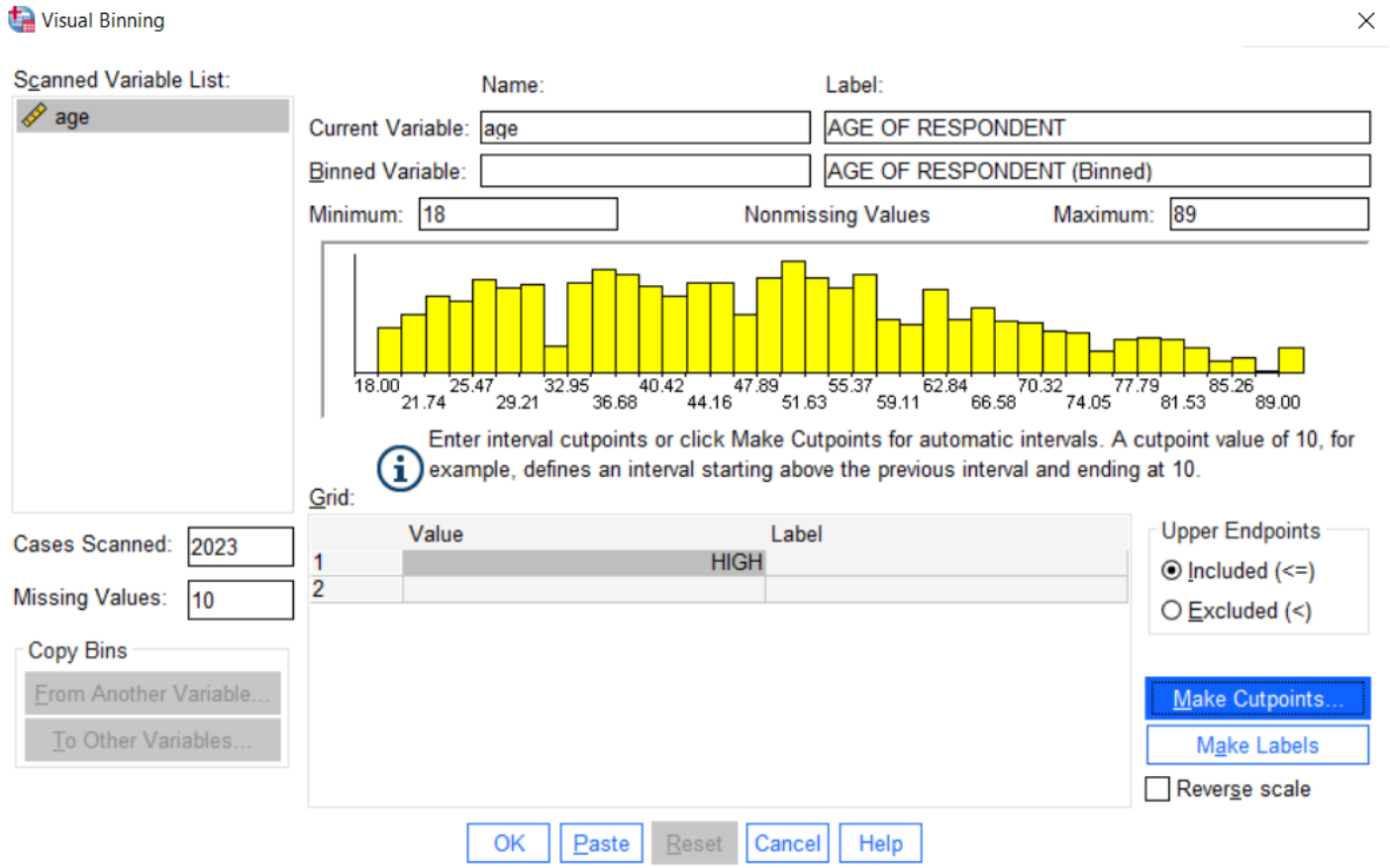
Histogram of age



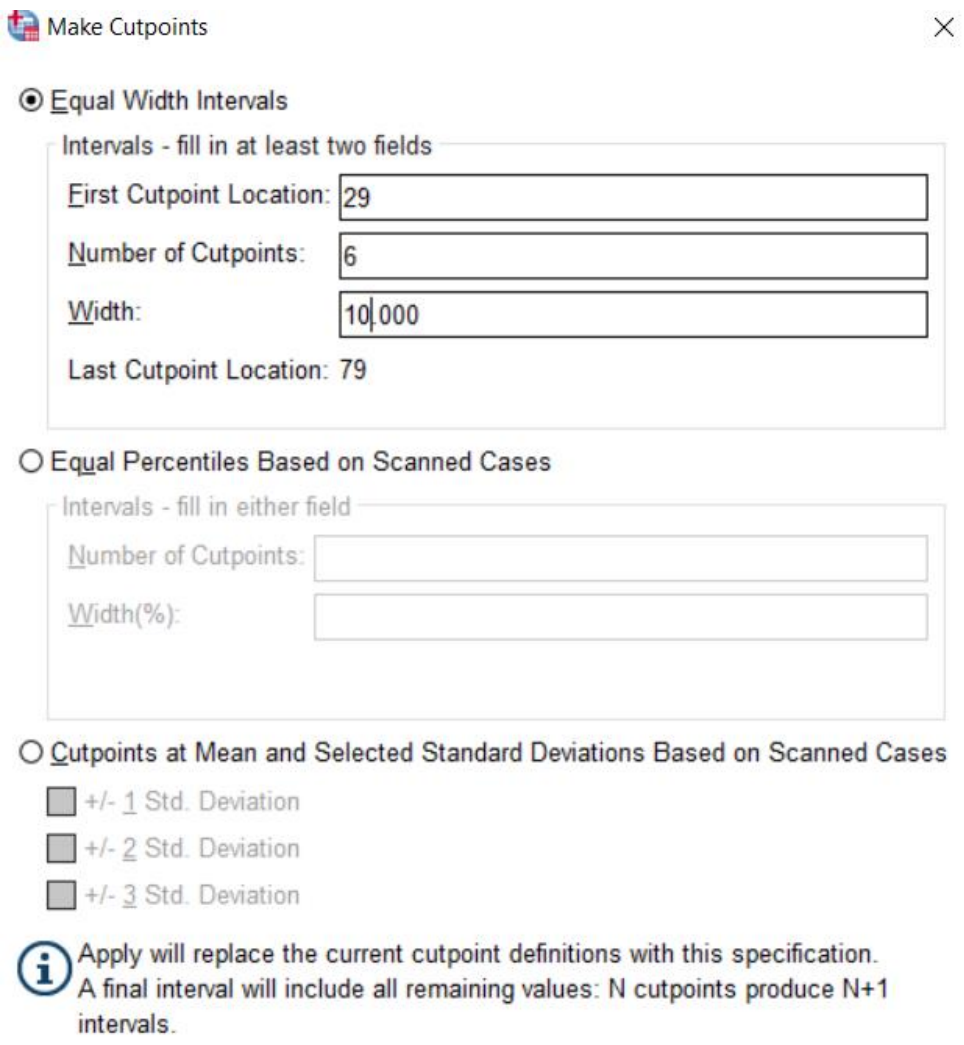
Data : "GSS2008.SAV"

Transform >> Visual Binning





Make Cut points



Equal width Interval

Equal percentile intervals

Mean and standard deviation interval

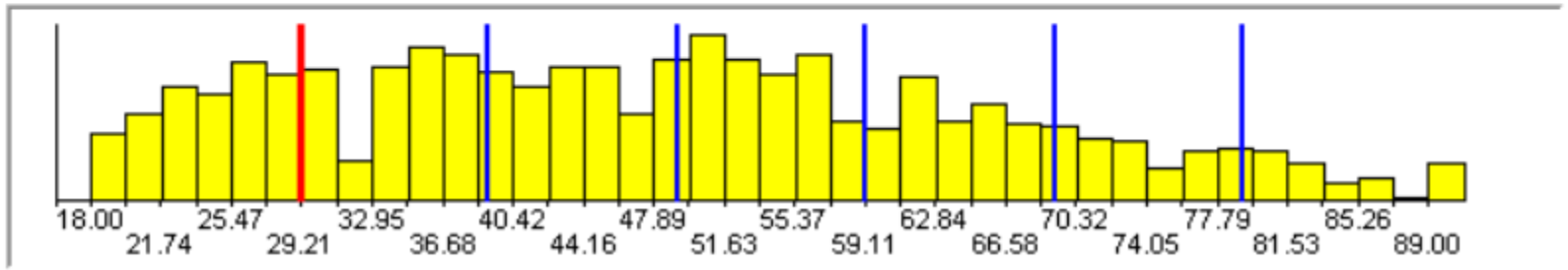
Visual Binning

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



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

Upper Endpoints

- (•) Included (<=)
- ( ) Excluded (<)

Make Cutpoints...

Make Labels

[ ] Reverse scale

Cases Scanned: 2023

Missing Values: 10

Copy Bins

From Another Variable...

To Other Variables...

3. Reliability Analysis

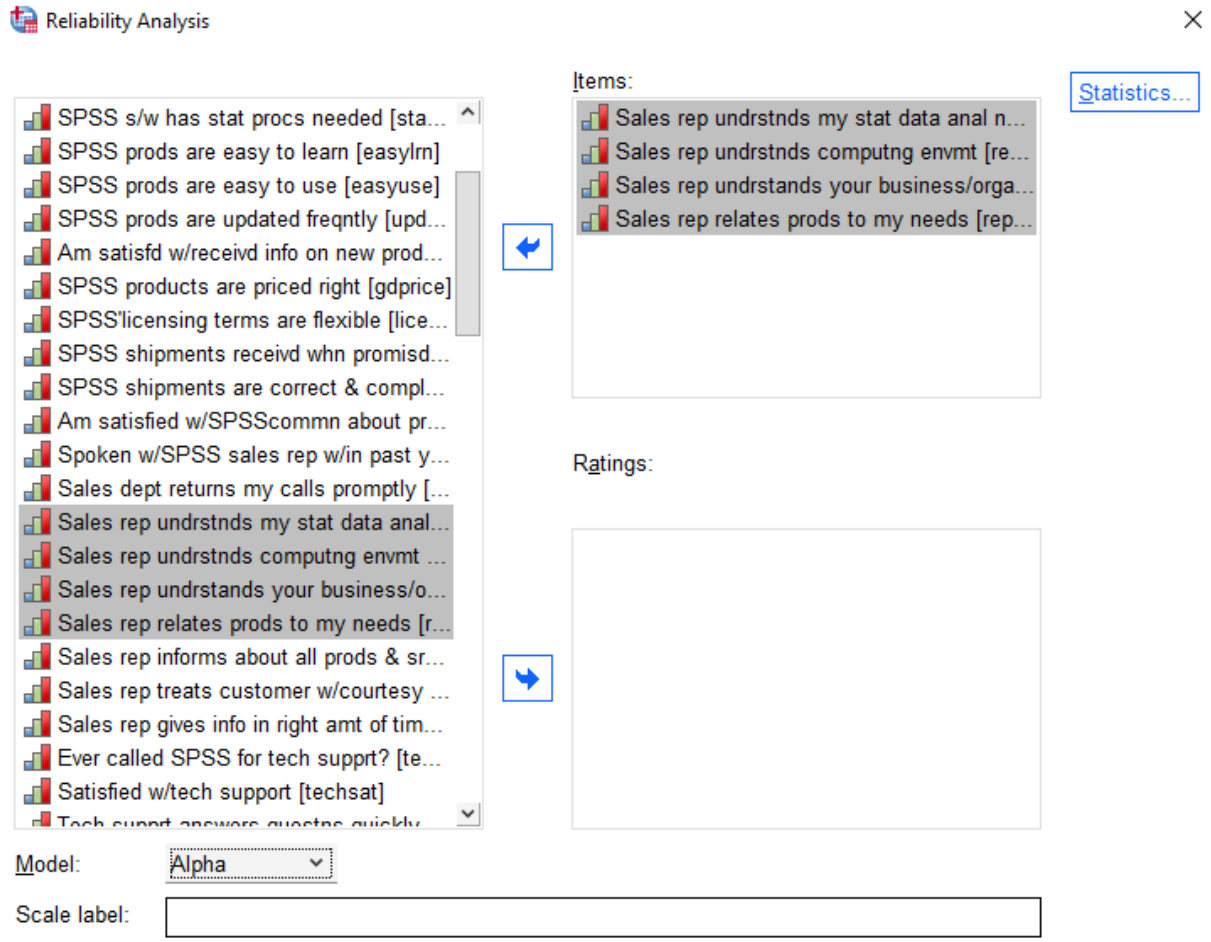
Measure of the consistency of a measurement overtime

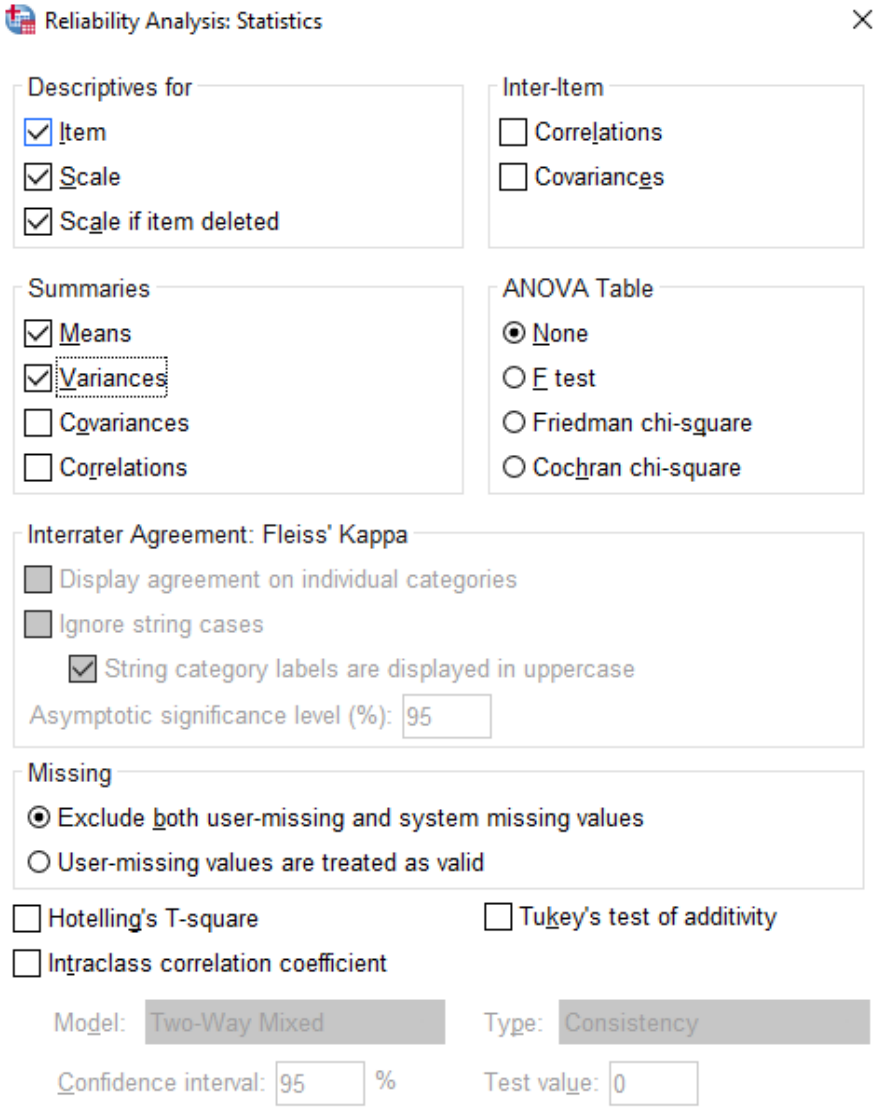
“with questions about the performance of sales reps for SPSS software, if reliability is high, respondents who strongly agree that the sale rep understand their business are also likely to agree that the sales rep understand their computing environment”

Cronbach’s Alpha : varies from 0 to 1

Analyze >> Scale >> Reliability Analysis

Data : "SPSS_CUST.SAV"





Reliability Statistics

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

Item Statistics

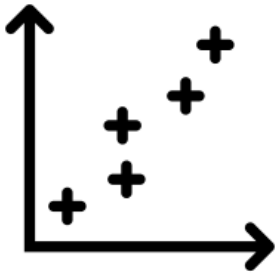
	Mean	Std. Deviation	N
Sales rep undrstnds my stat data anal needs	2.30	.878	500
Sales rep undrstnds computng envmt	2.35	1.082	500
Sales rep undrstands your business/organ	2.55	1.034	500
Sales rep relates prods to my needs	2.63	1.056	500

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 undrstands your business/organ	7.28	6.895	.795	.634	.835
Sales rep relates prods to my needs	7.20	7.054	.734	.599	.860

4. Analyzing Categorical Data

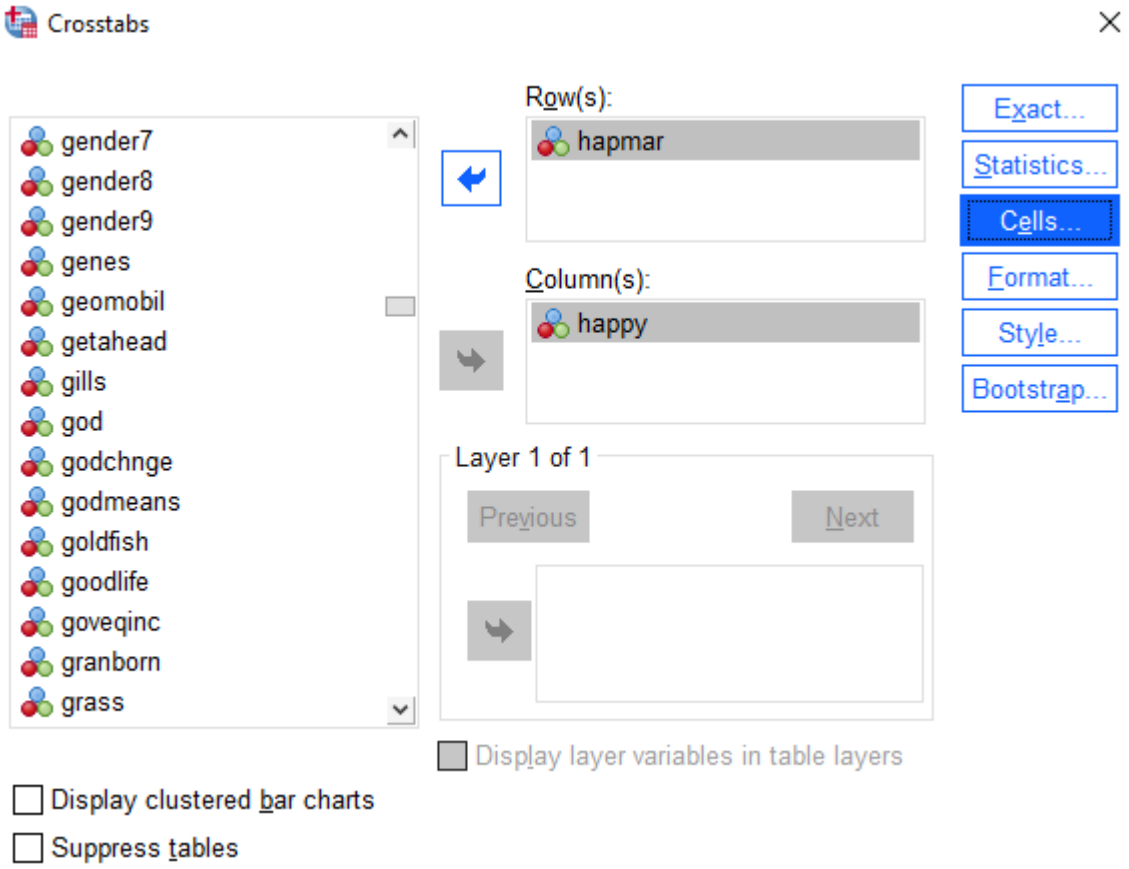
- Crosstabs are commonly used to explore how demographic characteristics are related to attitudes and behaviors



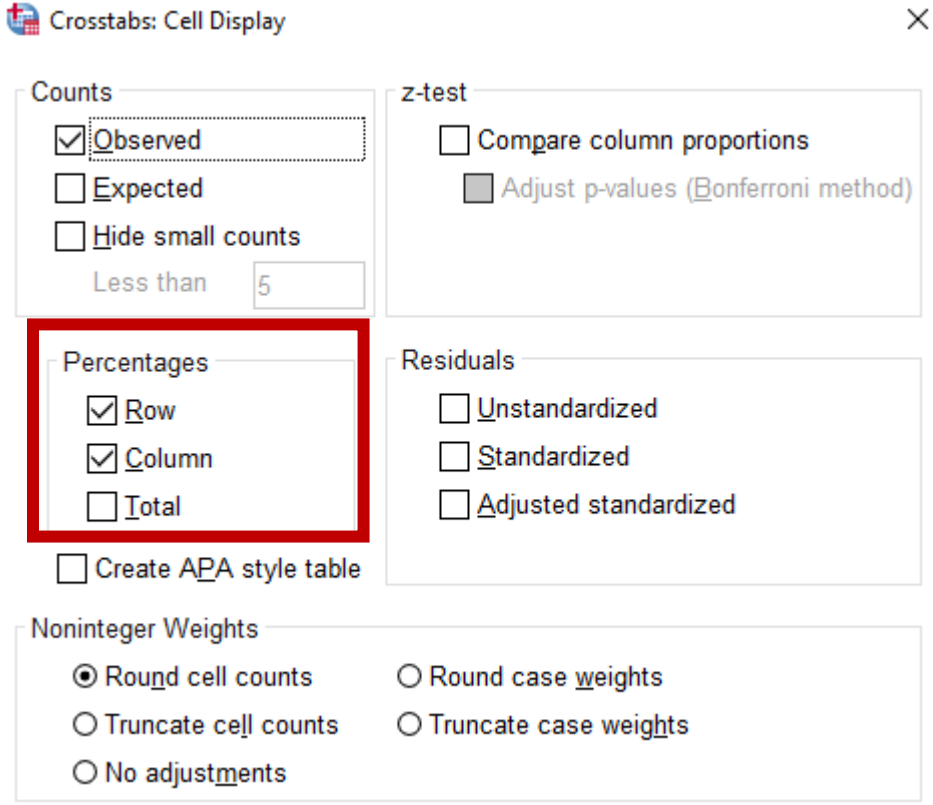
- Use to study the relationships between two, or more categorical variables

Data : “GSS2008.SAV”

Analyze >> Descriptive >> Crosstabs



Cells

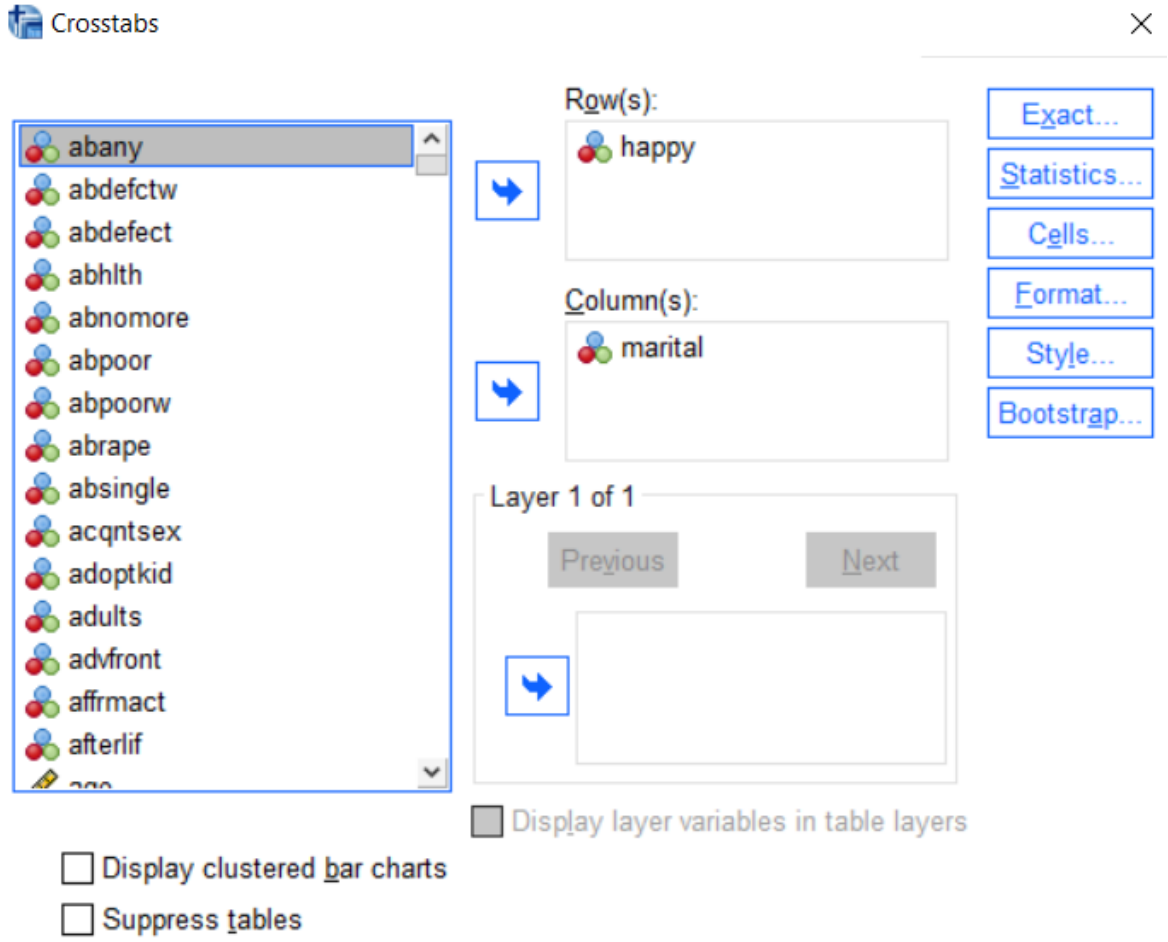


HAPPINESS OF MARRIAGE * GENERAL HAPPINESS Crosstabulation

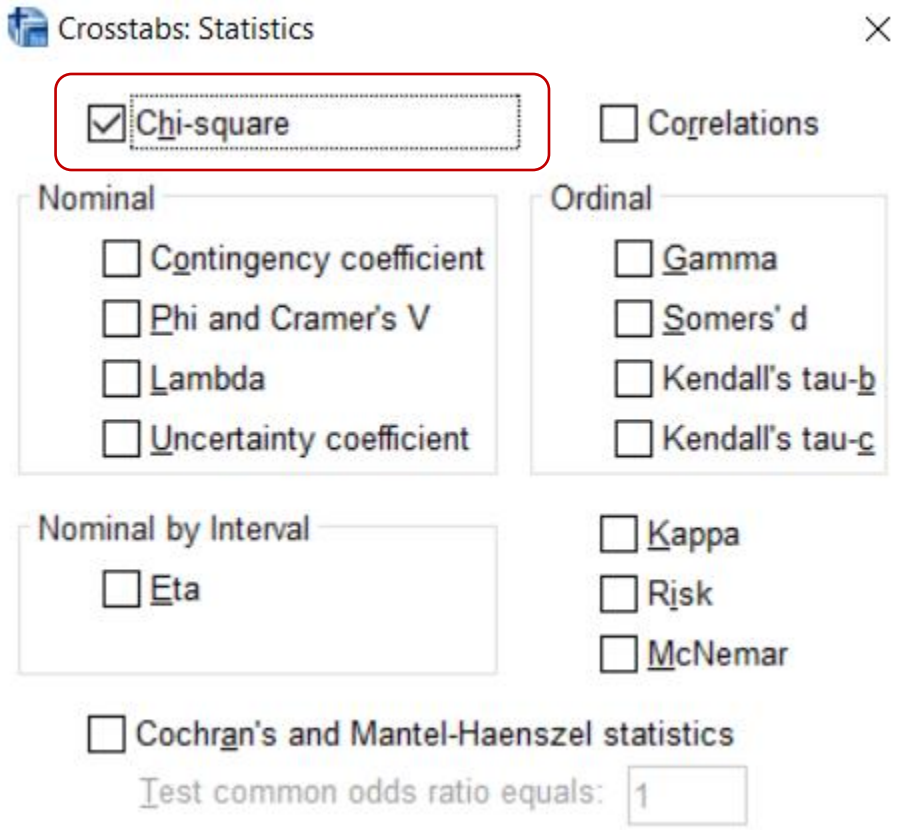
			GENERAL HAPPINESS			
			VERY HAPPY	PRETTY HAPPY	NOT TOO HAPPY	Total
HAPPINESS OF MARRIAGE	VERY HAPPY	Count	356	215	24	595
		% within HAPPINESS OF MARRIAGE	59.8%	36.1%	4.0%	100.0%
		% within GENERAL HAPPINESS	89.4%	43.9%	29.6%	61.4%
	PRETTY HAPPY	Count	40	259	44	343
		% within HAPPINESS OF MARRIAGE	11.7%	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%

Testing Relationships for Categorical Data

Analyze >> Descriptive >> Crosstabs



Statistics >> Chi-square



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%

5. Analysis of Variance

Univariate analysis of variance to test for mean difference

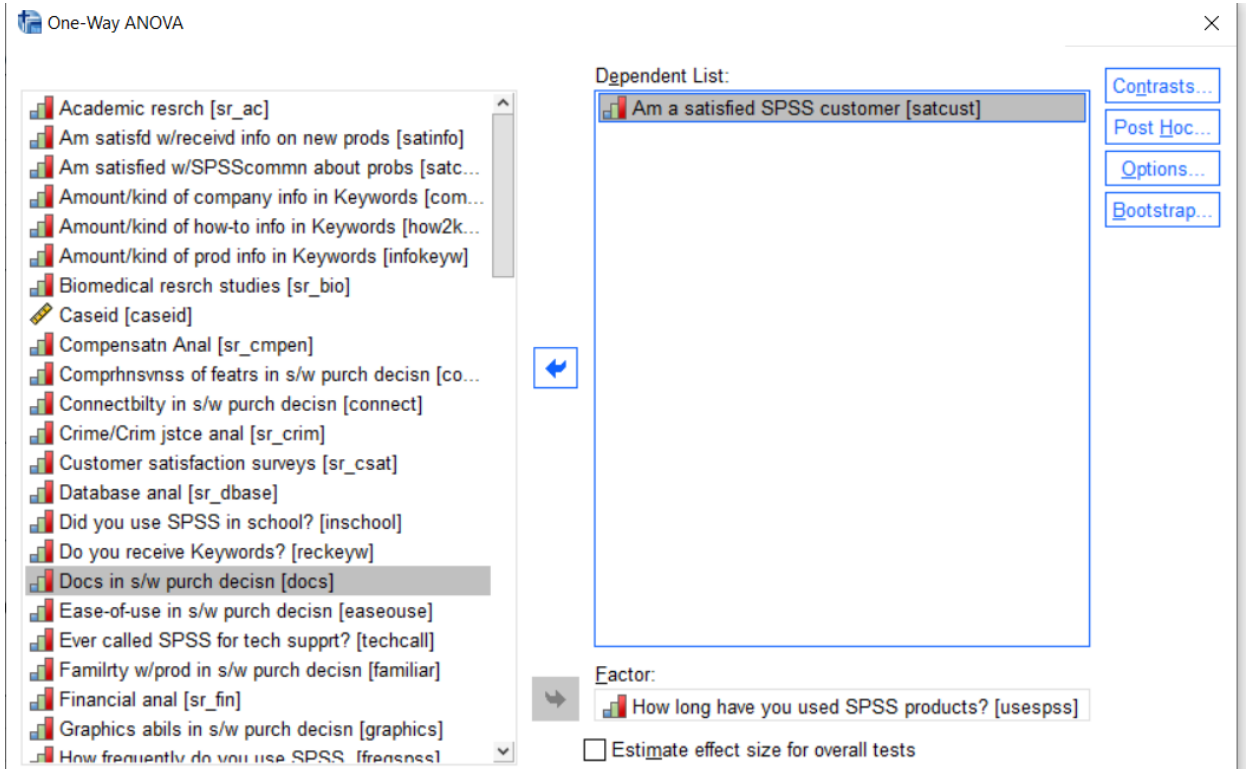
Business context

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"



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

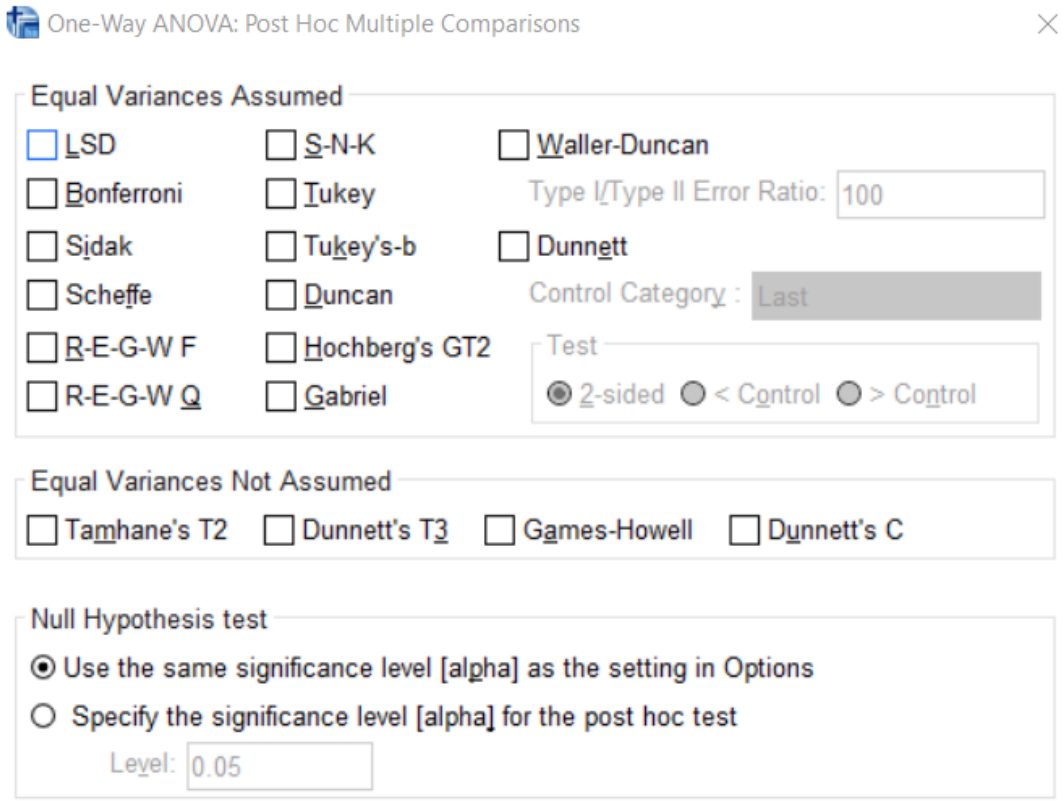
Am a satisfied SPSS customer

	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



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.

6. 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

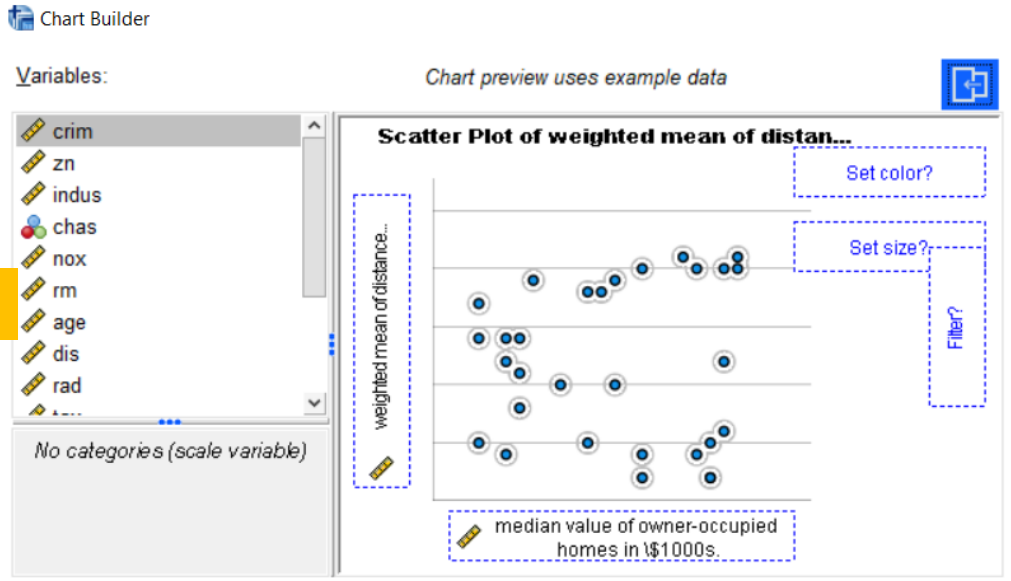
Data : “Boston House.SAV”

Graph >> Chart Builder

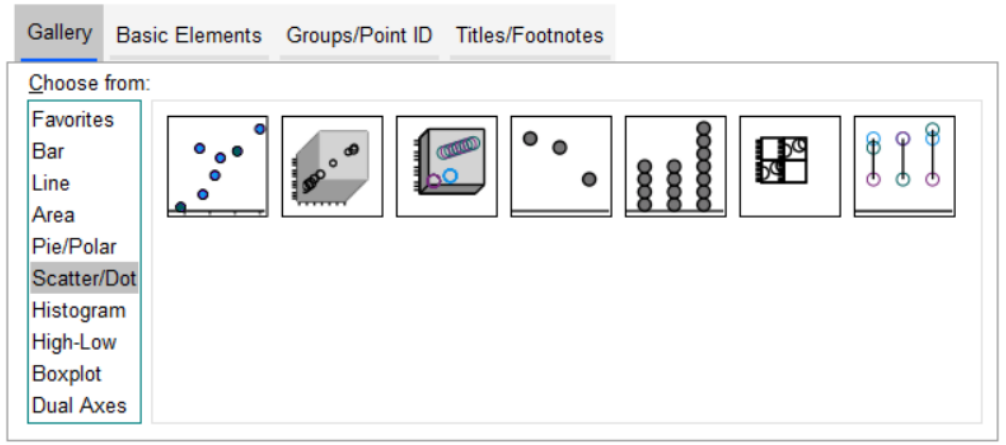
Select Variables

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

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

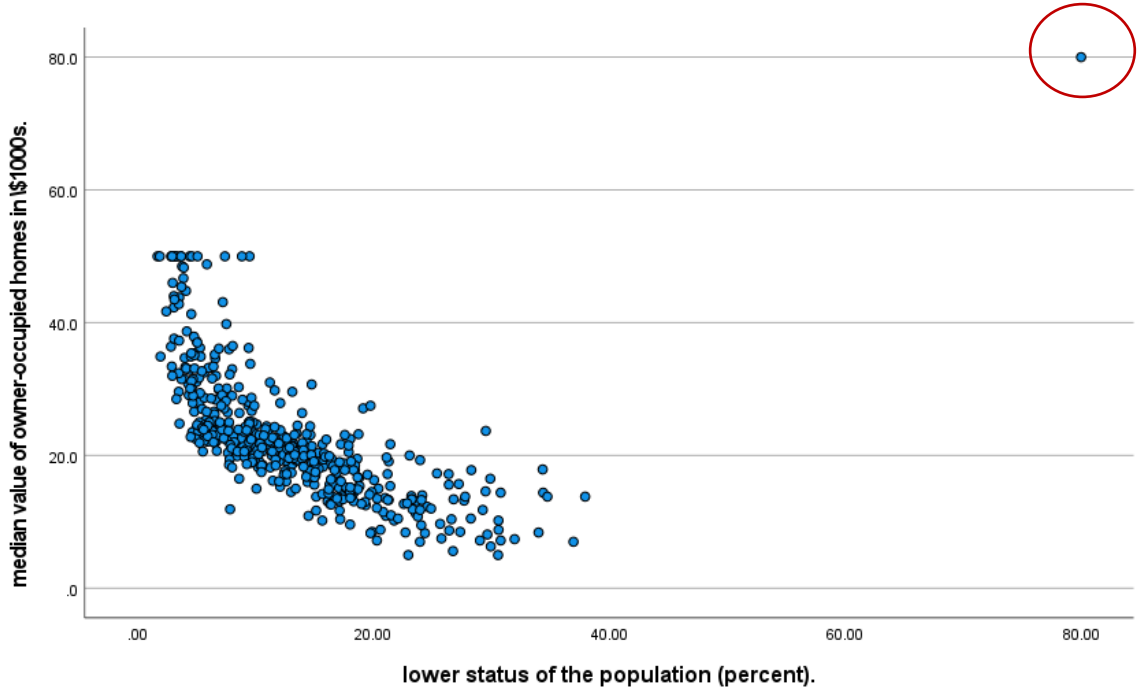


Select Chart Type

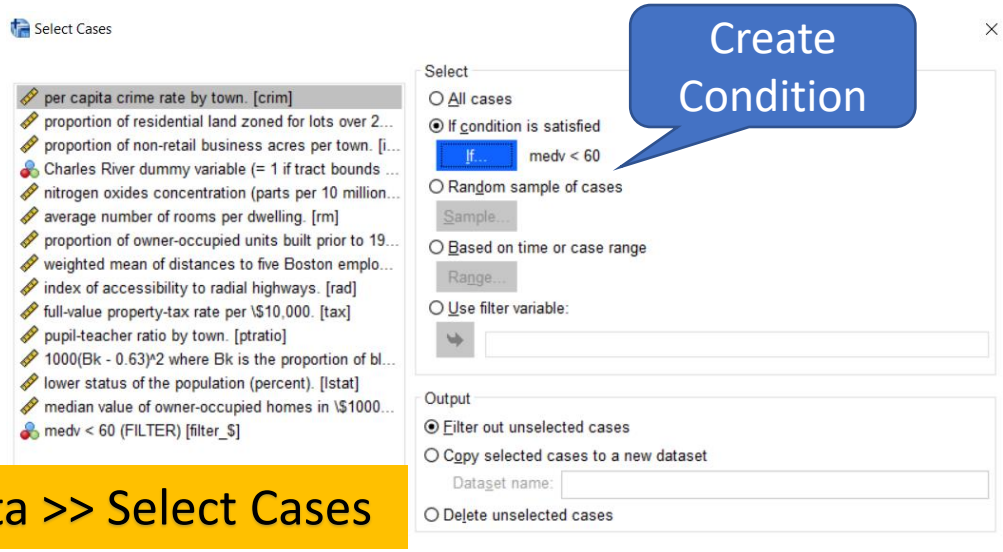


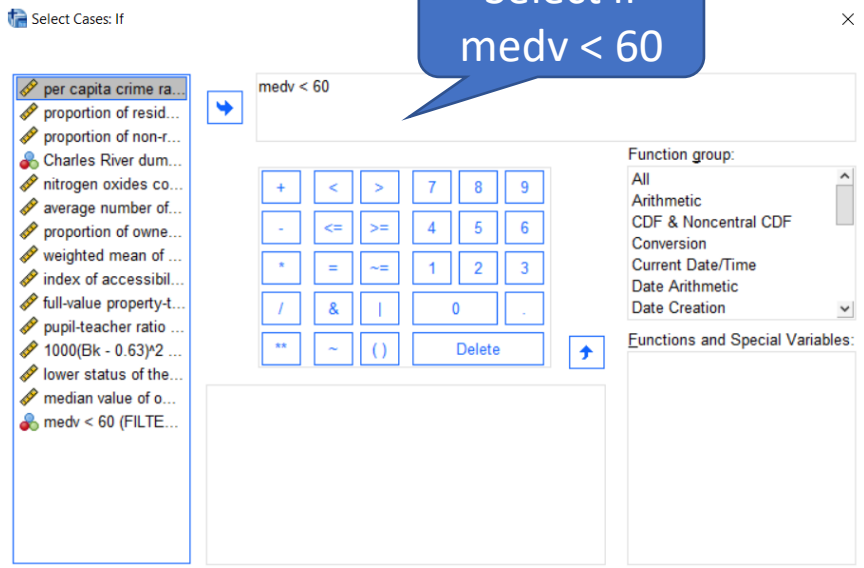
Remove Outlier and create chart again

Extreme Value



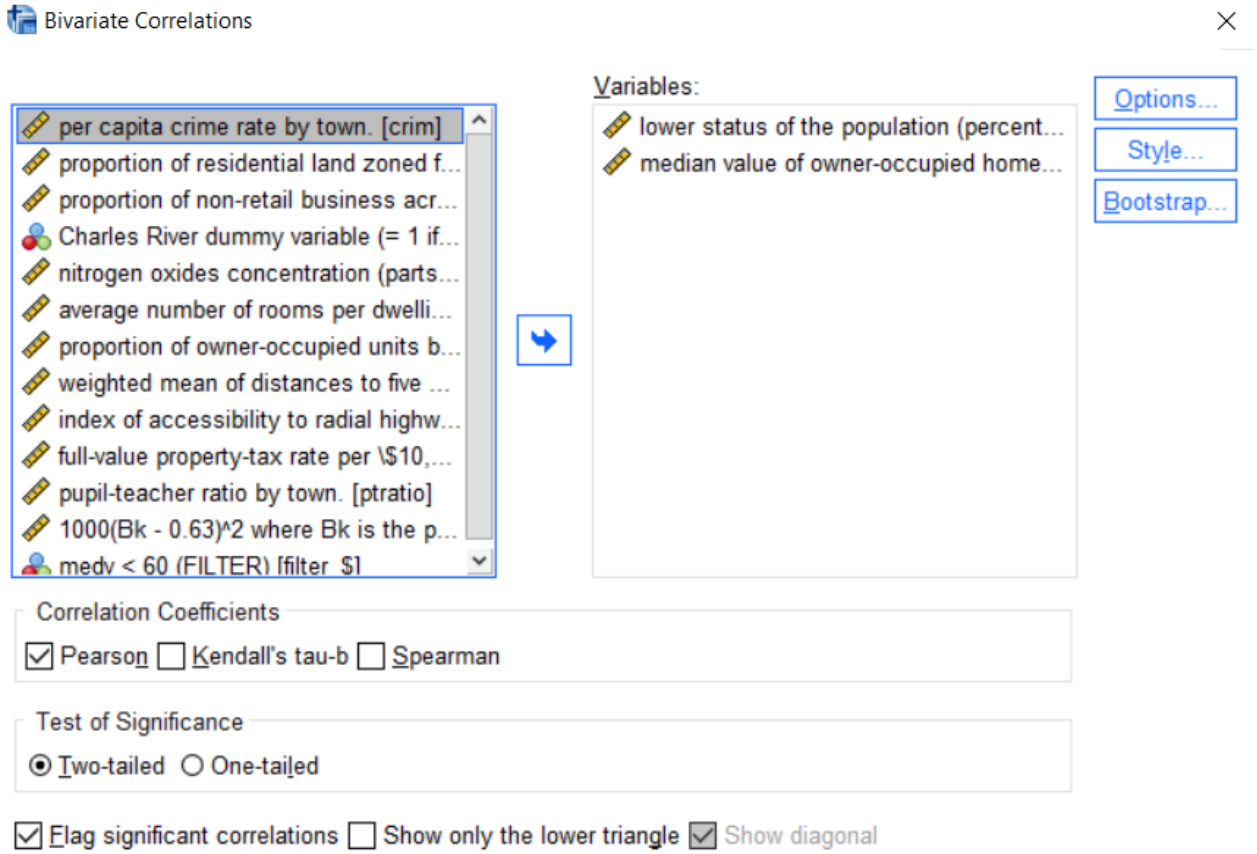
Data >> Select Cases





Correlations Coefficient

Analyze >> Correlate >> Bivariate Correlations



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).

7. 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

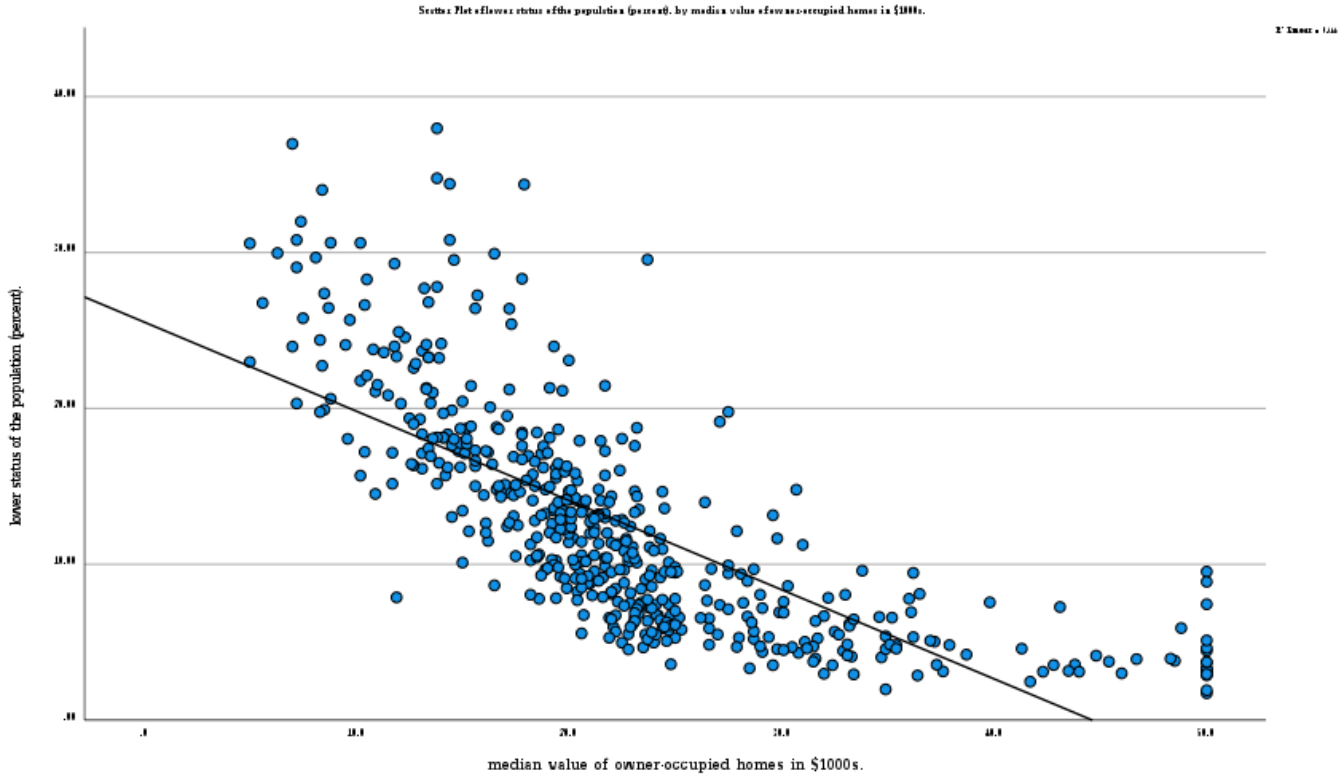
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.738[a]	.544	.543	6.2158

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

Coefficients[a]

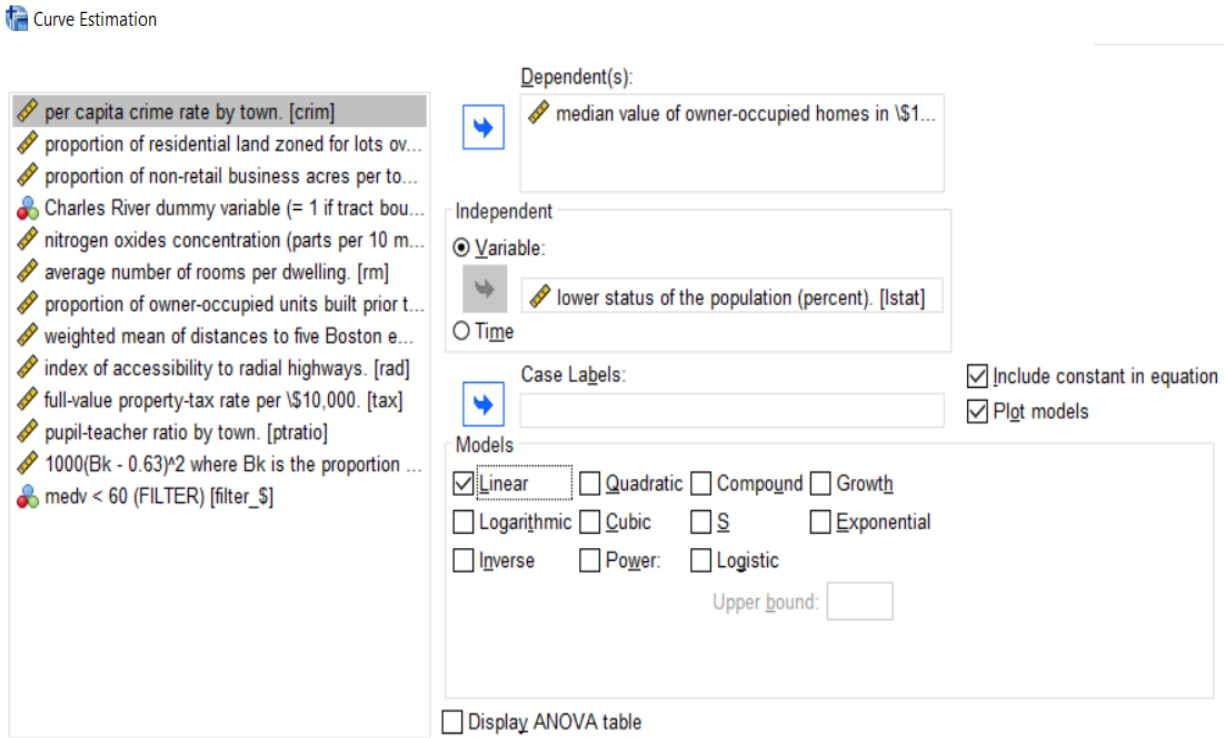
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	34.554	.563		61.415	.000
	lower status of the population (percent).	-.950	.039	-.738	-24.528	.000

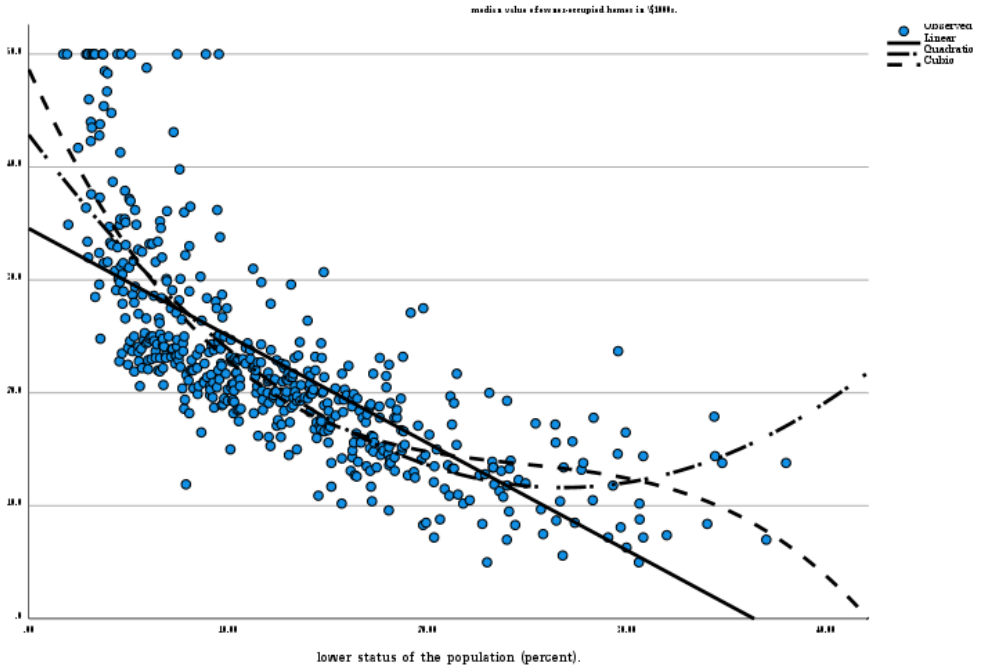
a. Dependent Variable: median value of owner-occupied homes in \$1000s.



Which line is the best fit?

Analyze >> Regression >> Curve Estimate





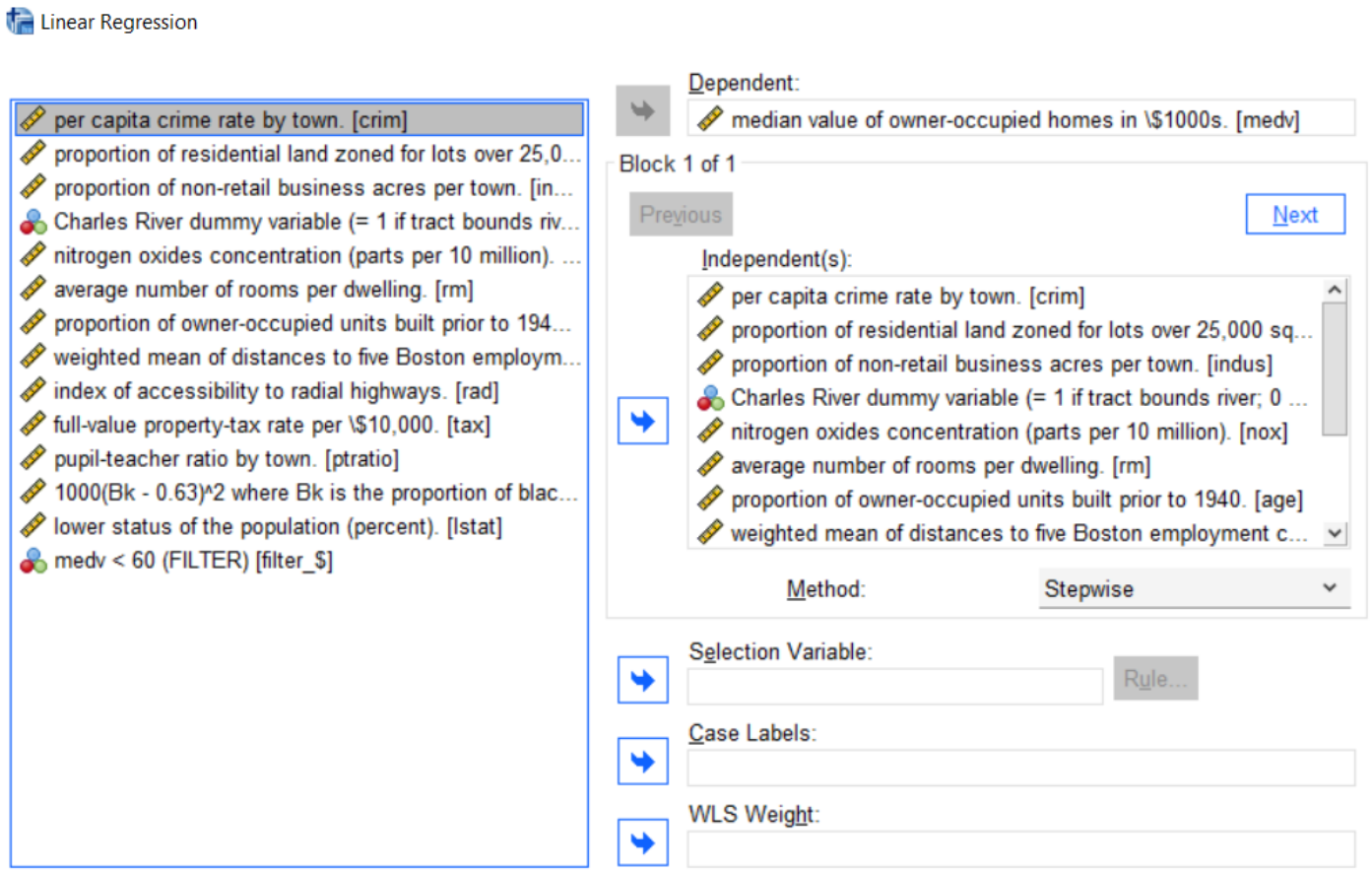
Model Summary and Parameter Estimates

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

	Model Summary					Parameter Estimates			
Equation	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



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

Q & A