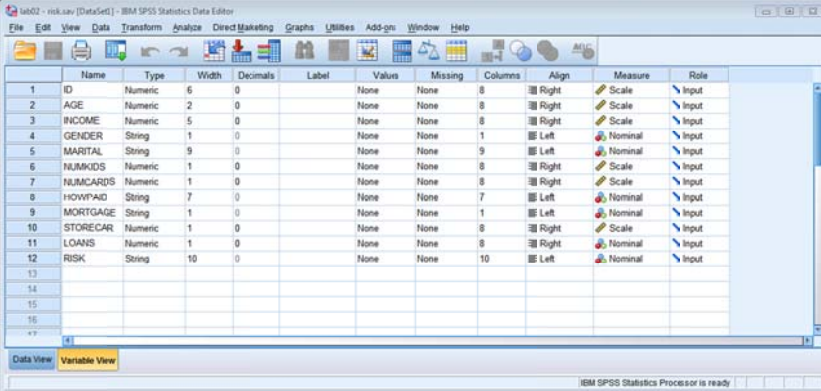


Lab 02 – SPSS

risk.csv data set

Import data

- start spss
- open an existing data source
- localize file on disk and set 'files of type' combo box to *.*
- wait for data to load
- check variable view tab and NUMKIDS,NUMCARDS,etc to scale

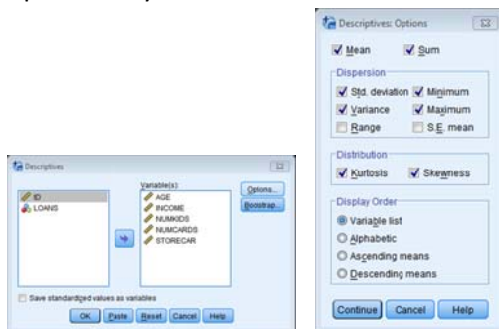


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	6	0		None	None	8	Right	Scale	Input
2	AGE	Numeric	2	0		None	None	8	Right	Scale	Input
3	INCOME	Numeric	5	0		None	None	8	Right	Scale	Input
4	GENDER	String	1	0		None	None	1	Left	Nominal	Input
5	MARITAL	String	9	0		None	None	9	Left	Nominal	Input
6	NUMKIDS	Numeric	1	0		None	None	8	Right	Scale	Input
7	NUMCARDS	Numeric	1	0		None	None	8	Right	Scale	Input
8	HOWPAID	String	7	0		None	None	7	Left	Nominal	Input
9	MORTGAGE	String	1	0		None	None	1	Left	Nominal	Input
10	STORECAR	Numeric	1	0		None	None	8	Right	Scale	Input
11	LOANS	Numeric	1	0		None	None	8	Right	Nominal	Input
12	RISK	String	10	0		None	None	10	Left	Nominal	Input

- save file

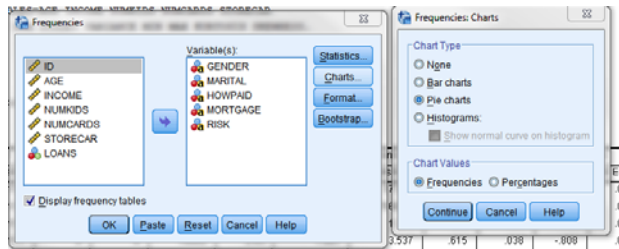
Summary statistics for numerical attributes

- Analyze/descriptive statistics/descriptives
- Select attributes (only numeric are shown)
- options let you choose the statistics to be computed



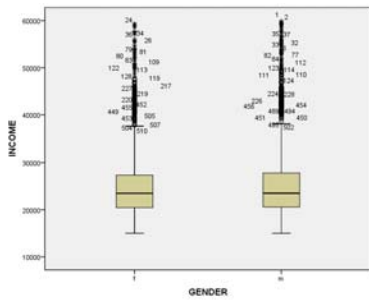
Summary of nominal attributes

- Analyze/descriptive statistics/frequencies
- select nominal variables
- Set 'display frequency tables' to true
- in 'charts' set 'Bar charts' or 'Pie charts'



nominal vs numeric

- analyze/descriptive statistics/explore
- dependent list: age; factor list: gender;



crosstabs

- Analyze/descriptive statistics/crosstabs
- Row(s): marital Columns(s): gender Layer: risk

MARITAL * RISK Crosstabulation

Count

		RISK			Total
		bad loss	bad profit	good risk	
MARITAL	divsepwid	297	397	179	873
	married	507	1182	400	2089
	single	102	828	225	1155
Total		906	2407	804	4117

Correlation matrix

- Analyze/correlate/bivariate (choose numeric variables)
-

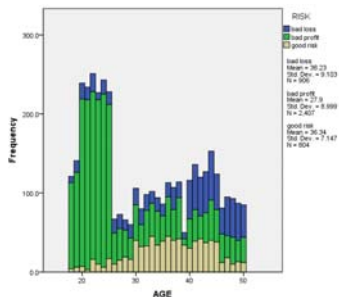
		ID	AGE	INCOME	NUMKIDS	NUMCARDS	STORECAR
ID	Pearson Correlation	1	.411**	-.463**	.626**	.744**	.733**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	4117	4117	4117	4117	4117	4117
AGE	Pearson Correlation	.411**	1	.295**	.435**	.600**	.478**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	4117	4117	4117	4117	4117	4117
INCOME	Pearson Correlation	-.463**	.295**	1	-.283**	-.139**	-.256**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	4117	4117	4117	4117	4117	4117
NUMKIDS	Pearson Correlation	.626**	.435**	-.283**	1	.667**	.663**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	4117	4117	4117	4117	4117	4117
NUMCARDS	Pearson Correlation	.744**	.600**	-.139**	.667**	1	.687**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	4117	4117	4117	4117	4117	4117
STORECAR	Pearson Correlation	.733**	.478**	-.256**	.663**	.687**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	4117	4117	4117	4117	4117	4117

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

-
-

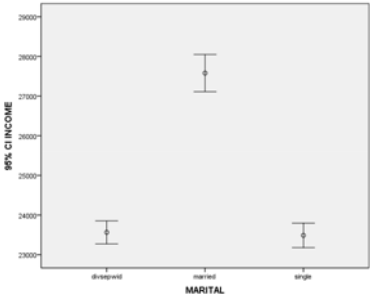
Compare means

- Analyze/compare means, dependent list: age; independent list: risk; options: anova table.
- Plot histogram of age stacked by risk
- do the same for other variables.



-

- Graphs/Legacy dialogs/Error Bar/simple



-