SPSS TRAINING SESSION 2

STATISTICAL ANALYSIS

(SPSS 16.0)

Sun Li
Centre for Academic Computing
lsun@smu.edu.sg

OUTLINE

- Elementary Data Analysis
- Group Comparison & One-way ANOVA
- Non-parametric Tests
- Correlations
- General Linear Regression
- Logistic Models
  - Binary Logistic Model
  - Ordinal Logistic Model
  - Multinomial Logistic Model

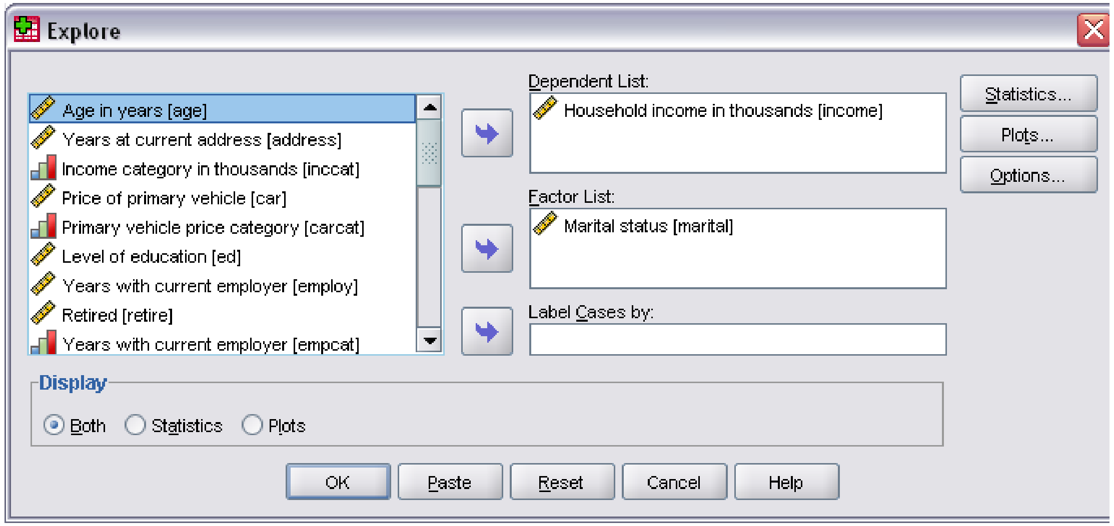
ELEMENTARY DATA ANALYSIS

The Explore procedure

- Exploratory data analysis
  - Summary statistics
  - Distribution plots
  - Normality plots with tests

Analyze
Descriptive Statistics
Explore

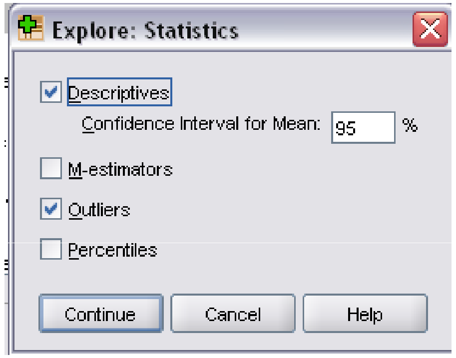
- The dependent variable must be a scale variable, while the grouping variables may be ordinal or nominal.

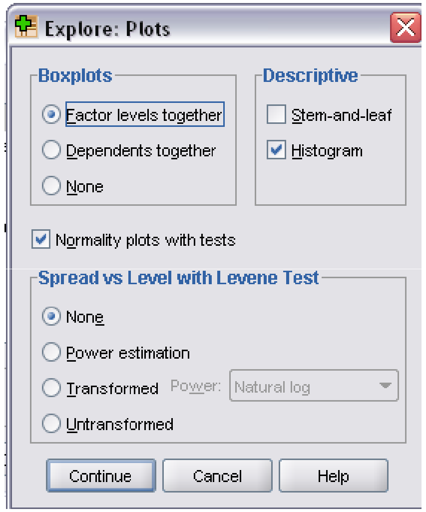


ELEMENTARY DATA ANALYSIS

E.g.: ***demo.sav***

Explore the household income for married and unmarried people.



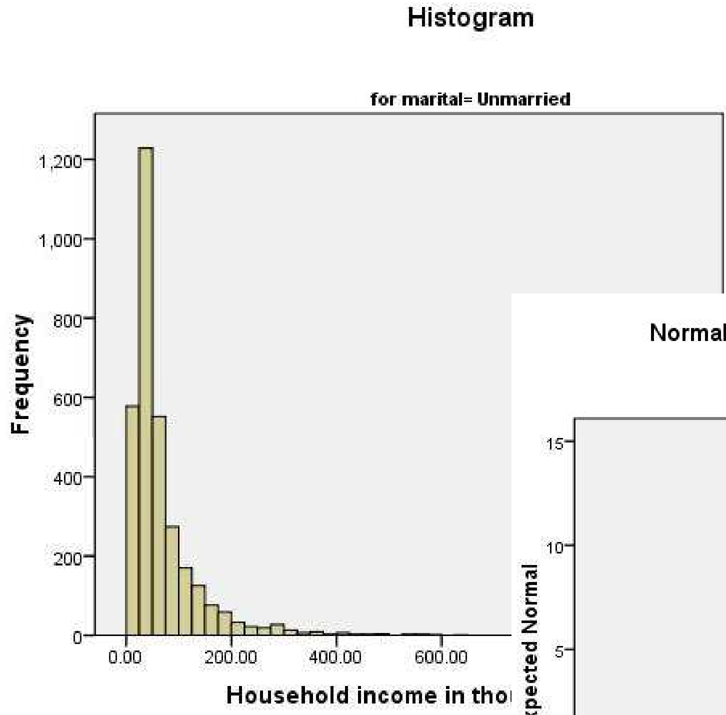


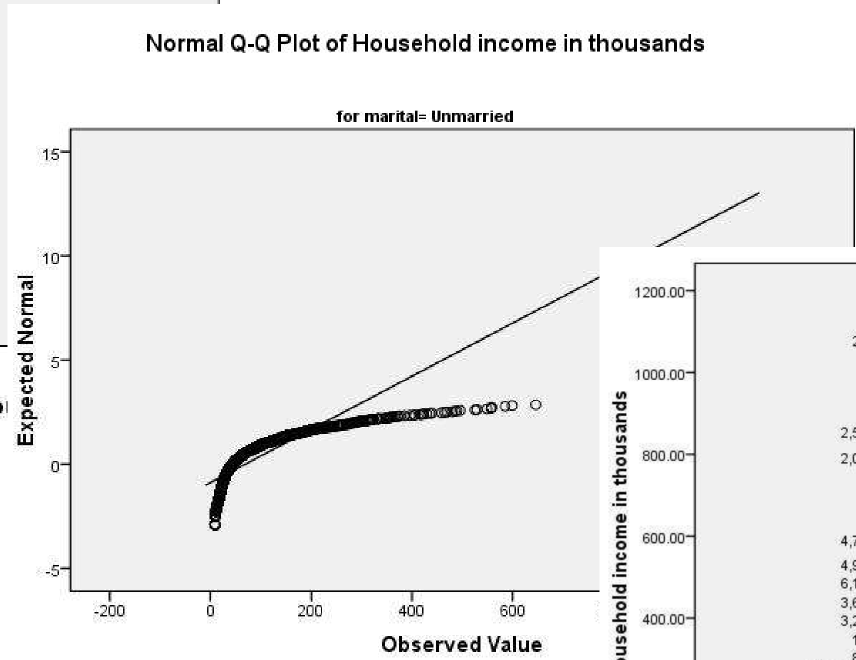
Tests of Normality

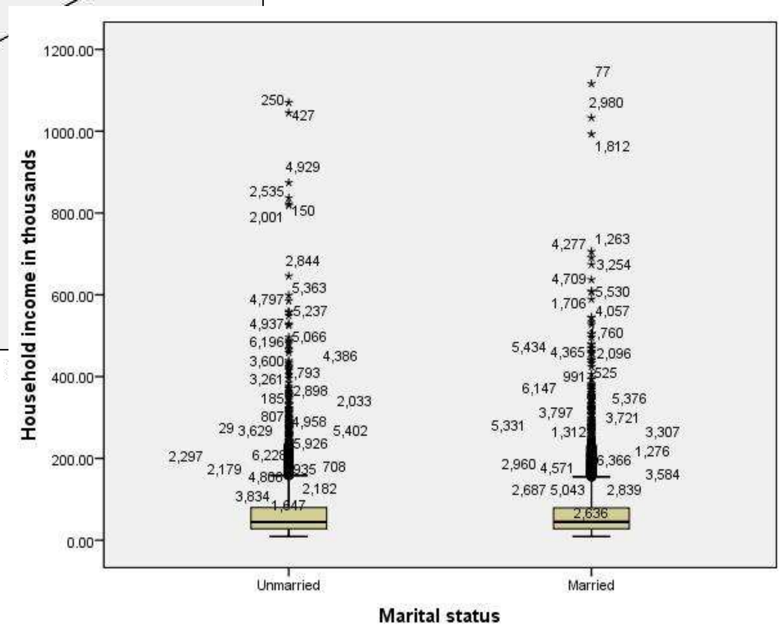
	Marital status	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Household income in thousands	Unmarried	.221	3224	.000	.600	3224	.000
	Married	.222	3176	.000	.602	3176	.000

a. Lilliefors Significance Correction

ELEMENTARY DATA ANALYSIS





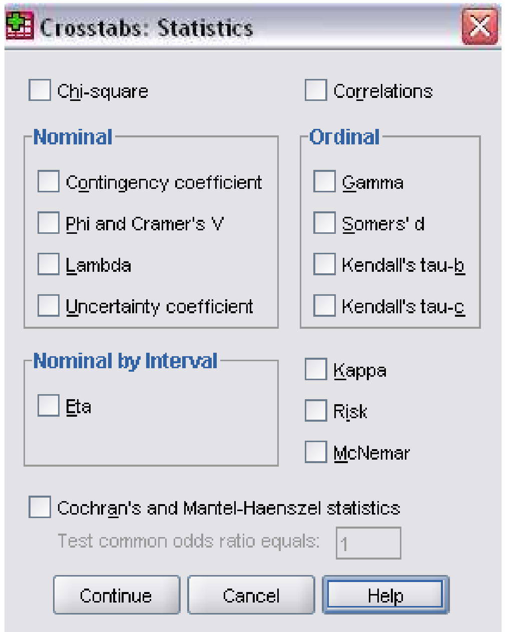


ELEMENTARY DATA ANALYSIS

The Crosstabulation table

Analyze
Descriptive Statistics
Crosstabs

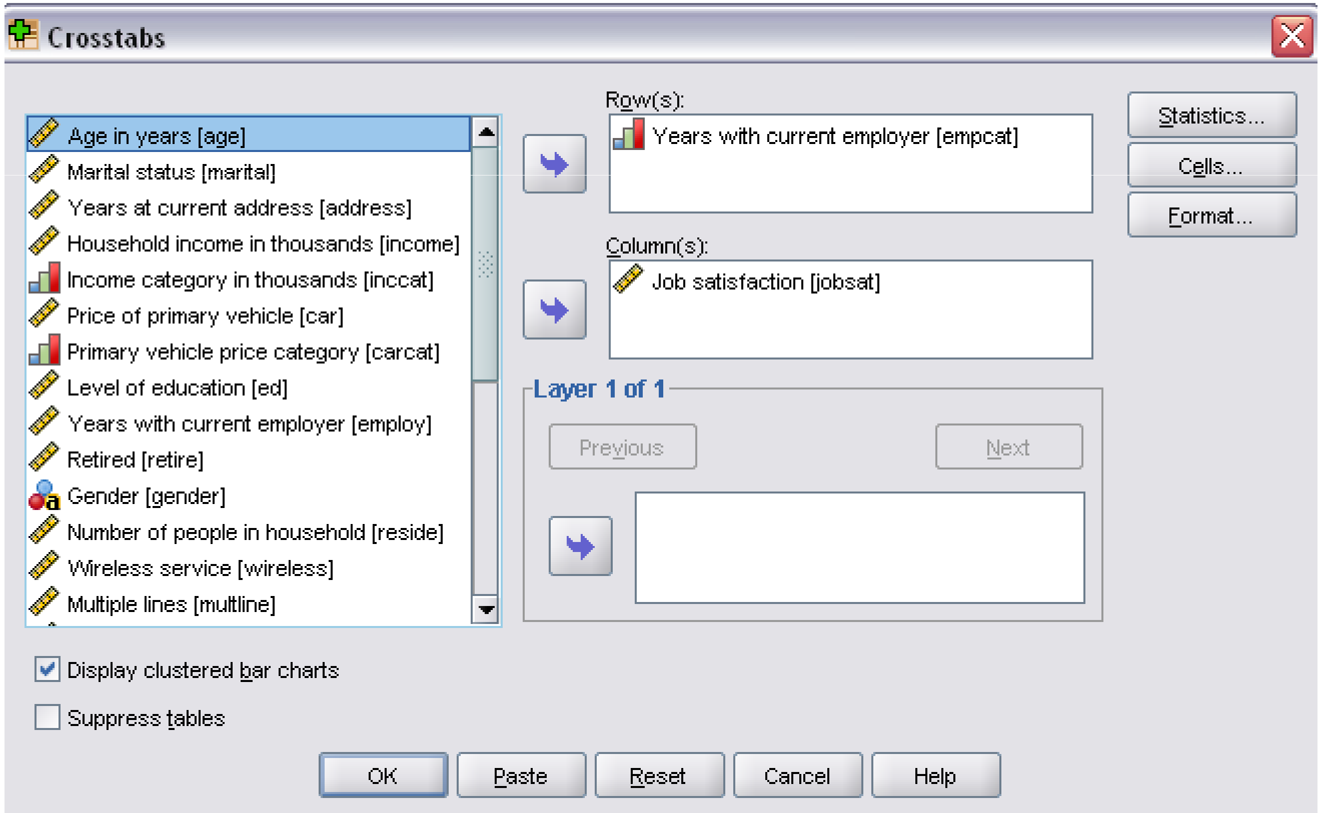
- Analysis of cross-classifications
- To examine the relationship btw two categorical variables
  - Nominal-by-nominal relationships
  - Ordinal-by-ordinal relationships
  - Nominal-by-interval relationships
  - Relative risk measurement
  - Agreement measurement



ELEMENTARY DATA ANALYSIS

E.g.:

Test and measure the relationship btw the job satisfactions and the number of year with current employer.



ELEMENTARY DATA ANALYSIS

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.690E3	8	.000
Likelihood Ratio	1747.380	8	.000
Linear-by-Linear Association	1525.767	1	.000
N of Valid Cases	6400		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 315.37.

Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T
Ordinal by Ordinal	Somers' d	Symmetric	.418	.009	47.6
		Years with current employer Dependent	.382	.008	47.6
		Job satisfaction Dependent	.461	.010	47.6

a. Not assuming the null hypothesis.

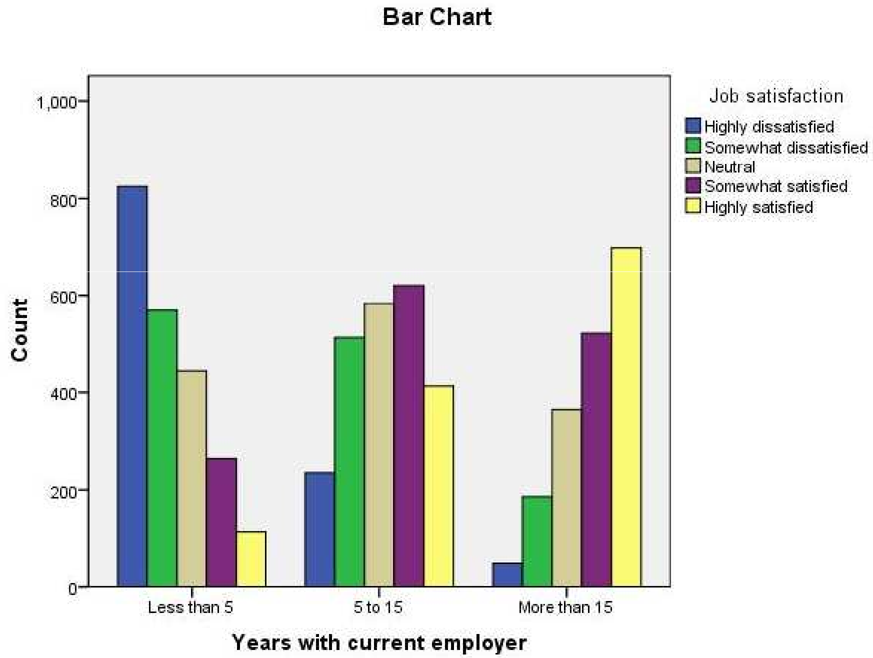
b. Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.420	.009	47.655	.000
	Kendall's tau-c	.458	.010	47.655	.000
	Gamma	.560	.011	47.655	.000
	N of Valid Cases	6400			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.



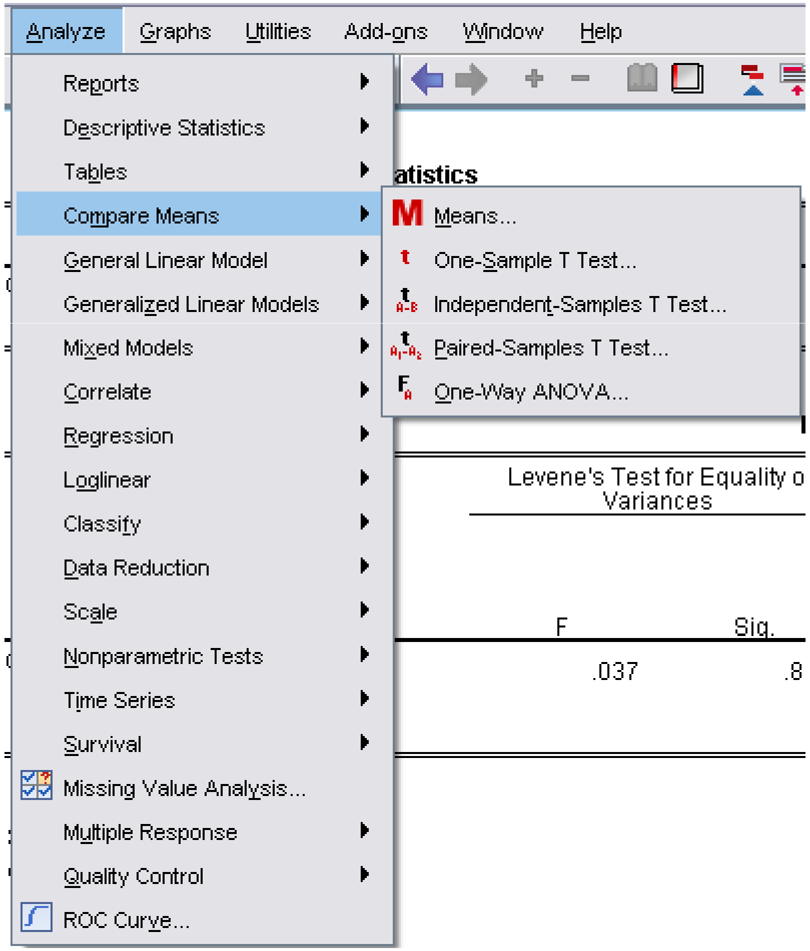
GROUP COMPARISON & ONE-WAY ANOVA

T Tests

- The one-sample T test
- The paired-samples T test (*skip*)
- The independent-samples T test

Analyze

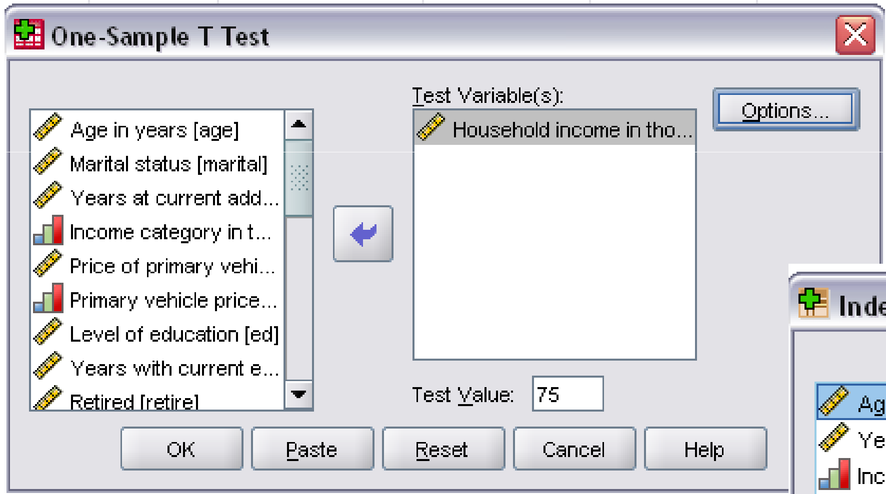
Compare Means

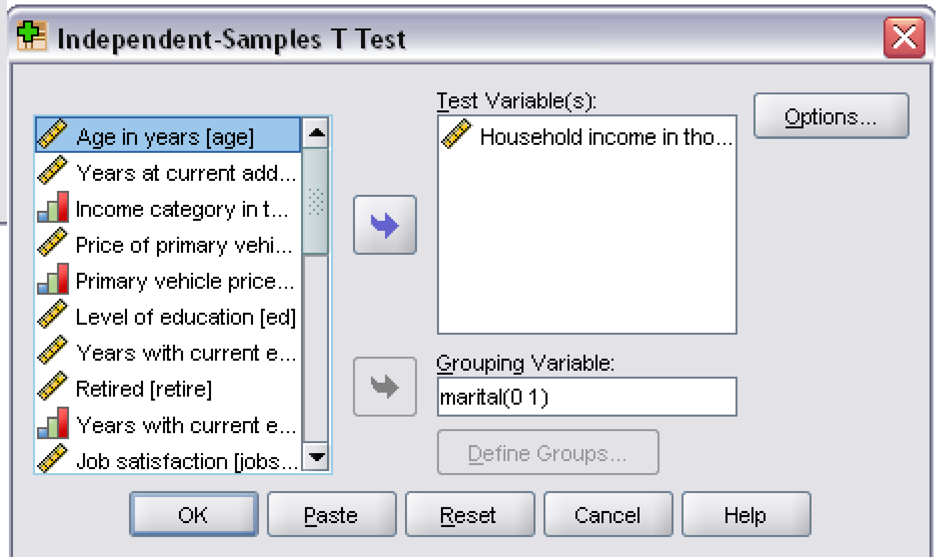


GROUP COMPARISON & ONE-WAY ANOVA

E.g.:

1. Test if the average household income equals to 75k.
2. Test if the average household income for married and unmarried people has no significant difference.



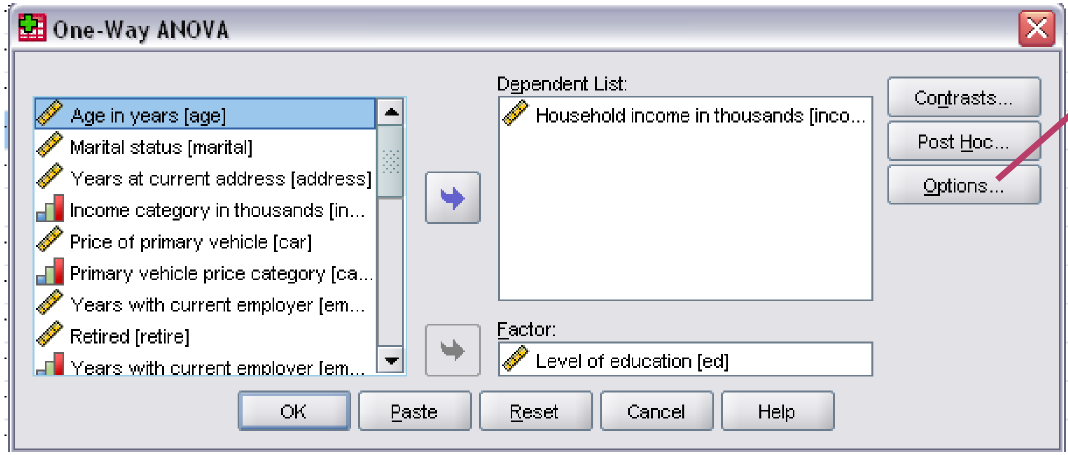


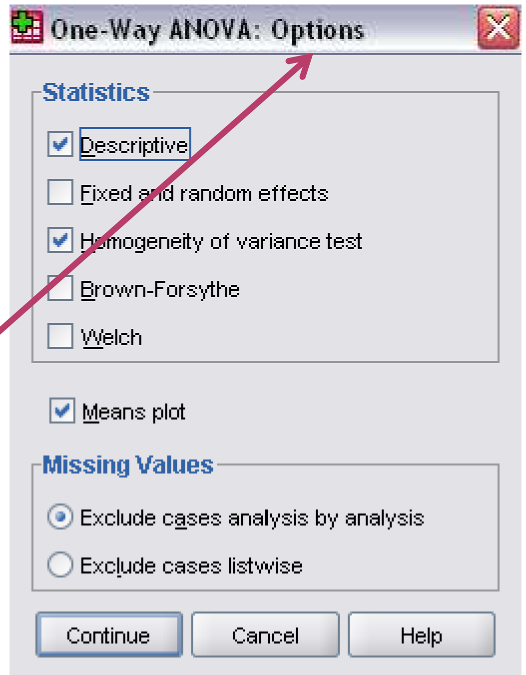
GROUP COMPARISON & ONE-WAY ANOVA

One-way analysis of variance

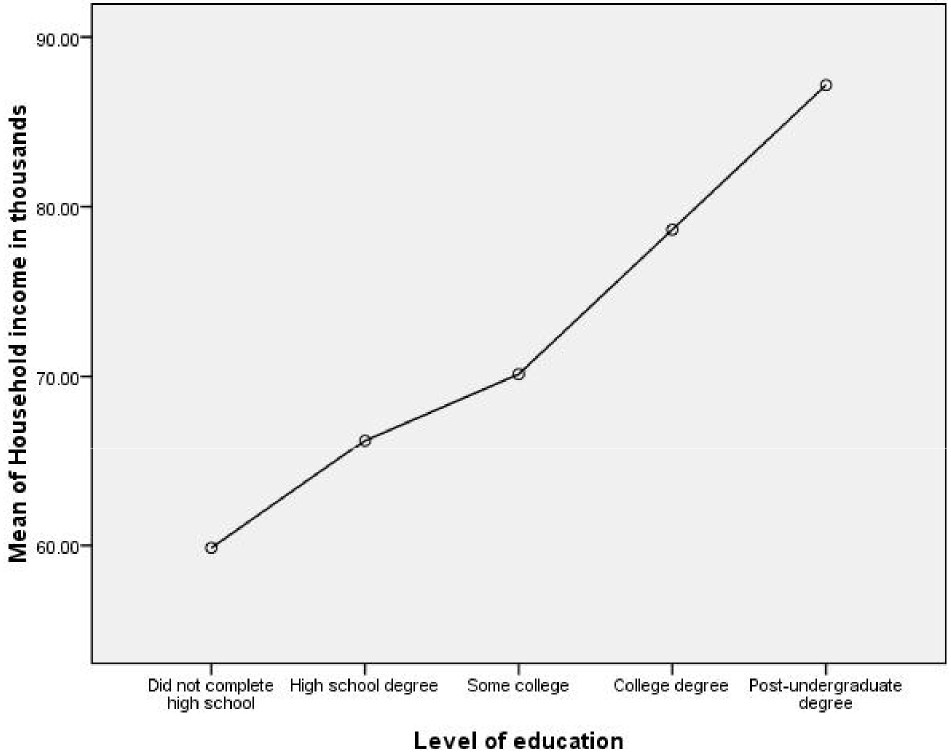
- to test the hypothesis that the means of two or more groups are not significantly different.

E.g.:
Test if the average household income for the five education groups are different. If there is significant difference, identify which groups have the major difference.





GROUP COMPARISON & ONE-WAY ANOVA



Test of Homogeneity of Variances

Household income in thousands

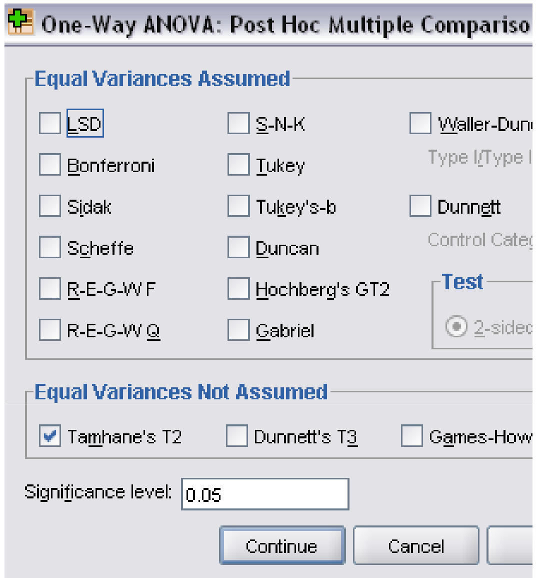
Levene Statistic	df1	df2	Sig.
14.766	4	6395	.000

ANOVA

Household income in thousands

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	376079.699	4	94019.925	15.309	.000
Within Groups	3.928E7	6395	6141.680		
Total	3.965E7	6399			

GROUP COMPARISON & ONE-WAY ANOVA



Multiple Comparisons

Household income in thousands
Tamhane

(I) Level of education	(J) Level of education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Did not complete high school	High school degree	-6.34094	2.33139	.064	-12.8724	.1905
	Some college	-10.26837*	2.74450	.002	-17.9587	-2.5781
	College degree	-18.78400*	3.01158	.000	-27.2233	-10.3447
	Post-undergraduate degree	-27.30373*	5.26659	.000	-42.1229	-12.4846
High school degree	Did not complete high school	6.34094	2.33139	.064	-.1905	12.8724
	Some college	-3.92743	2.74970	.811	-11.6318	3.7769
	College degree	-12.44306*	3.01632	.000	-20.8952	-3.9909
	Post-undergraduate degree	-20.96279*	5.26930	.001	-35.7894	-6.1362
Some college	Did not complete high school	10.26837*	2.74450	.002	2.5781	17.9587
	High school degree	3.92743	2.74970	.811	-3.7769	11.6318
	College degree	-8.51563	3.34591	.105	-17.8907	.8594
	Post-undergraduate degree	-17.03536*	5.46465	.019	-32.4016	-1.6691
College degree	Did not complete high school	18.78400*	3.01158	.000	10.3447	27.2233
	High school degree	12.44306*	3.01632	.000	3.9909	20.8952
	Some college	8.51563	3.34591	.105	-.8594	17.8907
	Post-undergraduate degree	-8.51973	5.60355	.749	-24.2704	7.2309
Post-undergraduate degree	Did not complete high school	27.30373*	5.26659	.000	12.4846	42.1229
	High school degree	20.96279*	5.26930	.001	6.1362	35.7894
	Some college	17.03536*	5.46465	.019	1.6691	32.4016
	College degree	8.51973	5.60355	.749	-7.2309	24.2704

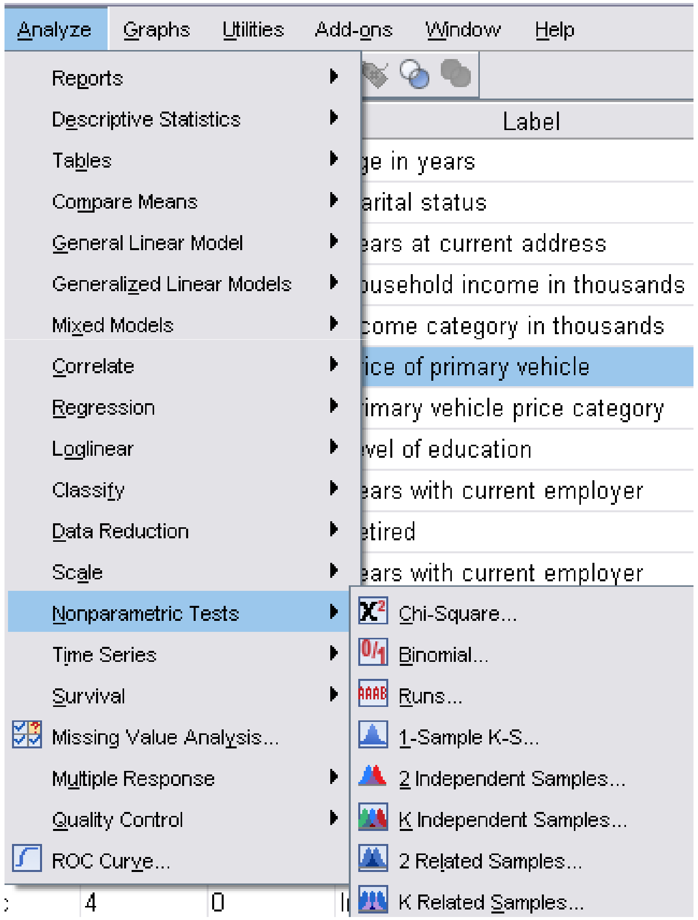
*. The mean difference is significant at the 0.05 level.

NON-PARAMETRIC TESTS

Non-parametric tests

- Two-independent samples tests
- Tests for several independent samples

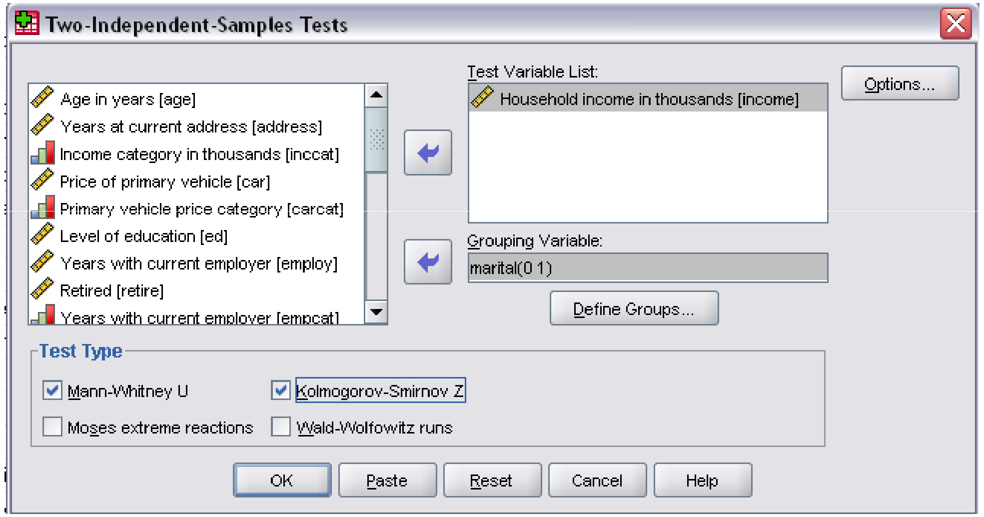
Analyze
Nonparametric Tests
2 Independent Samples
K Independent Samples



NON-PARAMETRIC TESTS

Two-independent samples tests:

Test if the household income varies btw married and unmarried people.



Mann-Whitney and Wilcoxon tests--

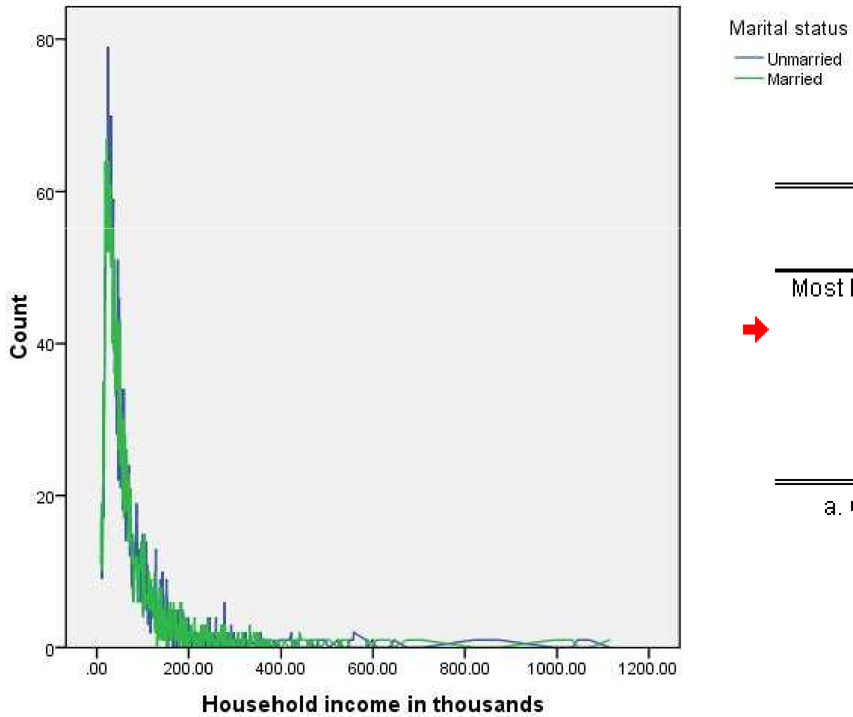
Test Statistics[a]

	Household income in thousands
Mann-Whitney U	5108032.500
Wilcoxon W	1.031E7
Z	-.158
Asymp. Sig. (2-tailed)	.874

a. Grouping Variable: Marital status

NON-PARAMETRIC TESTS

The two-sample Kolmogorov-Smirnov test



Test Statistics[a]

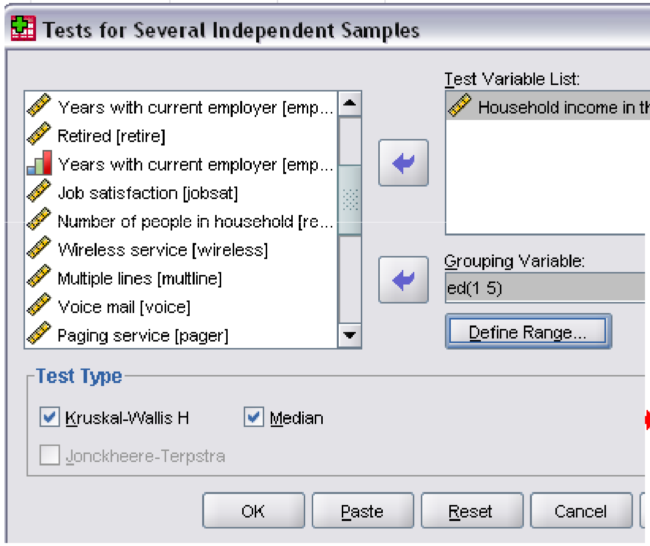
		Household income in thousands
Most Extreme Differences	Absolute	.016
	Positive	.016
	Negative	-.008
	Kolmogorov-Smirnov Z	.632
	Asymp. Sig. (2-tailed)	.820

a. Grouping Variable: Marital status

NON-PARAMETRIC TESTS

K-independent samples tests:

Test if the household income varies among people with different educations.



Using the median test to detect the difference:

Median Test

Frequencies

		Level of education				
		Did not complete high school	High school degree	Some college	College degree	Post-undergraduate degree
Household income in thousands	> Median	596	907	654	751	218
	<= Median	794	1029	706	604	141

Test Statistics[b]

	Household income in thousands
N	6400
Median	45.0000
Chi-Square	66.957[a]
df	4
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 175.3.

b. Grouping Variable: Level of education

NON-PARAMETRIC TESTS

Using Kruskal-Wallis to Test Ordinal Outcomes

Kruskal-Wallis Test

Ranks

	Level of education	N	Mean Rank
Household income in thousands	Did not complete high school	1390	2923.64
	High school degree	1936	3117.33
	Some college	1360	3195.49
	College degree	1355	3460.83
	Post-undergraduate degree	359	3757.41
	Total	6400	

Test Statistics[a,b]

	Household income in thousands
Chi-Square	94.672
df	4
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Level of education

CORRELATION

Correlation

Analyze

Correlation

- Bivariate correlation:

  Describe relationship btw two variables.
- Partial correlation:

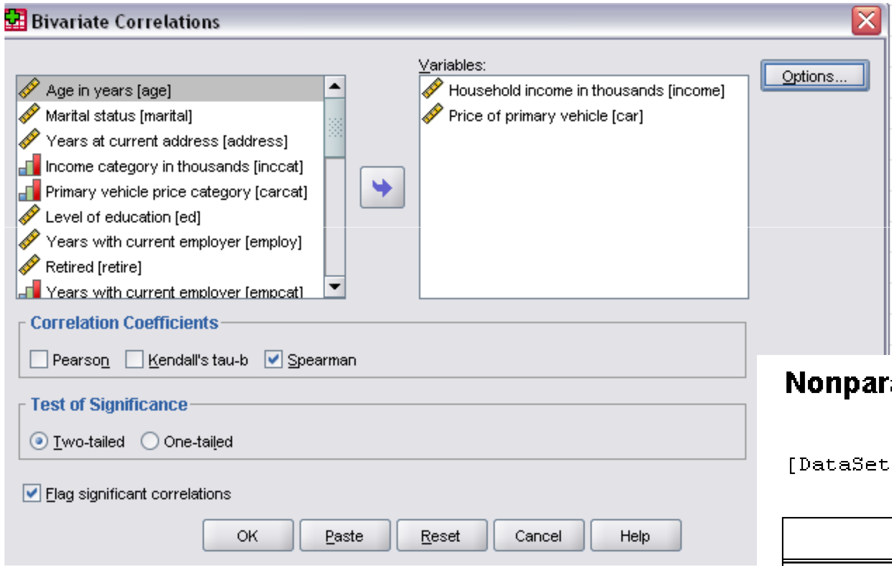
  Describe relationship btw two variables while controlling for the effects of one or more additional variables.
- Distances: (skip)

  Similarities/dissimilarities btw pairs of variables/cases.

CORRELATION

E.g.:

Compute the correlation coefficient btw the household income and the price of primary vehicle.



Nonparametric Correlations

```
[DataSet1] C:\Program Files\SPSSInc\SPSS16\Samples\demo.sav
```

Correlations

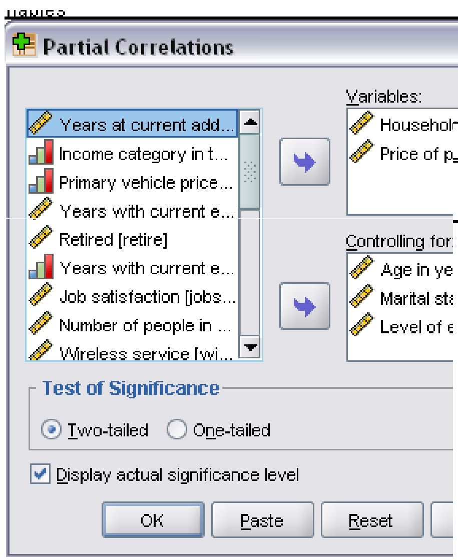
			Household income in thousands	Price of primary vehicle
Spearman's rho	Household income in thousands	Correlation Coefficient	1.000	.998**
		Sig. (2-tailed)	.	.000
		N	6400	6400
	Price of primary vehicle	Correlation Coefficient	.998**	1.000
		Sig. (2-tailed)	.000	.
		N	6400	6400

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

CORRELATION

E.g.:

Compute the correlation coefficient btw the household income and the price of primary vehicle after controlling factor age, marriage status and education.



Correlations

Control Variables			Household income in thousands	Price of primary vehicle	Age in years	Marital status	Level of education
-none-[a]	Household income in thousands	Correlation	1.000	.792	.335	.003	.096
		Significance (2-tailed)	.	.000	.000	.836	.000
		df	0	6398	6398	6398	6398
	Price of primary vehicle	Correlation	.792	1.000	.376	-.002	.102
		Significance (2-tailed)	.000	.	.000	.846	.000
		df	6398	0	6398	6398	6398
	Age in years	Correlation	.335	.376	1.000	.003	-.126
		Significance (2-tailed)	.000	.000	.	.812	.000
		df	6398	6398	0	6398	6398
	Marital status	Correlation	.003	-.002	.003	1.000	-.030
		Significance (2-tailed)	.836	.846	.812	.	.016
		df	6398	6398	6398	0	6398
	Level of education	Correlation	.096	.102	-.126	-.030	1.000
		Significance (2-tailed)	.000	.000	.000	.016	.
		df	6398	6398	6398	6398	0
Age in years & Marital status & Level of education	Household income in thousands	Correlation	1.000	.757			
		Significance (2-tailed)	.	.000			
		df	0	6395			
	Price of primary vehicle	Correlation	.757	1.000			
		Significance (2-tailed)	.000	.			
		df	6395	0			

a. Cells contain zero-order (Pearson) correlations.

GENERAL LINEAR MODELS

The GLM Univariate procedure

Analyze
General Linear Model
Univariate

Analyze
Regression
Linear

In SPSS, there are two menus about linear regressions under pull-down menu "Analyze". The difference btw the two menus is that "Linear" function under "Regression" treats every predictor to be continuous variables, while in "General Linear Model", there are options to define categorical predictors and continuous predictors.

GENERAL LINEAR MODELS

- Factors: Categorical predictors should be selected as factors in the model.
  - Fixed-effects factors are generally thought of as variables whose values of interest are all represented in the data file.
  - Random-effects factors are variables whose values in the data file can be considered a random sample from a larger population of values. They are useful for explaining excess variability in the dependent variable.
- Covariates: Scale predictors should be selected as covariates in the model.

E.g.: ***grocery_1month.sav***

Use the GLM Univariate procedure to fit a model with fixed and random effects to the amounts customers spent in grocery stores.

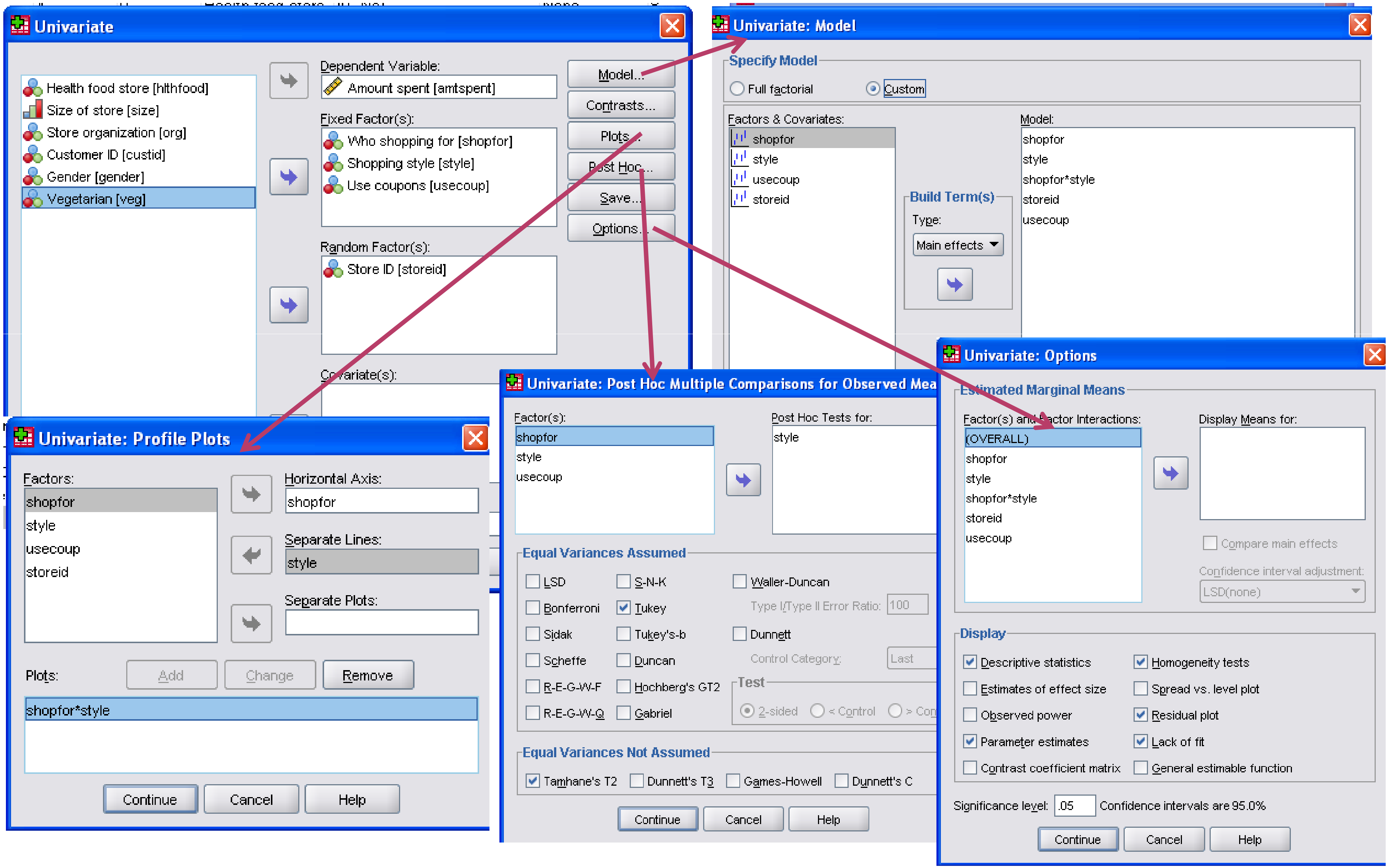
GENERAL LINEAR MODELS

E.g.:

A grocery store chain surveyed a set of customers concerning their purchasing habits. Given the survey results and how much each customer spent in the previous month, the store wants to see the factors.

Variable	Variable information
storeid	Store id
shop for	Who shop for: 1 = self 2 = self and spouse 3 = self and family
style	Shopping style: 1 = biweekly, in bulk 2 = weekly, similar items 3 = often, what's on sale
usecoup	Use coupons: 1 = no 2 = from newspaper 3 = from mailings 4 = from both
amtspent	Amount spent last month

GENERAL LINEAR MODELS



GENERAL LINEAR MODELS

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Amount spent

F	df1	df2	Sig.
.650	320	30	.961

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

a. Design: Intercept + shopfor + style + shopfor * style + storeid + usecoup

Tests of Between-Subjects Effects

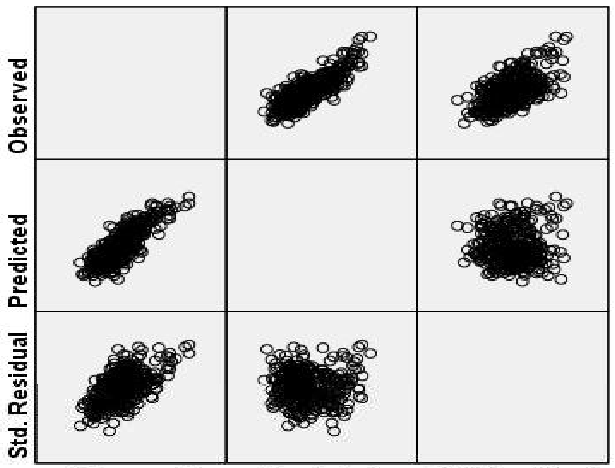
Dependent Variable:Amount spent

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	3.362E7	1	3.362E7	5219.104	.000
	Error	672713.222	104.418	6442.505[a]		
shopfor	Hypothesis	761142.660	2	380571.330	94.249	.000
	Error	1130619.470	280	4037.927[b]		
style	Hypothesis	64332.117	2	32166.059	7.966	.000
	Error	1130619.470	280	4037.927[b]		
shopfor * style	Hypothesis	12766.092	4	3191.523	.790	.532
	Error	1130619.470	280	4037.927[b]		
storeid	Hypothesis	479152.932	59	8121.236	2.011	.000
	Error	1130619.470	280	4037.927[b]		
usecoup	Hypothesis	191472.817	3	63824.272	15.806	.000
	Error	1130619.470	280	4037.927[b]		

a. .589 MS(storeid) + .411 MS(Error)

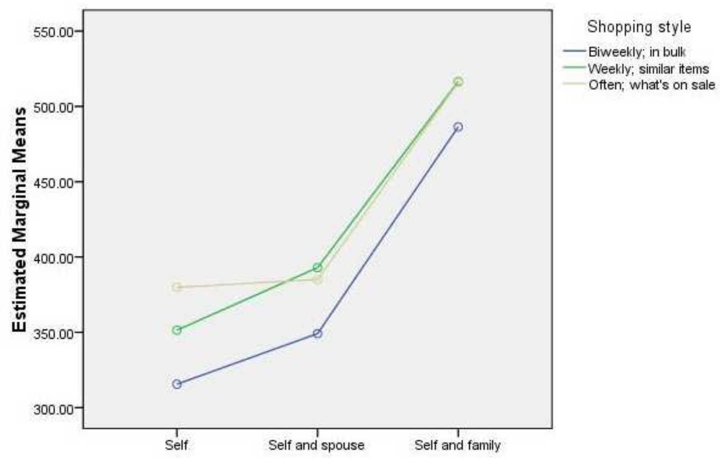
b. MS(Error)

Dependent Variable: Amount spent



Model: Intercept + shopfor + style + shopfor * style + storeid + usecoup

Estimated Marginal Means of Amount spent



GENERAL LINEAR MODELS

- Hays, W. L. 1981. Statistics, 3rd ed. New York: Holt, Rinehart, and Winston.
- Brown, M. B., and A. B. Forsythe. 1974b. Robust tests for the equality of variances. Journal of the American Statistical Association, 69:, 364-367.
- Milliken, G., and D. Johnson. 1992. Analysis of Messy Data: Volume 1. Designed Experiments. New York: Chapman & Hall.
- Neter, J., W. Wasserman, and M. H. Kutner. 1990. Applied Linear Statistical Models, 3rd ed. Homewood, Ill.: Irwin.
- Siegel, S., and N. J. Castellan. 1988. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, Inc..
- Conover, W. J. 1980. Practical Nonparametric Statistics, 2nd ed. New York: John Wiley and Sons.
- Horton, R. L. 1978. The General Linear Model. New York: McGraw-Hill, Inc..

LOGISTIC MODELS

Logistic Models

- Binary logistic model: dichotomous response outcomes

  e.g.: presence or absence of an event

$$\pi_i = E(y_i \mid x_i) \qquad \mathrm{logit}(\pi) = \log(\pi/(1-\pi)) = g(\pi) = \alpha + \beta' X$$

- Ordinal logistic model: ordinal response variable with more than two ordered categories

  e.g.: a 5-point Likert scale

$$g(\Pr(Y \le i \mid X)) = \alpha_i + \beta' X, \quad i = 1, \ldots\ldots, k$$

- Multinomial logistic model: nominal response variables with more than two categories

  e.g.: different types of programs in school

$$\log\left(\frac{\Pr(Y = i \mid X)}{\Pr(Y = k+1 \mid X)}\right) = \alpha_i + \beta'_i X, \quad i = 1, \ldots\ldots, k$$

BINARY LOGISTIC MODELS

Using binary logistic regression to assess credit risk

Analyze
Regression
Binary Logistic

E.g.: ***bankloan.sav***

If you are a loan officer at a bank, then you want to be able to identify characteristics that are indicative of people who are likely to default on loans, and use those characteristics to identify good and bad credit risks.

We have information on 850 past and prospective customers. The first 700 cases are customers who were previously given loans. Use these 700 customers to create a logistic regression model. Then use the model to classify the 150 prospective customers as good or bad credit risks.

BINARY LOGISTIC MODELS

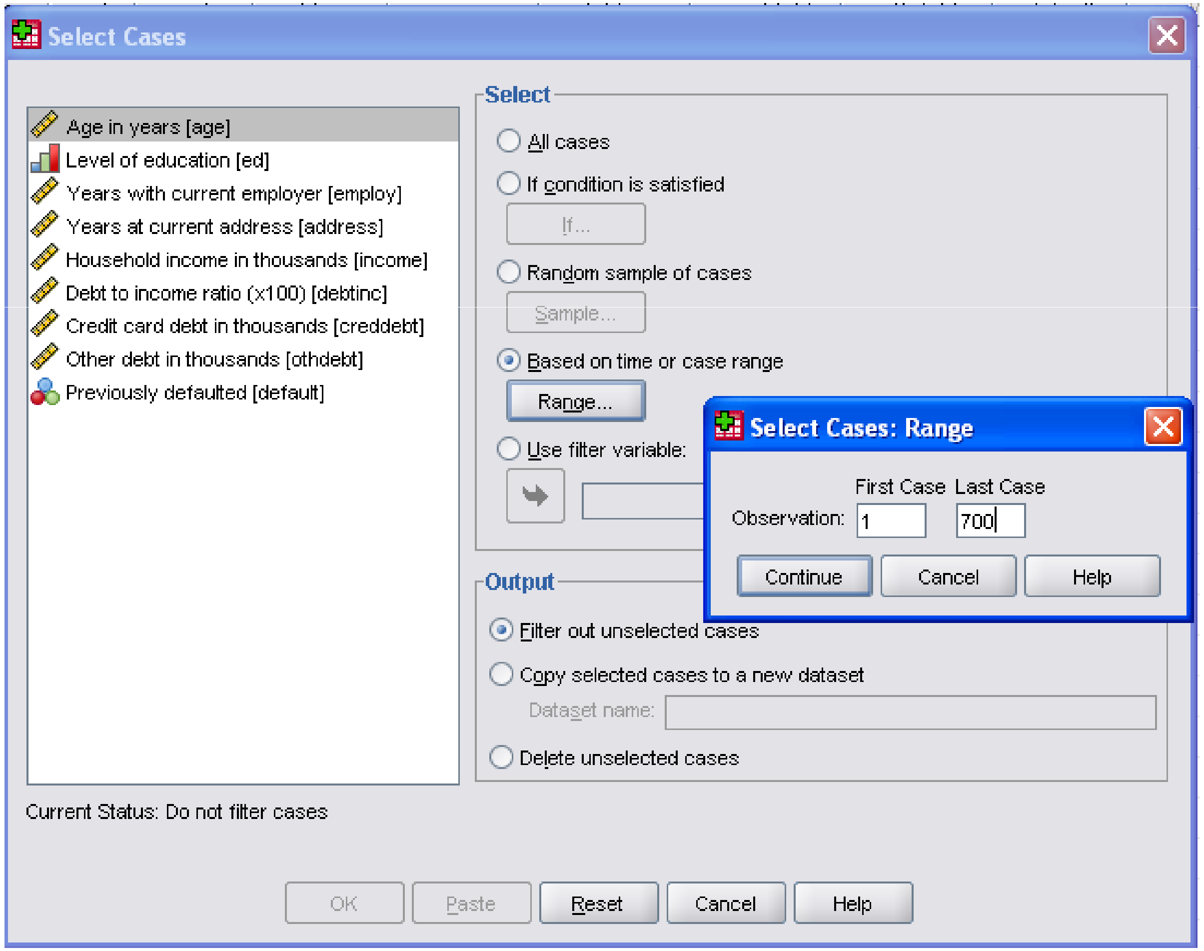
Variable name	Variable information
age	Age in years
ed	Level of education 1= didn't complete high school 2= high school degree 3= college degree 4= undergraduate 5= postgraduate
employ	Years with current employer
address	Years in current address
income	Household income in thousands
debtinc	Debt to income ratio (*100)
creddebt	Credit card debt in thousands
othdebt	Other debts in thousands
default	Previously defaulted 1= Yes 0 = No

BINARY LOGISTIC MODELS

Step 1: select the first 700 obs for logistic model

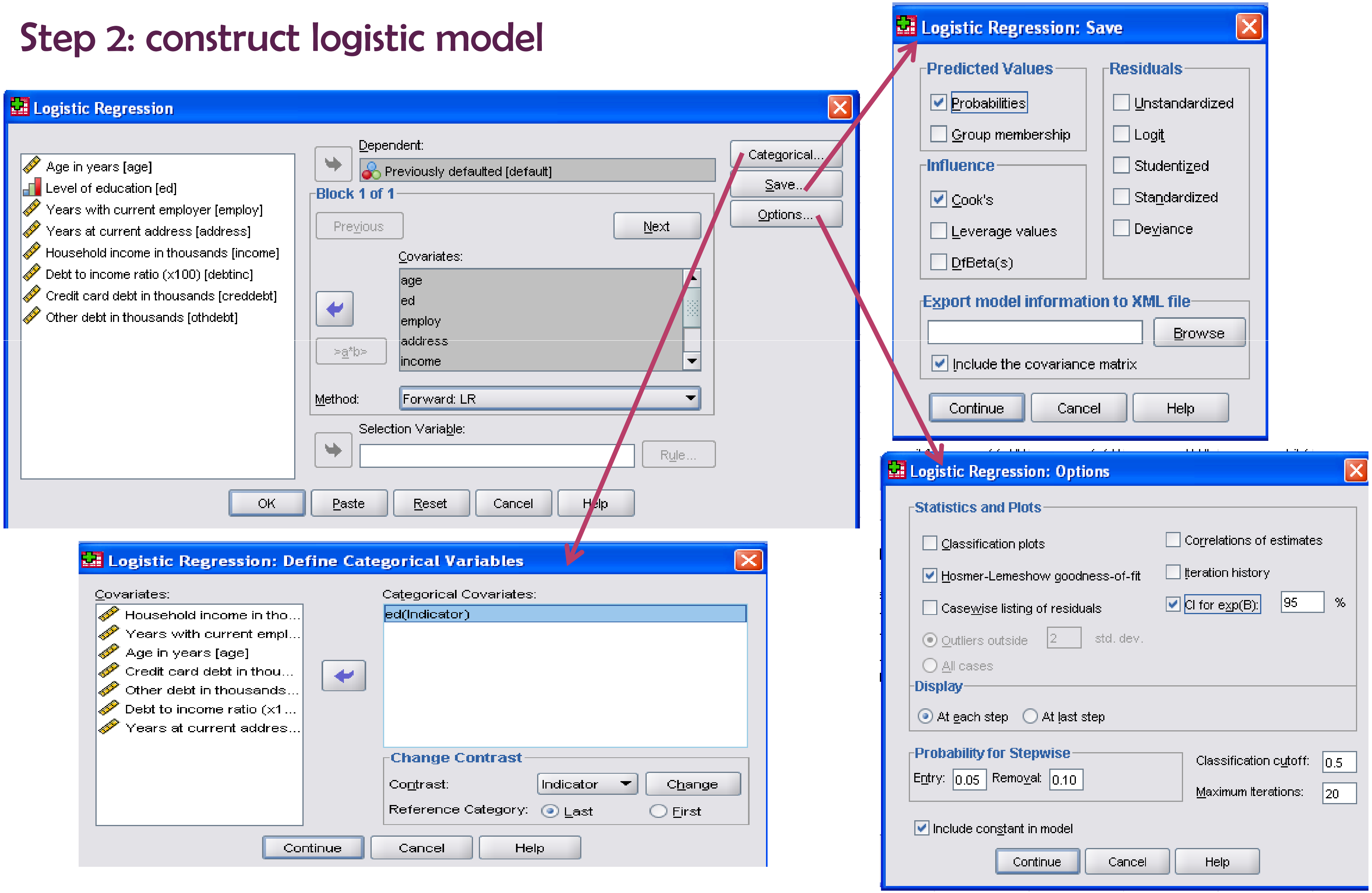
Data

Select Cases



BINARY LOGISTIC MODELS

Step 2: construct logistic model



BINARY LOGISTIC MODELS

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	701.429[a]	.137	.200
2	631.083[b]	.219	.321
3	575.636[b]	.279	.408
4	556.732[c]	.298	.436

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.160	8	.924
2	4.158	8	.843
3	6.418	8	.600
4	8.556	8	.381

Classification Table[a]

Observed			Predicted: Previously defaulted No	Predicted: Previously defaulted Yes	Percentage Correct
Step 1	Previously defaulted	No	490	27	94.8
		Yes	137	46	25.1
		Overall Percentage			76.6
Step 2	Previously defaulted	No	481	36	93.0
		Yes	110	73	39.9
		Overall Percentage			79.1
Step 3	Previously defaulted	No	477	40	92.3
		Yes	99	84	45.9
		Overall Percentage			80.1
Step 4	Previously defaulted	No	478	39	92.5
		Yes	91	92	50.3
		Overall Percentage			81.4

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	95.0% C.I.for EXP(B) Upper
Step 1[a]	debtinc	.132	.014	85.377	1	.000	1.141	1.109	1.173
	Constant	-2.531	.195	168.524	1	.000	.080		
Step 2[b]	employ	-.141	.019	53.755	1	.000	.868	.836	.902
	debtinc	.145	.016	87.231	1	.000	1.156	1.122	1.192
	Constant	-1.693	.219	59.771	1	.000	.184		
Step 3[c]	employ	-.244	.027	80.262	1	.000	.783	.743	.826
	debtinc	.088	.018	23.328	1	.000	1.092	1.053	1.131
	creddebt	.503	.081	38.652	1	.000	1.653	1.411	1.937
	Constant	-1.227	.231	28.144	1	.000	.293		
Step 4[d]	employ	-.243	.028	74.761	1	.000	.785	.743	.829
	address	-.081	.020	17.183	1	.000	.922	.887	.958
	debtinc	.088	.019	22.659	1	.000	1.092	1.053	1.133
	creddebt	.573	.087	43.109	1	.000	1.774	1.495	2.104
	Constant	-.791	.252	9.890	1	.002	.453		

a. Variable(s) entered on step 1: debtinc.

b. Variable(s) entered on step 2: employ.

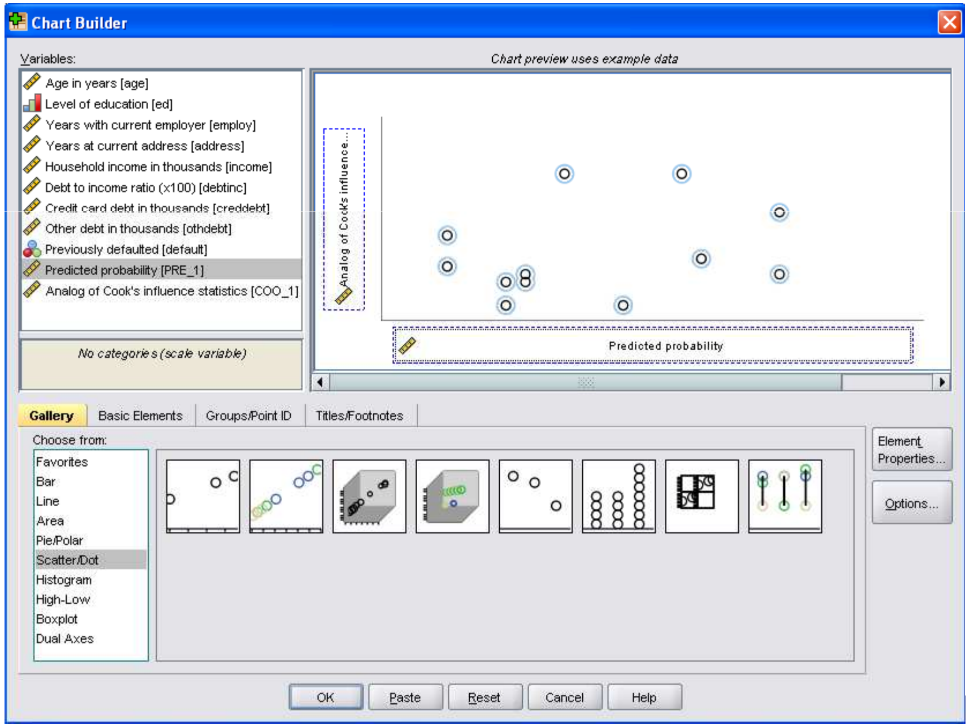
c. Variable(s) entered on step 3: creddebt.

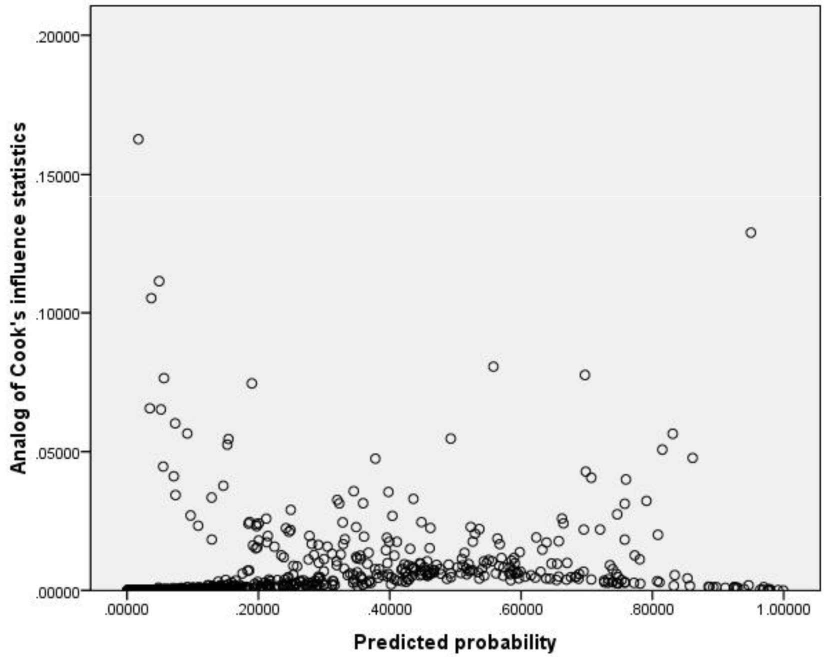
d. Variable(s) entered on step 4: address.

BINARY LOGISTIC MODELS

Step 3: identify possible outlying obs

Scatter plot: predicted probability vs Cook's D statistics

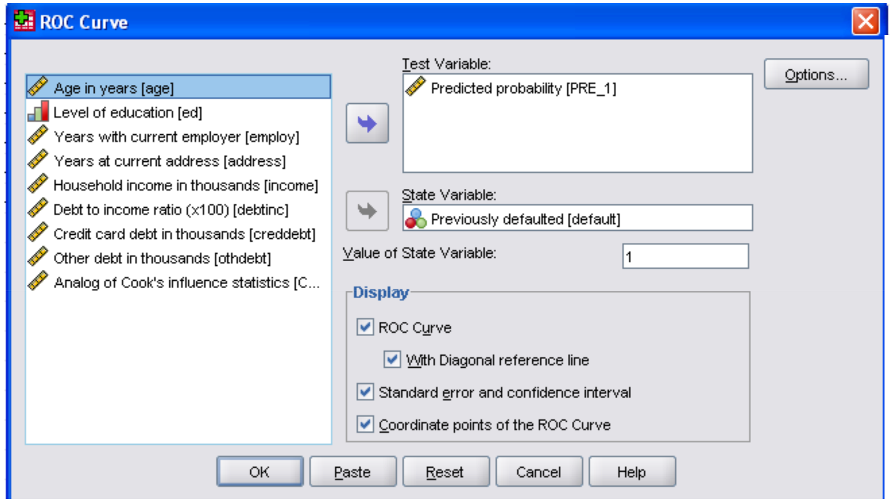




BINARY LOGISTIC MODELS

Step 4: draw ROC curve and find the optimal cut-off point

Analyze
ROC Curve



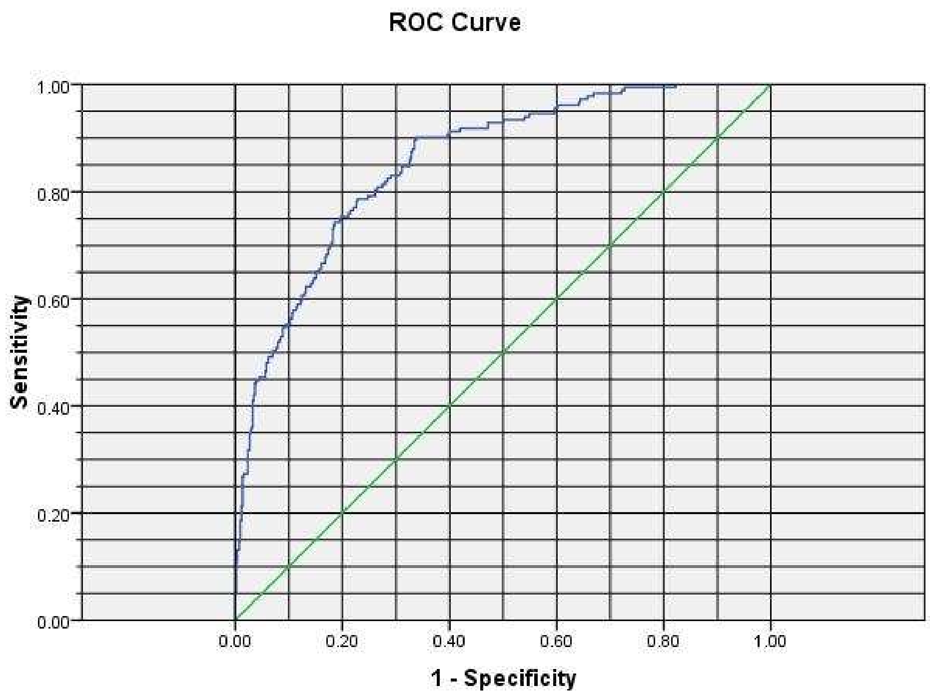
Area Under the Curve

Test Result Variable(s):Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.856	.016	.000	.825	.886

a. Under the nonparametric assumption

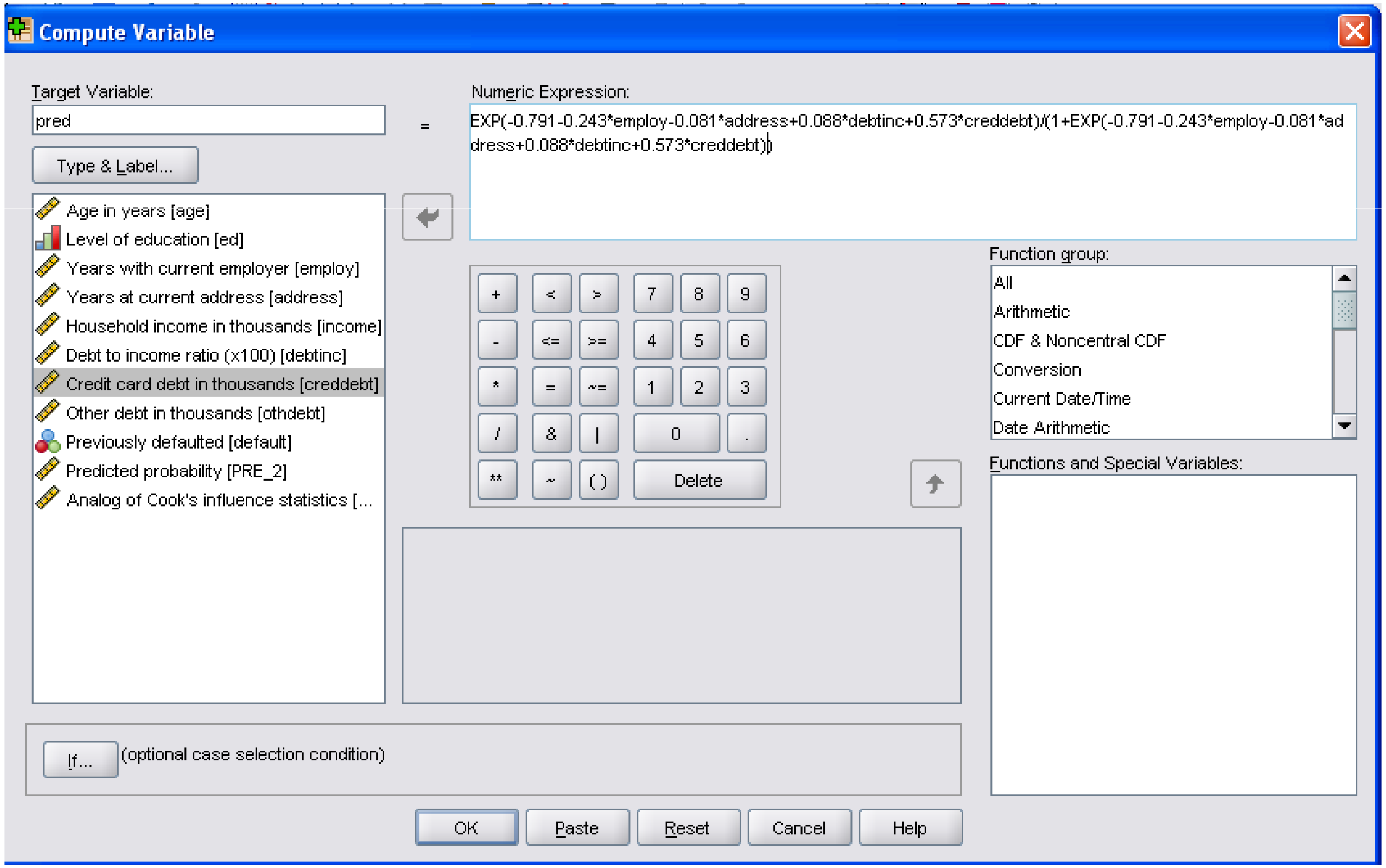
b. Null hypothesis: true area = 0.5



BINARY LOGISTIC MODELS

Step 5: predict credit risk

Select the rest 150 obs and compute predicted probability with model coefficients

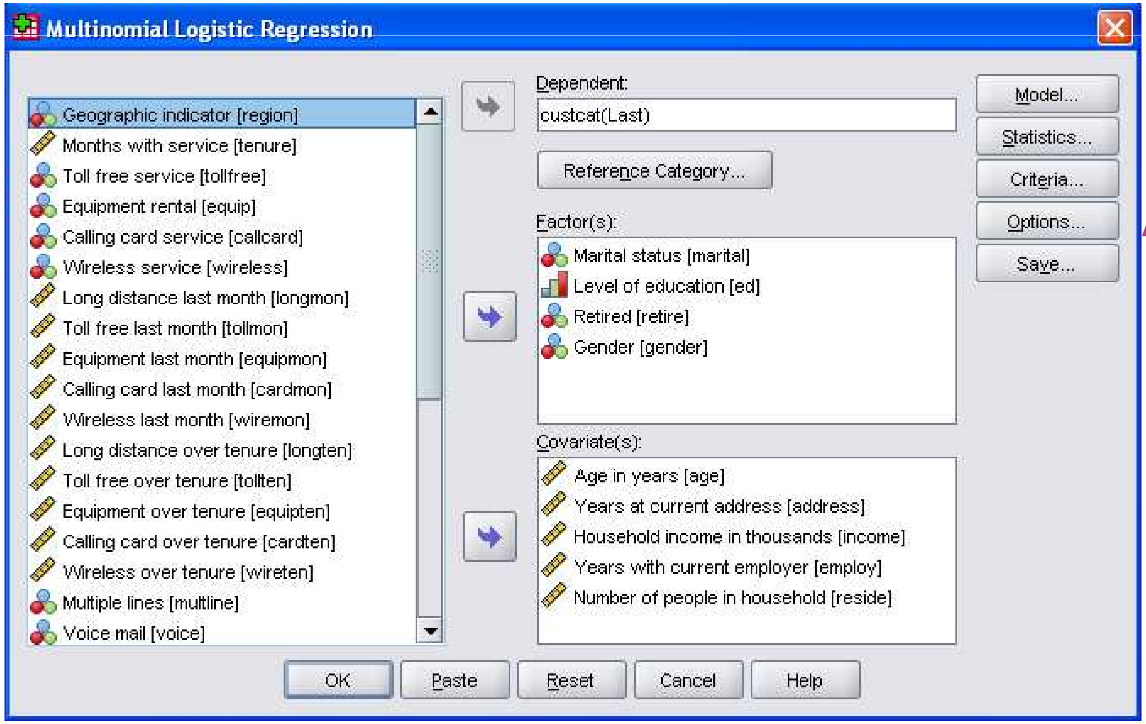


MULTINOMIAL LOGISTIC MODELS

Using Multinomial Logistic Regression to Classify Telecommunications Customers

E.g.: *telco.sav*

A telecommunications provider has segmented its customer base by service usage patterns, categorizing the customers into four groups. If demographic data can be used to predict group membership, you can customize offers for individual prospective customers.

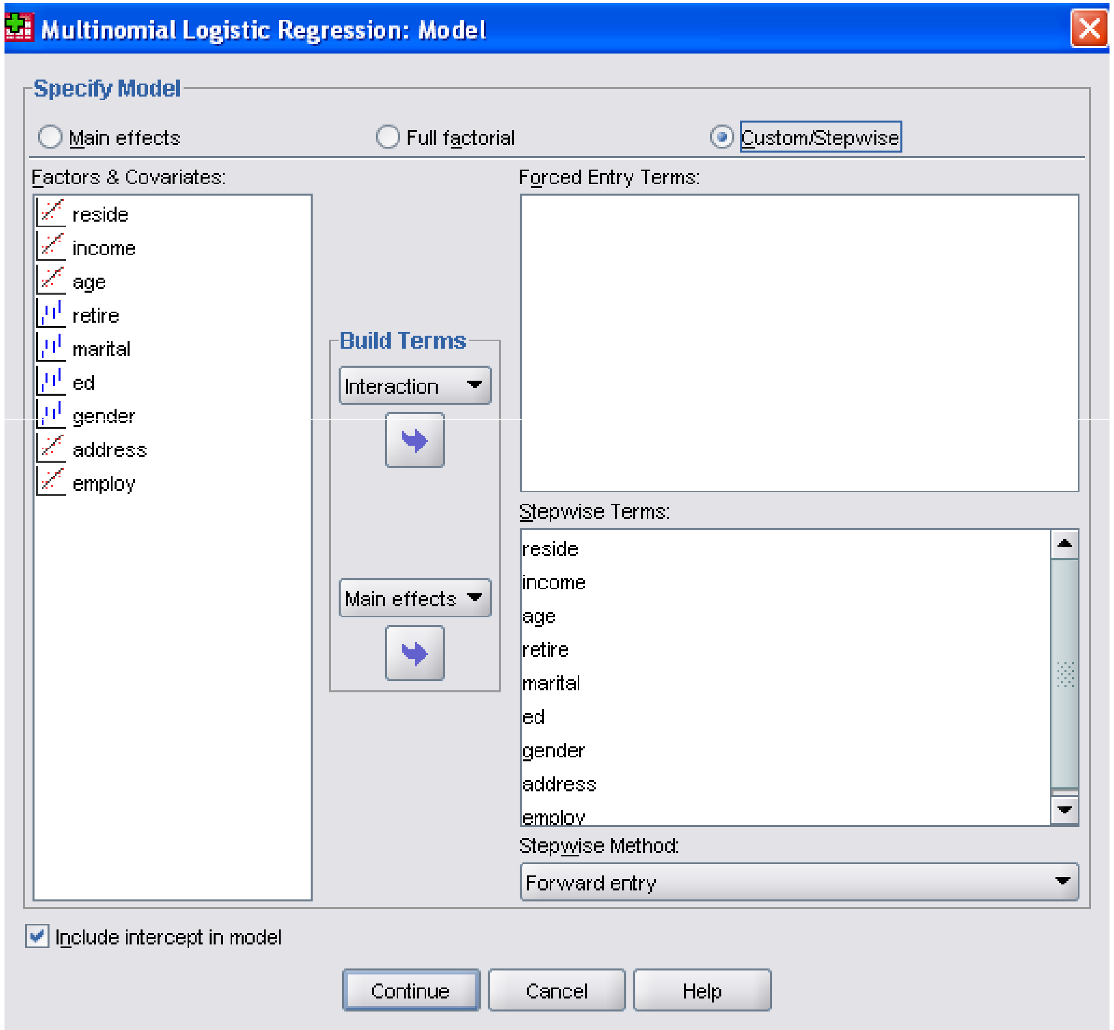


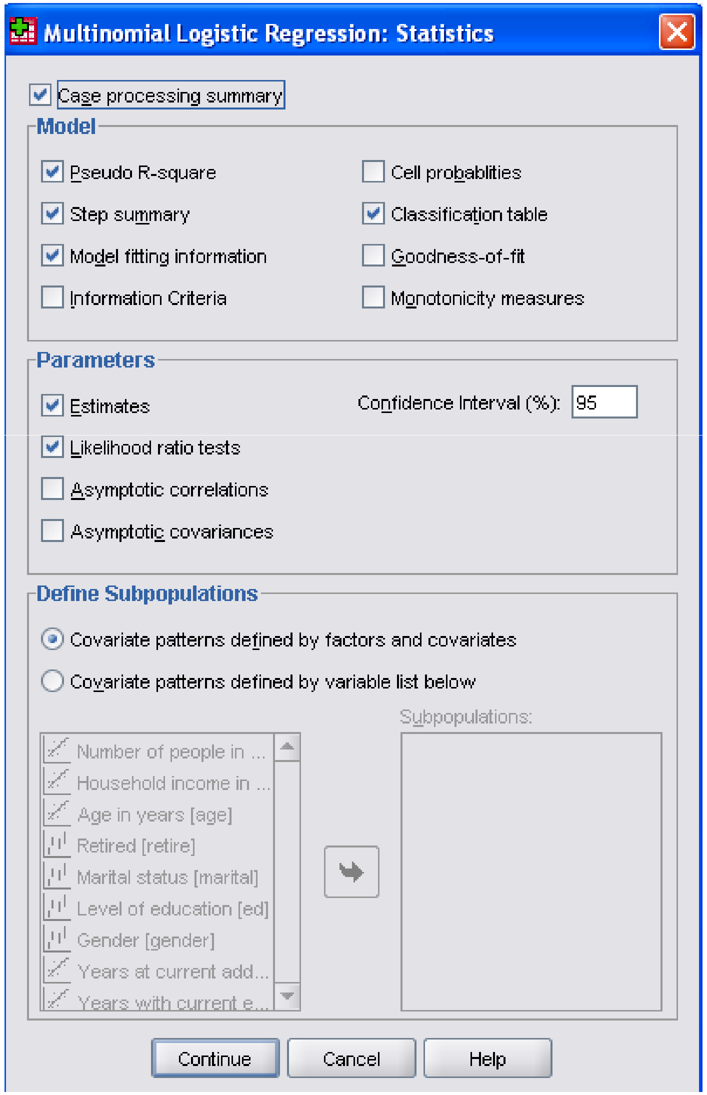
Analyze

Regression

Multinomial Logistic

MULTINOMIAL LOGISTIC MODELS





MULTINOMIAL LOGISTIC MODELS

Parameter Estimates

Customer category[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Basic service	Intercept	-.181	.431	.176	1	.675			
	reside	-.258	.068	14.418	1	.000	.773	.677	.883
	[ed=1]	3.762	.532	50.070	1	.000	43.047	15.183	122.044
	[ed=2]	1.959	.427	21.042	1	.000	7.095	3.072	16.390
	[ed=3]	1.453	.435	11.171	1	.001	4.276	1.824	10.025
	[ed=4]	.584	.425	1.893	1	.169	1.794	.780	4.123
	[ed=5]	0[b]	.	.	0	.	.	.	.
	address	-.022	.012	3.498	1	.061	.978	.956	1.001
	employ	-.042	.012	12.437	1	.000	.958	.936	.981
E-service	Intercept	-.132	.351	.141	1	.707			
	reside	-.110	.066	2.761	1	.097	.896	.787	1.020
	[ed=1]	1.592	.481	10.938	1	.001	4.913	1.913	12.619
	[ed=2]	.452	.345	1.717	1	.190	1.571	.799	3.087
	[ed=3]	.482	.345	1.948	1	.163	1.620	.823	3.188
	[ed=4]	-.092	.326	.080	1	.778	.912	.481	1.728
	[ed=5]	0[b]	.	.	0	.	.	.	.
	address	.015	.011	1.860	1	.173	1.015	.993	1.037
	employ	-.016	.011	1.969	1	.161	.984	.962	1.006
Plus service	Intercept	-1.732	.572	9.173	1	.002			
	reside	-.173	.067	6.583	1	.010	.841	.737	.960
	[ed=1]	4.318	.642	45.173	1	.000	75.032	21.301	264.298
	[ed=2]	2.678	.562	22.730	1	.000	14.554	4.840	43.763
	[ed=3]	2.126	.569	13.952	1	.000	8.381	2.747	25.573
	[ed=4]	1.049	.569	3.399	1	.065	2.855	.936	8.708
	[ed=5]	0[b]	.	.	0	.	.	.	.
	address	.000	.011	.001	1	.980	1.000	.979	1.021
	employ	.009	.011	.682	1	.409	1.009	.988	1.031

a. The reference category is: Total service.

b. This parameter is set to zero because it is redundant.

ORDINAL REGRESSION

Using Ordinal Regression to Build a Credit Scoring Model

$$g(\Pr(Y \leq i \mid X)) = \alpha_i + \beta' X, \qquad i = 1, \ldots\ldots, k$$

Link function. The link function is a transformation of the cumulative probabilities that allows estimation of the model. Five link functions are available, summarized in the following table.

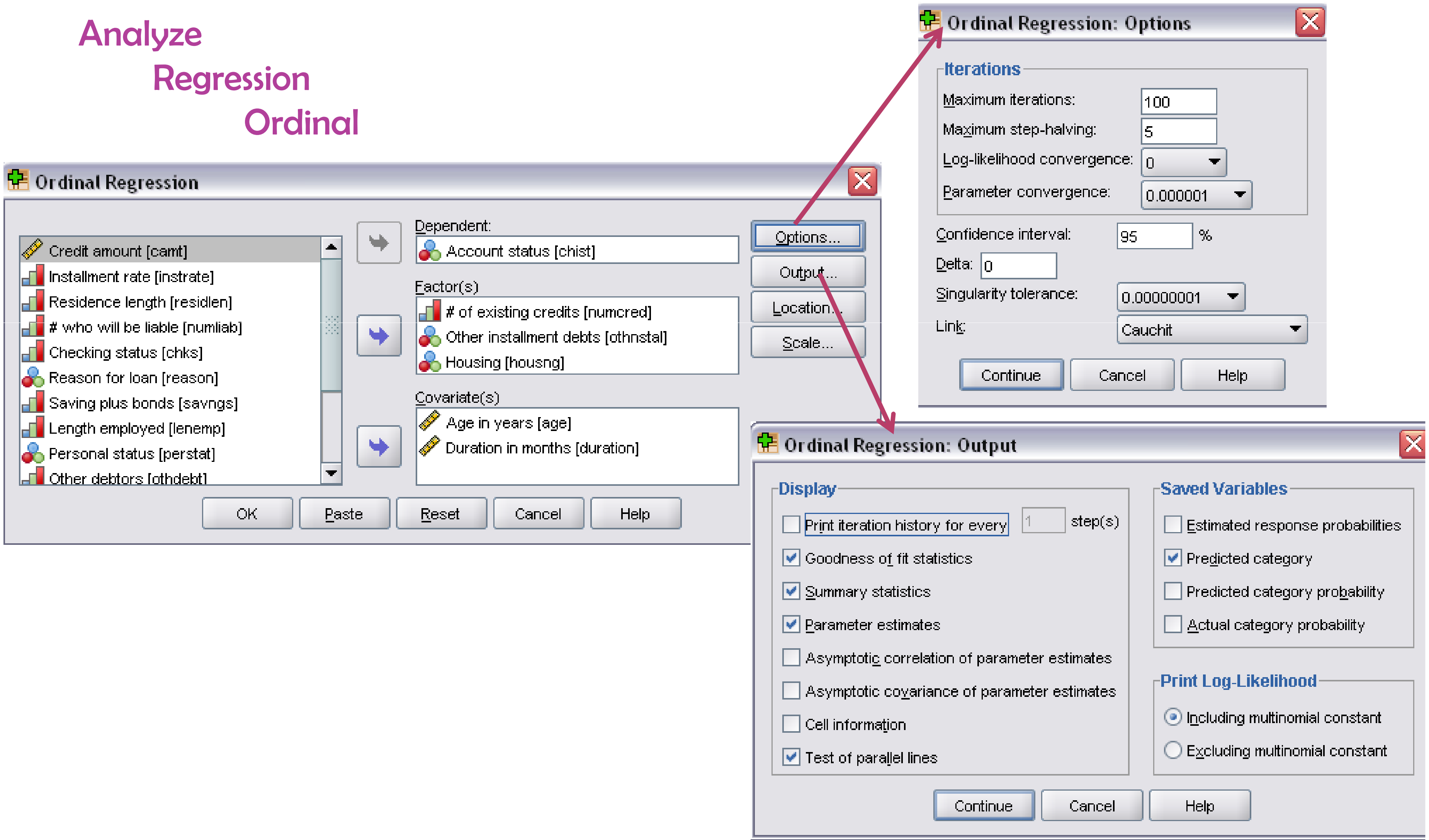
Function	Form	Typical application
Logit	$\log(\xi / (1-\xi))$	Evenly distributed categories
Complementary log-log	$\log(-\log(1-\xi))$	Higher categories more probable
Negative log-log	$-\log(-\log(\xi))$	Lower categories more probable
Probit	$\Phi^{-1}(\xi)$	Latent variable is normally distributed
Cauchit (inverse Cauchy)	$\tan(\pi(\xi-0.5))$	Latent variable has many extreme values

E.g.: ***german_credit.sav***

A creditor wants to be able to determine whether an applicant is a good credit risk, given various financial and personal characteristics. From their customer database, the creditor (dependent) variable is account status, with five ordinal levels: no debt history, no current debt, debt payments current, debt payments past due, and critical account. Potential predictors consist of various financial and personal characteristics of applicants, including age, number of credits at the bank, housing type, checking account status, and so on.

ORDINAL REGRESSION

Analyze
Regression
Ordinal



ORDINAL REGRESSION

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	2249.888			
Final	1790.028	459.860	9	.000

Link function: Cauchit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	3467.625	3131	.000
Deviance	1690.392	3131	1.000

Link function: Cauchit.

Pseudo R-Square

Cox and Snell	.369
Nagelkerke	.407
McFadden	.194

Link function: Cauchit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[chist = 1.00]	-9.356	1.432	42.689	1	.000	-12.163	-6.549
	[chist = 2.00]	-5.232	.989	28.010	1	.000	-7.169	-3.294
	[chist = 3.00]	-.552	.933	.350	1	.554	-2.382	1.277
	[chist = 4.00]	.432	.929	.216	1	.642	-1.389	2.254
Location	age	.016	.008	3.393	1	.065	.000	.032
	duration	-.013	.007	3.012	1	.083	-.028	.002
	[numcred=1.00]	-2.616	.729	12.867	1	.000	-4.046	-1.187
	[numcred=2.00]	.817	.702	1.353	1	.245	-.560	2.193
	[numcred=3.00]	2.002	.940	4.533	1	.033	.159	3.845
	[numcred=4.00]	0[a]	.	.	0	.	.	.
	[othnstal=1.00]	-1.257	.237	28.113	1	.000	-1.721	-.792
	[othnstal=2.00]	-1.031	.355	8.424	1	.004	-1.727	-.335
	[othnstal=3.00]	0[a]	.	.	0	.	.	.
	[housng=1.00]	-.275	.377	.533	1	.465	-1.014	.464
	[housng=2.00]	.049	.320	.024	1	.878	-.577	.676
	[housng=3.00]	0[a]	.	.	0	.	.	.

Link function: Cauchit.

a. This parameter is set to zero because it is redundant.

LOGISTIC MODELS

- Hays, W. L. 1981. Statistics, 3rd ed. New York: Holt, Rinehart, and Winston.
- McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall.
- Cox, D. R., and E. J. Snell. 1989. The Analysis of Binary Data, 2nd ed. London: Chapman and Hall.
- Hosmer, D. W., and S. Lemeshow. 2000. Applied Logistic Regression, 2nd ed. New York: John Wiley and Sons.
- Kleinbaum, D. G. 1994. Logistic Regression: A Self-Learning Text. New York: Springer-Verlag.
- Norusis, M. 2004. SPSS 13.0 Advanced Statistical Procedures Companion. Upper Saddle-River, N.J.: Prentice Hall, Inc..

END

Advanced Models

30 May Friday 10am-12pm

Training Room @ Library Level 5

- Curve Estimation
- Non-linear Regression
- Survival Analysis
- Linear Mixed Model
- Time-series Data Analysis