



Understanding Data Quickly and Easily



Understanding Data Quickly and Easily: Agenda

- Case Study – Understanding Data in ~ 270 seconds
- General Overview of IBM SPSS Statistics Standard
- To Understand Data Quickly and Easily
 - Save Time Getting to the Data
 - Leverage Metadata
 - Start with a Picture
 - Use Words and Code
 - Quick Access to Analytics
 - Automate Tasks
- Case Study – Who Will Churn and When?
- Recap & Questions

Understanding Data Quickly and Easily: A Case Study in ~ 270 Seconds

- Motivating Question

- Is age or income level an risk indicator at my bank? If not, what are the top two or three attributes about my customers that would help us understand (and predict) default on a loan?

- Inputs

- Age
 - Level of Education
 - Years at Current Employer
 - Years at Current Address
 - Income
 - Debt to Income Ratio
 - Credit Card Debt
 - Other Debt (e.g. Auto Loan)

Understanding Data Quickly and Easily: A General Overview of IBM SPSS Statistics Standard

- Spreadsheet-like Look and Feel
- General Environment for Predictive Analytics and Statistical Analysis
- Well-Suited for Ad-hoc Analysis and Hypothesis Testing (Top Down Analysis)

*stroke_survival.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

14 : gender 0 Visible: 42 of 42 Variables

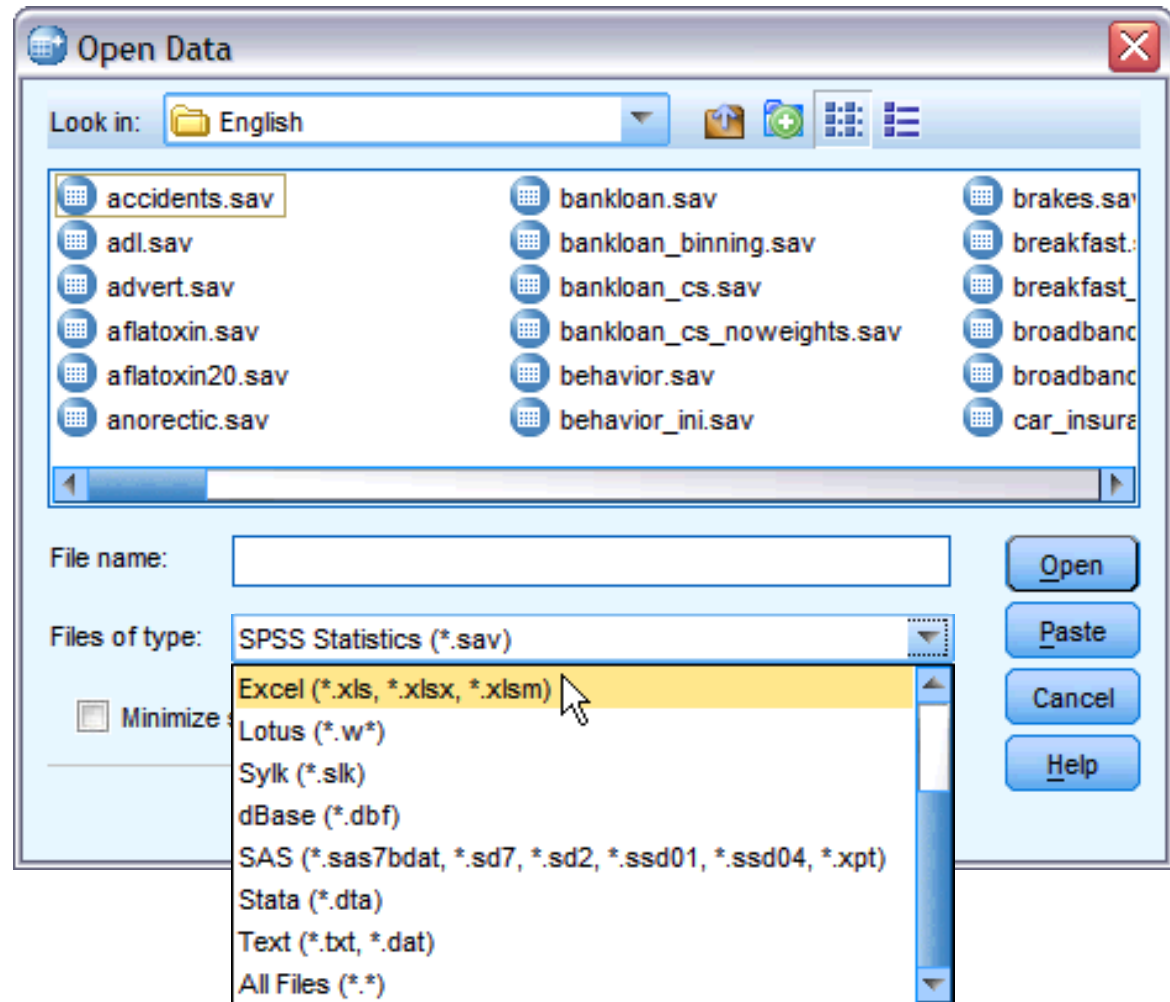
	patid	physid	age	agecat	gender	active	obesity	diabetes
14	5357069859	822229	81	75+	Male	No	Yes	No
15	5132742071	297378		45-54	Male	Yes	Yes	Yes
16	2660586207	297378	53	45-54	Male	No	Yes	No
17	5408312498	799998	62	55-64	Female	No	No	No
18	9069087682	297378	64	55-64	Male	Yes	No	No
19	8173197592	799998	58	55-64	Female	No	No	No
20	8808732689	822229	83	75+	Male	Yes	No	No
21	5666440246	822229	67	65-74	Female	Yes	Yes	No
22	8483212117	822229	59	55-64	Male	No	No	No
23	0674952107	297378	60	55-64	Male	Yes	No	No
24	9726045305	799998	71	65-74	Female	No	No	No
25	6197593400	297378	52	45-54	Male	Yes	No	No
26	2088521482	799998	52	45-54	Male	No	Yes	No
27	4775117168	822229	54	45-54	Male	Yes	No	No

Data View Variable View

IBM SPSS Statistics Processor is ready

Understanding Data Quickly and Easily: Save Time Getting to the Data

- Use Analytics Tools that can Open (and Export to) a Variety of Data Formats
- Tap Directly into Databases when Possible
- Utilize “Wizards”



Understanding Data Quickly and Easily: Leverage Metadata

- Metadata Provides Information about the Data that Accelerates Analyses

bankloan.sav [DataSet2] - IBM SPSS Statistics Data Editor

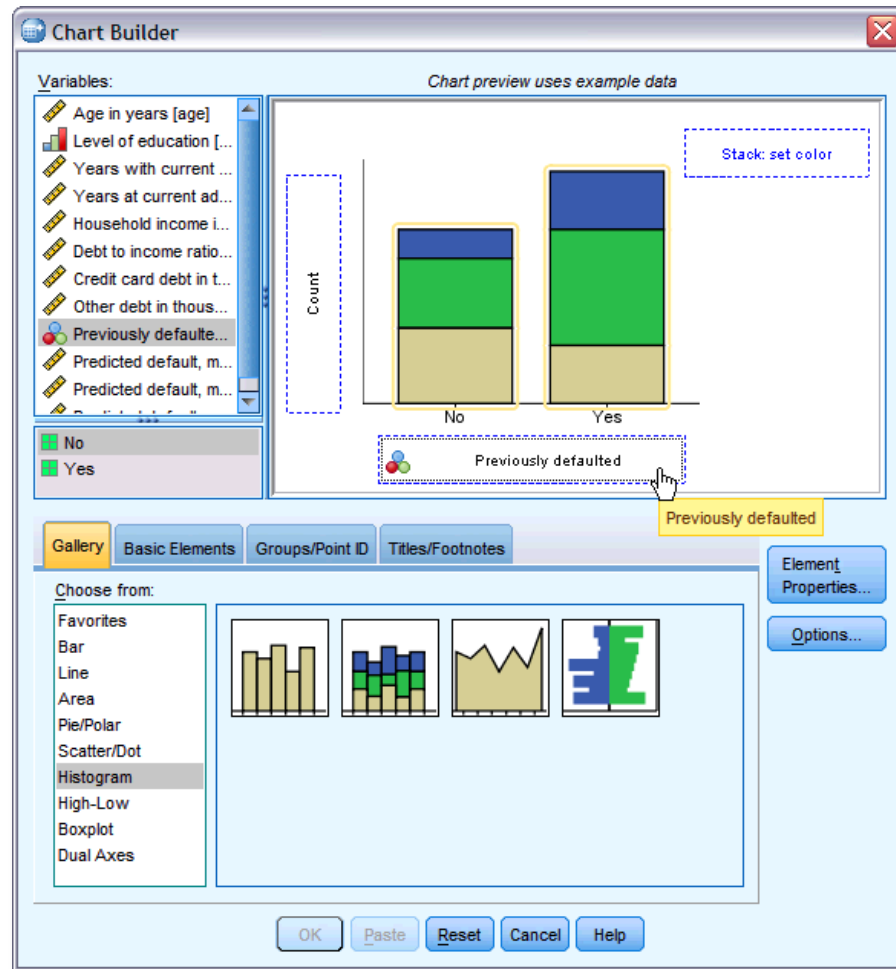
	Name	Type	Width	Decim...	Label	Values	Missing	Columns	Align	Measure	Role
1	age	Numeric	4	0	Age in years	None	None	4	Right	Scale	Input
2	ed	Numeric	4	0	Level of education	{1, Did not c...	None	15	Right	Ordinal	Input
3	employ	Numeric	4	0	Years with current employer	None	None	6	Right	Scale	Input
4	address	Numeric	4	0	Years at current address	None	None	7	Right	Scale	Input
5	income	Numeric	8	2	Household income in thousands	None	None	8	Right	Scale	Input
6	debtinc	Numeric	8	2	Debt to income ratio (x100)	None	None	8	Right	Scale	Input
7	creddebt	Numeric	8	2	Credit card debt in thousands	None	None	8	Right	Scale	Input
8	othdebt	Numeric	8	2	Other debt in thousands	None	None	8	Right	Scale	Input
9	default	Numeric	4	0	Previously defaulted	{0, No}...	None	7	Right	Nominal	Target
10	preddef1	Numeric	11	5	Predicted default, model 1	None	None	11	Right	Scale	None
11	preddef2	Numeric	11	5	Predicted default, model 2	None	None	11	Right	Scale	None
12	preddef3	Numeric	11	5	Predicted default, model 3	None	None	11	Right	Scale	None
13											

Data View Variable View

IBM SPSS Statistics Processor is ready

Understanding Data Quickly and Easily: Start with a Picture

- Dragging and Dropping using Variable Names Speeds Time-to-Results
- Provides a Quick and Holistic “Understanding” of the Data



Understanding Data Quickly and Easily: Use Words and Code

- Creating New Variables in Words
- Sorting and Filtering using Variable Names

The screenshot displays the PASW Statistics Data Editor interface. The main window shows a dataset named 'bankloan.sav [DataSet2]' with a table of variables including 'age', 'debtinc', 'creddebt', and 'othdebt'. The 'Compute Variable...' menu is open, showing options like 'Count Values within Cases...', 'Recode into Same Variables...', and 'Recode into Different Variables...'. The 'Compute Variable' dialog box is also visible, showing the 'Target Variable' field, a 'Numeric Expression' field, and a list of functions and special variables.

Compute Variable Dialog Box:

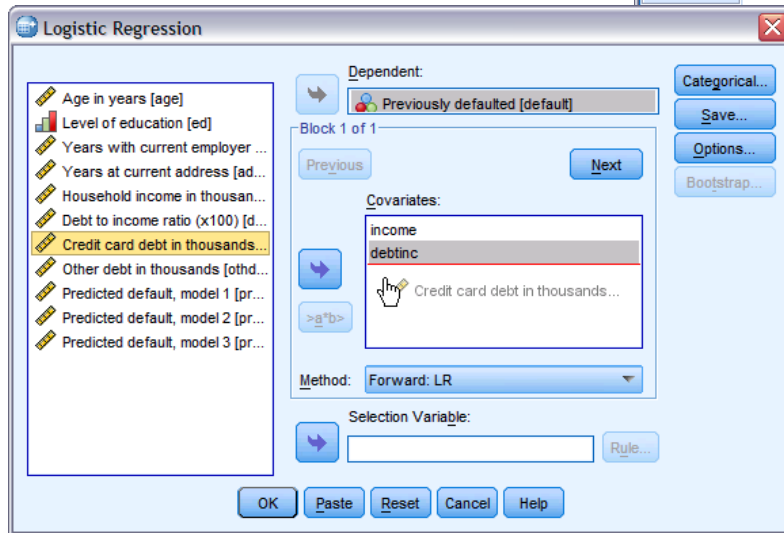
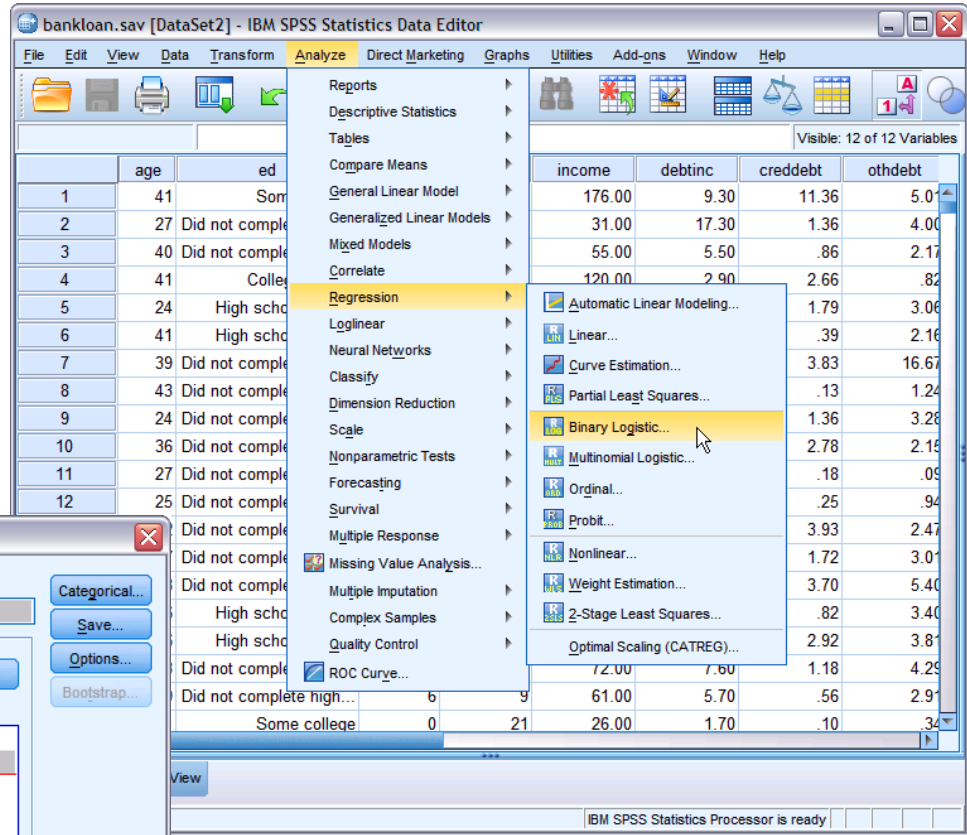
- Target Variable: [Empty]
- Numeric Expression: [Empty]
- Function group: [All]
- Functions and Special Variables: [Empty]
- if... (optional case selection condition): [Empty]

Main Data Editor Table (Visible):

	age	debtinc	creddebt	othdebt
1	41	9.30	11.36	5.01
2	27	17.30	1.36	4.00
3	40	5.50	.86	2.17
4	41	2.90	2.66	.82
5	24	17.30	1.79	3.06
...
39	41	10.20	.39	2.16
43	27	3.60	.13	1.24
24	41	24.40	1.36	3.28
36	27	19.70	2.78	2.15
27	41	1.70	.18	.09
25	24	5.20	.25	.94
52	Did not complete h...	24	14	64.00
		10.00	3.93	2.47

Understanding Data Quickly and Easily: Quick Access to Analytics

- Menus Provide Quick Access to and Logical Organization of Analytics
- Dialog Boxes Provide Intuitive Set-up of Analysis
- Both are Configurable



Understanding Data Quickly and Easily: Automate Tasks (or Use Words and Code Revisited)

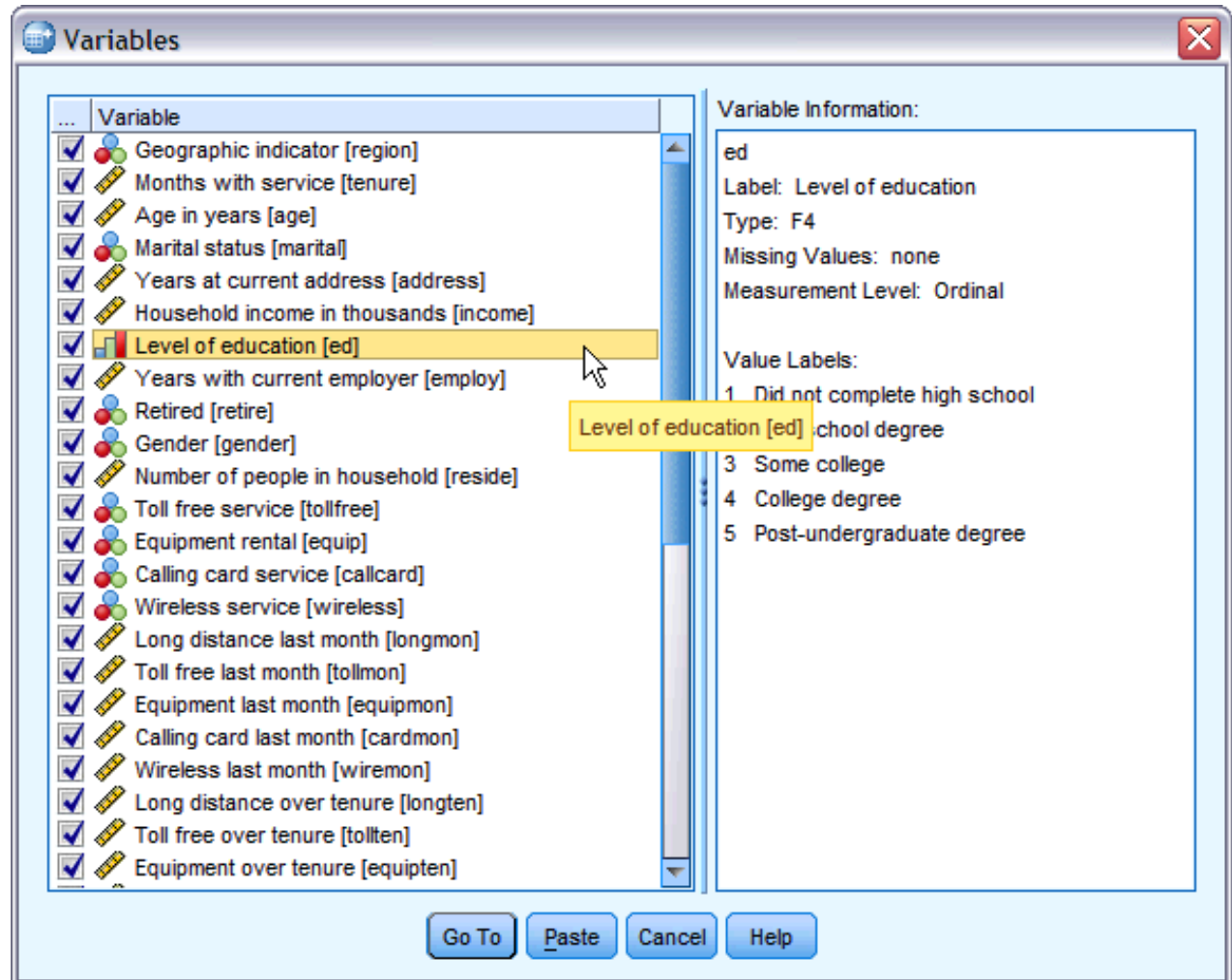
- Command Syntax Helps Automate Routine Tasks
- Enables a Focus on Analysis

```

1  DATASET ACTIVATE DataSet2.
2
3  LOGISTIC REGRESSION VARIABLES default
4  /SELECT=validate EQ 1
5  /METHOD=FSTEP(LR) age ed employ address income debttinc creddebt othdebt
6
7  /CONTRAST (ed)=Indicator
8  /SAVE=PRED COOK SRESID
9  /PRINT=GOODFIT
10 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
    
```

Understanding Data Quickly and Easily: Concluding Example: Who is Likely to Churn and When?

- Data from Telecommunications Service Provider
- Goal: Understand the Who and When of Customer Churn



Recap & Questions

- Case Study – Understanding Data in ~ 270 seconds
- General Overview of IBM SPSS Statistics Standard
- To Understand Data Quickly and Easily
 - Save Time Getting to the Data
 - Leverage Metadata
 - Start with a Picture
 - Use Words and Code
 - Quick Access to Analytics
 - Automate Tasks
- Case Study – Who Will Churn and When?
- Recap & Questions