

## Exercise No. 5: Credit Status of Consumers

New Stata Commands		Old Commands Reviewed	
label values	label define	graph box <i>varname</i>	graph twoway <i>varname</i>
tabulate	generate	sort	graph twoway histogram
tabstat	“discrete” option		histogram
replace	ttest		

*{Be sure to open a log and keep it for your records. You'll attach it to this exercise.}*

Download the Stata data set “**credit.dta**” from the course web site.

### Introduction

Until now in the course we've been learning how to do various things in Stata without paying too much attention to the *substance* of the data with which we're working or the *purpose* of the data analysis itself. This exercise seeks to help you consolidate what you've learned so far and to give you the opportunity to try your hand at actually using the techniques you've learned.

We will be using mainly commands that you have seen before, either in the previous computer exercises or in lectures. Remember, you have the Stata Help file to refer to, along with Hamilton's *Statistics with Stata* and the handouts on Stata syntax and Stata commands that I've given you previously. Use *all* those resources to help you complete the exercise. The point of this exercise is to get you started working with Stata on your own. I've not provided a lot of guidance, because a big part of doing statistical work on the computer is figuring out how to do things. Don't wait until the night before it's due or you may be sorry.

### The Scenario:

You are newly appointed as the loan officer of a small bank and you're required to write a memo to your boss describing the characteristics of customers who apply for loans. You acquire a random sample of 100 loan applicants upon which to base your memo's discussion. Here's a description of the data set (**credit.dta**):

```
Contains data from \your_path\credit.dta
  obs:      100 (max=      1,125)
  vars:      7 (max=      99)      30 Sep 2002 13:26
  width:    13 (max=     200)
```

---

variable name	storage type	display format	value label	variable label
mdr	byte	%8.0g		Number of Derogatory Reports
acc	byte	%8.0g		Credit card applications accepted (1=yes)
age	byte	%8.0g		Age in years + 12ths of a year
income	float	%9.0g		Income, divided by 10,000
avgexp	float	%9.0g		Avg. monthly credit card expenditure
ownrent	byte	%8.0g		OwnRent, individual owns (1) or rents (0) home
selfempl	byte	%8.0g		Self employed (1=yes, 0=no)

---

## The Memo

Use the regular Word exercise template for this exercise; however, after you have filled in the heading as usual, the remainder of the exercise should look like a regular business memorandum with a heading (just below the exercise heading) that looks something like:

Memorandum to: Boone Turchi  
From: {your name}  
Subject: Analysis of Loan Customers  
Date

Of course, you can make the memo heading as fancy as you'd like (perhaps using one of Word's memo formats). The rest of the memo should be exactly like a memo you'd really write to your boss. It should contain the statistics and graphs requested below *and* it should contain your written discussion of the statistics that you present. Remember, you're putting your best "foot" forward here; you want your boss to think you have good analytical, organizational and presentation skills. So, use *Statistics with Stata* and any other help (e.g., Word's Help file, etc.) to put together a good looking and cogent report.

The memo should contain the following sections:

1. Description of the bank's loan applicants including graphics showing:

- ◆ Age distribution of loan customers (5 year age groups)
- ◆ Income distribution of loan customers (income categories of \$10,000 each)
- ◆ Proportion of owners and renters (use the "discrete" option with histogram {see Hamilton})
- ◆ Proportion of self-employed versus employees (use the "discrete" option with histogram {see Hamilton})

Use histograms to show each of these variables and provide a short discussion with each graph. (Figure out how to label values "own" and "rent" for the "ownrent" variable. Tip: You must first use the *label values* command and then use the *label define* command. Carry out the same value labeling procedure for the "selfempl" variable)

2. Credit History

- ◆ Use *tabulate* to produce a frequency tabulation of the number of derogatory credit reports for each credit customer.
- ◆ Then also include a frequency histogram (appropriately titled and labeled) showing the frequency distribution of negative credit reports. (Hint: check out the "discrete" option to get the bars to line up with the x-axis labels properly.)
- ◆ Use the *tabulate* command to produce a two-way table comparing the number of derogatory reports to whether or not a credit card application was accepted.

Again, provide a brief discussion of your results so that your (somewhat dimwitted) boss can have the results explained to him.

3. Marketing

Finally, present some analysis of the customers to see if their credit use makes them particularly good customers for other kinds of loans:

- ◆ Create a two-way scatter plot showing the relationship between average monthly credit card expenditures and the age of the individual *for those individuals who actually use credit cards. (i.e., ignore folks who don't use credit cards)* Discuss whether you think the plot shows that credit card use and age are related in this sample. How do you justify your conclusions?
- ◆ Create a two-way scatter plot showing the relationship between income and credit card use, *again for credit card users only.* Discuss whether income seems to be related to the use of credit cards and, hence, whether higher (or lower) income people might be good targets for the bank's other loan products.
- ◆ Create a new variable "user" (using the *generate* and *replace* commands) that takes on two values: 0 if not a credit card user and 1 if a credit card user. How would you figure this out from the data at your disposal? Then create a table showing how subjects' income differs depending upon whether or not they are credit card users. Use the command *tabstat* in conjunction with the variable "user" in order to create this table which should contain information on the mean income, standard deviation, median, interquartile range, and the number of people in each credit card group. Finally, create a boxplot showing how income distributions differ according to whether or not a person is a credit card user. (Remember to sort the data by "user" before trying to do the boxplot.) What would you say is the difference in income between credit card users and nonusers?
- ◆ Now, report the results of a hypothesis test in which you test the null hypothesis that credit card users have the same average income as nonusers against all three alternatives (one- and two-tailed alternatives). Describe how you did the significance tests and what assumptions you made, showing the results both of the test *and* the checks you made about the underlying assumptions. Describe your results and compare them with the results from the boxplot exercise you just completed. Do the results support each other or not? Why? Make your discussion clear enough so that even the average idiotic boss could understand it.

Wrap up the memo with a short paragraph in which you summarize your findings and flatter your boss for his/her infinite wisdom, good looks, etc.

- ◆ Finally, attach your log to the back of the memo so that we can see how you went about your various activities.