

## Extra data analysis problems

CHS Casebook:

- (1) “Predicting the sales and airplay of popular music” (p. 23)

Follow the guidelines given in the case, but do **not** use the `rock.mtp` data set from the `chs` directory. Use instead the data set `rock2.mtp` from the `js` directory, which contains data corresponding to the week ending June 28, 1997.

- (2) “Another look at the ‘Old Faithful’ geyser and adoption visas” (p. 24)

Follow the guidelines given in the case, but do **not** use the data sets from the `chs` directory. Use instead the two variables in the data set `stocks97.mtp` called `Weekly price change` and `Weekly sales volume`. These are the weekly percentage change in stock price and weekly sales volume, respectively, for a sample of stocks in the week ending June 27, 1997.

(Note: the anchor position of a histogram *must* be less than or equal to the smallest value in the data set)

- (3) Consider again the stock data from problem (2). You can see that each stock given comes from one of three stock markets: The New York Stock Exchange (NYSE), the NASDAQ Over the Counter (OTC) market, or the American Stock Exchange (ASE). Does it appear that the weekly stock price change differs across different stock markets? What about P/E ratio? What about the weekly sales volume? How might you account for the patterns that you see?
- (4) The following table is based on data from the Bureau of Labor Statistics. It gives the deseasonalized unemployment rate by month for 1995 and 1996 in New York State and New Jersey (note: “months” are actually consecutive four-week periods, which accounts for the existence of 13 months in the year).

<i>Month</i>	<b>New York</b>			<b>New Jersey</b>		
	<i>1995</i>	<i>1996</i>	<i>1997</i>	<i>1995</i>	<i>1996</i>	<i>1997</i>
M1	6.2	6.3	?	6.4	6.5	?
M2	6.3	6.3	?	6.4	6.3	?
M3	6.3	6.4	?	6.2	6.3	?
M4	6.3	6.4	?	6.3	6.3	?
M5	6.4	6.3	?	6.4	6.1	?
M6	6.3	6.3		6.5	6.1	
M7	6.3	6.3		6.5	6.1	
M8	6.5	6.1		6.5	6.1	
M9	6.5	6.1		6.4	6.2	
M10	6.4	6.1		6.4	6.2	
M11	6.3	6.2		6.5	6.1	
M12	6.3	6.2		6.4	6.1	
M13	6.3	6.2		6.4	6.2	

Describe what you think these data are saying. What are the interesting patterns that you see? Do you have any hypotheses to explain the observed patterns? In particular, try to create statistical graphics that highlight the patterns that you're trying to discuss (you can do this by hand, or using MINITAB). You will note that the entries for the first five months of 1997 are entered as “?” in the table. The data source actually gave those values; what do you think they were? That is, give a forecast for the deseasonalized unemployment rate for the first five months of 1997 for New York and New Jersey, respectively. How did you come up with your forecast?

CHS Casebook:

- (5) “Predicting the sales and airplay of popular music” (p. 23)

Follow the guidelines given in the case, but do **not** use the `rock.mtp` data set from the `chs` directory. Use instead the data set `rock98.mtp` from the `js` directory, which contains data corresponding to the week ending April 11, 1998.

- (6) “Another look at the ‘Old Faithful’ geyser and adoption visas” (p. 24)

Follow the guidelines given in the case, but do **not** use the data sets from the `chs` directory. Use instead the two variables in the data set `stocks98.mtp` called `Weekly price change %` and `Market capitalization`. These are the weekly per-

centage change in stock price and market capitalization in billions of dollars, respectively, for a sample of stocks in the week ending April 17, 1998.

(Note: the anchor position of a histogram *must* be less than or equal to the smallest value in the data set)

- (7) Consider again the stock data from problem (6). You can see that each stock given comes from one of three stock markets: The New York Stock Exchange (NYSE), the NASDAQ Over the Counter market, or the American Stock Exchange (AMEX). Does it appear that the weekly stock price change differs across different stock markets? What about P/E ratio? What about market capitalization? How might you account for the patterns that you see?
- (8) Consider one more time the stock data from problems (6) and (7). Calculate the average market capitalization for the stocks in each of the three stock exchanges separately, and calculate the average of these three numbers (call this A). Calculate also the average market capitalization for all of the stocks considered as one group (call this B). Does A equal B? That is, does the average of the averages equal the overall average? How do you account for this result?
- (9) The file `bwrank98.mpj` is a Minitab project file created by Lena Salam that contains information from the *Business Week* 1998 survey on business schools, which appeared in the October 19, 1998, issue. The file gives the following information for each of the top 25 schools: the overall rank given by *Business Week*, the rank given by the 1998 graduates of the school, the rank given by corporate recruiters, the average amount of outstanding loans of 1998 graduates in dollars (ouch!), the average number of job offers to 1998 graduates, the percentage of 1998 graduates with starting compensation (including bonus and other compensation) over \$100,000, and the percentage of 1998 graduates with job offers.

Investigate the relationships between the variables in this data set. You can look at any relationships that interest you; possibilities that occur to me include seeing how different factors relate to the overall ranking or how different factors relate to the percentage of graduates with compensation over \$100,000, but feel free to look at anything that seems interesting to you. Describe what you find, why the patterns

might be what they are, and what the implications of your findings are.

- (10) Moody's Investors Service is in the business of rating bonds for credit risk (that is, the risk that a bond will default and an investor will lose money as a result). Corporate bonds are rated on the scale Aaa, Aa, A, Baa, Ba, and B, in descending order of rating (and hence increasing order of Moody's perception of credit risk). Each year Moody's conducts a corporate bond default study, where they examine the actual default experience of bonds separated by Moody's rating. The data in the file `moodys.mpj` come from the March 1999 study, and were provided to me by C. Gus Harris. For each year from 1970 through 1984, all bonds of each rating class were followed for 15 years to see if they defaulted; for each year and rating class (a so-called bond cohort) the default rate is then recorded. Similarly, cohorts from 1970 through 1989 were followed for 10 years, and cohorts from 1970 through 1994 were followed for 5 years. Are default rates different, based on rating class? (Moody's certainly hopes so!) Is there evidence to suggest that Moody's scale is too fine (that is, there aren't important differences in default risk for some rating classes)? Are there differences in patterns for the 5, 10, and 15 year time periods?