

## Case-based Social Statistics II

CSSS 321

Professor: Mark S. Handcock

### Solutions to Homework 7

Due Friday, May 31, 2002

*Problems to be handed in:*

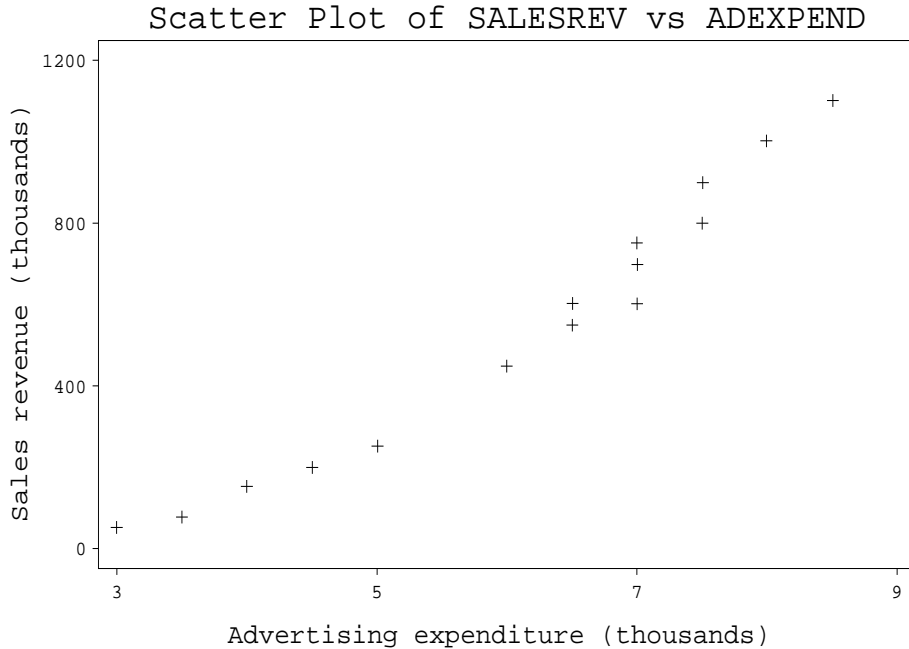
- 1) Submit electronically exercises 9 through 16 from Unit D-2 of *CyberStats* entitled “Simple Linear Regression.”
- 2) Advertising is often touted as the key to success in marketing a product. In seeking to determine just how influential advertising is, the management of a recently set-up retail chain has collected data over the previous 15 weeks on sales revenue and advertising expenditures on PBS. The data are available on the course website under “Data” and are given below:

ADEXPEND	SALESREV
3.0	50
5.0	250
7.0	700
6.0	450
6.5	600
8.0	1000
3.5	75
4.0	150
4.5	200
6.5	550
7.0	750
7.5	800
7.5	900
8.5	1100
7.0	600

- a) Plot the data in “DataTools”. Does it appear that advertising expenditures and sales are linearly related?

**Solution**(1): The first step is to plot the data in DataTools using the “Scatterplot” command.

```
Commands:-----  
Summary statistics  
Scatter plot  
SALESREV ADEXPEND  
Output:-----
```



Clearly a linear relationship is reasonable here.

b) Using DataTools, fit the linear regression model of sales based on advertising expenditure.

**Solution(1):** We now fit the simple linear regression model.

```

Commands:-----
Stat
  Multiple regression
    Dependent variable: SALESREV  Independent variable: ADEXPEND
Output:-----
LEAST SQUARES LINEAR REGRESSION OF SALESREV  Sales revenue (thousands)

PREDICTOR
VARIABLES      COEFFICIENT      STD ERROR      STUDENT'S T      P
-----
CONSTANT      -644.951         72.8973        -8.85            0.0000
ADEXPEND      195.074          11.5381        16.91            0.0000

R-SQUARED      0.9565          RESIDUAL MEAN SQUARE (MSE)  5404.98
ADJUSTED R-SQUARED  0.9532          STANDARD ERROR OF ESTIMATE  73.5186

SOURCE      DF      SS      MS      F      P
-----
REGRESSION  1      1.545E+06  1.545E+06  285.84  0.0000
RESIDUAL    13     70264.8   5404.98
TOTAL      14     1.615E+06

CASES INCLUDED 15  MISSING CASES 0

```

a) Give an economic interpretation of the slope,  $\hat{\beta}_1$ , in the model.

**Solution**(1): The interpretation of the slope is an increase of \$195.07 in sales for each dollar increase in advertising expenditure. Note that both the original units are in \$1,000s.  $\times$

- b) If the sign of the slope were negative, what would that say about the relationship between advertising and sales?

**Solution**(1): The interpretation of a negative slope is that for each dollar increase in advertising expenditure sales would decrease. That is increased advertising decreases sales!  $\times$

- c) Give an economic interpretation of the intercept,  $\hat{\beta}_0$ , in the model. What does the value of the intercept tell us?

**Solution**(1): The interpretation of the intercept is the sales if the advertising expenditure was zero. Here the value is negative \$644.95. This is not reasonable and indicates that extrapolation is unreasonable for this model.  $\times$

- d) Calculate a 99% confidence interval for  $\beta_0$ . Do you think that there will be no sales if there was no advertising?

**Solution(3)**: A 99% confidence interval for  $\beta_0$  is

$$\begin{aligned} & \hat{\beta}_0 \pm t_{\alpha/2}(n-2) \times \widehat{SE}(\hat{\beta}_0) \\ & = -644.95 \pm t_{0.005}(13) \times 72.897 \\ & = -644.95 \pm 3.012 \times 72.897 \\ & = (-\$864.52, -\$425.38) \text{ sales dollars} \end{aligned}$$

⊗

- e) Calculate a 95% confidence interval for  $\beta_1$ .

**Solution(3)**: A 95% confidence interval for  $\beta_1$  is

$$\begin{aligned} & \hat{\beta}_1 \pm t_{\alpha/2}(n-2) \times \widehat{SE}(\hat{\beta}_1) \\ & = 195.07 \pm t_{0.025}(13) \times 11.538 \\ & = 195.07 \pm 2.160 \times 11.538 \\ & = (170.15, 219.99) \text{ sales dollars per advertizing dollars} \end{aligned}$$

⊗

*Extra Credit Problem:*

- 3) Complete the open case “The Pattern in Unemployment Rates over Time” from page 190 of **CHS**. Brief answers suffice.

**Solution(4)**: The rate of change of the employment rate can be estimated using a semilog model. The fitted regression model is

$$\widehat{\text{Log(Employment rate)}} = 10.2922 - .00424 \times \text{Year}.$$

Thus, the estimated rate of change of the employment rate equals  $10^{-.00424} - 1 = -.00972$ , or a decrease of about 1% per year.

A naive estimate of the rate of change satisfies

$$\left( \frac{\text{Employment rate}_{1990}}{\text{Employment rate}_{1971}} \right)^{\frac{1}{19}} - 1 = \left( \frac{75.9}{87.9} \right)^{\frac{1}{19}} - 1 = -.00770.$$

The naive estimate is less negative than the regression-based estimate because the employment rate actually increased from 1971 to 1974; since the initial value for 1971 is below the regression line, the naive estimate corresponds to a “flattening out” of the estimated slope.

Note, by the way, that the assumption of independence seems to be in trouble here as there is autocorrelation apparent in the residuals from the regression model.

⊗