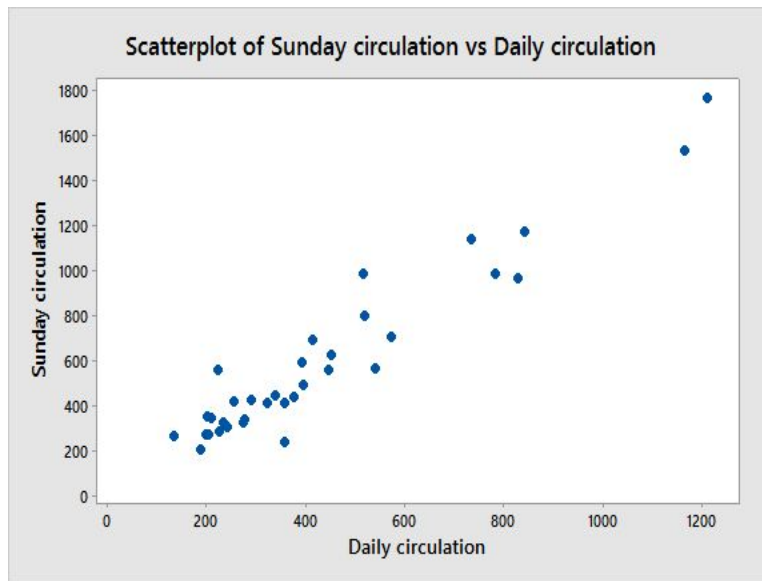


Answers to Homework 1

1. (a) Here is a scatter plot:



There does appear to be a strong linear relationship. This is not surprising, since one would expect that most of the people who purchase the Sunday edition also purchase the daily edition, augmented by people who only want the Sunday paper.

(b) Here is the regression output:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	4292653	4292653	358.53	0.000
Daily circulation	1	4292653	4292653	358.53	0.000
Error	32	383136	11973		
Total	33	4675788			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
109.421	91.81%	91.55%	90.71%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	13.8	35.8	0.39	0.702	
Daily circulation	1.3397	0.0708	18.93	0.000	1.00

Regression Equation

Sunday circulation = 13.8 + 1.3397 Daily circulation

- (c) A 95% confidence interval for β_j has the form

$$\hat{\beta}_j \pm t_{.025}^{32} s.e.(\hat{\beta}_j),$$

where $t_{.025}^{32}$ is the t -based critical value for $\alpha/2$ with $\alpha = .05$ on 32 degrees of freedom. Thus, the two requested confidence intervals have the form

$$\beta_0 : 13.84 \pm (2.04)(35.80) = 13.84 \pm 72.925 = (-59.085, 86.765)$$

and

$$\beta_1 : 1.34 \pm (2.04)(.071) = 1.34 \pm .144 = (1.196, 1.484).$$

- (d) This question is a hypothesis testing question. The hypotheses being tested are

$$H_0 : \beta_1 = 0$$

versus

$$H_1 : \beta_1 \neq 0.$$

The appropriate test is the F -test, which equals 358.53, and is highly statistically significant. So, yes, Daily circulation is significantly related to Sunday circulation (the t -test for $\beta_1 = 0$ is equivalent here).

- (e) This is estimated by the R^2 , which is approximately 92%.
(f) This output gives the answer:

Settings

Variable	Setting
Daily circulation	600

Prediction

Fit	SE Fit	95% CI	95% PI
817.664	22.2529	(772.337, 862.992)	(590.219, 1045.11)

(remember to enter “600” for the value of the predictor, not “600000”). The question asks for the 95% confidence interval for the average Sunday circulation of newspapers with Daily circulation equaling 600,000; that is, the interval based on fitted bounds. Thus, the interval is (772300, 863000).

- (g) This now asks for a prediction interval. It is given above as (590200, 1045200). It is much wider than the one in part (f) because it also reflects the inherent variability of Sunday circulation values off the regression line.
(h) This output gives the answer:

Settings

Variable	Setting
Daily circulation	2500

Prediction

Fit	SE Fit	95% CI	95% PI	
3363.12	147.590	(3062.49, 3663.75)	(2988.88, 3737.36)	XX

XX denotes an extremely unusual point relative to predictor levels used to fit the model.

This is again a prediction interval, and is given as (2988800, 3737500). This interval is about 65% wider than the one given in part (g), which reflects the fact the prediction away from the mean of the given predictor values is more difficult than prediction closer to the mean. Indeed, there are no predictor values anywhere near 2,500,000, casting great doubt on the usefulness of this prediction. This sort of prediction is called *extrapolation*, and it is, in general, a bad idea.