OpenIntro Statistics Exercise 8.3

Keith Folsom

November 18, 2015

8.3 Baby weights, Part III.

We considered the variables smoke and parity, one at a time, in modeling birth weights of babies in Exercises 8.1 and 8.2. A more realistic approach to modeling infant weights is to consider all possibly related variables at once. Other variables of interest include length of pregnancy in days (gestation), mother's age in years (age), mother's height in inches (height), and mother's pregnancy weight in pounds (weight). Below are three observations from this data set.

(a) Write the equation of the regression line that includes all of the variables.

Variable	Estimate
(Intercept)	-80.41
gestation	0.41
parity	-3.33
age	-0.01
height	1.15
weight	0.05
smoke	-8.40

$$\widehat{babyWeight} = -80.41 + 0.44 * gestation - 3.33 * parity - 0.01$$
$$+ age + 1.15 * height + 0.05 + weight - 8.40 * smoke$$

(b) Interpret the slopes of gestation and age in this context.

With each additional day of gestation, we can expect the baby's weight to increase by 0.44 ounces on average.

For each additional year of the mother's age, the model predicts a 0.01 ounce decrease in the baby's birth weight.

(c) The coefficient for parity is different than in the linear model shown in Exercise 8.2. Why might there be a difference?

Parity might be different because this linear model considers more variables than the model in Exercise 8.2. A relationship between parity and another variable could be affecting the coefficient determined in the model.

(d) Calculate the residual for the first observation in the data set.

obs	bwt	gestation	parity	age	height	weight	smoke
1	120	284	0	27	62	100	0

$$\widehat{babyWeight} = -80.41 + 0.44 * gestation - 3.33 * parity - 0.01 + age + 1.15 * height + 0.05 + weight - 8.40 * smoke$$

Calculate the baby's weight using the variables from the first observation.

```
gestation <- 284; parity <- 0;   age <- 27
height <- 62;     weight <- 100; smoke <- 0

babyWeight <- -80.41 + (0.44 * gestation) +
                (-3.33 * parity) +
                (-0.01 * age) + (1.15 * height) +
                 (0.05 * weight) + (-8.40 * smoke)
```

$\widehat{babyWeight} = 120.58$

The residual can be calculated using:

$$e_i = y_i - \hat{y}_i$$

e $=$ 120 - $\widehat{babyWeight}$ where 120 is the observed bwt

The residual for the first observation is -0.58. The model has overestimated the birth weight of the baby.

(e) The variance of the residuals is 249.28, and the variance of the birth weights of all babies in the data set is 332.57. Calculate the R2 and the adjusted R2. Note that there are 1,236 observations in the data set.

$\underline{R^2:}$

R^2 = 1 - (variability in residuals)/(variability in the outcome)

R^2 = 1 - 249.28/332.57
R^2 = 0.25044

$\underline{R^2\text{adj:}}$

R^2adj = Var(e_i)/Var(y_i) * n - 1/n - k - 1
n = number of cases = 1236
k = number of predictor variables = 6

R^2adj = 1 - 249.28/332.57 * (1236 - 1)/(1236 - 6 - 1)
R^2adj = 0.24678