Spring 2018 DATA606 Presentation Exercise 8.3

Jonathan Hernandez

April 29, 2018

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.

	bwt	gestation	parity	age	height	weight	smoke
1	120	284	0	27	62	100	0
2	113	282	0	33	64	135	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1236	117	297	0	38	65	129	0

The summary table below shows the results of a regression model for predicting the average birth weight of babies based on all of the variables included in the data set.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-80.41	14.35	-5.60	0.0000
gestation	0.44	0.03	15.26	0.0000
parity	-3.33	1.13	-2.95	0.0033
age	-0.01	0.09	-0.10	0.9170
height	1.15	0.21	5.63	0.0000
weight	0.05	0.03	1.99	0.0471
smoke	-8.40	0.95	-8.81	0.0000

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

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

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

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

(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 R^2 and the adjusted R^2. Note that there are 1,236 observations in the data set.

a) By looking at the Estimate column in the summary table, we can create the regression line as follows:

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

b) For gestation, we can interpret the slope of it by holding all other variables constant and note that for every day the mother is in pregnancy, the baby weight goes up by a factor of 0.44.
Likewise for the age variable, the older the mother, the baby weight is expected to go down by a factor of -0.01.
c) Since we have added new variables to the model, the model is adjusted and re-fitted and the parameter estimates change to create the model.
d) The residual of the first observation is calculated as follows:

$$\$\hat{e}_1 = bwt_1 - \widehat{bwt}_1 = 120 - (-80.41 + 0.44 * 284 - 3.33 * 0 - 0.01 * 27 + 1.15 * 62 + 0.05 * 100 - 8.40 * 0) = -0.58$$

e) We can calcuate R^2 and R^2_{adj} by using the information given and the formulas below:

$R^2 = 1 - (var(e_i)/var(y_i))$ and

$R^2_{adj} = R^2 * (n - 1/n - k - 1)$

In this case, n = 1236, k = 6 (6 predictors)

$var(e_i) = 249.28$ and $var(y_i) = 332.57$

Plugging in the numbers gives us

$R^2 = 1 - (249.28/332.57) = 0.25$ (rounded to two decimal places)

$R^2_{adj} = 0.25 * (1235/1229) = 0.25$

Loading [MathJax]/jax/output/HTML-CSS/jax.js