# DATA606 Homework Presentation

Jai Jeffryes

11/18/2019

# Exercise 9.3 Baby weights, Part III

We considered the variables `smoke` and `parity`, one at a time, in modeling birth weights of babies in Exercises 9.1 and 9.2. 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 |

# Continued

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 model that includes all of the variables.

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

| | Estimate |
|---|---|
| (Intercept) | -80.41 |
| gestation | 0.44 |
| parity | -3.33 |
| age | -0.01 |
| height | 1.15 |
| weight | 0.05 |
| smoke | -8.40 |

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

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

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

- The gestation coefficient of 0.44 says that an additional day in a pregnancy increases a baby's birth weight, on average, by 0.44 ounces.
- The age coefficient of -0.01 says that an additional year in a mother's age reduces a baby's birth weight, on average, by 0.01 ounces.

## (c) The coefficient for parity is different from the linear model shown in Exercise 9.2. Why might there be a difference?

### Single-variable model from Exercise 9.2

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 120.07 | 0.60 | 199.94 | 0.0000 |
| parity | -1.93 | 1.19 | -1.62 | 0.1052 |

### Parity variable from this multiple regression model

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| parity | -3.33 | 1.13 | -2.95 | 0.0033 |

A difference like this may be due to correlation between predictor variables absent from the single-variable regression model, which are therefore confounding variables. Correlated predictor variables are said to be collinear.

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

Use the equation from answer (a). Supply the variables with values from the first observation.

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

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

```
bwt_pred = -80.41 + 0.44*284 - 3.33*0 - 0.01*27 + 1.15*62 -
bwt_obs <- 120
bwt_residual <- bwt_obs - bwt_pred
```

- Observed baby weight: 120 ozs.
- Predicted: 120.58 ozs.
- Residual: -0.58 ozs.

(e) The variance of the residuals is 249.28, the variance of the birth weights is 332.57. Calculate $R^2$ and adjusted $R^2$. There are 1,236 observations.

$$R^2 = 1 - \frac{\text{variability in residuals}}{\text{variability in the outcome}}$$

$$R^2_{adj} = 1 - \frac{\text{variability in residuals}}{\text{variability in the outcome}} \times \frac{n-1}{\text{degrees of freedom}}$$

Degrees of freedom: $n$ - $predictor\_count$ - 1

```
r2 <- 1 - (249.28 / 332.57)
n <- 1236; df <- n - 6 - 1
r2_adj <-  1 - ((249.28 / 332.57) * ((n - 1) / df))
```

- $R^2$: 0.2504435
- $R^2_{adj}$: 0.2467842