

Lab 8 Regression

Chen Wang, Junke Wang, Zhuozhi Xiong

due 11/27/19

Together as a class:

We are first going to install and load packages that we will need.

1. Install and load both the `modelr` package and the `openintro` package. Load the `tidyverse` package.

For this problem we will be using the data set called `gpa` from the `openintro` package. We are interest in association between the number of hours of sleep a student gets and their gpas.

2. Make a scatter plot and describe the association that you see.
3. Create the linear regression model for this relationship between hours of sleep and GPA. We are regressing `gpa` onto `hours of sleep` Call this model `gpa_model`.

4. Find the correlation of this regression line and give the estimated β values.
5. Interpret the $\hat{\beta}$ values. There are 2.

6. Add the predicted values of GPA and the Residuals to the data frame `gpa` using the `add_predictions()` and `add_residuals()` functions from the `modelr` package.
7. Now we want to check the conditions needed to use the least squares regression line. Create a qqplot and a residual plots in order to check the conditions. Are the conditions meet? Are there any outliers?

8. Conduct a hypothesis test to see if there is an association between how much sleep a student gets and their GPA. What can we conclude?

On your own:

In this problem, we will be using the `babies` data set from the `openintro` package. information about the data set can be found by running the following code in an `r` chunk: `?babies`. We are interested in finding out how the length of gestation(`gestation`), the age of the mother(`age`), the height of the mother(`height`), the mother's weight(`weight`), and whether or not this was the mother's first pregnancy (`parity`) are related to how much the baby will weight at birth (`bwt`).

1. Load the data and save it as `baby`.
2. Create the linear regression model that regresses `bwt` the variables mentioned above. Print out the summary.
3. Interpret the coefficients for `gestation` and `parity`.
4. Does the intercept value have any meaning in context?
5. What is the Multiple R^2 value for this model.

6. Add the residuals and predicted values to the data frame.
7. Create a residual plot and a qqplot. Comment on whether or not the conditions are met to use the model you found in part 2.

8. Obtain a 95% confidence interval for the coefficient on **gestation** and **age**. Interpret both confidence intervals.