The dataset Blood1 contains information on the systolic blood pressure for 500 randomly chosen adults. One of the variables recorded for each subject, Overwt, classifies weight as 0 = Normal, 1 = Overweight, or 2 = Obese. Fit two regression models to predict SystolicBP, one using Overwt as a single quantitative predictor and the other using indicator variables for the weight groups. Compare the results for these two models.

The data in BaseballTimes2017 contain information from 15 Major League Baseball games played on August 11, 2017. In Exercise 1.45 on page 57, we considered models to predict the time a game lasts (in minutes). One of the potential predictors is the number of runs scored in the game. Use a randomization procedure to test whether there is significant evidence to conclude that the correlation between Runs and Time is greater than zero.

In Section 4.7, we generated bootstrap samples by sampling with replacement from the original sample of Honda Accord prices. An alternate approach is to leave the predictor values fixed and generate new values for the response variable by randomly selecting values from the residuals of the original fit and adding them to the fitted values.

image

Run a regression to predict the Price of cars based on Mileage using the AccordPrice data. Record the coefficients for the fitted model and save both the fits and residuals.

Construct a new set of random errors by sampling with replacement from the original residuals. Use the same sample size as the original sample and add the errors to the original fitted values to obtain a new price for each of the cars. Produce a scatterplot of NewPrice versus Mileage.

Run a regression model to predict the new prices based on mileages. Compare the slope and intercept coefficients for this bootstrap sample to the original fitted model.

Repeat the process 1000 times, saving the slope from each bootstrap fit. Find the mean and standard deviation of the distribution of bootstrap slopes.

Use each of the three methods discussed in Section 4.7 to construct confidence intervals for the slope in this regression model. Compare your results to each other and the intervals constructed in Section 4.7.