Lin-HW8

Bin Lin

2016-12-6

8.2 Baby weights, Part II. (a) Write the equation of the regression line.

babyweight = 120.07 - 1.93 * parity

b. Interpret the slope in this context, and calculate the predicted birth weight of first borns and others.

The estimated body weight of babies who were not first born is 1.93 onces lower than babies who were first born. First born weight = 120.07 - 1.93 * 0 = 120.07 onces Non-first-born weight = 120.07 - 1.93 * 1 = 118.77 onces

c. Is there a statistically significant relationship between the average birth weight and parity?

There is no a statistically significant relationship between the average birth weight and parity, because p-value is 0.1052 which is higher than 0.05.

8.4 Absenteeism. (a) Write the equation of the regression line.

absence = 18.93 - 9.11 * eth + 3.10 * sex + 2.15 * lrn

b. Interpret each one of the slopes in this context.

Students who are non-aboriginal have 9.11 less absent days than the students who are original Male students have 3.1 days more absent day than female students Slow learner have 2.15 more absent day than average learner.

c. Calculate the residual for the first observation in the data set: a student who is aboriginal, male, a slow learner, and missed 2 days of school.

Estimated absence = 18.93 + 3.10 + 2.15 = 24.18 Residual = 2- 24.18 = -22.18

d. The variance of the residuals is 240.57, and the variance of the number of absent days for all students in the data set is 264.17. Calculate the R2 and the adjusted R2. Note that there are 146 observations in the data set.

R2 = 1 - 240.57 / 264.17 = 0.089 adjusted R2 = 1 - (240.57 / 264.17 * (146 - 1) / (146 - 3 - 1)) = 0.070

8.8 Absenteeism, Part II. Which, if any, variable should be removed from the model first?

Learner status should be removed from the model first, because this explanatory variable is associated with highest p-value, which makes it non-statistically significant.

8.16 Challenger disaster, Part I.

a. Each column of the table above represents a di???erent shuttle mission. Examine these data and describe what you observe with respect to the relationship between temperatures and damaged O-rings. As the temperature goes up, there is less occurence of damaged O-rings.
b. Failures have been coded as 1 for a damaged O-ring and 0 for an undamaged O-ring, and a logistic regression model was fit to these data. A summary of this model is given below. Describe the key components of this summary table in words.

Intercept: When temperatur is 0 degree Fahrenheit, there is estimated 11.663 O-ring damages Slope of temperature: All else held constant, there will be 0.2132 less O-ring damages, whenever the temperature goes up by 1 degree Fahrenheit.

c. Write out the logistic model using the point estimates of the model parameters. log(p / (1 - p)) = 11.663 - 0.2162 * temperature

d. Based on the model, do you think concerns regarding O-rings are justified? Explain.

Yes, because the p-value is 0, which means the model is statistically significant. Therefore the concern is justified.

8.18 Challenger disaster, Part II.

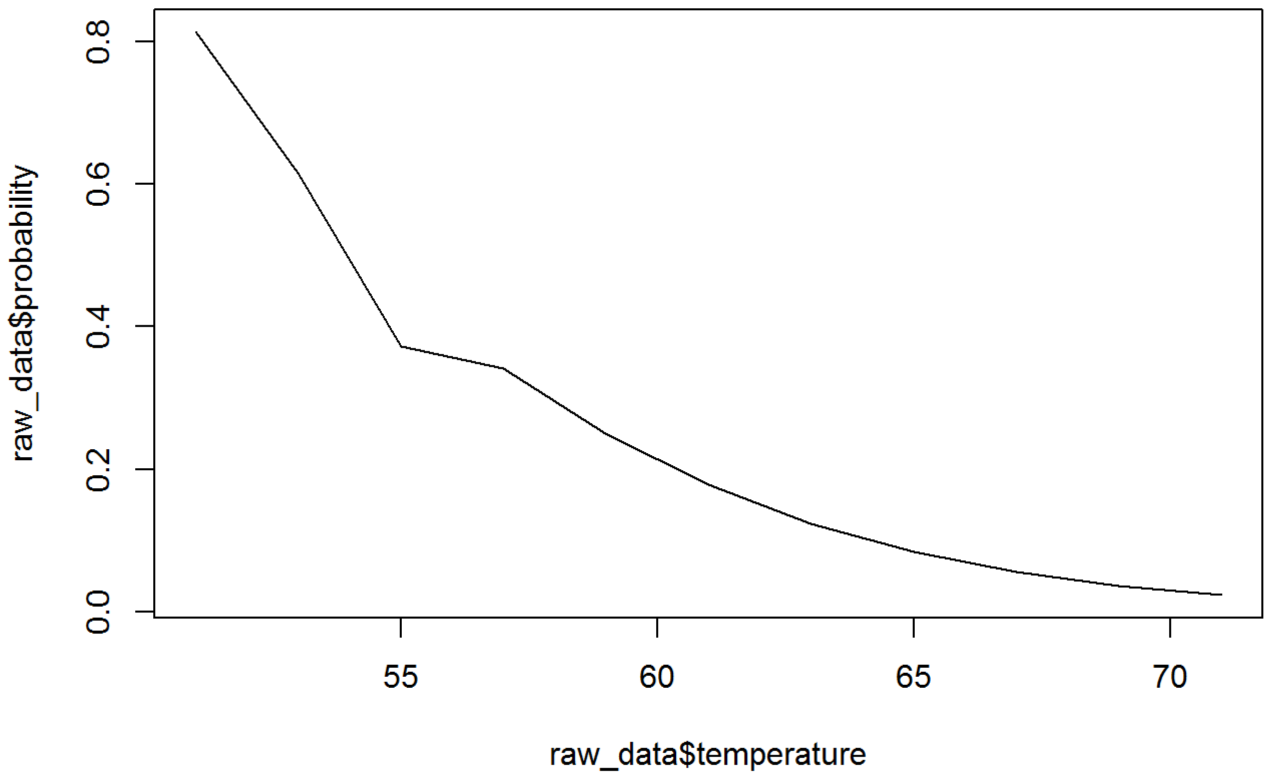
a. The data provided in the previous exercise are shown in the plot. The logistic model fit to these data may be written as log(p / (1 - p)) = 11.663 - 0.2162 * temperature where ^p is the model-estimated probability that an O-ring will become damaged. Use the model to calculate the probability that an O-ring will become damaged at each of the following ambient temperatures: 51, 53, and 55 degrees Fahrenheit. The model-estimated probabilities for several additional ambient temperatures are provided below, where subscripts indicate the temperature: ^p57 = 0.341 ^p59 = 0.251 ^p61 = 0.179 ^p63 = 0.124 ^p65 = 0.084 ^p67 = 0.056 ^p69 = 0.037 ^p71 = 0.024

p51 = 0.813 p53 = 0.616 p55 = 0.372

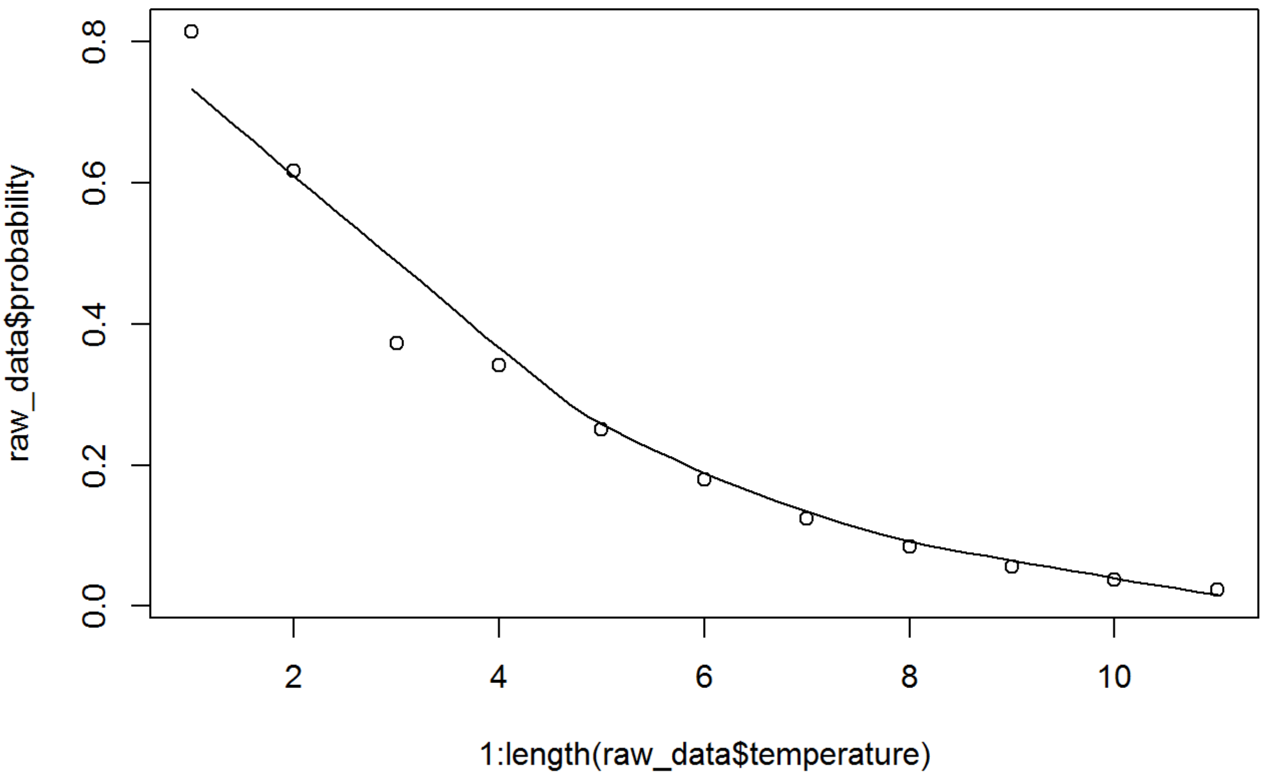
b. Add the model-estimated probabilities from part (a) on the plot, then connect these dots using a smooth curve to represent the model-estimated probabilities.

```
raw_data <- data.frame(temperature = c(51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71), probability
= c(0.813, 0.616, 0.372, 0.341, 0.251, 0.179, 0.124, 0.084, 0.056, 0.037, 0.024))

plot(raw_data$temperature, raw_data$probability, type='l')
```




```
scatter.smooth(x=1:length(raw_data$temperature), y=raw_data$probability)
```



c. Describe any concerns you may have regarding applying logistic regression in this application, and note any assumptions that are required to accept the model’s validity.

The dataset only contains 23 observations, which is not sufficient for making an accurate model. We assume a binomial distribution (damaged or non-damaged) produced the outcome variable and we therefore want to model p the probability of success for a given set of predictors.