homework_ch8_JimmyNg

Jimmy Ng

November 6, 2018

Question 8.2

(a) equation: y = -1.93 * x + 120.07
(b) the slope is negative, meaning there's a negative effect for non-first born. For example, first born would be equal to (-1.93 * 0) + 120.07 = 120.07, whereas non-first born would be equal to (-1.93 * 1) + 120.07 = 118.14
(c) not really, the p-value associated with parity (0.1052) in this case is above the standard threshold of 0.05

Question 8.4

(a) equation: y = (-9.11 * x1) + (3.1 * x2) + (2.15 * x3) + 18.93
(b) the estimate of eth is negative, meaning the ethnic background (1 = not aboriginal) would have negative impact of average number of days absent by 9.11. That means they will have less number of absence by 9.11; sex is positive, meaning (1 = male) boys will be more likely to have more days of absence by 3.1; and finally, lrn (1 = slow learner) is also positive, meaning slow learner will be more likely to have 2.15 more days of absence.
(c) plug this in the above equation: y-actual = 2 y-hat = (-9.11 * 0) + (3.1 * 1) + (2.15 * 1) + 18.93 residual = y-actual - y-hat = -22.18
(d) R^2 = 1 - (240.57 / 264.17) = 0.08933641 n = 146; k = 3 Adjusted R^2 = 1 - ((240.57 / 264.17) * ((n - 1) / (n - k - 1))) = 0.07009704

Question 8.8

JN: We should remove the lrn (learner status) because dropping it would result in higher adjusted R^2 than the full model, i.e. 0.0723 > 0.0701

Question 8.16

(a) higher temperature corresponds to less damage
(b) same observation as above, i.e. higher temperature corresponds to less chance/probability of damage
(c) p = exp(1)^(b0 + b1 * x1) / (1 + exp(1)^(b0 + b1 * x1))
(d) yes the concern is justified, using the min and max temperature values from the table and subsequently plugging them in the equation, we see the stunning comparison:

```
b0 = 11.6630
b1 = -0.2162

x1 = 53
x2 = 81

print(paste("the chance of O-rings damange with temperature 53F is approximately ",
            round(100 * exp(1)^(b0 + b1 * x1) / (1 + exp(1)^(b0 + b1 * x1)), 1),
            "%.",
            sep = ""))
```

```
## [1] "the chance of O-rings damange with temperature 53F is approximately 55.1%."

print(paste("the chance of O-rings damange with temperature 81F is approximately ",
            round(100 * exp(1)^(b0 + b1 * x2) / (1 + exp(1)^(b0 + b1 * x2)), 1),
            "%.",
            sep = ""))
## [1] "the chance of O-rings damange with temperature 81F is approximately 0.3%."
```

Question 8.18

(a) see below calculation for the corresponding temperature 51, 53, 55

```
b0 = 11.6630
b1 = -0.2162

x1 = 51
x2 = 53
x3 = 55

print(paste("the chance of O-rings damange with temperature 51F is approximately ",
            round(100 * exp(1)^(b0 + b1 * x1) / (1 + exp(1)^(b0 + b1 * x1)), 1),
            "%.",
            sep = ""))
## [1] "the chance of O-rings damange with temperature 51F is approximately 65.4%."

print(paste("the chance of O-rings damange with temperature 53F is approximately ",
            round(100 * exp(1)^(b0 + b1 * x2) / (1 + exp(1)^(b0 + b1 * x2)), 1),
            "%.",
            sep = ""))
## [1] "the chance of O-rings damange with temperature 53F is approximately 55.1%."

print(paste("the chance of O-rings damange with temperature 55F is approximately ",
            round(100 * exp(1)^(b0 + b1 * x3) / (1 + exp(1)^(b0 + b1 * x3)), 1),
            "%.",
            sep = ""))
## [1] "the chance of O-rings damange with temperature 55F is approximately 44.3%."
```

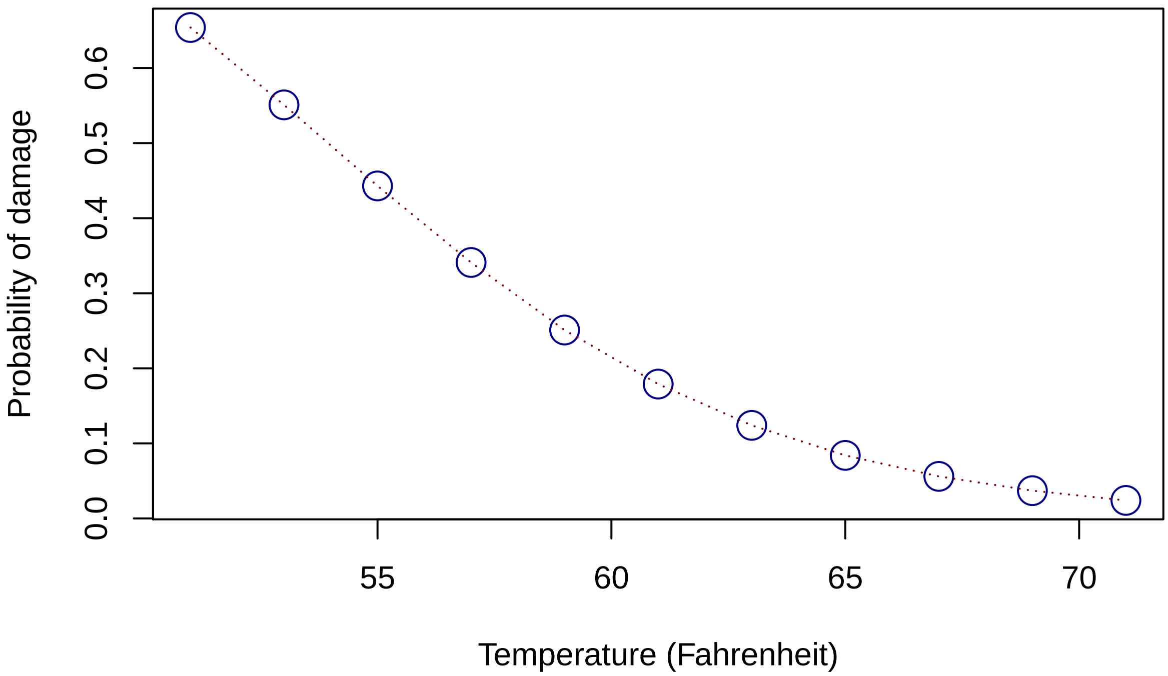
(b) see below for plot

```
temp <- c(51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71)
p <- c(0.654, 0.551, 0.443, 0.341, 0.251, 0.179, 0.124, 0.084, 0.056, 0.037, 0.024)
df <- data.frame(temp, p)

with(df,
     plot(temp, p,
          type = "p",
          col = "dark blue",
          cex = 2,
          xlab = "Temperature (Fahrenheit)",
          ylab = "Probability of damage"))

with(df,
     lines(temp, p, type = "l",
```

```
        lty = 3,
        col = "dark red"))
```



(c) there are two conditions: first, the predictor xi is linearly related to logit(pi), which we can confirm from observing the above chart (a strong negative correlation); second, each outcome Yi is independent of the other outcomes, which we cannot verify here but assuming that is the case when data is collected. Each trial should be independent from each other. The only concern for this model is applying a suitable range for the predictor, i.e. DO NOT go above or below certain temperature range. For instance, that makes no sense to go below 32F (freezing point of water) or above 212F (boiling point) at standard atmospheric pressure. Future replication of the study or application of the model must follow the same setting that the experiment(s) was carried out. Otherwise, the model will not be applicable at all.