

# DATA 606 Homework 8

*Ilya Kats*

*May 6, 2017*

## 8.2 Baby weights, Part II.

### Part (a)

$$\widehat{baby\ weight} = 120.07 - 1.93 \times parity$$

### Part (b)

The estimated weight of babies that are not first born is 1.93 ounces lower than the estimated weight of first born babies.

Not first born:  $120.07 - 1.93 * 1 = 118.14$

First born:  $120.07 - 1.93 * 0 = 120.07$

### Part (c)

According to the summary table, the  $p$ -value is about 0.1 and greater than 0.05. We fail to reject the null hypothesis that the slope parameter is different than 0. There is no significant relationship between the average birth weight and parity.

## 8.4 Absenteeism, Part I.

### Part (a)

$$\widehat{absenteeism} = 18.93 - 9.11 \times eth + 3.10 \times sex + 2.15 \times lrn$$

### Part (b)

**eth:** The model predicts that the average number of days absent by non-aboriginal students is 9.11 days lower than by aboriginal students.

**sex:** The model predicts that the average number of days absent by male students is 3.1 days higher than by female students.

**lrn:** The model predicts that the average number of days absent by slow learners is 2.15 days higher than by average learners.

### Part (c)

The residual for the first observation is -22.18.

$$\hat{y}_1 = 18.93 - 9.11 * 0 + 3.10 * 1 + 2.15 * 1 = 24.18$$

$$e_1 = y_1 - \hat{y}_1 = 2 - 24.18 = -22.18$$

### Part (d)

$Var(e_i) = 240.57$ ,  $Var(y_i) = 264.17$ ,  $n = 146$ ,  $k = 3$

$$R^2 = 1 - \frac{Var(e_i)}{Var(y_i)} = 1 - \frac{240.57}{264.17} \approx 0.0893$$

$$R^2_{adj} = 1 - \frac{Var(e_i)}{Var(y_i)} \times \frac{n-1}{n-k-1} = 1 - \frac{240.57}{264.17} * \frac{146-1}{146-3-1} \approx 0.0701$$

### 8.8 Absenteeism, Part II.

The fourth model (no `learner status`) has the highest adjusted  $R^2_{adj} = 0.0723$ , and it is higher than the adjusted  $R^2_{adj} = 0.0701$  of the full model. Variable `learner status` should be eliminated from the model first.

### 8.16 Challenger disaster, Part I.

#### Part (a)

Out of 23 shuttle missions, all but 1 had either no damaged O-rings or just 1 damaged O-ring. The mission with the most damaged O-rings (5) is also the mission that launched on the coldest day (53 F). All missions launched in temperature below 65 F had at least one damaged O-ring. Out of 13 missions launched in temperature 70 F or higher, only 3 show damaged O-rings. There may be some significant relationship between two variables and further analysis is necessary.

#### Part (b)

Based on the summary table, there is a negative relationship between temperature and O-ring failures. Increase in temperature decreases the probability of an O-ring failure. Additionally, the  $p$ -value is close to 0, so the relationship between temperature and O-ring failure has significance.

#### Part (c)

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = 11.6630 - 0.2162 \times \text{Temperature}$$

#### Part (d)

As mentioned in part b, the  $p$ -value shows that the model parameter is not due to chance and has significance. It is justified to be concerned.

### 8.18 Challenger disaster, Part II.

#### Part (a)

```
temp <- c(53, 57, 58, 63, 66, 67, 67, 67, 67, 68, 69, 70, 70,
          70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81,
          53, 57, 58, 63, 66, 67, 67, 67, 67, 68, 69, 70, 70,
          70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81,
          53, 57, 58, 63, 66, 67, 67, 67, 67, 68, 69, 70, 70,
          70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81,
          53, 57, 58, 63, 66, 67, 67, 67, 67, 68, 69, 70, 70,
          70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81,
```

```

53, 57, 58, 63, 66, 67, 67, 67, 68, 69, 70, 70,
70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81)
failure <- c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0,
1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

glm.missions <- glm(failure ~ temp, family = binomial)

summary(glm.missions)

##
## Call:
## glm(formula = failure ~ temp, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2646  -0.3395  -0.2472  -0.1299   3.0216
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 11.66299    3.29616   3.538 0.000403 ***
## temp        -0.21623    0.05318  -4.066 4.77e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 76.745  on 137  degrees of freedom
## Residual deviance: 54.759  on 136  degrees of freedom
## AIC: 58.759
##
## Number of Fisher Scoring iterations: 6

prob <- data.frame(temp = c(50:85), prob = rep(0, length(c(50:85))))
prob[, 2] <- predict(glm.missions,
                     newdata = data.frame(temp = prob[, 1]),
                     type = "response")

prob[prob$temp == 51, 2]

## [1] 0.6536388

prob[prob$temp == 53, 2]

## [1] 0.5504788

```

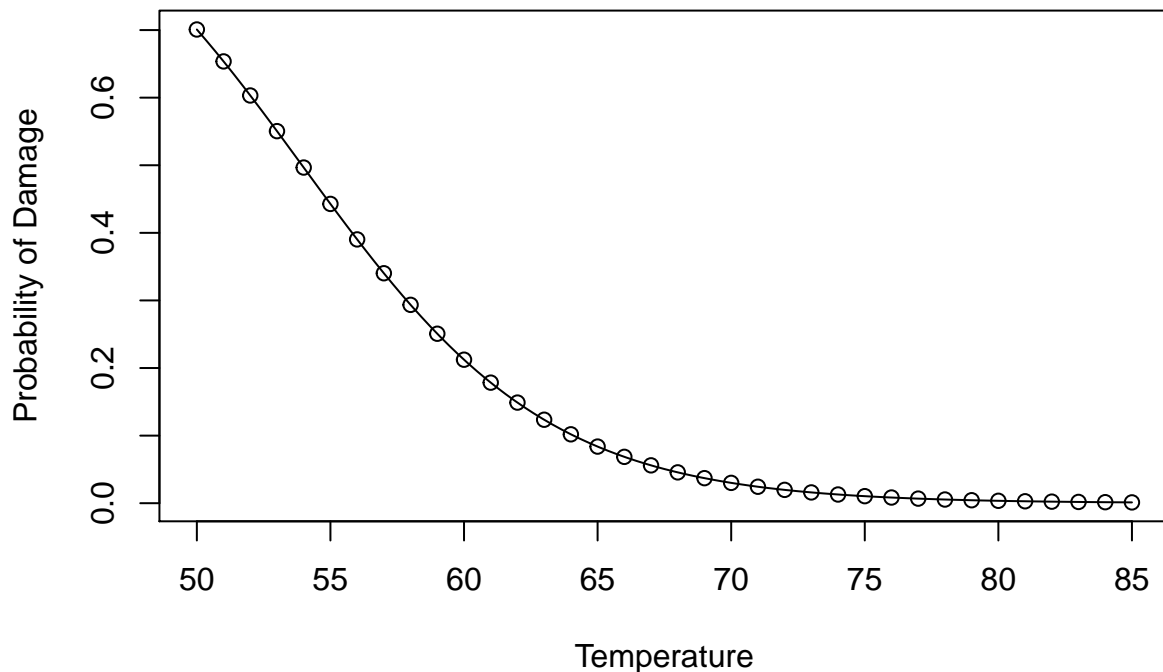
```
prob[prob$temp == 55, 2]
```

```
## [1] 0.4427862
```

The probability that O-ring will become damaged at 51 F is 65.36%. At 53 F it is 55.05%; and, at 55 F it is 44.28%. Alternatively, I could have plugged the temperatures into the model equation and solved for  $\hat{p}$ .

#### Part (b)

```
plot(prob$temp, prob$prob, xlab = "Temperature", ylab = "Probability of Damage")  
curve(predict(glm.missions, data.frame(temp=x), type="response"), add=TRUE)
```



#### Part (c)

The model assumes that an O-ring damage is a binary event - it is either damaged (1) or not (0). However, it may be the case that degree of damage carries significant information. It also assumes that observations are independent. However, they may not be independent due to the manufacturing process of O-rings and other factors or if the O-rings are reused for different missions. There may be some connection between failures on the same mission. We have 138 observations with 11 failures, but they come from only 23 missions. The sample size may not be large enough. Finally, there may be other significant factors that contribute to O-ring damage, so the model could be improved by adding other significant variables.