Class Activity#12

Sunny Lee

2021-03-24

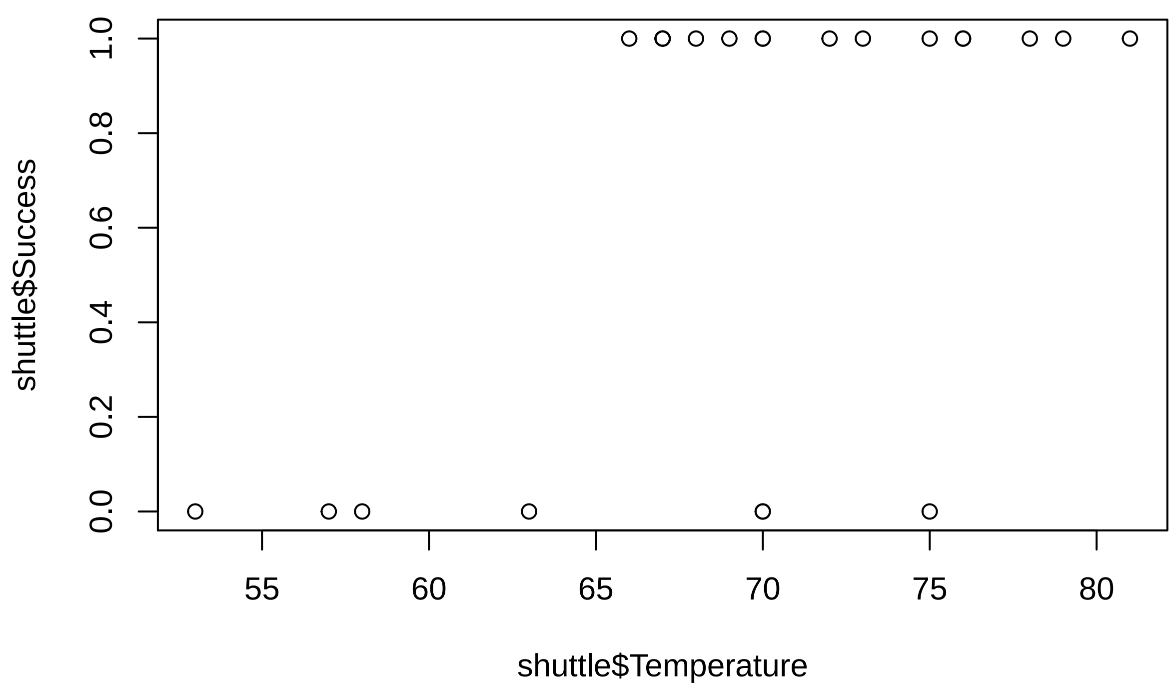
Data Set: C7 Shuttle.csv 1) Based on the description of the Challenger disaster O-ring concerns shuttle data, identify which variable should be the explanatory variable and which should be the response variable.

```
shuttle <- read.csv("C7 Shuttle.csv")
```

The Success should be the response variable while temperature should be the explanatory variable.

2) Imagine you were an engineer working for Thiokol Corporation prior to January 1986. Create a few graphs of the data. Is it obvious that temperature is related to the success of the O-rings? Produce any charts or graphs you have created that show a potential relationship between temperature and O-ring damage.

```
plot(shuttle$Success~shuttle$Temperature)
```



From the plot above, we see that there does seem to be some kind of correlation between the temperature and the success.

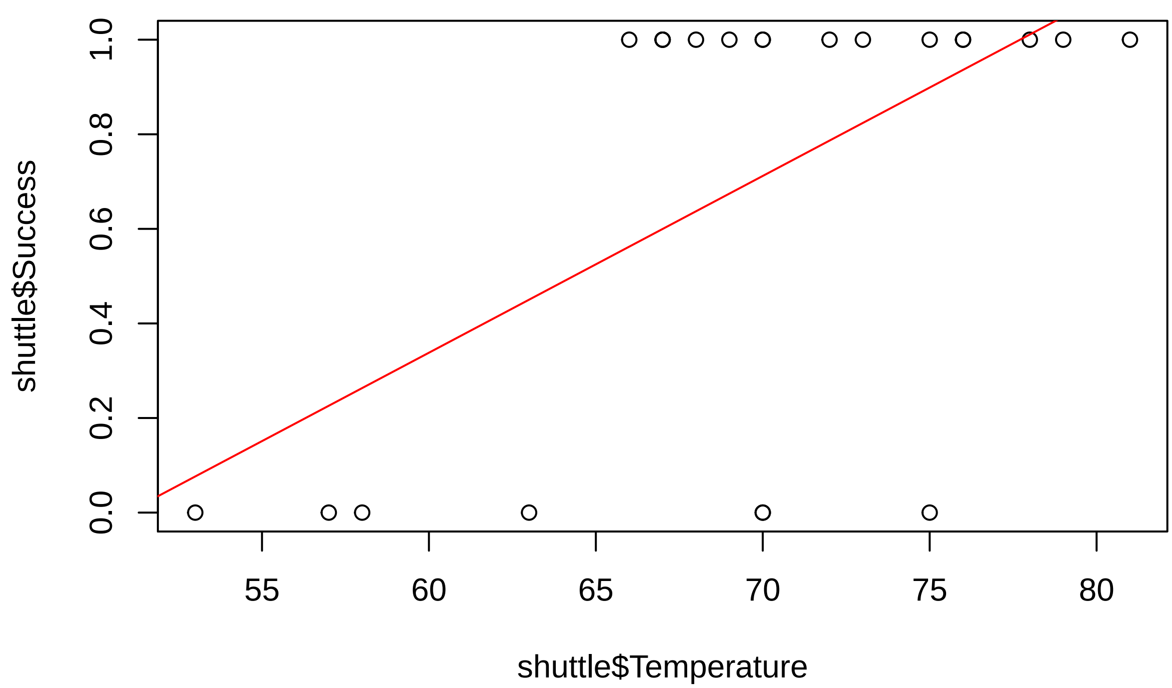
3) Use the shuttle data to create a scatterplot with a least squares regression line for the space shuttle

data. Calculate the predicted response values ($\hat{y} = b_0 + b_1 x$) when the temperature is $60^0 F$ and when the temperature is $85^0 F$.

```
lin.model <- lm(Success~Temperature, data = shuttle)
summary(lin.model)
```

```
##
## Call:
## lm(formula = Success ~ Temperature, data = shuttle)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89881 -0.17452  0.06381  0.30679  0.43762
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.90476    0.84208  -2.262  0.03443 *
## Temperature  0.03738    0.01205   3.103  0.00538 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3987 on 21 degrees of freedom
## Multiple R-squared:  0.3144, Adjusted R-squared:  0.2818
## F-statistic:  9.63 on 1 and 21 DF,  p-value: 0.005383
```

```
plot(shuttle$Success~shuttle$Temperature)
abline(lin.model, col="red")
```



```r
lin.model$coefficients[1]+lin.model$coefficients[2]*60
```

```
## (Intercept)
##   0.3380952
```

```r
lin.model$coefficients[1]+lin.model$coefficients[2]*85
```

```
## (Intercept)
##    1.272619
```

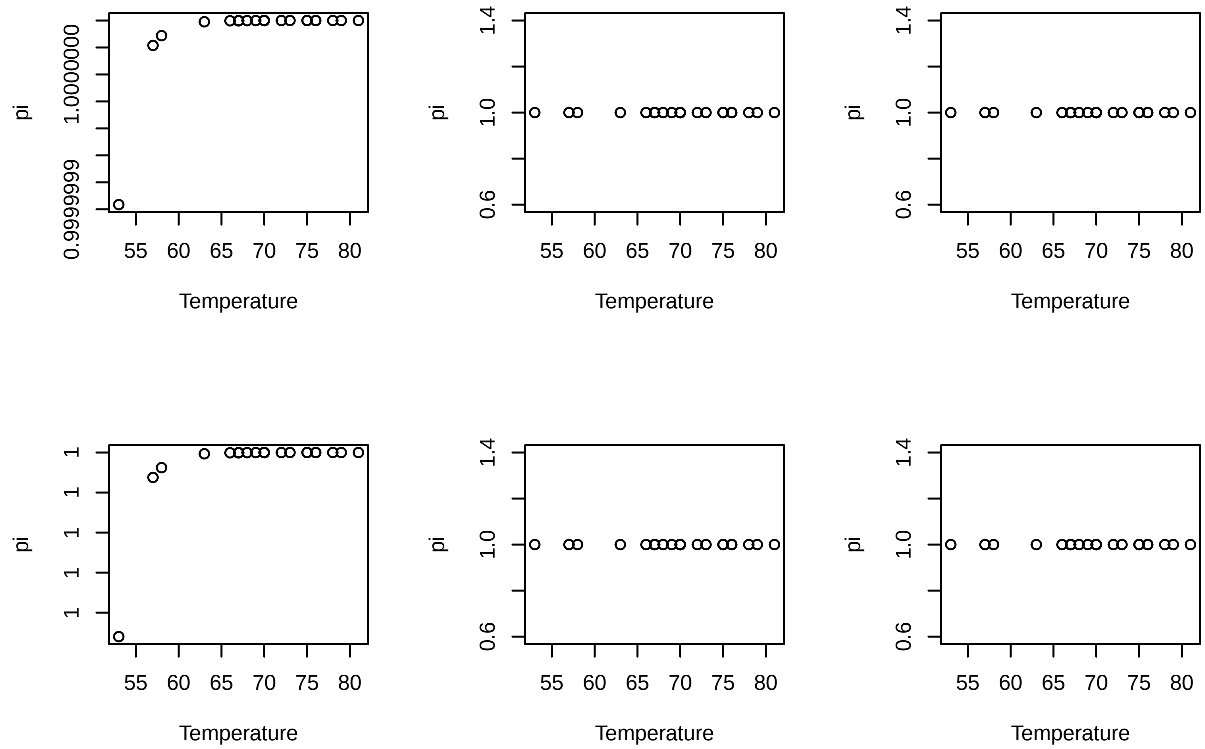
When the temperature is $60^0 f$, we see a success of .34 and when the temperature is $85^0 F$, we see a success of 1.27.

4) Solve for π_i if $\ln(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i$. $\frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1 x_i}$ $\pi_i = e^{\beta_0+\beta_1 x_i} - \pi_i(e^{\beta_0+\beta_1 x_i})$ $\pi_i + \pi_i(e^{\beta_0+\beta_1 x_i}) = e^{\beta_0+\beta_1 x_i}$ $\pi_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$

5) Use the equation of π_i to create six plots. In each graph, plot the explanatory variable (x) versus the expected probability of success(π) using $\beta_0 = -10 \quad \text{and} -5$ and $\beta_1 = 0.5, 1, \quad \text{and} 1.5$. explain the impact of changing β_0 and β_1.

```r
par(mfrow=c(2,3))
beta_0 = c(-10,-5)
beta_1 = c(0.5,1,1.5)

for(i in 1:2){
  for(j in 1:3){
    pi <- (exp(beta_0[i]+beta_1[j]*shuttle$Temperature))/(1+exp(beta_0[i]+beta_1[j]*shuttle$Temperature
    plot(pi~shuttle$Temperature, xlab="Temperature")
  }
}
```



We see from the top row and the bottom row of plots that increasing β_1 automatically forces our model to predict 1 for all of our temperatures. Increasing β_0 also seems to move our model to predicting a success more often.

6) Use statistical software to calculate the maximum likelihood estimates of β_0 and β_1. Compare the maximum likelihood estimates to the least square estimates.

```
logit <- glm(Success~Temperature, data = shuttle, family = "binomial")
summary(logit)
```

```
##
## Call:
## glm(formula = Success ~ Temperature, family = "binomial", data = shuttle)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2175  -0.4524   0.3783   0.7613   1.0611
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -15.0429     7.3786  -2.039   0.0415 *
## Temperature   0.2322     0.1082   2.145   0.0320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
```

```
## Residual deviance: 20.315  on 21  degrees of freedom
## AIC: 24.315
##
## Number of Fisher Scoring iterations: 5
```

From our linear regression model, we find $\beta_0 = -1.90476$ and $\beta_1 = 0.03738$. However, for our logistic regression model, we find $\beta_0 = -15.0429$ and $\beta_1 = 0.2322$.

7) Use the regression out put to predict the probabilty that a launch has no O-ring damage when the temperature is 31^0F, 50^0F, and 75^0F.

```
exp(logit$coefficients[1] + logit$coefficients[2]*31) / (1 + exp(logit$coefficients[1] + logit$coeffici
```

```
##  (Intercept)
## 0.0003912171
```

```
exp(logit$coefficients[1] + logit$coefficients[2]*50) / (1 + exp(logit$coefficients[1] + logit$coeffici
```

```
## (Intercept)
##  0.03122648
```

```
exp(logit$coefficients[1] + logit$coefficients[2]*75) / (1 + exp(logit$coefficients[1] + logit$coeffici
```

```
## (Intercept)
##   0.9144564
```

Using our β_i from our logistic regression model, we find a a probability of $0.0003912171, 0.03122648, 0.9144564$ for the temperatures of $31, 50, and 75$ degrees celsius respectively.

8) Calculate the odds of a launch with no O-ring damage when the temperature is 60^0F and when the temepetature is 70^0F.

```
exp(logit$coefficients[1] + logit$coefficients[2] * 60)
```

```
## (Intercept)
##   0.3285268
```

```
exp(logit$coefficients[1] + logit$coefficients[2] * 70)
```

```
## (Intercept)
##    3.348426
```

The odds of launching with no O-ring damage when the temperature is 60^0F is .33 times higher than launching with O-ring damage. The odds of launching with no 0-ring damage when the temperature is 70^0F is 3.35 times higher than launching without O-ring damage.