# Marah Notes.

## A good start is with the data display.

```
data = read.csv("module2_exo5_shuttle.csv",header=T)
data
```

```
##        Date Count Temperature Pressure Malfunction
## 1   4/12/81     6          66       50           0
## 2  11/12/81     6          70       50           1
## 3   3/22/82     6          69       50           0
## 4  11/11/82     6          68       50           0
## 5   4/04/83     6          67       50           0
## 6   6/18/82     6          72       50           0
## 7   8/30/83     6          73      100           0
## 8  11/28/83     6          70      100           0
## 9   2/03/84     6          57      200           1
## 10  4/06/84     6          63      200           1
## 11  8/30/84     6          70      200           1
## 12 10/05/84     6          78      200           0
## 13 11/08/84     6          67      200           0
## 14  1/24/85     6          53      200           2
## 15  4/12/85     6          67      200           0
## 16  4/29/85     6          75      200           0
## 17  6/17/85     6          70      200           0
## 18  7/29/85     6          81      200           0
## 19  8/27/85     6          76      200           0
## 20 10/03/85     6          79      200           0
## 21 10/30/85     6          75      200           2
## 22 11/26/85     6          76      200           0
## 23  1/12/86     6          58      200           1
```

```
data_without_malfunction_0 = data[data$Malfunction>0,]
data_without_malfunction_0
```

```
##        Date Count Temperature Pressure Malfunction
## 2  11/12/81     6          70       50           1
## 9   2/03/84     6          57      200           1
## 10  4/06/84     6          63      200           1
## 11  8/30/84     6          70      200           1
## 14  1/24/85     6          53      200           2
## 21 10/30/85     6          75      200           2
## 23  1/12/86     6          58      200           1
```

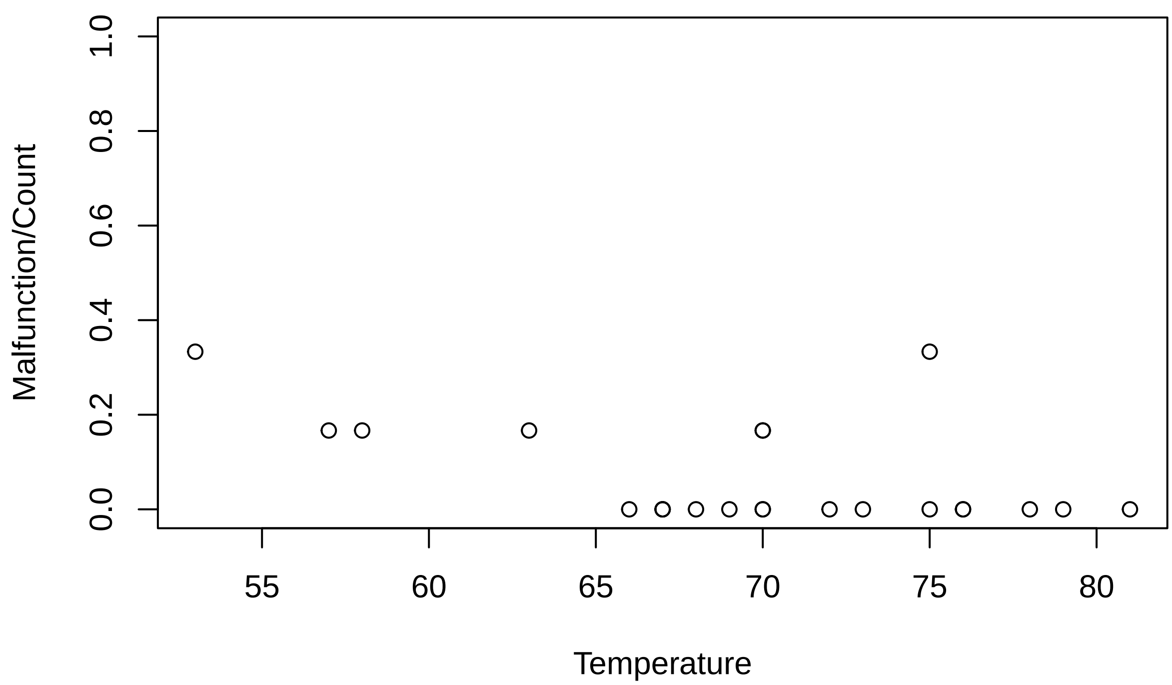
**The data is too small, which will affect the results , and the estimate.**

**I think we shouldn't remove the 0 malfuntion , that will be missing some good information about the data and its distribution.**

**I will ty the both with and without 0 .**

**Ploting the data without remove anthing.**

```
plot(data=data, Malfunction/Count ~ Temperature, ylim=c(0,1))
```




**Also, I think the graph from above was enough to understand that there is not a significant impact between temperature and the malfunction , although .**

**Note: I should determine the audience for my computational document to know what I should explain and what I shouldn't.**

Suppose that each of the six O-rings is damaged with the same probability and independently of the others and that this probability depends only on the temperature. If $p(t)$ is this probability, the number $D$ of malfunctioning O-rings during a flight at temperature $t$ follows a binomial law with parameters $n = 6$ and $p = p(t)$. To link $p(t)$ to $t$, we will therefore perform a logistic regression.

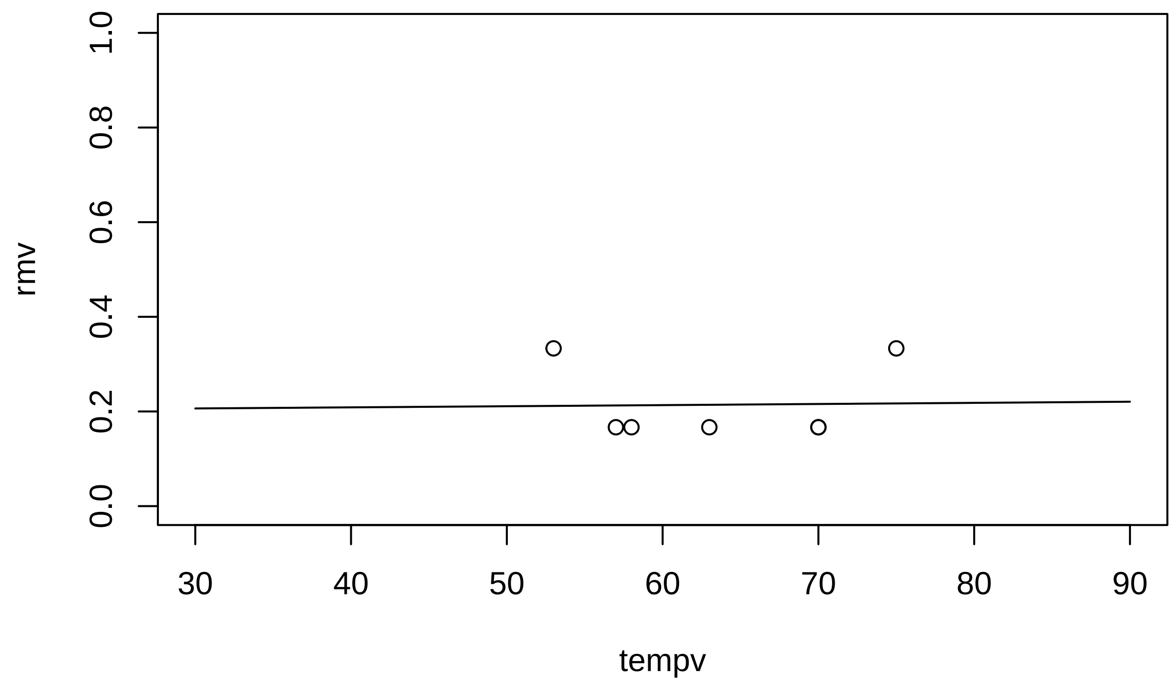
**Here as they did without the malfunction equated to 0.**

```r
logistic_reg1 = glm(data=data_without_malfunction_0, Malfunction/Count ~ Temperature, weights=Count,
                   family=binomial(link='logit'))
summary(logistic_reg1)
```

```
##
## Call:
## glm(formula = Malfunction/Count ~ Temperature, family = binomial(link = "logit"),
##     data = data_without_malfunction_0, weights = Count)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.389528   3.195752  -0.435    0.664
## Temperature  0.001416   0.049773   0.028    0.977
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1.3347  on 6  degrees of freedom
## Residual deviance: 1.3339  on 5  degrees of freedom
## AIC: 18.894
##
## Number of Fisher Scoring iterations: 4
```

**We can see the Pr(>|z|) for Temperature more than 0.05 which indicates a nonsignificant impact.**

```r
tempv = seq(from=30, to=90, by = .5)
rmv <- predict(logistic_reg1,list(Temperature=tempv),type="response")
plot(tempv,rmv,type="l",ylim=c(0,1))
points(data=data_without_malfunction_0, Malfunction/Count ~ Temperature)
```




**We can see how the estimator doesn't work very well here (which means we cant tell if the temperature has impact or not) .**

**New test :to see from where the problem of under estimation: Here we will see the estimation when we dont remove the malfunction 0.**

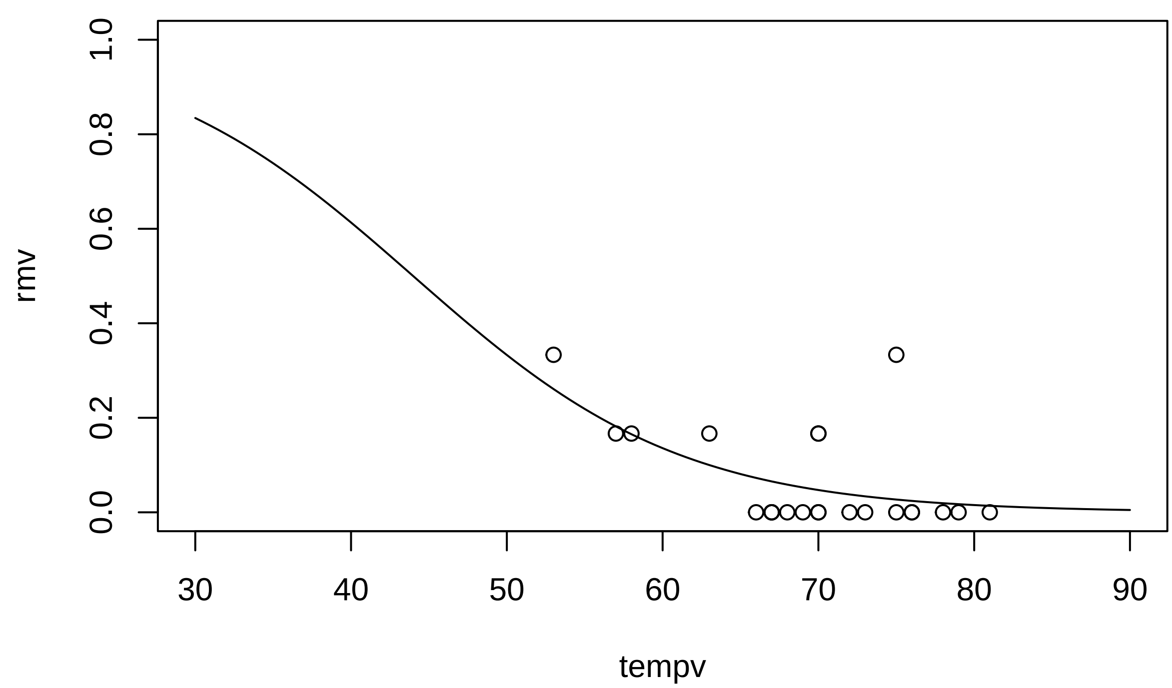
```
logistic_reg2 = glm(data=data, Malfunction/Count ~ Temperature, weights=Count,
                    family=binomial(link='logit'))
summary(logistic_reg2)
```

```
##
## Call:
## glm(formula = Malfunction/Count ~ Temperature, family = binomial(link = "logit"),
##     data = data, weights = Count)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.08498    3.05247   1.666   0.0957 .
## Temperature -0.11560    0.04702  -2.458   0.0140 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 24.230  on 22  degrees of freedom
## Residual deviance: 18.086  on 21  degrees of freedom
## AIC: 35.647
##
## Number of Fisher Scoring iterations: 5
```

**We can there is little impact from the temperature because the P(>|z|) 0.01 is less than.**

```r
# shuttle=shuttle[shuttle$r!=0,]
tempv = seq(from=30, to=90, by = .5)
rmv <- predict(logistic_reg2,list(Temperature=tempv),type="response")
plot(tempv,rmv,type="l",ylim=c(0,1))
points(data=data, Malfunction/Count ~ Temperature)
```



**we can see how with the low degree for tempreture and with malfunction 0 we can get a good estimator where with low degree give us a high probablity for failure .**

**The data is so small and they removed the malfunction when it is equal to zero, also the temperature there are not enough experiments when the temperature low degree all of them affect the estimation.**

**In conclusion, if there are human souls, we must do more experiments and not limit ourselves to experimenting with biased data, despite the change in the estimator and its giving different results when we added the 0 for the malfunction column and did not remove it, which enriched the distribution, but we still need well distributed Especially with the variables we mentioned in the prediction.**