Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Faillure

In this document we reperform some of the analysis provided in

Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure by *Siddhartha R. Dalal, Edward B. Fowlkes, Bruce Hoadley* published in *Journal of the American Statistical Association*, Vol. 84, No. 408 (Dec., 1989), pp. 945-957 and available at http://www.jstor.org/stable/2290069.

On the fourth page of this article, they indicate that the maximum likelihood estimates of the logistic regression using only temperature are: $\hat{\alpha} = 5.085$ and $\hat{\beta} = -0.1156$ their asymptotic standard errors are $s_{\hat{\alpha}} = 3.052$ and $s_{\hat{\beta}} = 0.047$. The Goodness of fit indicated for this model was $G^2 = 18.086$ with 21 degrees of freedom. Our goal is to reproduce the computation behind these values and the Figure 4 of this article, possibly in a nicer looking way.

Technical information on the computer on which the analysis is run

We will use Matlab.

```
% Peut afficher des informations personnelles
%system('systeminfo')
version
```

```
ans =
'9.6.0.1072779 (R2019a)'
```

Loading and inspecting data

Let's start by reading data.

```
data = readtable("./shuttle.csv")
```

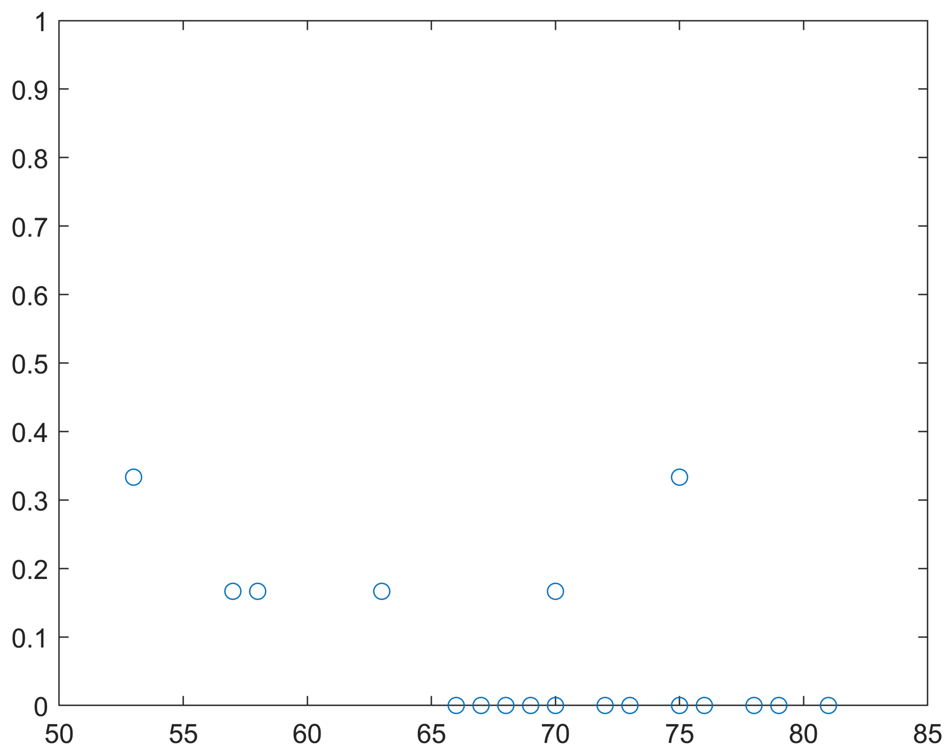
data = 23×5 table

	Date	Count	Temperature	Pressure	Malfunction
1	04/12/0...	6	66	50	0
2	11/12/0...	6	70	50	1
3	03/22/0...	6	69	50	0
4	11/11/0...	6	68	50	0
5	04/04/0...	6	67	50	0
6	06/18/0...	6	72	50	0
7	08/30/0...	6	73	100	0
8	11/28/0...	6	70	100	0
9	02/03/0...	6	57	200	1
10	04/06/0...	6	63	200	1

	Date	Count	Temperature	Pressure	Malfunction
11	08/30/0...	6	70	200	1
12	10/05/0...	6	78	200	0
13	11/08/0...	6	67	200	0
14	01/24/0...	6	53	200	2
15	04/12/0...	6	67	200	0
16	04/29/0...	6	75	200	0
17	06/17/0...	6	70	200	0
18	NaT	6	81	200	0
19	08/27/0...	6	76	200	0
20	10/03/0...	6	79	200	0
21	10/30/0...	6	75	200	2
22	11/26/0...	6	76	200	0
23	01/12/0...	6	58	200	1

We know from our previous experience on this data set that filtering data is a really bad idea. We will therefore process it as such.

```
data.Frequency = data.Malfunction./data.Count;
figure()
plot(data.Temperature,data.Frequency,"o")
ylim([0 1])
```



Logistic regression

Let's assume O-rings independently fail with the same probability which solely depends on temperature. A logistic regression should allow us to estimate the influence of temperature.

```
data.Success = data.Count - data.Malfunction;
data.Intercept = ones(size(data.Success))
```

```
data = 23×8 table
```

...

	Date	Count	Temperature	Pressure	Malfunction	Frequency	Success
1	04/12/0...	6	66	50	0	0	6
2	11/12/0...	6	70	50	1	0.1667	5
3	03/22/0...	6	69	50	0	0	6
4	11/11/0...	6	68	50	0	0	6
5	04/04/0...	6	67	50	0	0	6
6	06/18/0...	6	72	50	0	0	6
7	08/30/0...	6	73	100	0	0	6
8	11/28/0...	6	70	100	0	0	6
9	02/03/0...	6	57	200	1	0.1667	5
10	04/06/0...	6	63	200	1	0.1667	5

	Date	Count	Temperature	Pressure	Malfunction	Frequency	Success
11	08/30/0...	6	70	200	1	0.1667	5
12	10/05/0...	6	78	200	0	0	6
13	11/08/0...	6	67	200	0	0	6
14	01/24/0...	6	53	200	2	0.3333	4
15	04/12/0...	6	67	200	0	0	6
16	04/29/0...	6	75	200	0	0	6
17	06/17/0...	6	70	200	0	0	6
18	NaT	6	81	200	0	0	6
19	08/27/0...	6	76	200	0	0	6
20	10/03/0...	6	79	200	0	0	6
21	10/30/0...	6	75	200	2	0.3333	4
22	11/26/0...	6	76	200	0	0	6
23	01/12/0...	6	58	200	1	0.1667	5

```
logmodel = fitglm([data.Intercept,data.Temperature],data.Frequency,'Distribution','binomial','\
```

```
Warning: Regression design matrix is rank deficient to within machine precision.
logmodel =
Generalized linear regression model:
    logit(Frequency) ~ 1 + Intercept + Temperature
    Distribution = Binomial

Estimated Coefficients:
                   Estimate      SE        tStat       pValue
                   ________    _______    _______    _______

    (Intercept)          0           0        NaN        NaN
    Intercept        5.085       7.477    0.68008    0.49645
    Temperature    -0.1156     0.11518    -1.0036    0.31556


23 observations, 21 error degrees of freedom
Dispersion: 1
Chi^2-statistic vs. constant model: 1.02, p-value = 0.312
```

The maximum likelihood estimator of the intercept and of Temperature are thus $\hat{\alpha} = 5.0849$ and $\hat{\beta} = -0.1156$. This **corresponds** to the values from the article of Dalal *et al.* The standard errors are $s_{\hat{\alpha}} = 7.477$ and $s_{\hat{\beta}} = 0.115$, which is **different** from the 3.052 and 0.04702 reported by Dallal *et al.* The deviance is 3.01444 with 21 degrees of freedom. I cannot find any value similar to the Goodness of fit ($G^2 = 18.086$) reported by Dalal *et al.* There seems to be something wrong. Oh I know, I haven't indicated that my observations are actually the result of 6 observations for each rocket launch. Let's indicate these weights (since the weights are always the same throughout all experiments, it does not change the estimates of the fit but it does influence the variance estimates).

```
logmodel = fitglm([data.Intercept,data.Temperature],data.Frequency,'Distribution','binomial','L
```

```
Warning: Regression design matrix is rank deficient to within machine precision.
logmodel = 
Generalized linear regression model:
    logit(Frequency) ~ 1 + Intercept + Temperature
    Distribution = Binomial

Estimated Coefficients:
                   Estimate       SE        tStat       pValue 
                   ________    ________    _______    ________

    (Intercept)          0            0        NaN         NaN
    Intercept        5.085       3.0525     1.6658    0.095744
    Temperature    -0.1156     0.047024    -2.4584    0.013958


23 observations, 21 error degrees of freedom
Dispersion: 1
Chi^2-statistic vs. constant model: 6.14, p-value = 0.0132
```

Good, now I have recovered the asymptotic standard errors $s_{\hat{\alpha}} = 3.052$ and $s_{\hat{\beta}} = 0.047$. The Goodness of fit (Deviance) indicated for this model is $G^2 = 18.086$

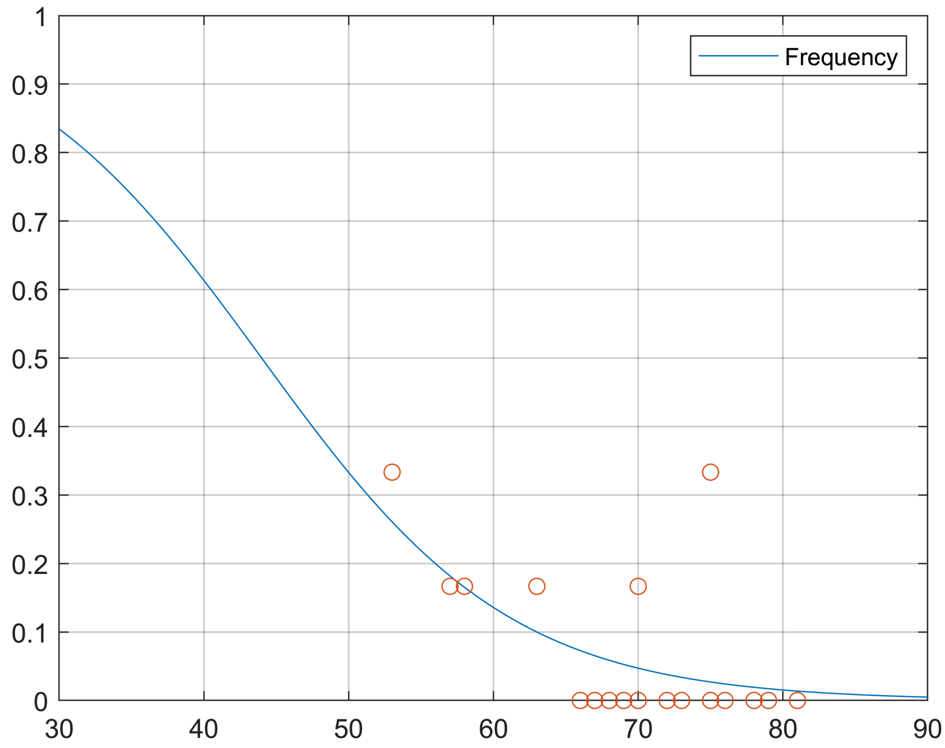
with 21 degrees of freedom (Df Residuals).

I have therefore managed to fully replicate the results of the Dalal *et al.* article.

Predicting faillure probability

The temperature when launching the shuttle was 31°F. Let's try to estimate the failure probability for such temperature using our model.:

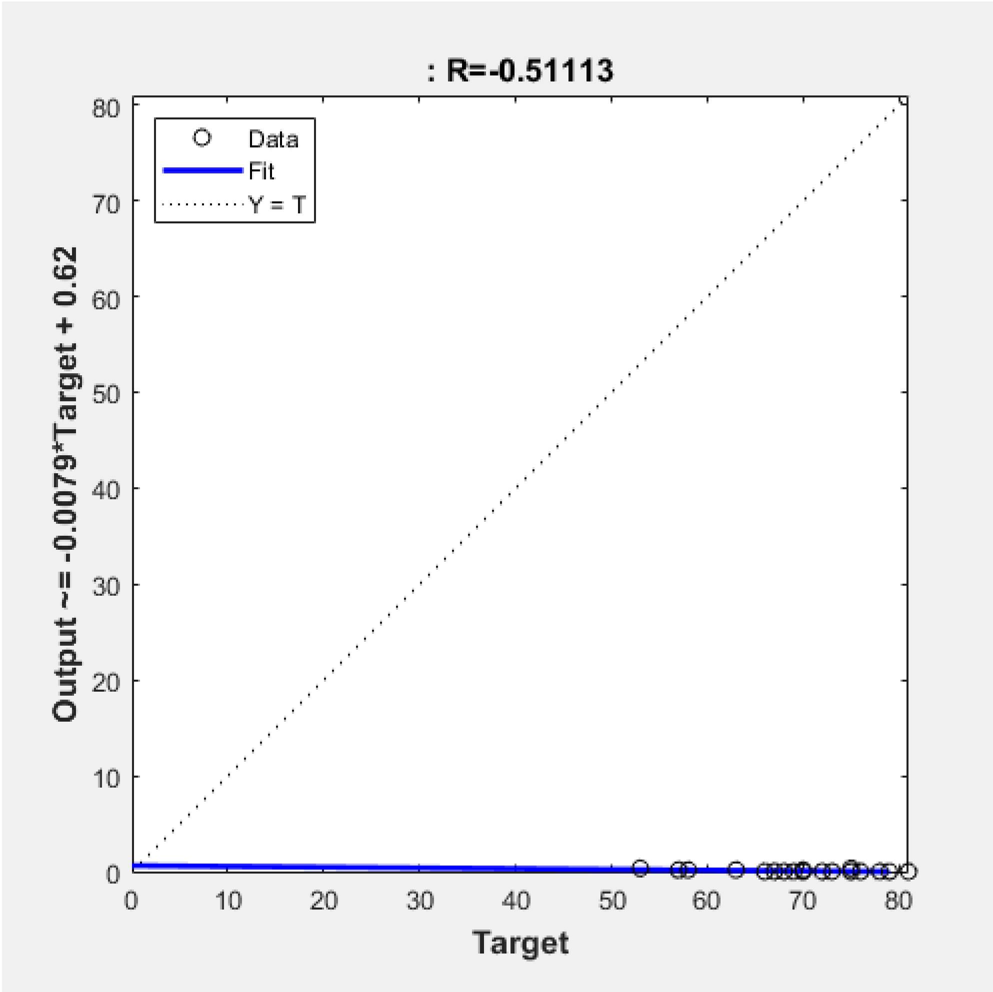
```
data_pred = table();
data_pred.Temperature = linspace(30,90,121)';
data_pred.Intercept = ones(size(data_pred.Temperature));
data_pred.Frequency = logmodel.predict(data_pred);
figure()
plot(data_pred.Temperature,data_pred.Frequency)
ylim([0 1])
hold on
scatter(data.Temperature,data.Frequency,"o")
hold off
grid
legend("Frequency")
```



This figure is very similar to the Figure 4 of Dalal *et al.* **I have managed to replicate the Figure 4 of the Dalal *et al.* article.**

Computing and ploting uncertainty

```
figure()
xlim([30 90])
ylim([0 1])
plotregression(data.Temperature,data.Frequency)
```



: R=-0.51113
Data
Fit
Y = T
Output ~= -0.0079*Target + 0.62
Target