

Analyse du risque de défaillance des joints toriques de la navette Challenger

Arnaud Legrand & Un élève

28 juin 2018 & septembre 2020

Le 27 Janvier 1986, veille du décollage de la navette *Challenger*, eu lieu une télé-conférence de trois heures entre les ingénieurs de la Morton Thiokol (constructeur d'un des moteurs) et de la NASA. La discussion portait principalement sur les conséquences de la température prévue au moment du décollage de 31°F (juste en dessous de 0°C) sur le succès du vol et en particulier sur la performance des joints toriques utilisés dans les moteurs. En effet, aucun test n'avait été effectué à cette température.

L'étude qui suit reprend donc une partie des analyses effectuées cette nuit là et dont l'objectif était d'évaluer l'influence potentielle de la température et de la pression à laquelle sont soumis les joints toriques sur leur probabilité de dysfonctionnement. Pour cela, nous disposons des résultats des expériences réalisées par les ingénieurs de la NASA durant les 6 années précédant le lancement de la navette Challenger.

Chargement des données

Nous commençons donc par charger ces données:

```
data = read.csv("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
```

Le jeu de données nous indique la date de l'essai, le nombre de joints toriques mesurés (il y en a 6 sur le lanceur principal), la température (en Fahrenheit) et la pression (en psi), et enfin le nombre de dysfonctionnements relevés.

Inspection graphique des données

Les vols où aucun incident n'est relevé n'apportant aucune information sur l'influence de la température ou de la pression sur les dysfonctionnements, nous nous concentrons sur les expériences où au moins un joint a été défectueux.

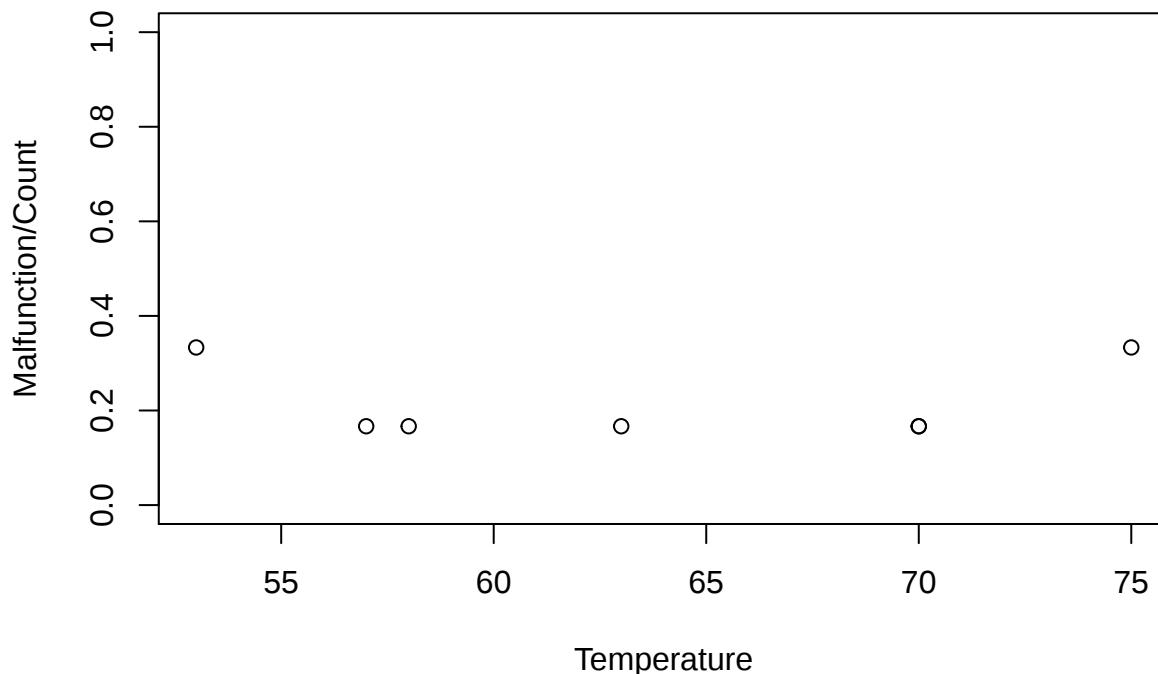
```
data = data[data$Malfunction>0,]  
data
```

##	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

Très bien, nous avons une variabilité de température importante mais la pression est quasiment toujours égale à 200, ce qui devrait simplifier l'analyse.

Comment la fréquence d'échecs varie-t-elle avec la température ?

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



À première vue, ce n'est pas flagrant mais bon, essayons quand même d'estimer l'impact de la température t sur la probabilité de dysfonctionnements d'un joint.

Estimation de l'influence de la température

Supposons que chacun des 6 joints toriques est endommagé avec la même probabilité et indépendamment des autres et que cette probabilité ne dépend que de la température. Si on note $p(t)$ cette probabilité, le nombre de joints D dysfonctionnant lorsque l'on effectue le vol à température t suit une loi binomiale de paramètre $n = 6$ et $p = p(t)$. Pour relier $p(t)$ à t , on va donc effectuer une régression logistique.

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

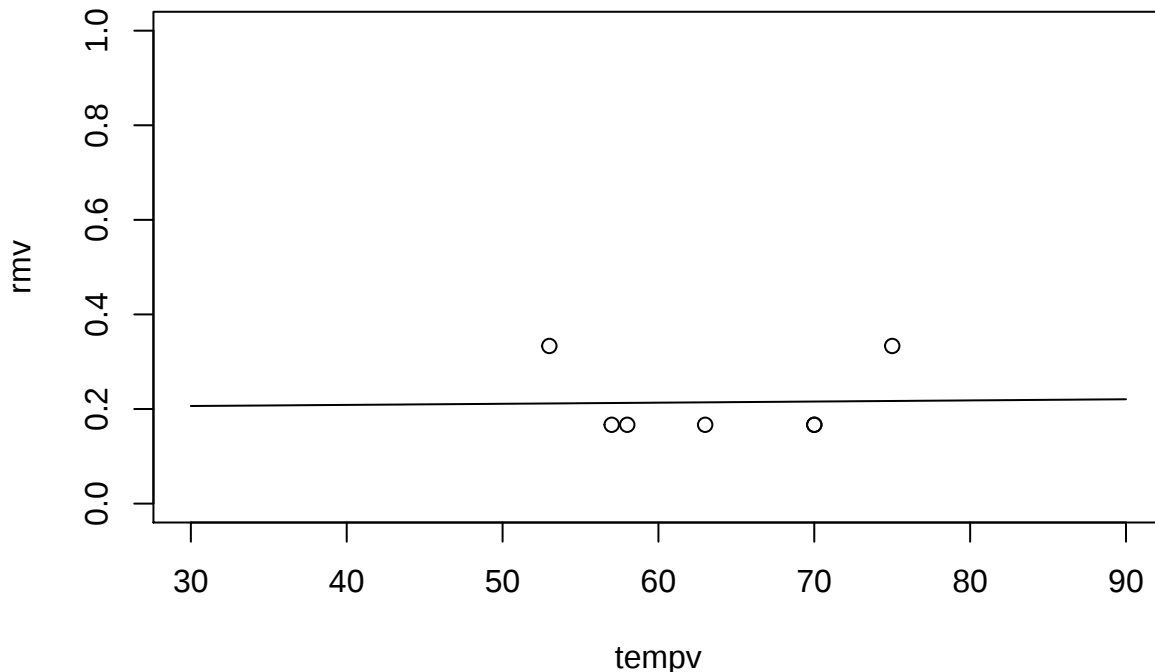
##
## Call:
## glm(formula = Malfunction/Count ~ Temperature, family = binomial(link = "logit"),
##      data = data, weights = Count)
##
## Deviance Residuals:
##      2      9     10     11     14     21     23
## -0.3015 -0.2836 -0.2919 -0.3015  0.6891  0.6560 -0.2850
##
## Coefficients:
##              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
```

L'estimateur le plus probable du paramètre de température est 0.001416 et l'erreur standard de cet estimateur est de 0.049, autrement dit on ne peut pas distinguer d'impact particulier et il faut prendre nos estimations avec des pincettes.

Estimation de la probabilité de dysfonctionnant des joints toriques

La température prévue le jour du décollage est de 31°F. Essayons d'estimer la probabilité de dysfonctionnement des joints toriques à cette température à partir du modèle que nous venons de construire:

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



Comme on pouvait s'attendre au vu des données initiales, la température n'a pas d'impact notable sur la probabilité d'échec des joints toriques. Elle sera d'environ 0.2, comme dans les essais précédents où nous il y a eu défaillance d'au moins un joint. Revenons à l'ensemble des données initiales pour estimer la probabilité de défaillance d'un joint:

```
data_full = read.csv("shuttle.csv",header=T)
sum(data_full$Malfunction)/sum(data_full$Count)
```

```
## [1] 0.06521739
```

Cette probabilité est donc d'environ $p = 0.065$, sachant qu'il existe un joint primaire un joint secondaire sur chacune des trois parties du lanceur, la probabilité de défaillance des deux joints d'un lanceur est de $p^2 \approx 0.00425$. La probabilité de défaillance d'un des lanceur est donc de $1 - (1 - p^2)^3 \approx 1.2$. Ça serait vraiment pas de chance... Tout est sous contrôle, le décollage peut donc avoir lieu demain comme prévu.

Seulement, le lendemain, la navette Challenger explosera et emportera avec elle ses sept membres d'équipages. L'opinion publique est fortement touchée et lors de l'enquête qui suivra, la fiabilité des joints toriques sera directement mise en cause. Au delà des problèmes de communication interne à la NASA qui sont pour beaucoup dans ce fiasco, l'analyse précédente comporte (au moins) un petit problème... Saurez-vous le trouver ? Vous êtes libre de modifier cette analyse et de regarder ce jeu de données sous tous les angles afin d'expliquer ce qui ne va pas.

Contre expertise

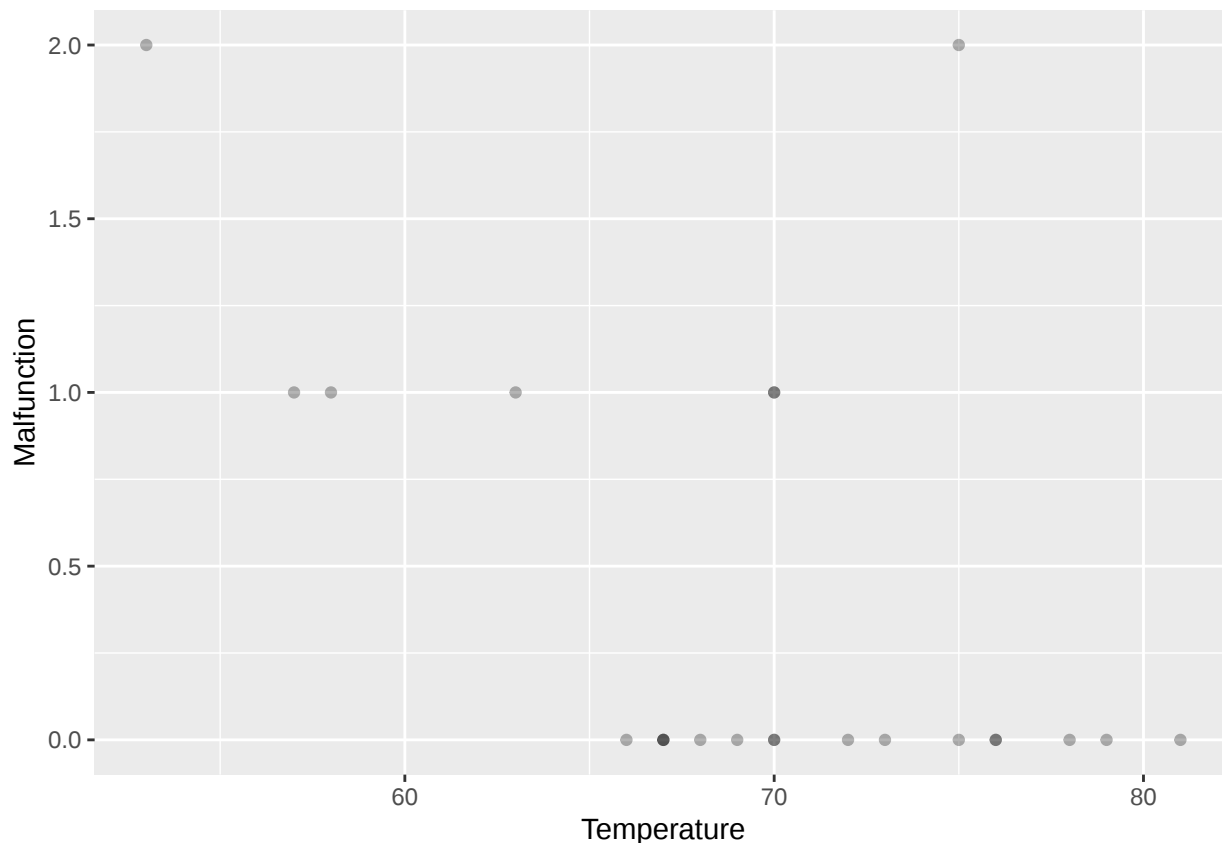
```
library("ggplot2")
```

Commençons par charger et visualiser les données brut, sans aucun filtre ou pré traitement.

```
ce_data = read.csv("shuttle.csv",header=T)
```

On commence par le plus évident, une graphie qui met en relation la température et les défaillances.

```
ggplot(ce_data, aes(x=Temperature, y=Malfunction)) + geom_point(alpha=.3)
```



On remarque que en dessous de ~65°F, aucun test n'est exempt de défaillance. Ce premier graphe est prometteur.

Essayons de reproduire la régression logistique de l'analyse initiale.

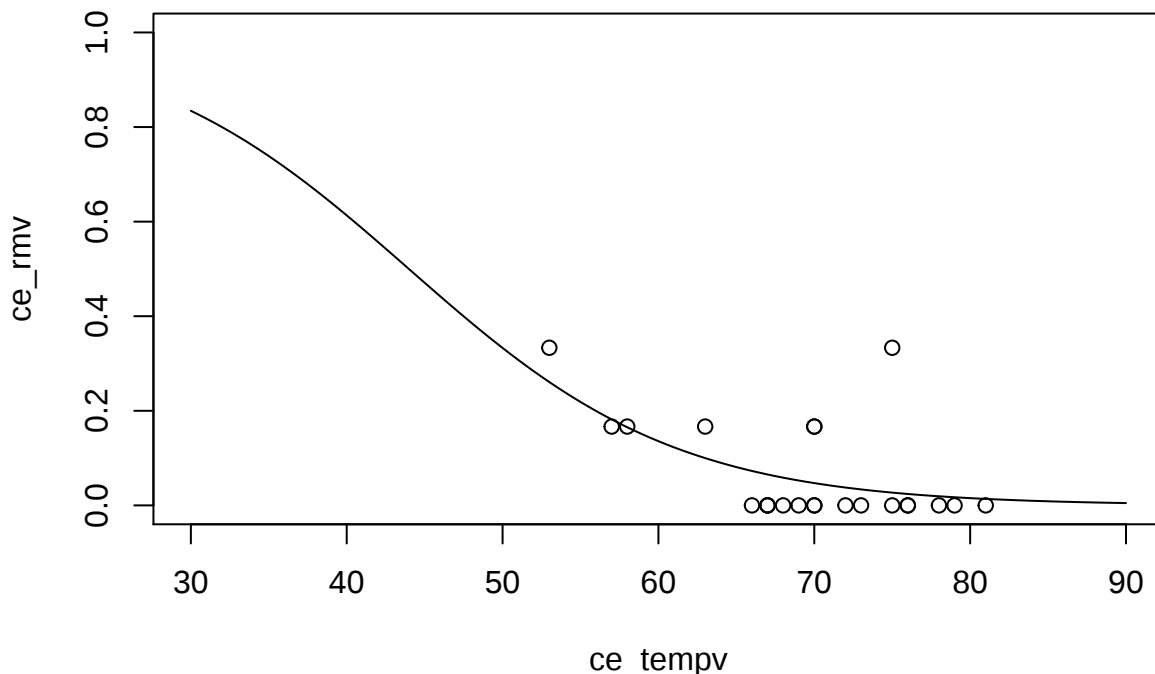
```
ce_logistic_reg = glm(
  data=ce_data,
  Malfunction/Count ~ Temperature,
  weights=Count,
  family=binomial(link='logit')
)
summary(ce_logistic_reg)
```

```
##
## Call:
## glm(formula = Malfunction/Count ~ Temperature, family = binomial(link = "logit"),
##      data = ce_data, weights = Count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95227  -0.78299  -0.54117  -0.04379   2.65152
##
## 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
```

Comme on est une grosse bille en statistique, on a aucune idée de ce que ce rapport nous indique. Nous allons donc refaire la même opération que dans l'analyse initiale, et on verra bien ce que ça raconte.

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



Bon, c'est cool ça quand même. Une belle courbe qui tend vers 1 à l'approche de 30°F. Mais on va quand même s'arrêter là, par ce que j'ai aucune idée de l'indice de confiance de la prédiction. Je suppose que je devrais suivre un MOOC sur les statistiques.

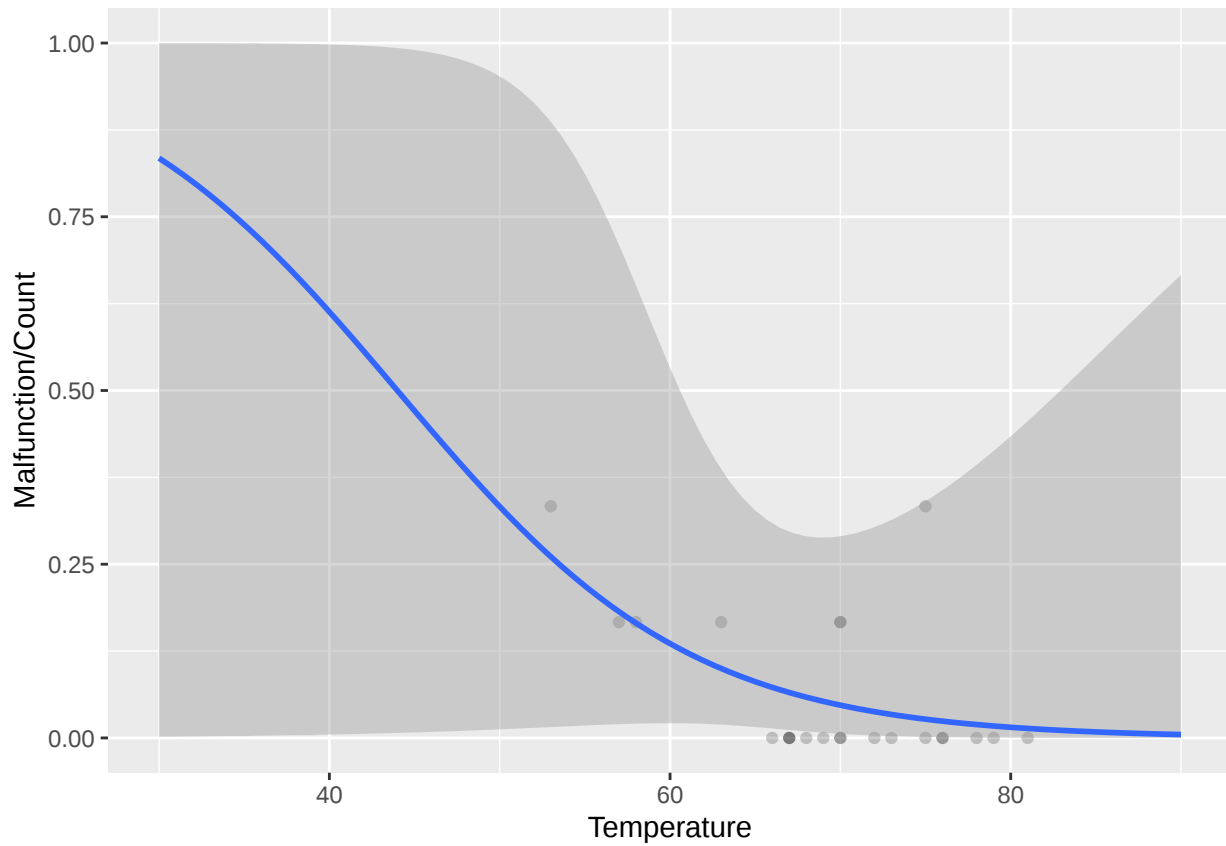
Avec geom_smooth

Je viens de repenser au document exemple sur la régression logistique et je me souviens que ça affichait l'incertitude. On va essayer de faire pareil.

```
ggplot(ce_data, aes(x=Temperature, y=Malfunction/Count)) +
  geom_point(alpha=.2) +
  geom_smooth(method="glm", method.args=list(family="binomial"), fullrange=TRUE) +
  xlim(30, 90)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```



Eh beh voila ! C'est ce dont j'avais peur. On a aucune vrai info à par qu'on sait pas. Alors, si je me souviens bien de mes cours de master on est sensé pouvoir enlever les outliers avec une loi normale. Si c'est abordé dans le module 4, je reviendrais peut-être ajouter des choses ici.