exo5_en

January 4, 2025

1 Analysis of the risk of failure of the O-rings on the Challenger shuttle

On January 27, 1986, the day before the takeoff of the shuttle *Challenger*, had a three-hour tele-conference was held between Morton Thiokol (the manufacturer of one of the engines) and NASA. The discussion focused on the consequences of the temperature at take-off of 31°F (just below 0°C) for the success of the flight and in particular on the performance of the O-rings used in the engines. Indeed, no test had been performed at this temperature.

The following study takes up some of the analyses carried out that night with the objective of assessing the potential influence of the temperature and pressure to which the O-rings are subjected on their probability of malfunction. Our starting point is the results of the experiments carried out by NASA engineers during the six years preceding the launch of the shuttle Challenger.

1.1 Loading the data

We start by loading this data:

```
[1]: import numpy as np
     import pandas as pd
     data = pd.read_csv("shuttle.csv")
     data
```

```
[1]:          Date  Count  Temperature  Pressure  Malfunction
     0      4/12/81      6           66        50            0
     1     11/12/81      6           70        50            1
     2      3/22/82      6           69        50            0
     3     11/11/82      6           68        50            0
     4      4/04/83      6           67        50            0
     5      6/18/82      6           72        50            0
     6      8/30/83      6           73       100            0
     7     11/28/83      6           70       100            0
     8      2/03/84      6           57       200            1
     9      4/06/84      6           63       200            1
     10     8/30/84      6           70       200            1
     11    10/05/84      6           78       200            0
     12    11/08/84      6           67       200            0
```

13	1/24/85	6	53	200	2
14	4/12/85	6	67	200	0
15	4/29/85	6	75	200	0
16	6/17/85	6	70	200	0
17	7/29/85	6	81	200	0
18	8/27/85	6	76	200	0
19	10/03/85	6	79	200	0
20	10/30/85	6	75	200	2
21	11/26/85	6	76	200	0
22	1/12/86	6	58	200	1

The data set shows us the date of each test, the number of O-rings (there are 6 on the main launcher), the temperature (in Fahrenheit) and pressure (in psi), and finally the number of identified malfunctions.

1.2 Graphical inspection

Flights without incidents do not provide any information on the influence of temperature or pressure on malfunction. We thus focus on the experiments in which at least one O-ring was defective.

```
[2]: data = data[data.Malfunction>0]
     data
```

```
[2]:         Date  Count  Temperature  Pressure  Malfunction
     1   11/12/81      6           70        50            1
     8    2/03/84      6           57       200            1
     9    4/06/84      6           63       200            1
     10   8/30/84      6           70       200            1
     13   1/24/85      6           53       200            2
     20  10/30/85      6           75       200            2
     22   1/12/86      6           58       200            1
```

We have a high temperature variability but the pressure is almost always 200, which should simplify the analysis.

How does the frequency of failure vary with temperature?

```
[3]: %matplotlib inline
     pd.set_option('mode.chained_assignment',None) # this removes a useless warning
      ↪from pandas
     import matplotlib.pyplot as plt

     data["Frequency"]=data.Malfunction/data.Count
     data.plot(x="Temperature",y="Frequency",kind="scatter",ylim=[0,1])
     plt.grid(True)
```
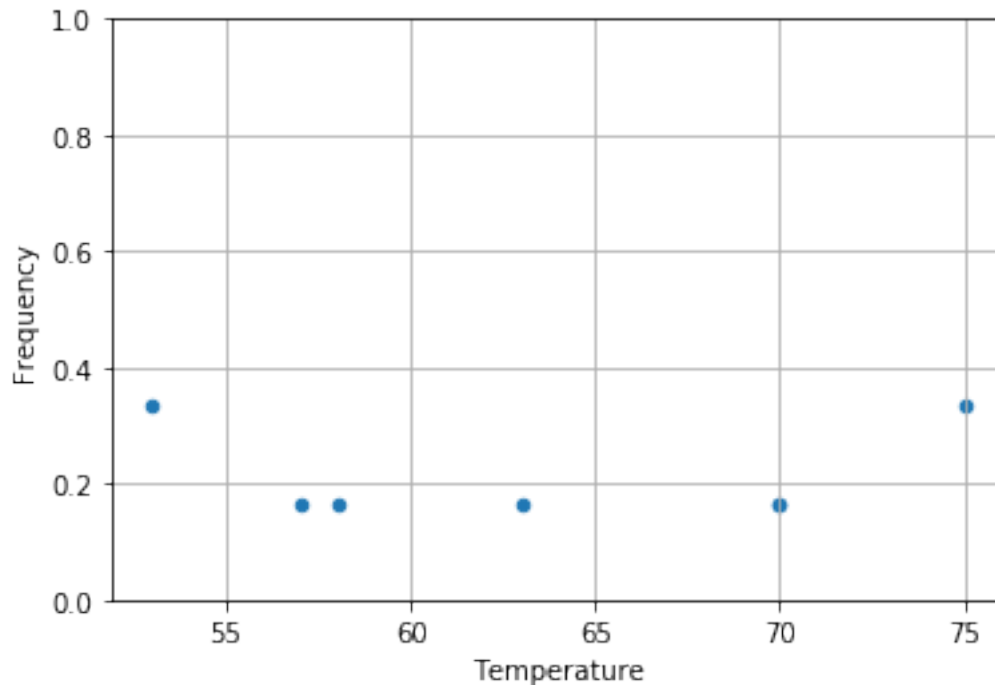
At first glance, the dependence does not look very important, but let's try to estimate the impact of temperature t on the probability of O-ring malfunction.

1.3 Estimation of the temperature influence

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.

```
[4]: import statsmodels.api as sm

data["Success"]=data.Count-data.Malfunction
data["Intercept"]=1

logmodel=sm.GLM(data['Frequency'], data[['Intercept','Temperature']], family=sm.
 ↪families.Binomial(sm.families.links.logit)).fit()

logmodel.summary()
```

```
[4]: <class 'statsmodels.iolib.summary.Summary'>
     """
                     Generalized Linear Model Regression Results
     ==============================================================================
```

```
Dep. Variable:              Frequency   No. Observations:              7
Model:                            GLM   Df Residuals:                  5
Model Family:                Binomial   Df Model:                      1
Link Function:                  logit   Scale:                    1.0000
Method:                          IRLS   Log-Likelihood:          -2.5250
Date:               Sat, 13 Apr 2019   Deviance:                0.22231
Time:                        19:12:05   Pearson chi2:              0.236
No. Iterations:                     4   Covariance Type:       nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      -1.3895      7.828     -0.178      0.859     -16.732      13.953
Temperature     0.0014      0.122      0.012      0.991      -0.238       0.240
==============================================================================
"""
```

The most likely estimator of the temperature parameter is 0.0014 and the standard error of this estimator is 0.122, in other words we cannot distinguish any particular impact and we must take our estimates with caution.

1.4 Estimation of the probability of O-ring malfunction

The expected temperature on the take-off day is 31°F. Let's try to estimate the probability of O-ring malfunction at this temperature from the model we just built:

```python
[5]: %matplotlib inline
data_pred = pd.DataFrame({'Temperature': np.linspace(start=30, stop=90,
  ↪num=121), 'Intercept': 1})
data_pred['Frequency'] = logmodel.
  ↪predict(data_pred[['Intercept','Temperature']])
data_pred.plot(x="Temperature",y="Frequency",kind="line",ylim=[0,1])
plt.scatter(x=data["Temperature"],y=data["Frequency"])
plt.grid(True)
```
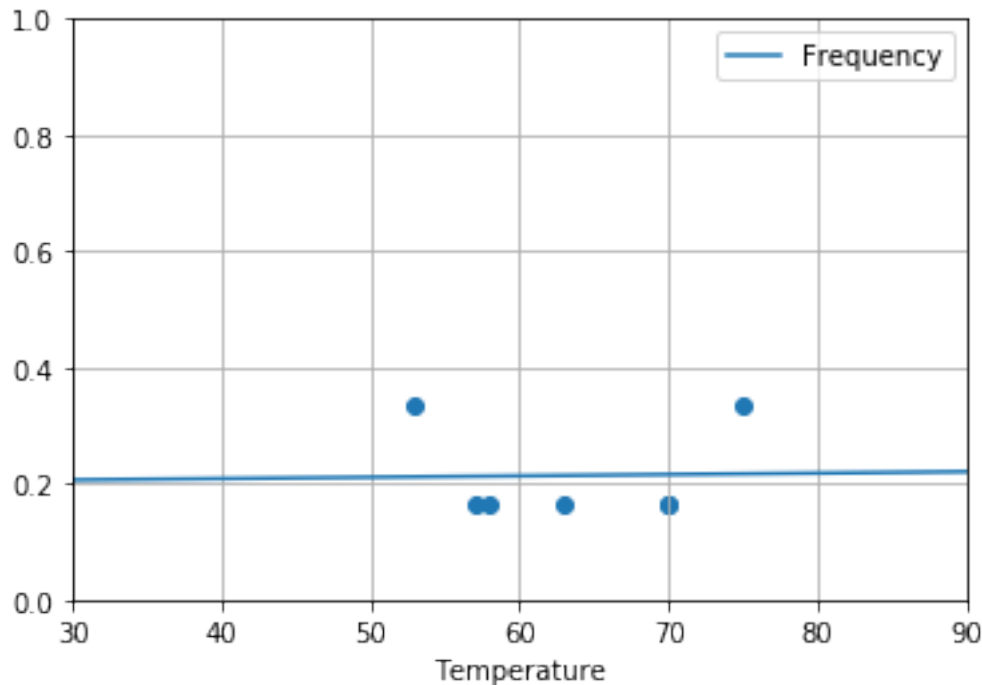
As expected from the initial data, the temperature has no significant impact on the probability of failure of the O-rings. It will be about 0.2, as in the tests where we had a failure of at least one joint. Let's get back to the initial dataset to estimate the probability of failure:

```
[6]: data = pd.read_csv("shuttle.csv")
     print(np.sum(data.Malfunction)/np.sum(data.Count))
```

0.06521739130434782

This probability is thus about $p = 0.065$. Knowing that there is a primary and a secondary O-ring on each of the three parts of the launcher, the probability of failure of both joints of a launcher is $p^2 \approx 0.00425$. The probability of failure of any one of the launchers is $1 - (1 - p^2)^3 \approx 1.2$. That would really be bad luck…. Everything is under control, so the takeoff can happen tomorrow as planned.

But the next day, the Challenger shuttle exploded and took away with her the seven crew members. The public was shocked and in the subsequent investigation, the reliability of the O-rings was questioned. Beyond the internal communication problems of NASA, which have a lot to do with this fiasco, the previous analysis includes (at least) a small problem…. Can you find it? You are free to modify this analysis and to look at this dataset from all angles in order to to explain what's wrong.

```
[ ]:
```