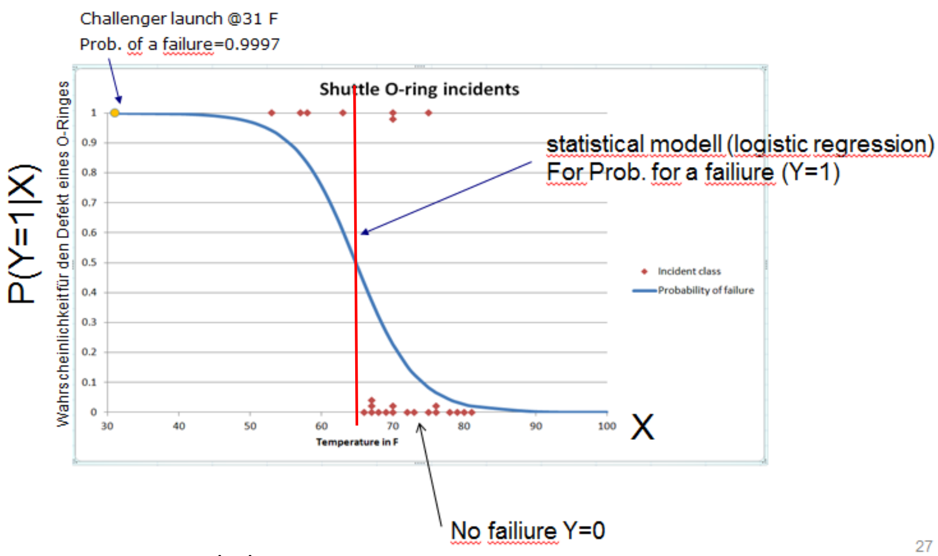
In-class exercise week 10
Topic: Using a logistic regression for binary classification

1) Challenger accident

We want to predict if a O-ring will break (Y=1) depending on the start temperature x during takeoff of the challenger. The probability for Y=1 can be estimated by a logistic regression model which is visualized below.



$$\hat{\pi}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X}}$$

a) Lets assume β_1 is given with -1. Guess an appropriate value for β_0.

Hint: at which x value should $p(y_i = 1|x_i)$ be 0.5? Look at the data!
At this x value the denominator must be twice as big as the nominator.

At x ≈ 65 we expect to be p=0.5,
to get 0.5 we require:

$$e^{\hat{\beta}_0 - 65} \overset{!}{=} 1$$

$$\hat{\beta}_0 - 65 \overset{!}{=} 0$$

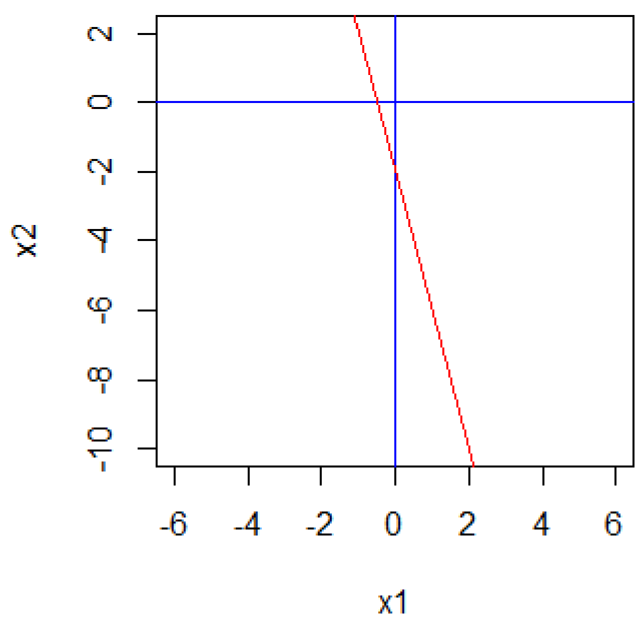
$$\hat{\beta}_0 \overset{!}{=} 65$$

Remark: if we would fit these data in R, we would get ≈65 for the intercept and ≈-1 for the coefficient β_1.

b) In the example above we only had one predictor leading to a cutoff at about 65° Fahrenheit, indicating that for temperatures below this cutoff we would predict a damage at the o-rings.
Assume we have a second predictor x_2 in the logistic regression model and have the following estimated model:

$$\ln\left(\frac{p}{1-p}\right) = 1 + 2x_1 + 0.5x_2$$

Determine the separation curve between Y=1 and Y=0 in the room which is spanned by x_1 and x_2 and draw it in the following plot x_2 and x_2.
Hint: on the separation curve should hold: $p(y_i = 1|x_i)=0.5$
-> plug in 0.5 for p and solve for x_2.



$$\ln\left(\frac{0.5}{1-0.5}\right) = 0 = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$x_2 = -\frac{\beta_0}{\beta_2} - \frac{\beta_1}{\beta_2} \cdot x_1$$

$$x_2 = -2 - 4 \cdot x_1$$

Remark: with a logistic regression model we only can model linear separation bounderies (lines or hyper-planes in case of >2 predictors). Depending on the values of the predictors x1 and x2 we have an observation on one or die other side of the boundery. On one side of the boundary the model prdicts outcome y=1 (p>0.5) and on the other side the model predicts y=0 (p<0.5).