

40.016: Analytics Edge

Week 3 Lecture 1

RISK ANALYSIS: CHALLENGER DISASTER

Term 5, 2022



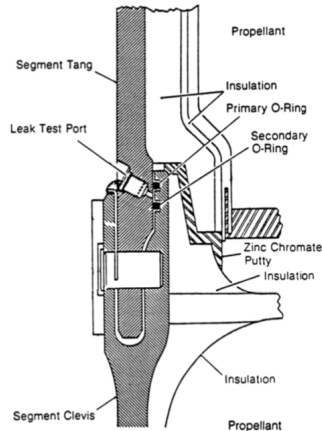
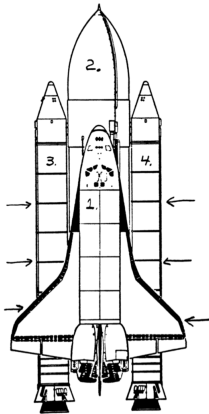
SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

The Challenger space shuttle

- January 28, 1986.
- 7 crew members.
- Included a teacher (Teacher in space program).
- One of the objectives: observe Halley's comet.

If you are interested in this topic, check out: *Challenger: The Final Flight* – Netflix

Challenger schematic and the O-Ring



Left: Schematic of Challenger shuttle. Subsystem 1 houses crew and system control, subsystem 2 is the external liquid-fuel tank, and subsystems 3 and 4 are the solid rocket motors (Morton Thiokol).

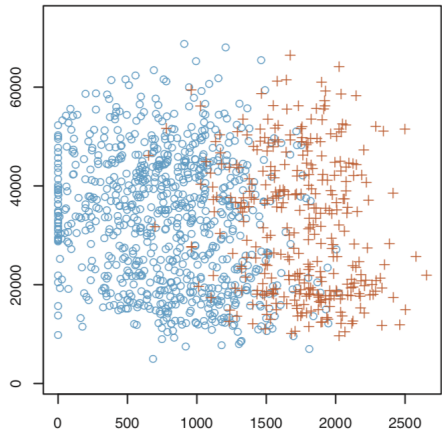
Right: Solid Rocket Motor Subsection (Morton Thiokol).

Challenger disaster:

Rogers commission report *excerpt*

The report of the fateful series of NASA and Morton Thiokol meetings, telephone conferences, notes and facsimile transmissions on January 27, the night before the launch of 51-L, shows that only limited consideration was given to the past history of O-ring damage in terms of temperature. The managers compared as a function of temperature, the flights for which thermal distress of O-rings had been observed - not the frequency of occurrences based on all flights. In such a comparison, there is nothing irregular in the distribution of O-ring distress over the spectrum of joint temperatures at launch between 53° and 75° Fahrenheit.

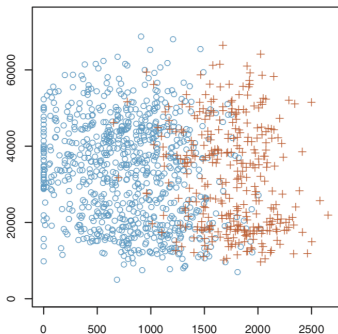
Classification



Classification

- Objective: Predict binary/ categorical variables with some respondent, explanatory variables.
- O-Ring failure \sim Temperature, Pressure
- Many methods: Linear discriminant analysis, perceptrons, even linear regression.
- We concentrate on **Logistic Regression**.

Logistic Regression



❶ n = number of observations

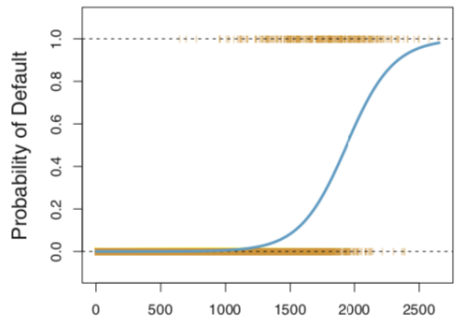
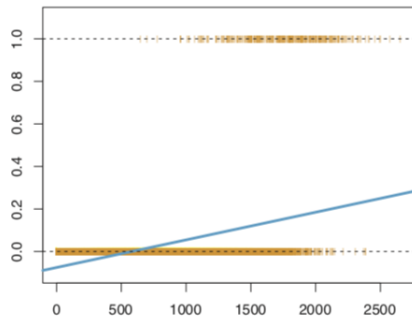
❷ p = number of predictor variables

❸ $y \in \{0, 1\}$: dependent variable (binary outcome)

❹ x_1, \dots, x_p = independent variables (predictors)

We want to model $y \sim x_1, \dots, x_p$.

Linear regression vs. Logistic regression



Logistic regression

Define

- $\boldsymbol{\beta}^\top = (\beta_0, \dots, \beta_p)$.
- $\boldsymbol{x}^\top = (1, x_1, \dots, x_p)$.

Logistic regression

Data: (y_i, \mathbf{x}_i) for $i = 1, \dots, n$ where $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$.

Quality of fit: deviance and AIC

- **Null deviance:**
- **Residual deviance:**
- **AIC:**

Quality of fit: confusion matrix

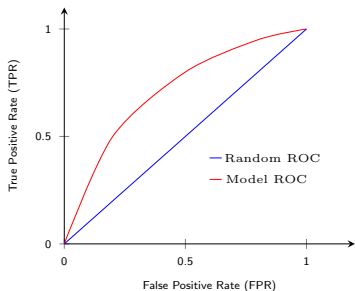
Predict based on a threshold t and the estimated $\hat{\beta}$.

	Actual = 0	Actual = 1
Predict = 0	True Negative (TN)	False Negative (FN)
Predict = 1	False Positive (FP)	True Positive (TP)

- True negative rate: TNR =
- False positive rate: FPR =
- True positive rate: TPR =
- False negative rate: FNR =
- Overall Accuracy: Accuracy =

Quality of fit:

Receiver operating characteristic (ROC) curve



Quality of fit: Area under the curve (AUC)

