Advanced STATS and R
GLM's

The University of Oklahoma

The Exponential Family

- We have been looking at the linear model of the form:

$$E(Y_i) = \mu_i = x_i'\beta$$

Or $E(Y) = X\beta$ where X is the design matrix and x_i' is the ith row of the design matrix X

Where

$$Y_i \overset{\text{iid}}{\sim} N(\mu_i, \sigma^2)$$

This would be our typical multiple regression model.

Modification

- We can modify the form of this model to make it a little more general.
- Distribution no longer simply Normal but Exponential family
- $g(\mu_i) = x_i'\beta$
- g is called the link function

Exponential Family

- Suppose that the distribution of Y depends on only one parameter θ
- Then we could define Y to be distributed within the exponential family of distributions if:

$$f(y|\theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

- Or equivalently

$$f(y|\theta) = \exp\Big(\big(a(y)b(\theta) + c(\theta) + d(y)\big)\Big)$$

Another form of the exponential family called the exponential dispersion family (EDF).

$$f(y|\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

θ is called the canonical parameter – represents location

ϕ is the dispersion parameter and represents the scale

This will have some consequences

Mean and Variance for EDF

$$E(Y) = \mu = b'(\theta)$$
$$V(Y) = b''(\theta)a(\phi)$$

$b''(\theta)$ describes how the variance of the response relates to the mean.

In the case of the normal $b''(\theta) = 1$ and hence the variance is independent of the mean.

We shall do some more algebra later on the EDF

Modeling the GLM - 3 parts

The three components:

1. **A random response** Y that is a member of the exponential family (usually EDF)
2. **A linear predictor** $\eta = X\beta$
3. **A link function**

$$g(\mu_i) = \sum_j \beta_j x_{ij}\,, i = 1, \dots, n$$

$$\mu_i = E(Y_i)$$

$$g(\mu_i) = \eta_i$$

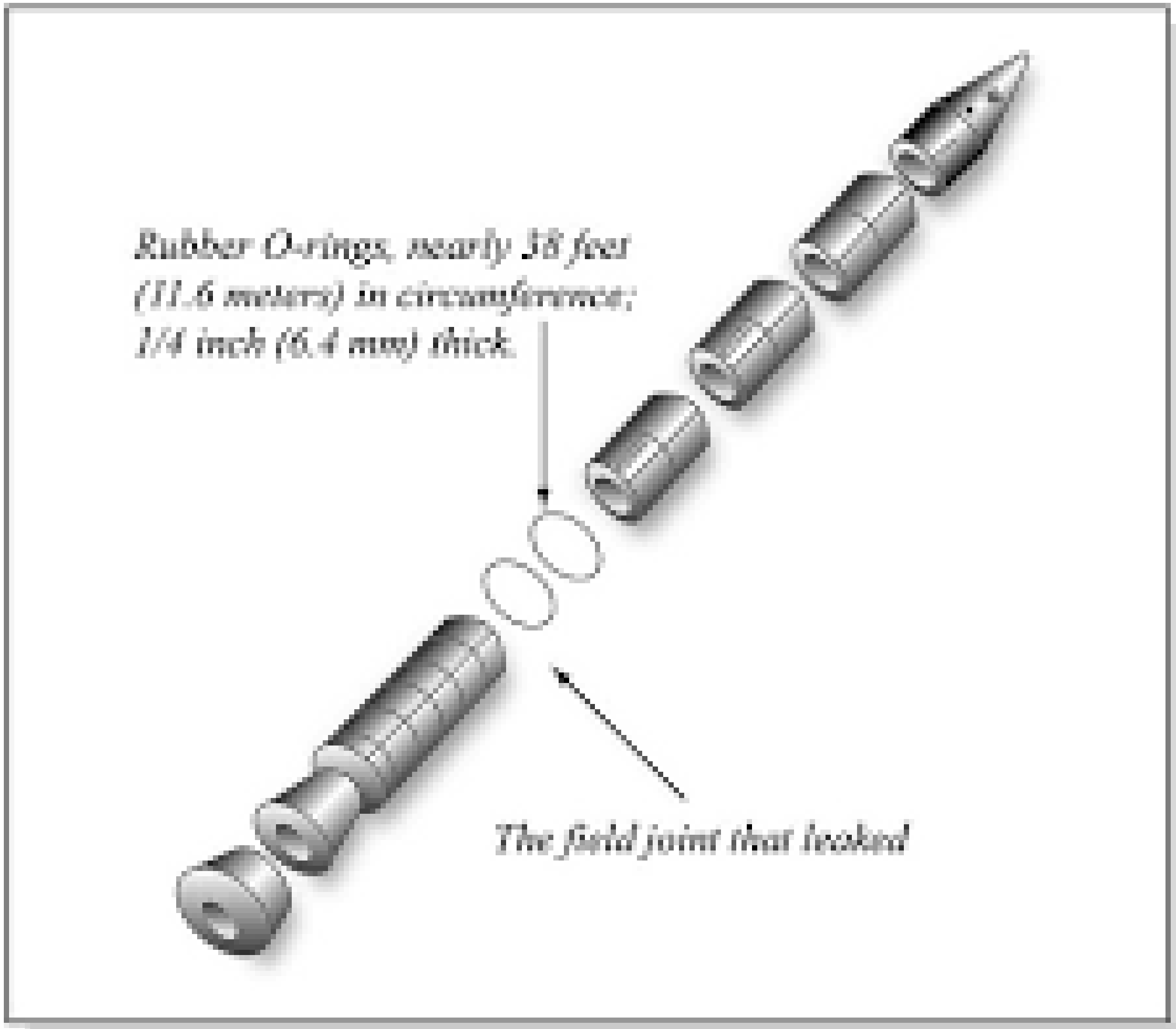
Canonical link

$$\eta = g(\mu) = \theta$$

Challenger disaster


FEBRUARY 10, 1986
$1.95
TIME
SPACE SHUTTLE
CHALLENGER
JANUARY 28, 1986

O Rings


Rubber O-rings, nearly 38 feet
(11.6 meters) in circumference;
1/4 inch (6.4 mm) thick.
The field joint that leaked

A major malfunction

Challenger's brief flight

.678 seconds

Following Challenger's liftoff, a puff of black smoke — seen only by automatic launch cameras — indicates a problem with one of the O-ring seals at the joint between segments of the shuttle's right-hand solid rocket booster.

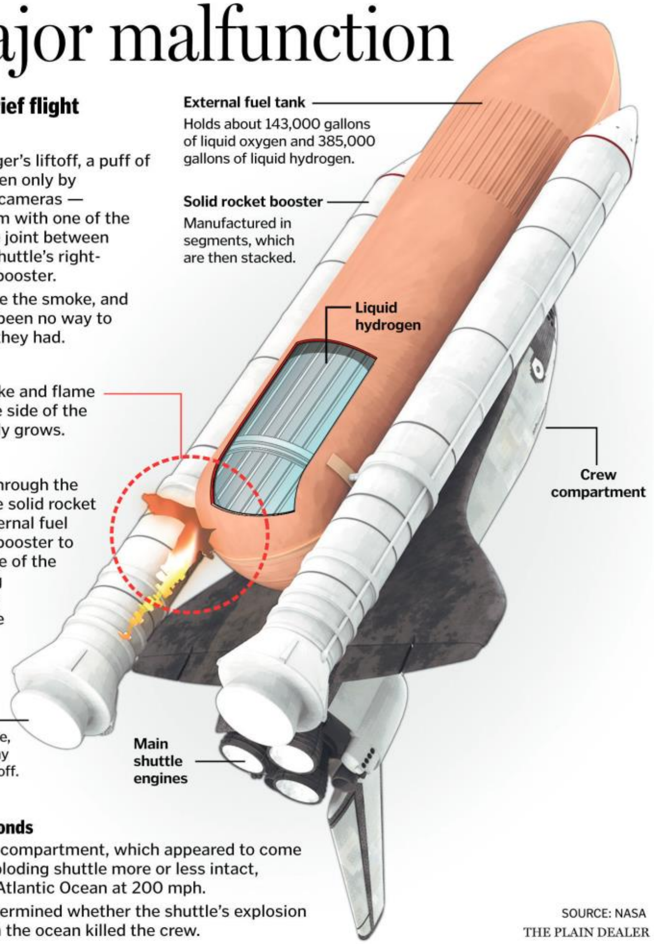
No human eyes see the smoke, and there would have been no way to abort the flight if they had.

58 seconds

A small jet of smoke and flame bursts through the side of the booster and quickly grows.

73 seconds

The flame burns through the strut attaching the solid rocket booster to the external fuel tank, causing the booster to swivel into the side of the tank. The resulting massive explosion destroys the space shuttle.



3 minutes, 58 seconds

Challenger's crew compartment, which appeared to come away from the exploding shuttle more or less intact, smashes into the Atlantic Ocean at 200 mph.

Officials never determined whether the shuttle's explosion or the impact with the ocean killed the crew.

SOURCE: NASA

THE PLAIN DEALER

Data

> head(orings)

```
   temp damage
1    53      5
2    57      1
3    58      1
4    63      1
5    66      0
6    67      0
```

Bring your laptop or use Network.

The University of Oklahoma

Get the Data and Book

- Link below:
- statsandr.oucreate.com

Courses

- **Bayesian Stats MATH 4803/5803**
- Advanced Applied STATS MATH 4793/5793 (Next Year)
- **Applied Statistical Methods MATH 4753**