

22S:105
Statistical Methods and Computing

Introduction

Lecture 1
 January 19, 2011

Kate Cowles
 374 SH, 335-0727
 kcowles@stat.uiowa.edu

What is statistics?

- Statistics is the science of using data to make decisions and answer questions.
- Statistics involves
 - designing studies
 - collecting data
 - organizing and analyzing data
 - interpreting and reporting results

**The Challenger:
 How understanding of statistical
 methods might have prevented a
 tragedy**

References:

Dalal, SR, Fowlkes, EB, Hoadley, B. (1989) "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure." *Journal of the American Statistical Association*, **84**, 945-957.

Tufte, Edward R. (1997) "The Decision to Launch the Space Shuttle Challenger," in *Visual and Statistical Thinking: Displays of Evidence for Making Decisions*, Graphics Press

On 1/28/86 space shuttle Challenger exploded during launch

- 7 astronauts killed
- reason: gas leak through a joint that should have been sealed by two rubber O-rings
 - O-rings had lost resiliency due to cold temperature

On the previous day, extensive discussions of whether or not it would be safe to launch

- predicted temperature for launch time: 26-29°
- no shuttle had ever been launched at temperature lower than 53°
- engineers who designed rocket faxed to NASA a recommendation not to launch due to risk of O-ring failure at low temperatures
- NASA officials pointed out weaknesses of engineers' evidence
- after lengthy discussion, managers of rocket-making company changed their minds and recommended launch

The engineers' plot of data from previous shuttle launches: joint temperature vs. number of O-rings having some temperature-related problems

The engineers' evidence

- history of serious but non-catastrophic O-ring damage during previous cool-weather launches
- physics of resiliency of rubber
- experimental data

What was missing from the engineers' argument?

- quantification of the relationship between joint temperature and O-ring failure
- prediction of the probability of O-ring failure at 29°, with assessment of degree of uncertainty

an appropriate statistical method: logistic regression

- Dalal et al. carried out such an analysis (after the fact) using data from the 23 shuttle launches prior to the Challenger
- found strong statistical evidence of a temperature effect on O-rings
- we will analyze these data later in the semester

A plot showing data from all 23 previous launches, including those in which no O-rings were damaged

Subjects, observations, and variables

In statistical studies, we generally choose a set of **individuals** or **subjects** on whom data is collected.

We usually are interested in collecting a number of different kinds of information to describe each subject.

A **variable** is a particular characteristic that may take on different values for different subjects. For example,

- age
- gender
- diagnosis

are three variables that might be included in a study of length of hospital stays of hospital patients.

For analysis by a computer, a set of data collected for a study is often organized as a table with a row for each subject and a column for each variable.

Pat id	age	sex	diagnosis
101	25	F	hepatitis A
102	38	F	cirrhosis
103	76	M	hepatitis C

Each row in such a table, corresponding to the data for a single subject, is called an **observation**.

Types of variables

- Qualitative (textbook calls this “categorical”)
 - **Nominal**
 - * values fall into *unordered* categories
 - * numbers may be used to represent categories, but they are just labels
 - * example: variable called “occupational area” coded as
 - 1 = education
 - 2 = business
 - 3 = service
 - 4 = industry
 - etc., etc.
 - * special case: **binary** data, which can take on only 2 possible values
 - **Ordinal**
 - * data representing *ordered* categories
 - * example: variable called “prognosis” taking on possible values “poor,” “fair,” “good”