

Source:

Whitlock MC and Schluter, D

The Analysis of Biological Data, 2nd

Overheads 01

Introduction to statistics and Sampling

Statistics are "a quantitative technology for empirical science; it is a logic and methodology for the measurement of uncertainty and for an examination of that uncertainty."

The key word here is "uncertainty."
Statistics become necessary when
observations are variable.

Goals of statistics

- ! Estimate the values of important parameters
- ! Test hypotheses about those parameters

Statistics is also about good scientific practice

Feline High-Rise Syndrome (FHRS)

The injuries associated with a cat falling out of a window.

“The diagnosis of high-rise syndrome is not difficult. Typically, the cat is found outdoors, several stories below, and a nearby window or patio door is open.”



High falls show lower injury rates

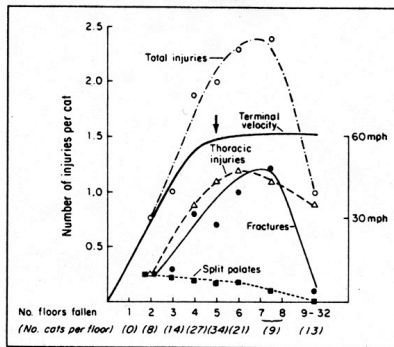


Figure 2—Relationship of injuries to distance fallen and velocity in 132 cats with high-rise syndrome: ↑ points to terminal velocity (—); total number of injuries/cat (○, —); number of thoracic injuries (pulmonary contusions + pneumothorax)/cat (△, - - -); number of fractures/cat (●, —); number of split palates/cat (■, - - -).

Whitney and Mehloff, Journal of the American Veterinary Medicine Association, 1987

Why?



- 1.! Cats have high surface-to-volume ratios
- 2.! Cats have excellent vestibular systems
- 3.! Cats reach terminal velocity quickly, relax, and therefore absorb impact better
- 4.! Cats land on their limbs and absorb shock through soft tissue

Jared Diamond, Nature 1988

Or not...



A sample of convenience is a collection of individuals that happen to be available at the time.

A newer study reports more injuries with longer falls

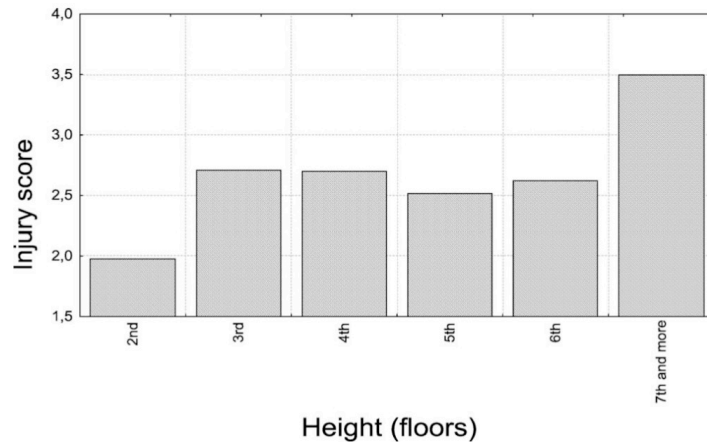


Figure 5 Graph showing the relationship between injury score and height of fall.

Vnuk et al. 2004. Feline high-rise syndrome: 119 cases (1998-2001). J. Fel. Med. Surg. 6:305-312.

Read: Chapters 1 & 2

Variable

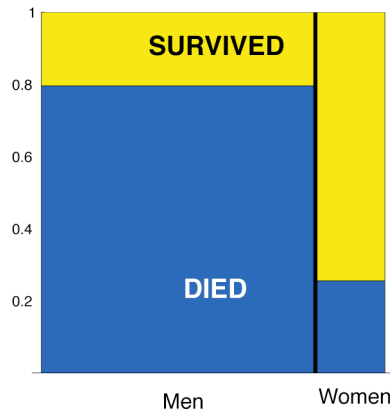
- ! A **variable** is a characteristic measured on individuals drawn from a population under study.
- ! **Data** are measurements of one or more variables made on a collection of individuals.

Explanatory and response variables

We try to predict or explain a **response variable** from an **explanatory variable**.

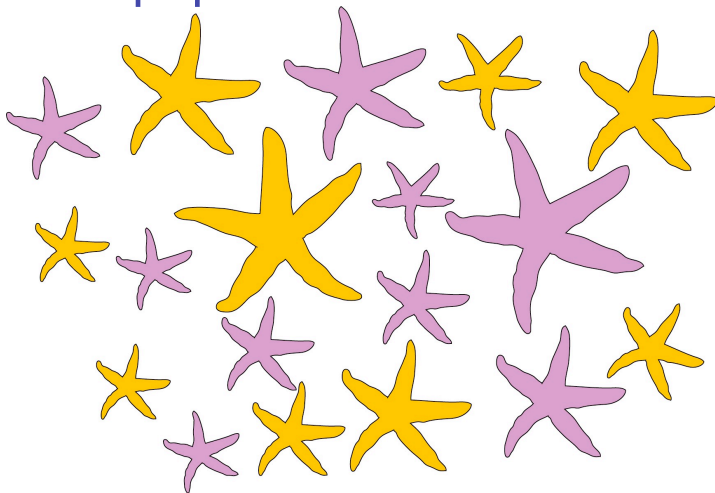
Older terminology:
dependent variable and independent variable

Mortality on the Titanic, as predicted by sex



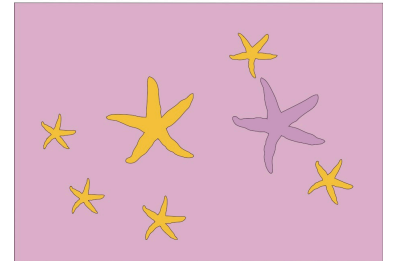
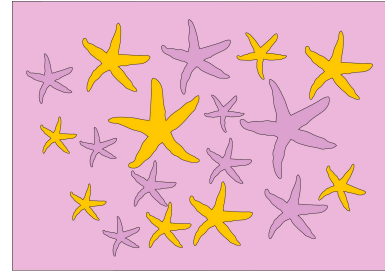
Populations and samples

A population of starfish



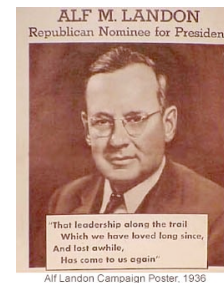
A biased sample

Populations \leftrightarrow Parameters;
Samples \leftrightarrow Estimates



Bias is a systematic discrepancy between estimates and the true population characteristic.

The 1936 US presidential election



Alf Landon
Republican

VS.



Franklin Roosevelt
Democrat

1936 Literary Digest Poll

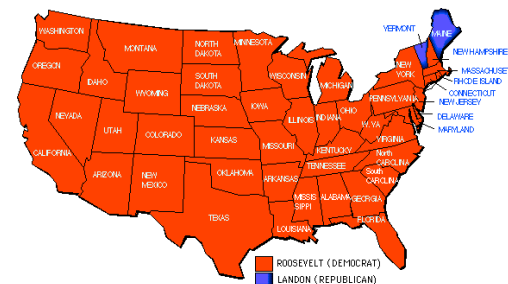
- ! 2.4 million respondents
- ! Based on questionnaires mailed to 10 million people, chosen from telephone books and club lists
- ! Predicted Landon wins: Landon 57% over Roosevelt 43%

What went wrong?

Subjects given the questionnaire were chosen from telephone books and clubs, biasing the respondents to be those with greater wealth

Voting and party preference is correlated with personal wealth

1936 election results



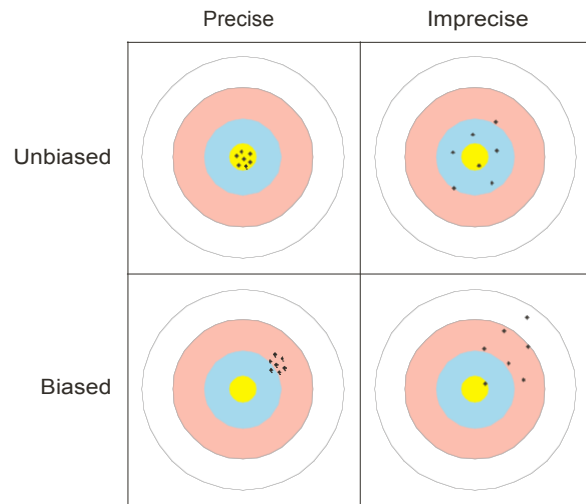
Roosevelt won with 62% of the vote

Volunteer bias

Volunteers for a study are likely to be different, on average, from the population

For example:

- !Volunteers for sex studies are more likely to be open about sex
- !Volunteers for medical studies may be sicker than the general population



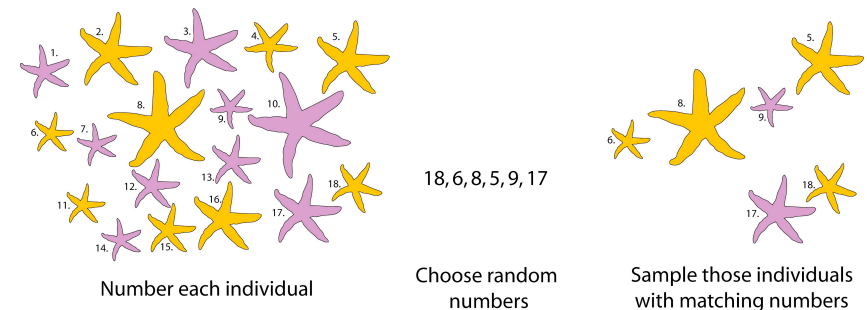
Each point represents an estimate of a parameter.

Properties of a good sample

- ! Independent selection of individuals
- ! Random selection of individuals
- ! Sufficiently large

In a random sample, each member of a population has an equal and independent chance of being selected.

One procedure for random sampling



Population parameters are constants whereas estimates are random variables, changing from one random sample to the next from the same population.

Sampling error

- ! The difference between the estimate and average value of the estimate

Larger samples on average will have smaller sampling error