Inductive inference

Contents

4 Inductive inference

- Probabilistic reasoning can be used to quantify the potential magnitude of a sample error when a sample is used to describe a population
  - provided that a **probabilistic** sampling mechanism is used.

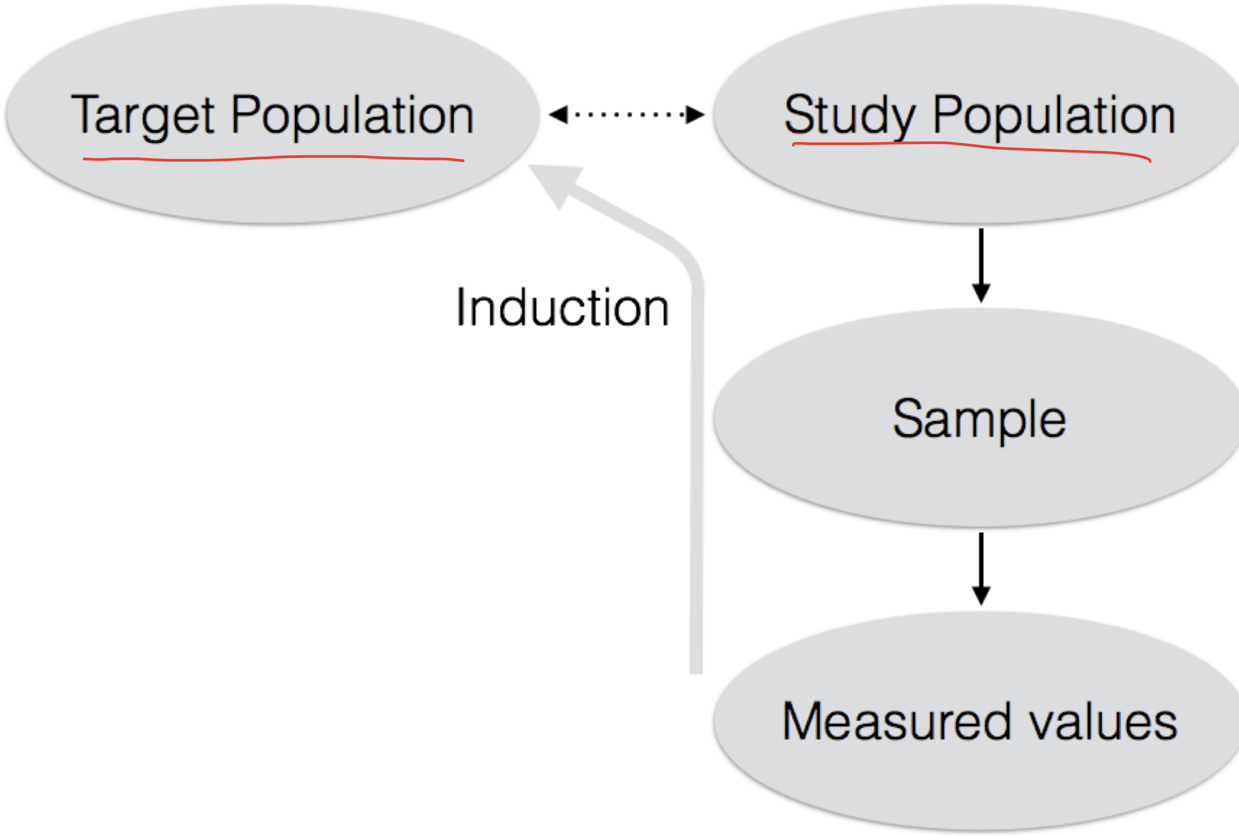
- The sampl**ing** behaviour of any population attribute is examined by repeatedly drawing samples according to the sampling mechanism and calculating the attribute on the samples.

- The sampling behaviour of that attribute can be summarized by its
  - **sampling bias**, and
  - **sampling variability**
- In particular, probabilistic sampling allows us to quantify the relative frequency in which any sample attribute value might be realized.
  - This provides an insurance policy in what is learned about the attributes based on the sample, and
  - How it compares the attributes based on the population.
  - While there is no guarantee that this application will be without error, by careful planning the probability that the error is small can be made large.
- When the sampling is not probabilistic, the insurance policy no longer applies.
- Of course, the nearer the sampling mechanism is to being probabilistic, the more it might be argued that the benefits of probabilistic sampling apply.

In practice however, when using sample information to make inference for a population, there exist other types of error beyond those already discussed.

4.1 Sources of Error

The **Path of Inductive Inference** is shown below.



4.1.1 Target and Study populations

- Carefully planned sampling designs
  - can provide considerable assurance that the conclusions drawn from a sample will not likely be that different from those which would have been drawn were we able to access the entire population.
  - However there is almost always another source of error in our inferences that is not resolved by probabilistic sampling.
- The problem is that in most applications
  - *the population which we are able to draw samples from is not the population about which we would like to draw inferences.*

- The **target population** is the population about which we would like to draw an inference, but the **study population** is the population from which samples are taken.

- The difference between the attribute evaluated on the two populations is the **study error**:

$$\boxed{\text{Study Error} = a(\mathcal{P}_{study}) - a(\mathcal{P}_{target}).}$$

- Then if we obtain a sample $\mathcal{S}$ to draw inferences about the target population $\mathcal{P}_{target}$, the error for a given attribute $a(\cdot)$ is

$$\begin{aligned} a(\mathcal{S}) - a(\mathcal{P}_{target}) &= (a(\mathcal{S}) - a(\mathcal{P}_{study})) + (a(\mathcal{P}_{study}) - a(\mathcal{P}_{target})) \\ &= (\text{Sample error}) \;+\; (\text{Study error}) \,. \end{aligned}$$

- We use probabilistic sampling to control the sample error but not the study error.
- Making the case that the study error is small remains a challenge, and sometimes, it is not even within the domain of statistics (e.g. neuroscience research and studying the brain).

Medical Example

- In medical studies interest often lies in the progression of a disease or the efficacy of its treatment *in humans*.
  - The **target** population is the set of all humans,
  - However, for ethical and other reasons, the study cannot be conducted on humans but must instead be conducted on some other animal, such as mice, which serve as a model for humans.

- The population of mice available from which we can select is a **study** population.
  - Sampling from this population provides some assurance about the quality and uncertainty of our inferences about **study** population's attributes,
  - but this assurance does **not** carry over to inferences about the **target** population.

- Mice, after all, might be fundamentally different from humans for these particular attributes.
  - The quality of the inference depends on how close are the target and study populations.

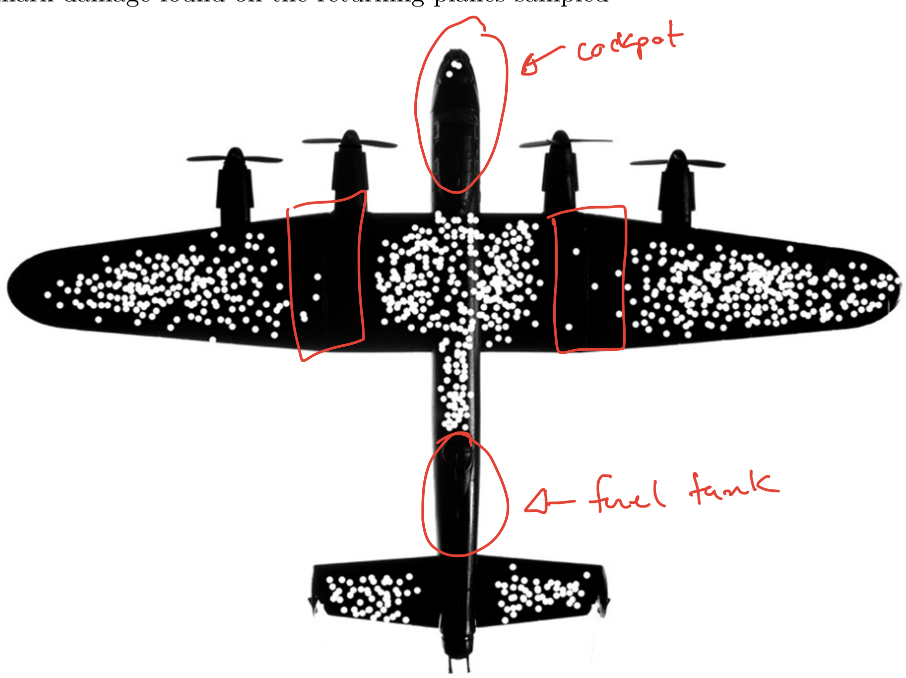
Forecasting Examples

- Ofteh our target population includes **future realizations** of units which are not available at the time of study. e.g.,
  - A natural phenomenon such as the monthly tidal patterns.
  - The daily closing price of a stock such as AMZN.
- In either case, arguing that the study error must be small requires that the future should be much like the past (at least for these attributes).

Bomber Example

- During World War II the statistician Abraham Wald fled persecution in Hungary to immigrate to the United States where he helped U.S. military determine how to minimize bomber losses.
- The military wanted to *know where they should add heavy armour plating to the bomber*.
- The planes which *returned* from bombing missions are the **study population**.
  - Unfortunately, these are not the planes of interest because they survived whatever damage they received.

    study error in this case is referred to as "survivor bias"
- In contrast, the **target population** are those planes which *did not return*; whatever damage they received caused them to not survive the mission.
- White dots mark damage found on the returning planes sampled



- Knowing the nature of the study error, the recommendation becomes clear:

- Armour *only* those areas which show no damage in the returning planes – the cockpit and fuel tank.

David Hume and the Uniformity of Nature

- David Hume famously wrote in 1748:

  For all inferences from experience suppose, as their foundation, that the future will resemble the past, and that similar powers will be conjoined with similar sensible qualities. If there be any suspicion that the course of nature may change, and that the past may be no rule for the future, all experience becomes useless, and can give rise to no inference or conclusion. It is impossible, therefore, that any arguments from experience can prove this resemblance of the past to the future; since all these arguments are founded on the supposition of that resemblance.

  From Hume (1748) "An Enquiry Concerning Human Understanding" (Section IV, Part II) Also accessible as an ebook including here on page 37 of the archive edition

- Many writers after Hume (including John Stuart Mill, Immanuel Kant, John Venn, and Bertrand Russell) have appealed to the uniformity of nature, either as observed by experience (Mill), or as an a fundamental truth (Kant), or as a postulate that is simply required to make progress (Venn, Russell).
  - While all others might agree on its necessity, as Hume pointed out centuries ago, there is no forceful argument that it must hold in general.

- e.g. does uniformity of nature apply to
  - The monthly tidal patterns of Georgian Bay?
  - The daily closing price the AMZN stock?

- In practice, the principle needs to be applied on a case by case basis.
  - For our purposes, it needs to be argued that the study error be small and
  - such argument is typically beyond any statistical argument.

Shark encounters

Example 1

- The target population is the 65 worldwide shark encounters.
  - The attribute of interest might be the average length of sharks involved in these encounters worldwide.
  - Suppose our **study population** are those encounters in Australian waters.

- The difference between the average shark length in Australian encounters and the average shark length of all encounters worldwide is the study error.
  - How would you argue that the study error is small?
  - The study error is $a(\mathcal{P}_{study}) - a(\mathcal{P}_{target}) \approx -4.03$

Example 2

- Suppose we are interested in the average length of sharks in great white shark encounters with humans in US waters but have access only to those encounters which were in Australian waters.
  - The **target population** is the set of all encounters in US waters and
  - the **study population** the set of all encounters in Australian waters.

- The difference between the average shark length in Australian encounters and the average shark length in US encounters is the study error.
  - How would you argue that the study error is small?
  - The study error is $a(\mathcal{P}_{study}) - a(\mathcal{P}_{target}) \approx -5.52$ inches.

Example 3

- Suppose we are interested in the average length of great white sharks in all **future encounters**.
  - The target population can not be observed at all, since it consists of all encounters yet to be realized anywhere in the world.
  - Our entire data set of 65 encounters is the **study population**

- How would you argue that the study error is small?
- Here the study error **cannot be determined** because the target population involves the future.

Example 4

- Suppose we are interested in the average length of great white sharks while fishing.
  - The target population cannot be observed because we only know if the person was scuba diving or not.

- Our entire data set of 65 encounters is the **study population**.

- How would you argue that the study error is small?
- Here the study error **cannot be determined**.

4.1.2 Measurement Error

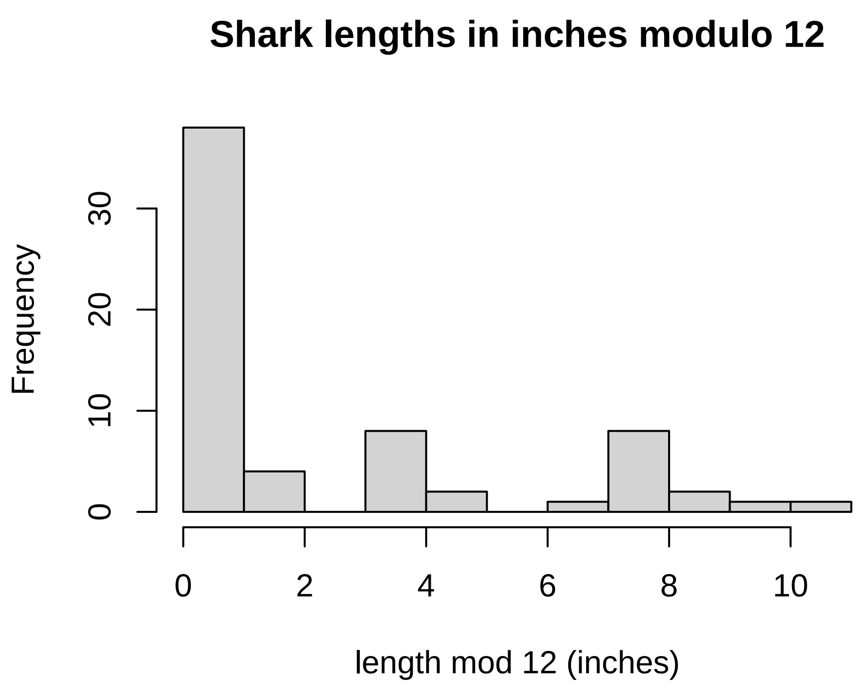
- The inductive path of inference includes the set of measured values.
  - Errors made in measurement can also affect conclusions drawn about attributes.

- For example, in the population of shark encounters, the measuring of the shark length is one place where **measurement error** could occur.
  - How this measurement was taken is never described.
  - Was the measurement taken vertically or horizontally?
  - Was the measurement taken by the victim while under attack?

- The great bulk of measurements are divisible by 12. What does this tell you?

```r
sharks = read.csv("/Users/nstevens/Dropbox/Teaching/STAT_341/Lectures/Data/sharks.csv",
                  header=TRUE)

hist(sharks[,"Length"] %% 12, main="Shark lengths in inches modulo 12",
     xlab="length mod 12 (inches)", col = "lightgrey")
```



Measurement systems

- Every measuring system has at least three sources of potential error:

1. the measuring device (sometimes called the gauge)
2. the person reading or recording the measurement (some called the operator)
3. the method followed to take the measurement

The third source of measurement error includes anything independent of the gauge and the operators

- e.g., shark length measurements taken on sharks hanging vertically could differ systematically from those taken on sharks laying horizontally on the ground.
- e.g., even the measurement of fatality, which should be straightforward, could vary.
  - is a fatality recorded if the person died immediately due to the severity of the attack? Or does a death later on, due to complications, also count?

Note that an entire discipline known as **measurement system analysis** is devoted to the practical and academic study of measurement systems and the methods by which their adequacy and comparability is determined.