Scientific Method & Statistical Reasoning

Paul Gribble

http://www.gribblelab.org/stats/

Winter, 2017

MD Chapters 1 & 2

- The idea of pure science
- Philosophical stances on science
- Historical review
- Gets you thinking about the logic of science and experimentation

Assumptions

Lawfulness of nature

- Regularities exist, can be discovered, and are understandable
- Nature is uniform

Causality

- events have causes; if we reconstruct the causes, the event should occur again
- can we ever prove causality?

Reductionism

- Can we ever prove anything? What is proof?

Assumptions

Finite Causation

- causes are finite in number and discoverable
- generality of some sort is possible
- We don't have to replicate an infinite # of elements to replicate an effect

Bias toward simplicity (parsimony)

- seek simplicity and distrust it
- start with simplest model: try to refute it; when it fails, add complexity (slowly)

Philosophy of Science

- Logical Positivism
- Karl Popper & deductive reasoning
- progress occurs by falsifying theories

Logical Fallacy

Fallacy of inductive reasoning (affirming the consequent)

- **Predict**: If theory T, then data will follow pattern P
- **Observe**: data indeed follows pattern P
- **Conclude**: therefore theory T is true

example

- A sore throat is one of the symptoms of influenza (the flu)
- I have a sore throat
- Therefore, I have the flu

Of course other things besides influenza can cause a sort throat. For example the common cold. Or yelling a lot. Or cancer.

Falsification is better

Falsification

- **Predict**: If theory T is true, then data will follow pattern P
- **Observe**: data do not follow pattern P
- **Conclude**: theory T cannot be true

We cannot prove a theory to be true.
We can only prove a theory to be false.

Karl Popper

- Theories must have concrete predictions
- constructs (measures) must be valid
- empirical methodology must be valid

Basis of Interpreting Data

the Fisher tradition

- statistics is not mathematics
- statistics is not arithmetic or calculation
- statistics is a **logical framework** for:
  - making decisions about theories
  - based on data
  - defending your arguments
- Fisher (1890-1962) was a central figure in modern approaches to statistics
- The F-test is named after him

The Fundamental Idea

THE critical ingredient in an inferential statistical test (in the frequentist approach):

- determining the **probability**, assuming the null hypothesis is true, of obtaining the observed data

The Fundamental Idea

Calculation of probability is typically based on probability distributions

- continuous (e.g. z, t, F)
- discrete (e.g. binomial)

We can also compute this probability without having to assume a theoretical distribution

- Use resampling techniques
- e.g. bootstrapping

Basis of Interpreting Data

- design experiments so that inferences drawn are fully justified and logically compelled by the data
- theoretical explanation is different from the statistical conclusion
- Fisher's key insight:
  - randomization
  - assures no uncontrolled factor will bias results of statistical tests

A Discrete Probability Example

- Last fall in my lab we were making espresso, and I claimed that I could taste the difference between Illy beans (which are expensive) and Lavazza beans (which are less expensive).
- Let's think about how to design a test to determine whether or not I actually have this ability

Testing Mr. EspressoHead

Many factors might affect his judgment

- temperature of the espresso
- temperature of the milk
- use of sugar
- precise ratio of milk to espresso

Prior to Fisher

- you must experimentally control for everything
- every latte must be **identical** except for the independent variable of interest

Testing Mr. EspressoHead

How to design your experiment?

- a single judgment?
- he might get it right just by guessing
- ⋆ this is the **null hypothesis**!
- H_0 is he does not have the claimed ability
- H_0 is that he is guessing

Testing Mr. EspressoHead

How many cups are required for a sufficient test?

- how about 8 cups (4 Illy, 4 Lavazza)
- present in random order
- tell subject that they have to separate the 8 cups into 2 groups: 4 Illy and 4 Lavazza
- is this a sufficient # of judgments?
- how do we decide how many is sufficient?

Testing Mr. EspressoHead

Key Idea

- consider the possible results of the experiment, and **the probability of each, given the null hypothesis that he is guessing**
- there are many ways of dividing a set of 8 cups into Illy and Lavazza
- Pr(correct by chance) = (# exactly correct divisions) / (total # possible divisions)

Testing Mr. EspressoHead

- only one division exactly matches the correct discrimination
- therefore numerator = 1
- what about the denominator?
- how many ways are there to classify 8 cups into 2 groups of 4?
- equals # ways of choosing 4 Illy cups out of 8 (since the other 4 Lavazza are then determined)

Testing Mr. EspressoHead

- 8 possible choices for first of 4 Illy cups
- for each of these 8 there are 7 remaining cups from which to choose the second Illy cup
- for each of these 7 there are 6 remaining cups from which to choose the third Illy cup
- for each of these 6 there are 5 remaining cups from which to choose the fourth and final Illy cup
- total # choices = 8 x 7 x 6 x 5 = 1680

Testing Mr. EspressoHead

- total # choices = 1680
- does order of choices matter? (no)
- any set of 4 things can be ordered 24 different ways (4 x 3 x 2 x1)
- each set of 4 Illy cups would thus appear 24 times in a listing of the 1680 orderings
- so total # of distinct sets (where order doesn't matter) = (1680 / 24) = **70 unique sets of 4 Illy cups**

Testing Mr. EspressoHead

- we can calculate this more directly using the formula for "# of combinations of **n** things taken **k** at a time"
- " 8 choose 4"

```
nCk = (n!) / (k! (n-k)! )
= 8! / (4! (8-4)! )
= (8x7x6x5x4x3x2x1) / (4x3x2x1)x(4x3x2x1)
= (8x7x6x5) / (4x3x2x1)
= 70
```

Testing Mr. EspressoHead

- we have now formulated a statistical test for our null hypothesis
- the probability of me choosing the correct 4 Illy cups **by guessing** is
  $(1\ /\ 70) = 0.014 = 1.4\ \%$
- so if **I do** pick the correct 4 Illy cups, then **it is much more likely (98.6 %) that I was not guessing**
- you **cannot** prove I wasn't guessing
- you can **only** say that the probability of the observed outcome, **if I was guessing**, is low (1.4 %)

Testing Mr. EspressoHead

from the Chapter

- Pr(perfect **or** 3/4 correct) = (1+16)/70 = 24 %
- nearly 1/4 of the time, just by guessing!
- so observed performance of 3/4 correct may not be sufficient to convince us of my claim

Logic of Statistical Tests

review

- to design a scientific test of Mr. EspressoHead's claim, we designed an experiment where the chances of him guessing correctly 4/4 were low
- so if he did get 4/4 correct then what can we conclude?
- we could choose to **reject the null hypothesis that he was guessing**, because we calculated that the chances of this happening, are low

How low should you go?

how low is low enough to reject the null hypothesis?

- 5 % (1 in 20) $p < .05$
- 2 % (1 in 50) $p < .02$
- 1 % (1 in 100) $p < .01$
- 0.0001 % (1 in 1,000,000) $p < .000001$

answer:

it is **arbitrary**, YOU must decide

but consider convention in:

your lab / *journal* / *field*

How low should you go?

what is the relative cost of making a wrong conclusion?

- concluding YES he has the ability when in fact he doesn't (type-I error)
- concluding NO he doesn't have the ability when in fact he does (type-II error)

costs may be different depending on the situation

- drug trial for a new, but very expensive (but potentially beneficial) cancer drug
- your thesis experiment, which appears to contradict a major accepted theory in neuroscience
- your thesis experiment, which appears to contradict your own previous study

Tests based on Distributional Assumptions

Instead of counting or calculating possible outcomes we typically rely on statistical tables

- give probabilities based on theoretical distributions of test statistics
- typically based on the assumption that the dependent variables are normally distributed
- allows generalization to population, not just a particular sample
- e.g. the t-test (next week)

We can however proceed without assuming particular theoretical distributions

- non-parametric statistical tests
- resampling techniques

for next week

catch up on readings

- MD 1 & 2 (today's class)
- Start in on readings for Hypothesis Testing (MC3, MC7, MCt, N14)