Medical Statistics, Informatics, and Telemedicine

Lecture 5

Introduction to Hypothesis Testing

8th October 2021

Gergely Agócs

Topics

- Understanding the process of scientific decision-making
- The steps of hypothesis testing demonstrated by an example
- How likely is our "discovery" true?
- How practically important is what we have discovered?
- Possible decision errors

Philosophical Background

The Evolution of Science

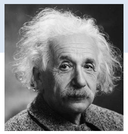
A statement is considered scientific, if it can be independently verified, reproduced.
But this does not tell us, how science is actually created.

Inductivism

Francis **Bacon**
1561–1626

How science is actually „created"

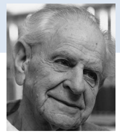
Philosophical Background



Albert **Einstein**
1879–1955

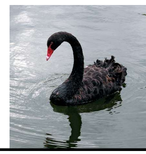
The Evolution of Science

"No amount of experimentation can ever prove me right;
a single experiment can prove me wrong."
after Albert Einstein: *Induction and Deduction*



Karl **Popper**
1902–1994

"**Induction is logically invalid**;
but refutation or falsification is
a logically valid way of arguing."
Karl Popper: *The Logic of Scientific Discovery*



Philosophical Background

Falsifiable (i.e. Scientific) Statements

Non-Falsifiable (i.e. Non-Scientific) Statements
(may eventually be verifiable)

4

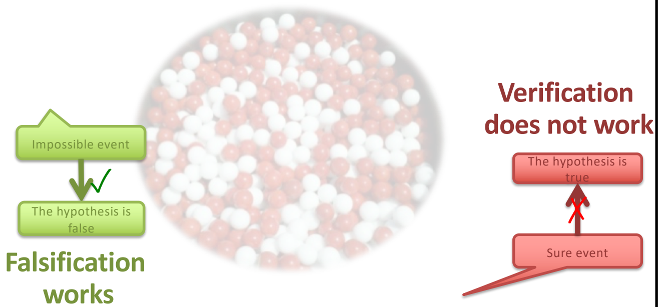
Philosophical background

The Burden of Proof (*onus probandi*)

5

Way of Thinking in Hypothesis Testing

Indirect Proof (*reductio ad absurdum*)



6

Way of Thinking in Hypothesis Testing

Indirect proof (*reductio ad absurdum*)

Mathematical Logic:

a hypothesis can only be rejected.

Statistical Logic:

a hypothesis cannot even be rejected with 100% certainty.

7

What Kind of Questions Can We Test?

The question...

...must be a yes/no (a.k.a. dichotomous or polar) question.

...must refer to a set of observations, not to individual cases.
(And the question is aimed at a population, not a sample.)

...must have at least one unambiguous answer.

8

What Kind of Answers Can We Test?

We have two answers for our question:

The null hypothesis (H_0) **The alternative hypothesis (H_1)**

9

A Worked-out Example

The current 5-years survival rate for myeloma patients is 50% (average between 2008–2012).

We have a new drug candidate that seems to be effective against myeloma in animal experiments. We want to test it on humans.

1.a **Physicians question:** Should we replace current treatment protocols with the new drug?
1.b **Clinical question: Is the effect clinically relevant?** Is the change in survival rate big enough?
1.c **Statistician's question:** Is there an effect?
2. **H_0: The drug has no effect:** Survival rate with the drug is same as with the conventional therapy.
3. **H_1: The drug has an effect:** Survival rate is different.
4. **Test design:** Select randomly 20 myeloma patients and treat them with the drug candidate. After 5 years, check the *number of patients still alive*. This number can be called here the **test statistic**.

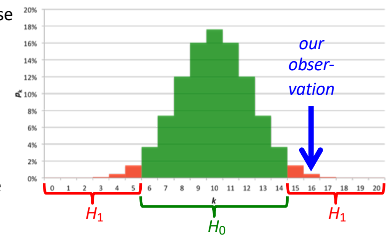
5. **Generate the H_0 distribution:** It is a binomial distribution with $p = 0.5$ and $n = 20$. Same as a coin tossing experiment.
6.a **Set up a significance level (α) which will also define your confidence level ($1–\alpha$):** Be $\alpha = 5\%$. (For historical reasons 5% is often used in health sciences but it is our choice.)
6.b **Define what change is clinically relevant:** We expect at least a 20% higher survival rate.

10

A Worked-out Example

7. **Determine your confidence interval using the H_0 distribution and α:** In our binomial distribution the range of outcomes from 6 to 14 have a combined probability of just above 95%. This is the set of those outcomes, which represent too weak evidence against H_0.
8. **Carry out the experiment:** (Note: ***this is the 8th step!***) Out of our 20 patients treated with the new drug 16 are still alive after 5 years.

...



10. **Make a decision:** Clinically speaking the effect is **relevant** (80% instead of 50% survival rate); Statistically speaking it is **significant** (our outcome is unlikely under the H_0 or, more precisely, it belongs to the set of 5% least probable outcomes).
11. **Answer the question:** Let's wait a little bit with this one...

11

A Worked-out Example

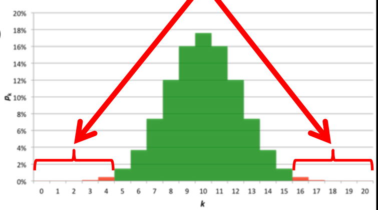
How significant is the outcome?

Obviously, a 16 out of 20 survival rate is more significant (= unlikely under H_0) than 14 out of 20 but less significant than 18 out of 20. How can we express this numerically?

9. **Determine the sample *p*-value:** The *p*-value is the Holy Grail of inferential statistics. It gives the **probability of your (or any less likely) observation** to occur ***given*** H_0 is true. In our case, it is the combined probability of all outcomes equally or less probable than getting 16 out of 20 survivals. The value is p_{sample} = 0.0118. This is shown in red in the graph.

11. **Answer the question:** "The group (n = 20) treated with the drug candidate had 80% survival rate that is significantly higher than that of the conventional treatment's 50% (p = 1,18%)"

So we are sort of done. But again, can we be sure about the correctness of our decision?



12

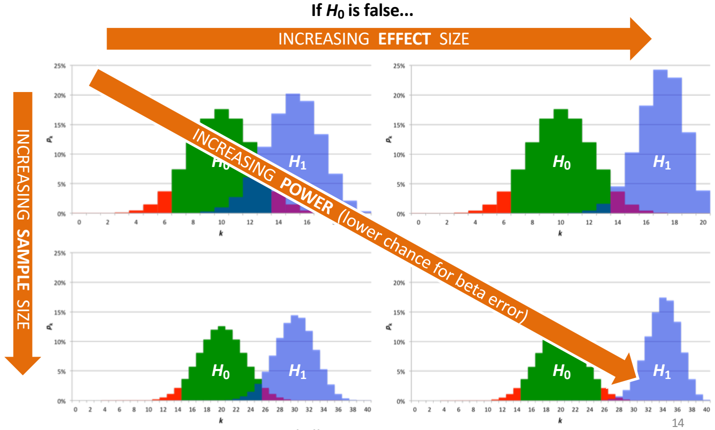
Decision Errors

		Our Decision	
		we do not reject H_0	we reject H_0
The Truth (never known)	H_0 true	**correct** $1 - \alpha$	**α or type I error** p(H_0 rejected | H_0 true) ≤ α
	H_0 false	**β or type II error** p(H_0 not rejected | H_0 false) ≤ β	**correct** $1 - \beta$

13

Decision Errors

If H_0 is false...



cf. Efficiency and Unbiasedness

14

p-Value Pitfalls

- **clinical relevance vs. statistical significance**
- **multiple testing**
- **H_0 not rejected ≠ H_0 proven**
- **correlation ≠ causation**
- **Should p-values be used at all?**

15