Sampling and hypothesis tests

Part 5: Hypothesis tests

By: Noureddin Sadawi, PhD

University of London

Hypothesis

- Imagine you conduct an experiment to check if a particular drug is effective at treating toothache.
- If you cannot repeat the experiment, your results might not be taken seriously.
- A **hypothesis** is a proposed explanation for something (e.g. a phenomenon).

Hypothesis tests

- ‘Whenever you see references to statistical significance, t-tests, or p-values, it is typically in the context of the classical statistical inference “pipeline”.
- This process starts with a hypothesis (“drug A is better than the existing standard drug,” or “price A is more profitable than the existing price B”).
- An experiment is designed to test the hypothesis—designed in such a way that it hopefully will deliver conclusive results.
- The data is collected and analyzed, and then a conclusion is drawn.
- The term inference reflects the intention to apply the experiment results, which involve a limited set of data, to a larger process or population.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Hypothesis tests

- ‘Hypothesis tests, also called significance tests, are ubiquitous in the traditional statistical analysis of published research.
- Their purpose is to help you learn whether random chance might be responsible for an observed effect.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Key terms for hypothesis tests

- ‘**Null hypothesis (H_0):**

  The hypothesis that chance is to blame.
- **Alternative hypothesis (H_a):**

  Counterpoint to the null (what you hope to prove).
- **One-way test:**

  Hypothesis test that counts chance results only in one direction.
- **Two-way test:**

  Hypothesis test that counts chance results in two directions.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

A/B testing

- ‘An A/B test is an experiment with two groups to establish which of two treatments, products, procedures, or the like is superior.
- Often one of the two treatments is the standard existing treatment, or no treatment.
- If a standard (or no) treatment is used, it is called the control.
- A typical hypothesis is that a new treatment is better than the control.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Key terms for A/B testing

- ‘**Treatment:** Something (drug, price, web headline) to which a subject is exposed.
- **Treatment group:** A group of subjects exposed to a specific treatment.
- **Control group:** A group of subjects exposed to no (or standard) treatment.
- **Randomization:** The process of randomly assigning subjects to treatments.
- **Subjects:** The items (web visitors, patients, etc.) that are exposed to treatments.
- **Test statistic:** The metric used to measure the effect of the treatment.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

A/B testing

- ‘A proper A/B test has subjects that can be assigned to one treatment or another.
- The subject might be a person, a plant seed, a web visitor; the key is that the subject is exposed to the treatment.
- Ideally, subjects are randomized (assigned randomly) to treatments.
- In this way, you know that any difference between the treatment groups is due to one of two things:
  - The effect of the different treatments.
  - Luck of the draw in which subjects are assigned to which treatments (i.e., the random assignment may have resulted in the naturally better-performing subjects being concentrated in A or B).’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

The null hypothesis (H_0)

- ‘Given the human tendency to react to unusual but random behavior and interpret it as something meaningful and real, in our experiments we will require proof that the difference between groups is more extreme than what chance might reasonably produce.
- This involves a baseline assumption that the treatments are equivalent, and any difference between the groups is due to chance.
- This baseline assumption is termed the **null hypothesis**.
- Our hope, then, is that we can in fact prove the null hypothesis wrong and show that the outcomes for groups A and B are more different than what chance might produce.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

The alternative hypothesis (H_a)

- ‘Hypothesis tests by their nature involve not just a null hypothesis but also an offsetting alternative hypothesis. Here are some examples:
  - Null = “no difference between the means of group A and group B”; alternative = “A is different from B” (could be bigger or smaller)
  - Null = “A ≤ B”; alternative = “A > B”
  - Null = “B is not X% greater than A”; alternative = “B is X% greater than A”
- Taken together, the null and alternative hypotheses must account for all possibilities.
- The nature of the null hypothesis determines the structure of the hypothesis test.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Examples

- We want to test if patients wait less than two hours at A&E, on the average. The null and alternative hypotheses are:
  - H_0: $\mu \geq 2$
  - H_a: $\mu < 2$
- A medical trial is conducted to test whether or not a new medicine reduces lung cancer risk by 30%. How do we state the null and alternative hypotheses?.
  - H_0 : The drug reduces lung cancer risk by 30%. $p = 0.30$
  - H_a : The drug does not reduce lung cancer risk by 30%. $p \neq 0.30$

One-way vs two-way hypothesis tests 1/2

- ‘Often in an A/B test, you are testing a new option (say, B) against an established default option (A), and the presumption is that you will stick with the default option unless the new option proves itself definitively better.
- In such a case, you want a hypothesis test to protect you from being fooled by chance in the direction favoring B.
- You don’t care about being fooled by chance in the other direction, because you would be sticking with A unless B proves definitively better.
- So you want a *directional* alternative hypothesis (B is better than A).
- In such a case, you use a *one-way* (or onetail) hypothesis test.
- This means that extreme chance results in only one direction count toward the p-value.’

  (Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

One-way vs two-way hypothesis tests 2/2

- ‘If you want a hypothesis test to protect you from being fooled by chance in either direction, the alternative hypothesis is *bidirectional* (A is different from B; could be bigger or smaller).
- In such a case, you use a two-way (or two-tail) hypothesis.
- This means that extreme chance results in either direction count toward the p-value.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).