Probability and Introduction to Statistics

ALY6010

Tom Breur

Week 5, 1-DEC-2020

Agenda

- Administrative notes
- Review Discussion board
- Hypothesis tests: one- or two-sided?
- Some “pop quiz” items
- Preparation week 6

Administrative notes

- Many students have pointed out errors, or suspected errors in several quiz items. Please, by all means continue to provide these
- At the moment, teacher is working with facukty to determine an equitable policy for adjusting the grades
  - For erroneous quiz items, an adjustment will be made on an item-by-item basis
  - For reference, in the past under similar circumstances the grading was always adjusted in favor of students' outcomes
- Everyone:
  please take note of the deadlines in the next two weeks, as grade submission deadlines allow teacher little or no leeway for leniency

Your TA for ALY6010 CRN 71709

Catherine Richard

Email: richard.ca@northeastern.edu

You can reach me via :

- Email
- Post on Canvas
- WhatsApp (781-526-6300)

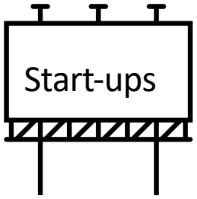
This week TA hours:

- Thursday 9-10 AM
- Friday 3-5 PM
- Saturday 10:30-11:30 AM

I've worked at and am interested in :

Healthcare

Deep Learning



Tufts University

Northeastern University
College of Professional Studies

MPS Analytics, Statistical Modeling Concentration
I'll graduate at the end of next quarte

Discussion Board

Critical review

Hypothesis Testing Refresher

- When constructing and implementing hypothesis tests, what reasoning is used behind the statement of the null and alternative hypotheses? Why are hypothesis tests set up in this way? Can a confidence interval obtained for estimating a population parameter be used to reject the null hypothesis? If your answer is yes, explain how. If your answer is no, explain why.
- When performing a hypothesis testing, two types of errors can be made: Type I and Type II. Explain in your opinion which of these errors would be a more serious error. Use specific examples to support your argument and reasoning.
- In your two replies to classmates, provide remedies to simultaneously minimize both types of errors mentioned in question 2 above.

Hypothesis Testing Refresher (a 1)

When constructing and implementing hypothesis tests, what reasoning is used behind the statement of the null and alternative hypotheses?

- Null and Alternative Hypothesis are framed as pairs that directly oppose each other
  - H_0 and H_1 belong together like a pair of Siamese twins
- The way hypotheses are framed, ***it should always hold*** that *either* one *or* the other (H_0 or H_1) has to be supported by the evidence
  - H_0 represents the current state of affairs
    (more on next slide)

Hypothesis Testing Refresher (a 2)

Why are hypothesis tests set up in this way?

- The choice of "which statement" should be represented by H0 and which statement should be H1 *is anything but "random"*:
  - H_0 represents the current state of affairs, the "status quo", or evidence as currently held for truth, generally accepted theory, etc.
  - H_1 is the alternative hypothesis (***not*** H_a), that would represent a "new finding", an outcome that innovates or updates existing knowledge
- Colloquially, one could say that the researcher "wants to prove H1", and will design his study with maximum power
  - Obviously all the while following the generally accepted principles of robust science

Hypothesis Testing Refresher (a 3)

H_0 versus H_1

- H_0 & H_1 need to be:

MECE

Mutually Exclusive & Collectively Exhaustive

- For a two-sided test:

$H_0 : \mu = X$ or: $H_0 : \mu_o - \mu_1 = 0$

$H_1 : \mu \neq X$ $H_1 : \mu_o - \mu_1 \neq 0$

- For a one-sided test:

$H_0 : \mu_o - \mu_1 > 0$

$H_1 : \mu_o - \mu_1 \leq 0$

Note that this most common (!) depiction of a one-sided pair of hypotheses is ***different*** from Bluman's notation !!

Hypothesis Testing Refresher (a 4)

Can a confidence interval obtained for estimating a population parameter be used to reject the null hypothesis?

- Yes. As per Bluman, p. 414, there are three possible methods that can be used to test hypotheses
  - Traditional method
  - p-value method
  - Confidence interval method
- Although not strictly *identical*, these three methods are *interchangeable*
- Given current availability of computers, rather than access via significance tables, most people would use the p-value method, since *for the majority of purposes* it is most informative
  - Preferred because it provides an estimate of Type I error

Hypothesis Testing Refresher (b 1)

When performing a hypothesis testing, two types of errors can be made: Type I and Type II. Explain in your opinion which of these errors would be a more serious error. Use specific examples to support your argument and reasoning

- This question is a bit of a “red herring”: as discussed in class last week, there is no better or worse outcome for Type I or Type II – which one is worse depends ***entirely*** on the context
- Concerns over Type I or Type II error should be evaluated in light of the relative costs of misclassification
- Sometimes Type I errors are more serious, in other settings a Type II error may be considered more serious

Hypothesis Testing Refresher (c 1)

Provide remedies to simultaneously minimize both types of errors (Type I & Type II)

- The simplest answer to this question, and a recipe that *always* works, is to improve sample size (N)
  - However, this approach of growing N always comes at additional cost of conducting research
- An alternative could be to change the design of the study in such a way that error variance is minimized
  - This can be done by using a “within subjects design”, or “matched pairs” dependent designs
- An other way to minimize both Type I and Type II errors is to elevate the measurement level, typically by treating an ordinal scale (like Likert) ***as if*** it is an interval scale – those tests have greater statistical power

Logical Positivism

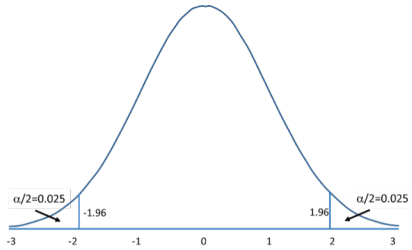
- Ever since the days of Karl Popper (1902-1994) "Logical Positivism" has been dominant Philosophy of Science paradigm in the Western world
- Essential element:
  Scientific theories need to be "falsifiable", i.e. stipulate what real-world evidence would refute the theory
- A "practical" consequence of this approach is that theories are never (ever!) proven to be "true"
  - A theory "holds" until evidence emerges that refutes it
- Consequently: a Null hypothesis can not be proven true
  - Null hypothesis represents the status quo
- Null hypothesis may be rejected, or the evidence will not allow you to reject it – **but no other options exist!!**

Hypothesis testing: one- or two-sided?

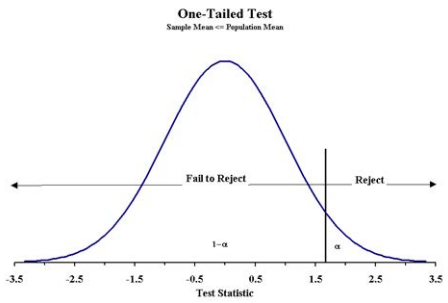
One vs. two-sided tests

- The research question ***determines*** whether a test should be conducted one- or two-sided

- Two-sided: ***is*** there a difference between groups?
  - $H_0 : \mu = X$
    $H_1 : \mu \neq X$

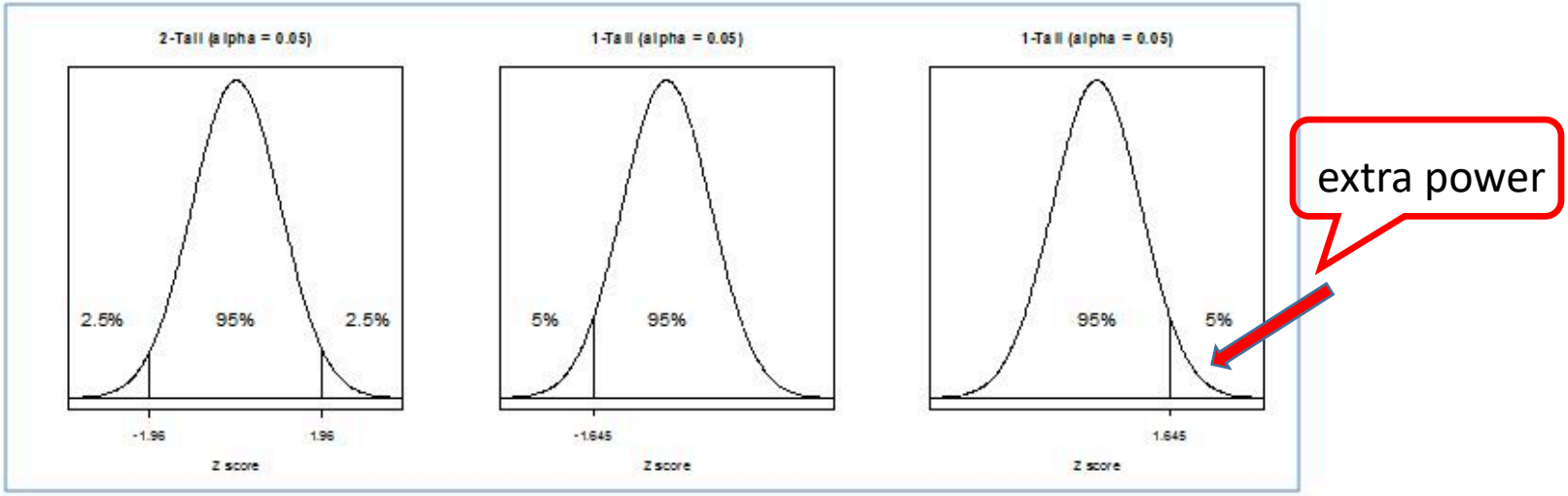


- One sided: is group A ***larger*** (or smaller) than group B?
  - $H_0 : \mu_o > \mu_1$ or $H_0 : \mu_o - \mu_1 > 0$
    $H_1 : \mu_o \leq \mu_1$ $H_1 : \mu_o - \mu_1 \leq 0$



One vs. two-sided & statistical power

- The same data, when tested against a one-sided hypothesis have greater statistical power
  - However, this choice needs to be made beforehand, and needs to be justified by the design and research question
- Note that the critical value "shrinks" from 1.96 to 1.65



Some “pop quiz” items

One versus two-sided tests (1)

Quiz statement:

"There are two possible options, one-sided versus two-sided tests;

therefore, the odds for each are 50%"

One versus two-sided tests (2)

How do you decide between the two possible options:

a one-sided versus a two-sided test?

Accepting versus Rejecting H_0/H_1

"What is the most likely outcome for H_0 & H_1?"

H_0

a) Accept

b) Reject

c) Same / can't tell / too scared to choose

H_1

a) Accept

b) Reject

c) Same / can't tell / too scared to choose

Preparation week #6

Requirements - REVIEW

- Discussion board:
  - Post contributions on *successive* (distinct!) days
  - *Minimum* of three posts, but this need not limit you
  - *First* post your primary contribution, *only then* will get access to other peoples' contributions
- Quizzes:
  - Bluman 10-1 to 10-4 quizzes
  - Week 5 R assignment
- Reading preparation week 6 (Chapter 10 Bluman)
- Reading preparation week 6 (Chapter 7 Kabacoff)

Discussion board: requirements

- You ***first*** (!) need to post an original contribution ("primary post") first, with a minimum (!) of 250 words
- This post needs to contain an academic reference to a reliable (!) and relevant source
  - The reference needs to be set in APA standard
- A minimum (!) of two responses are required, each 80+ words, and posted on distinct, successive dates
- All contributions need to be substantive
  - For clarity: "I agree", "I like your post", etc. do not count as substantive replies. *Instead* reason why you agree or disagree, and refer to outside soures to justify your position
- Referring to other sources or posts, or previous classes, earns "brownie points" towards top grades (100 points) for integrative learning

Discussion board: substance (1)

- #1
  Give an application of the two-sample t-test in the context of Six-Sigma

  In order for your response to qualify as a legitimate and valid example, it needs to pertain to a **real-world business application**. Ideally, gather empirical data from a *publicly available dataset* (which are plentiful), so that everyone can reproduce your conclusions.

  The t-test comes in various flavors, for different problem frames, make sure to justify your choice. Explain why *your* choice applies to this situation.

Type I & II errors

- Type I error: you find a significant effect, you conclude “there is something there” (reject H_0), but in reality there is not
- Type II error: you conclude there is no difference (fail to reject H_0), but in reality there is

- Type I error can be a statistical fluke, “chance” effect
- Type II error is often (usually) caused by using a test with (too) small power, and/or using a sample that is too small

- The relative severity is not (*ever!*) absolute, but merely and always a function of the relative costs of misclassification!

Discussion board: substance (2)

- #2
  Describe (in your own words…) what the p-value of this two sample t-test means

  The required "interpretation" is ***not*** a statistical definition or meaning, but rather the substantive interpretation. In other words: what are the business consequences or implication of this finding?

  In the context of Type I and Type II errors, and respective costs of misclassification, provide an explanation and justification for your proposed experimental design