



DATA ANALYSIS

Week 12: Midterm 2 review!

logistics: office hours

- Prof. Kumar
 - [calendly appointments](#) (Wednesday, Thursday, and Monday slots available)
- LAs
 - review session, 6-8 pm, Kanbar 101
 - no LA hours this weekend

midterm 2 format

- in-class **conceptual** (on Canvas): 40% of first midterm grade (6 out of 15 points)
 - multiple choice, matching, short answer (quiz-like)
 - bring flowchart + process sheet + ONE handwritten help sheet
 - **closed book (do NOT leave Canvas page once you begin)**
- take-home **computational**: 60% of first midterm grade (9 out of 15 points)
 - data analysis (problem set-like)
 - submissions will involve: (1) PDF of solution sheet + (2) downloaded worksheet
 - open book but **NOT open person (cannot ask LAs/friends)**

allowed formulas for midterm 2 computational

- AVERAGE
- SUM
- SQRT
- COUNT
- MEDIAN
- MODE
- CORREL
- STDEV.S, STDEV.P
- VAR.S, VAR.P

— lingering question

- Could we have access to a completed version of the hypothesis testing (with steps) document to make sure we have filled in the processes correctly with correct notation?

plan for today



matching question debrief



short answer question
debrief



lingering questions

matching question

Twenty boys watched 20 hours of violent television. The number of times each child was aggressive was recorded for 30 minutes. Is aggression after watching the violent television significantly different from the population average of 0 acts of aggression (sd = 1)?

one-way independent ANOVA, z-scores, one sample t-test, one-tailed t-test, t-test for independent samples, variance, regression, z-test, scatterplot, Pearson correlation, bar graph

matching question

The Director of Alumni wants to estimate the amount of money he can anticipate each alumni donating given the number of years since they've each graduated. What will allow him to make predictions about donations as accurately as possible?

one-way independent ANOVA, z-scores, one sample t-test, one-tailed t-test, t-test for independent samples, variance, regression, z-test, scatterplot, Pearson correlation, bar graph

matching question

Which is the best way to graphically display the marital status of 50 adults?

one-way independent ANOVA, z-scores, one sample t-test,
one-tailed t-test, t-test for independent samples, variance,
regression, z-test, scatterplot, Pearson correlation, bar graph

matching question

Using matched random assignment (based on SAT scores), different groups of students studied for a multiple-choice exam using only their class notes, only their textbook, or both. The professor recorded the number of errors on the test. Were errors significantly affected by study method?

one-way independent ANOVA, z-scores, one sample t-test, one-tailed t-test, t-test for independent samples, variance, regression, z-test, scatterplot, Pearson correlation, bar graph

matching question

The Fall-semester GPAs of 15 sophomores and 15 first-years were compared to see if one group performed better than another. Does class year significantly affect GPA?

one-way independent ANOVA, z-scores, one sample t-test, one-tailed t-test, t-test for independent samples, variance, regression, z-test, scatterplot, Pearson correlation, bar graph

matching question

Census data from the 1930's reveals that women, on average, had their first child at 18 years of age. Twenty randomly selected 50-year-old mothers reported having their first child at 25 (standard deviation = 1.5). Do women wait significantly longer before having children compared to the 1930's?

one-way independent ANOVA, z-scores, one sample t-test, one-tailed t-test, t-test for independent samples, variance, regression, z-test, scatterplot, Pearson correlation, bar graph

short answer

A cancer researcher comes up with a new treatment to treat advanced-stage lung cancer that is 10% more effective than the current treatment. However, the new treatment has potentially severe side effects, including immune system suppression and organ damage. To test its effectiveness, the researcher conducts a clinical trial comparing the new treatment to the current standard therapy.

In this context:

1. Define Type I and Type II errors in relation to this study.
2. Explain the potential consequences of each type of error, considering both effectiveness and side effects.
3. Which error do you think is more dangerous in this scenario, and why?

short answer

What is meant by the sampling distribution? Why does the sampling distribution change based on the data?

Why does statistical significance change based on the critical region? What is a likely outcome of a hypothesis test if the alpha level decreases from .05 to .01?

How is the z-test statistic similar to a t-test statistic? How is it different? Provide a conceptual explanation without using any mathematical notation/formulas.

— lingering question

- I'm having trouble differentiating between the critical region and alpha

lingering question

- Quiz 10 question Which of the following is expected if the null hypothesis cannot be rejected for an analysis of variance?

Which of the following is expected if the null hypothesis cannot be rejected for an analysis of variance?

- ☐ MS_{model} should be about the same size as MS_{total}
- ☐ SS_{model} should be about the same size as SS_{total}
- ☐ MS_{model} should be about the same size as MS_{error}
- ☐ SS_{model} should be about the same size as SS_{error}

lingering question

- How can we tell that we made a type II error in a study?
 - low power (low sample size or high measurement error)
 - prior research shows an effect
 - later studies find the effect
 - trends in data but not statistically significant

— lingering question

- Can we please have a list of all the key concepts for Midterm 2 Conceptual?

key concepts

sample statistics and population parameters

sampling distributions

standard errors

null and alternative hypotheses

p-value, alpha, critical region, critical values

type I and type II error

power and effect size

one vs. two-tailed tests

F, z, and t-tests (which-test-when + similarities and differences)

review: key concepts

- answer the following questions:
- what is it? how would I explain this concept without any math?
- do I have a computational/mathematical definition?
- what factors influence it?
- how is it similar to other terms/ideas and how is it different?

sample statistics and population parameters

sampling distributions

standard errors

null and alternative hypotheses

p-value, alpha, critical region, critical values

type I and type II error

power and effect size

one vs. two-tailed tests

F, z, and t-tests (which-test-when + similarities and differences)

terminology confusion

- F-test vs. ANOVA
- sample vs. sample statistic vs. population parameter
- sampling distribution vs. population distribution
- sample statistic vs. test statistic
- critical value and p-value