NYU

Advanced Statistics
Spring 2023

Instructor: gostab.jen@nyu.edu

Department of Applied Math and Data Science, New York University

Data Assignment 3

- Grade released(with feedback)
- Section Average: 95%, A
- Points off: 1) interpretation of your analysis
- 2) misspecification of hypothesis

NYU

The last Recitation

Some tasks to do:

Check your previous grades

Rough calculation of your final grade

Start working on your TWO **data reports**

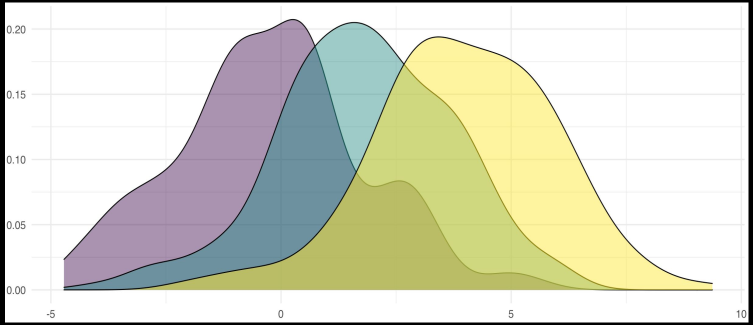
Week 14-15 Plans

Final Review

Topic 1: Regressions and Anova

Topic 2: Hypothesis Testing

Topic 3: Multivariate statistics and other non-computing topics



Week 14-15 Plans

Mann-Whitney U test/Wilcoxon Rank Sum Test
Margin values (Chi square test of independence)
Margin total
Mean of difference scores (paired t-test)
Mean Square (between, interaction, within)
Non-parametric statistical tests
Null Hypothesis testing strategy (steps involved in testing the Null Hypothesis assumption)
Observed frequency
Objectivity
Optional Stopping
Parametric statistical tests
P-hacking
Plots of possible two-way ANOVA outcomes
Power (also sample size and power)
Reliability
Sample size
Simple Main-effects (of the two-way ANOVA)
Split-half reliability
Standard deviation
Standard deviation of difference scores

Standard error of the mean
Statistical significance
Sum of squares (between, interaction, within, total)
T-test (one-sample, independent means, paired)
Test-retest reliability
Type I error
Type II error
Two-way ANOVA
Validity
Z-test vs. t-test

Regression and Anova

Regression will not be tested separately. However, that contains some important features linking to Anova table.

Regression and Anova

Regression will not be tested separately. However, that contains some important features linking to Anova table.

T-tests

One-way and two-way ANOVAs

Nonparametric tests

Effect size/power/null hypothesis testing

Test theory

Parametric Test-T test

- Degrees of freedom
-
- Mean of difference scores (paired t-test)
    - What does it mean for the mean of difference scores to be big? To be zero?
-
- Standard deviation of difference scores
-
- T-test (one-sample, independent means, paired)
    - What's the benefit of paired t-tests?

Non-parametric Test

- Chi Square test (Goodness of fit, test of independence)
  - Degrees of freedom
  - Expected frequency
  - Observed frequency
-
- Mann-Whitney U test/Wilcoxon Rank Sum Test
- Distribution-free test if there is no ties(same ranks)
-
- Parametric vs nonparametric statistical tests
-

Non-parametric Test

- **2. What measurement scale is associated with a chi square goodness** of fit test?

  a) Nominal b) Ordinal c) Interval d) Ratio

- **3. A non-parametric test**

  a) Is more powerful than a parametric test

  b) Has the advantage over a parametric test in that it requires only one parameter where as a parametric test requires at least two

  c) Is typically not as powerful a statistical test as a parametric test

- d) Does not require a critical value (for tests of statistical significance

Non-parametric Test

- **9. The Mann-Whitney U test/Wilcoxon Rank Sum test is**
-
- a) A test that lacks a determination of statistical significance
- **b) A non-parametric test**
- c) Uses the same degrees of freedom as a two-way ANOVA
- d) When the samples sizes are large (at least greater than 30 participants) the Mann-Whitney U test/Wilcoxon Rank Sum test used instead of the chi-square test of independence
-
- **10. If only frequency data is available (e.g. categorical data)**
-
- **a) A non-parametric statistical test can still be run**
- b) A two-way ANOVA can be run but without an interaction effect analysis
- c) Any parametric test can be run even without a standard deviation
- d) Internal validity cannot exist

Non-parametric Test

What other questions could you ask about nonparametric tests?

- Chi Square test (Goodness of fit, test of independence)
  - Degrees of freedom
  - Expected frequency
  - Observed frequency
- Mann-Whitney U test/Wilcoxon Rank Sum Test
- Parametric vs nonparametric statistical tests

effect size/power/null hypothesis testing

- Beta
- Effect: numerator of an inferential statistic
- Cohen's d
- Power
    - Beta, 1-beta, alpha
- Sample size and power
- Standard deviation
- Standard error of the mean
    - Think back to our CLT recitation: SEM is for the sampling distribution of sample means
- Statistical significance
- Type I error
- Type II error

effect size/power/null hypothesis testing

- **1. As the impact of an independent variable increases**
- **a) Beta decreases**
- b) Beta increases
- c) The effect size approaches zero
- d) Parametric tests become less reliable
-
- **4. Cohen's d is associated with**
- **a) The amount of overlap between two population distributions**
- b) Alpha (Cohen's d is a measure that determines what should be the optimal alpha level)
- c) Determining the shape of a population distribution (e.g. normal or skewed)
- d) Determining whether a test should be one-tailed or two-tailed

effect size/power/null hypothesis testing

- **6. Effect size involves**

  a) The differences between two (or more) populations (as measured by, for example, means)

  b) The size of a sample in a single cell (a specific condition) of a two-way ANOVA

  c) Two-tailed tests, but not one-tailed tests

  d) Determining whether alpha can be set to some value other than 0.05

Anova

- Bonferroni adjustment
    - Experiment-wise alpha: Bonferroni helps keep experiment-wise alpha down
- Degrees of freedom
- F-ratio
- Interaction effect (of the two-way ANOVA)
- Independent variables (especially as applied to the two-way ANOVA)
-
- Mean Square (between, interaction, within)
    - MS = SS/df
    - F = MSbetween/MSwithin
-
- Plots of possible two-way ANOVA outcomes
- Simple Main-effects (of the two-way ANOVA)
- Sum of squares (between, interaction, within, total)
-

Anova

- **8. An interaction effect of a two-way ANOVA**
-

a) Is seen when only one of the two other F ratios (of the two-way ANOVA) is statistically significant

b) Is seen when the dependent measure is expressed by different vectors (for example, increases across the levels of one independent variable, but decreases across the levels of the other)

c) Can't be indicated in a plot of the data (that is, when the means of the observed conditions are placed in a graph, an interaction effect wouldn't be suggested), it is only suggested by comparing the F values of the tests of the independent variables

d) Means that neither of the other two F-tests can be statistically significant

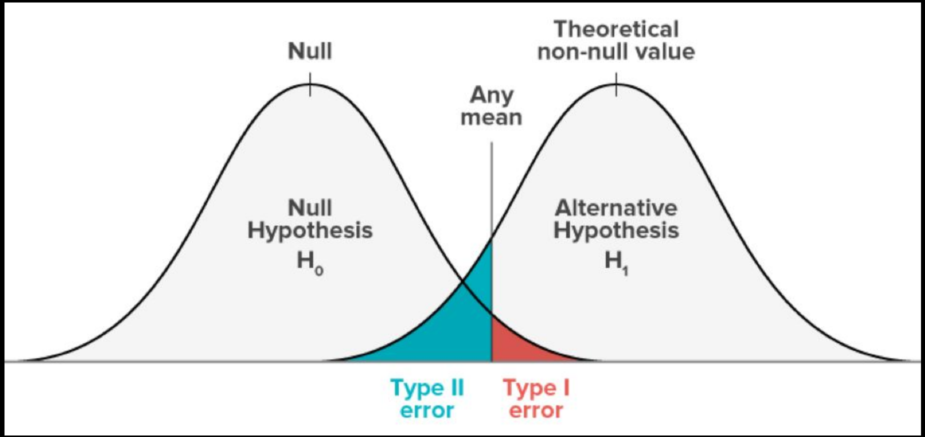
Hypothesis Testing

In this section I will walk you through all concepts of hypothesis testing and test of significance. This section is not designated for just doing manual calculations like statistics or critical value comparison **but for being a "deep thinker" of hypothesis testing:**

Build up a **strong link** to **conditional probability**

Null hypothesis testing: Type I and II errors

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β



Type I error: false positive

- You do find a significant result, but you shouldn't have (H_0 is true)

Type II error: miss

- You don't find a significant result, but you should have (H_0 is not true)

Which do you think is worse?

- Typically we think of Type I errors as worse (so we want to minimize Type I errors)

Central limit theorem

If you keep taking a bunch of samples (and take the mean of each sample), the **sample means will form a normal distribution, no matter what the original population distribution looks like**

No matter what the original distribution is, the sample means will form a normal distribution

Bigger sample size → less error, closer to true population mean