Small Sample Size

Reporting on and Analyzing Data with Small Samples in Institutional Data

Office of Assessment & Academic Program Review

Agenda

Goal: Promote proper handling of smaller datasets

- Tabulations
  - Removal
  - Aggregation
  - Imputation (not covered)
- Visualization
- Analysis
  - Two equal samples
  - Categorical
  - Pass/Fail
  - Proportions

- Where do you usually run into small numbers in your work?
- How many of you use statistical analysis regularly?

New Mexico State Guidelines for Education

- "The size required for reporting continues to be 10 or more students in a group, and publications of sensitive data follow uniform guidelines for avoiding disclosure of individual students."

  - NM State Plan 2017

For all public facing documents

Tabulations

Simple Solutions

Original

	2017-18	2018-19	2019-20	2020-21	2021-22
American Indian	2	4	4	4	4
Asian	2	0	0	3	0
Black or African American	12	12	11	13	10
Native Hawaiian	2	0	0	3	0
Hispanic	42	40	44	38	34
White	25	18	21	24	28
Unknown	34	25	21	20	20

Removal

	2017-18	2018-19	2019-20	2020-21	2021-22
American Indian	*	*	*	*	*
Asian	*	*	*	*	*
Black or African American	12	12	11	13	10
Native Hawaiian	*	*	*	*	*
Hispanic	42	40	44	38	34
White	25	18	21	24	28
Unknown	34	25	21	20	20

*Values below 10 have been suppressed to protect student privacy

*All data has been made up and does not reflect true values

Simple Solutions

Original

	2017-18	2018-19	2019-20	2020-21	2021-22
American Indian	2	4	4	4	4
Asian	2	0	0	3	0
Black or African American	12	12	11	13	10
Native Hawaiian	2	0	0	3	0
Hispanic	42	40	44	38	34
White	25	18	21	24	28
Unknown	34	25	21	20	20

Aggregation

	2017-18	2018-19	2019-20	2020-21	2021-22
Black or African American	12	12	11	13	10
Hispanic	42	40	44	38	34
White	25	18	21	24	28
Multiple Groups*	40	29	25	30	24

*Groups include American Indian, Asian, Native Hawaiian, and Unknown

*All data has been made up and does not reflect true values

Additional Solution

- Imputation:

“the assignment of a value to something by inference from the value of the products or processes to which it contributes.”

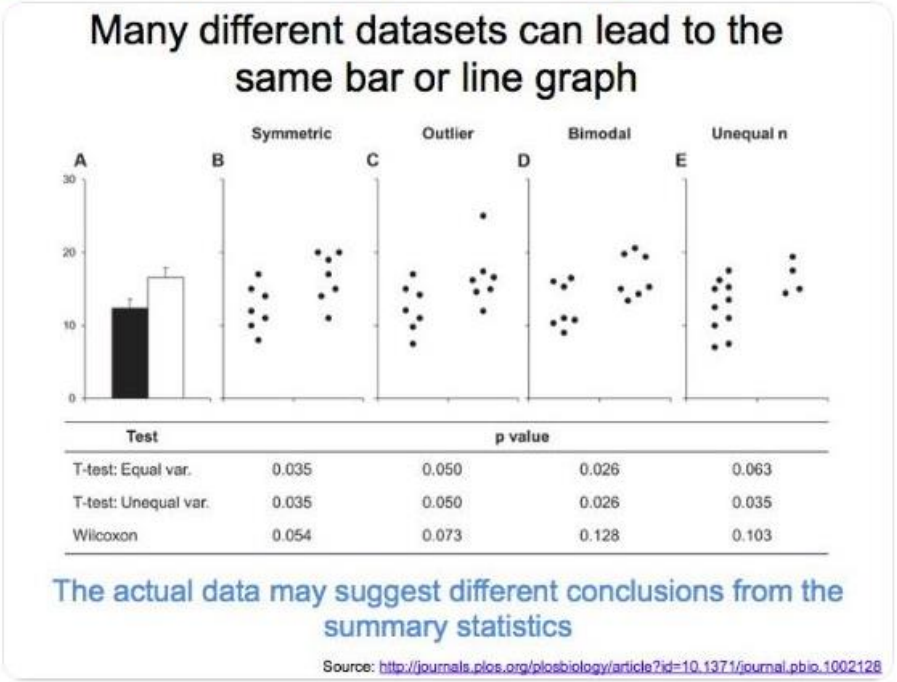
Complex solution that is not recommended unless you already know exactly what you are doing.

Visualization

Tracey Weissgerber
@T_Weissgerber

Follow

Designing better figures for small studies: Why you shouldn't use bar graphs for continuous data and what to do instead (A visual Q&A thread) journals.plos.org/plosbiology/ar ...



2:41 AM - 22 Jan 2019

Considerations

- Histograms and Line graphs are misleading
- Need to find alternatives that don’t mask the fact that samples are small
- When in doubt, tables are always ok

- Beyond Bar and Line Graphs; Weissgerber, Milic, Winham, Garovic

Analysis

Surveys

- Bias!
  - Inference based on limited data can be dangerous
  - Are there differences in respondents and non-respondents?
  - It's been proven those in the middle are less likely to respond
    - The good & the bad
    - If smaller populations are of interest, try to over sample
- Can you find ways to incentivize?
  - Be creative, it doesn't always have to be monetary
  - Sometimes removing a barrier is better than adding incentives

Meeting Benchmarks

- Think through how this impacts benchmarks
  - You may not be able to break down anymore when reporting!
  - If consistently 4 students, is 80% reasonable?
  - Move the mark?
  - Can you average over multiple assessment cycles?
  - Add a qualitative component?

Preliminary Conclusions

When presenting data, always focus on protecting your population

Create graphs in ways that don't over emphasize small trends

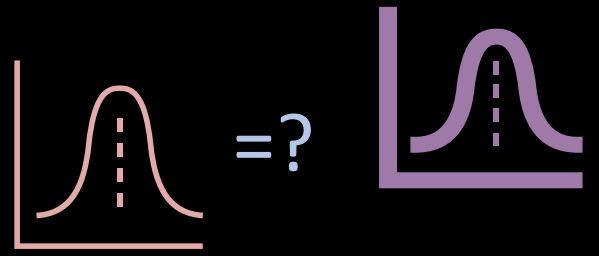
- For those of you who need more than to simply present data, what do you need to know about your populations?

Statistical Tests Have Assumptions!

- The test you chose depends on the distribution of the data being tested.
- Always make sure you know the assumptions and test them before you use an analysis

Nonparametric tests hold up better when you are unsure of the assumptions

T-Test



- Used to compare two populations to one another
  - Example) Compare test scores at beginning and end of course
    Students who engage with tutor or not in embedded course
- Using a t-test is highly common but has strict assumptions
  1) Normality
     - Must be looked at first (second assumption is dependent)
     - Normal probability plot
     - What if the sample is really small? - Institutional Data or Gold Standard Benchmarking
  2) Equal Variance
     - If sample sizes are unequal, just assume variance is not equal when sample is small
     - Ratio of smallest to largest should be <= 3

- If either assumption is violated, go nonparametric
  - Wilcoxon or Mann-Whitney

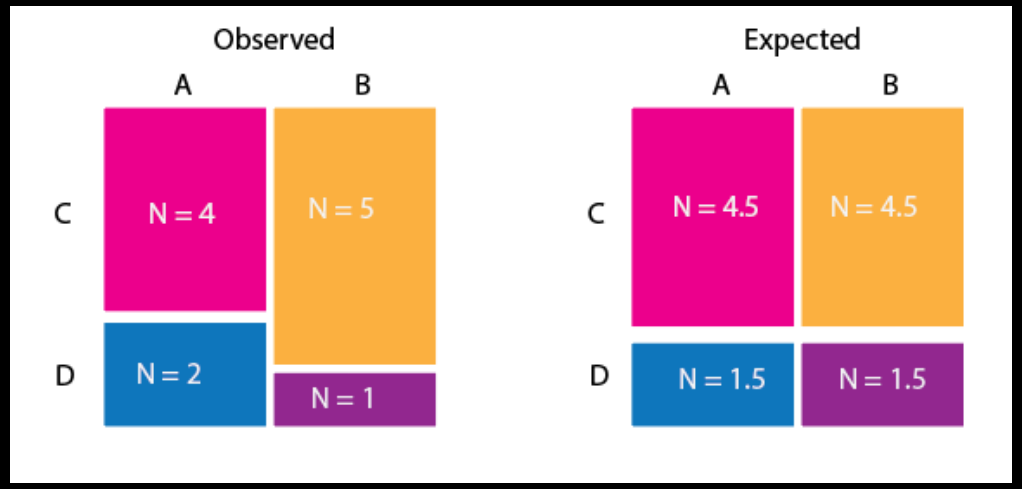
Use of proper statistical techniques for research studies with small samples, Morgan

Fisher Exact Test

- Chi-square alternative when expected values are 5 or less. Can handle expected values of 0.

- Good for comparing categorical groups

Example) A/B is course pass/fail
C/D is an intervention



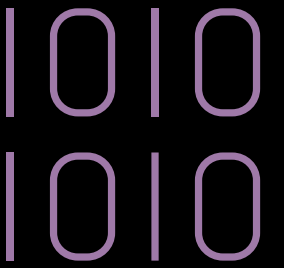
- Fisher's Exact Test, Wolfram MathWorld

Sign Test

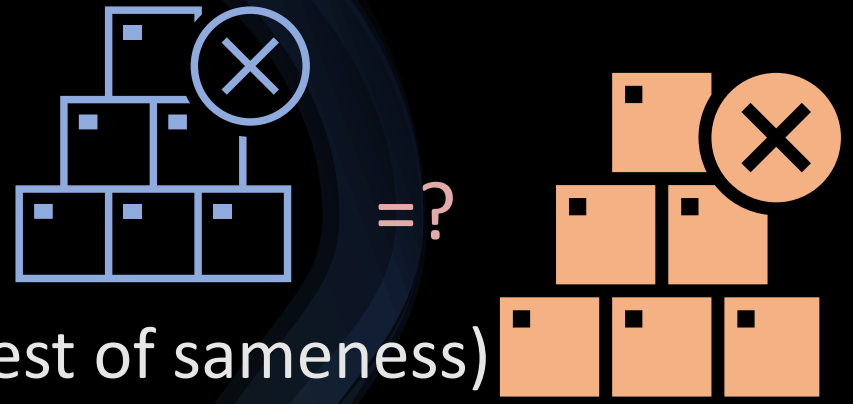
- Used to detect likelihood of success vs failure
  - Examples) course grade, retention, graduation

- Distribution free (no need to determine)
  1) Each trial comes from the same population
  2) Measurements are ordered (success/failure or another binomial)
  3) Trials are independent (one instance is not dependent on another)

- Small Samples with Meaningful Results, Pizur



Mantel –Haenszel Tests



- Good for comparing population distributions (test of sameness)
  Example) Does my small sample come from a known larger population?

- Can compare samples as small as 2 (though I wouldn't recommend it)

- Can also be used for survival analysis
  - We aren't looking at survival you say? If Death=Graduation, we want a quick curve to the end for all our students! The metrics still work... and are happier.

Mantel Haenszel Test, Encyclopedia of Social Measurement

Conclusions

When presenting data, always focus on protecting your population

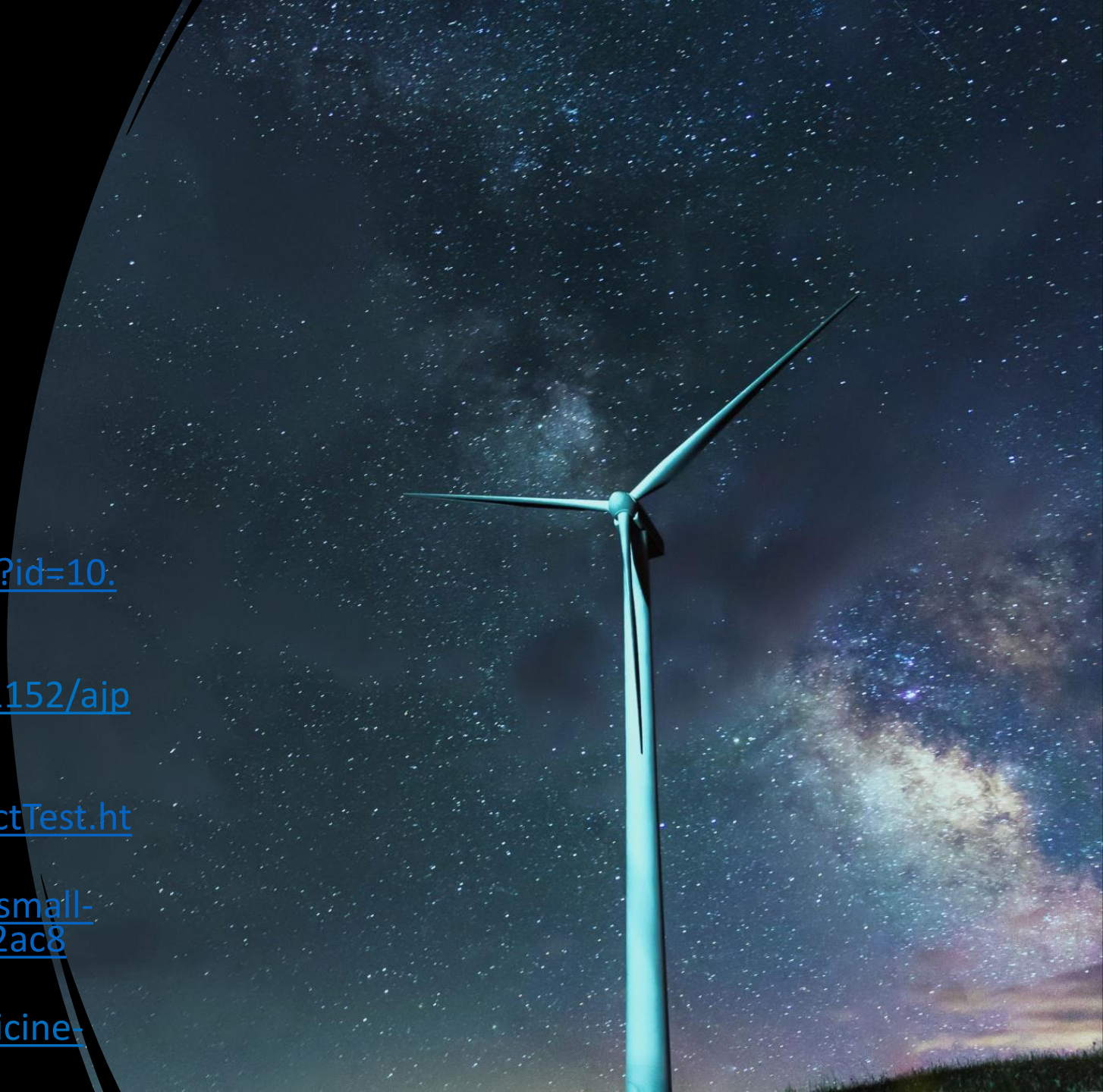
Create graphs in ways that don't over emphasize small trends

Make sure your analysis fits your data

Result will limit you to seeing large results only but miss the nuance so try not to over-generalize

Resources

- General Best Practices - https://measuringu.com/small-n/
- Visualizations - https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128
- Proper Parametric Tests - https://journals.physiology.org/doi/pdf/10.1152/ajplung.00238.2017
- Fisher Exact - https://mathworld.wolfram.com/FishersExactTest.html
- Sign Test - https://towardsdatascience.com/small-samples-with-meaningful-results-1a1b15052ac8
- Mantel-Hanezel - https://www.sciencedirect.com/topics/medicine-and-dentistry/mantel-haenszel-test



Thank you!

Office of Assessment & Academic Program Review