STATS 101B

M. Brooke Robertshaw, PhD
Office of Assessment, Purdue University
18 July 2014

ANOVA

What is it?

Analysis of Variance

Types of ANOVA

- One-way ANOVA (just one independent variable)
- Factorial ANOVA (more than one independent variable, one dependent variable)

ANOVA

How is it used?

To test for differences in means between groups (ex. URM/ NURM)

Data types for ANOVA

- Dependent variable: continuous, (GPA, test scores)
- Independent variable: nominal, ordinal, ratio
- Assumptions: Normality, equal cell sizes, equal variance between groups

Can be run in Excel, see: http://www.excel-easy.com/ examples/anova.html

ANOVA

Example for using ANOVA

Scenario: 3 instructors, one quiz, and an intervention. Testing for differences between instructors, to see if there is a difference:

- When highest grade earned is compared, there is a significant difference for instructor, $F(3)=8.632$, $p=.000$

Other examples:

- You want to know if GPA is higher for different races.
- You want to know what group has the highest mean usage of DRC resources
- You want to know if there is a difference in 4 year graduation rates between athletes in different sports

Other types of ANOVA

ANCOVA, MANOVA, Generalized Linear Modeling

Why use these?
Use of co-variates (ANCOVA), More than one dependent variable (MANOVA)

Data types

- Dependent variable: continuous, (GPA, test scores)
- Covariate: Continuous, nominal, ordinal, ratio (SPSS just continuous)
- Independent variable: nominal, ordinal, ratio

Assumptions: Normality, equal cell sizes, equal variance between groups

ANCOVA can be run in Excel (but it is easier in SPSS), see: http://www.real-statistics.com/analysis-of-covariance-ancova/

Effect Size

What is it?

- A way to measure magnitude of change, different than statistical significance (p values)

You can have statistical significance, but no magnitude of difference, and you can have no statistical sigificance but a high magnitude of change.

Particularly good with large sample sizes (above 500) and small sample sizes (below 100)

Effect Size

Example (from above)

When highest grade earned is compared, there is a significant difference for instructor, F(3)=8.632, p=.000

Effect size for pairwise comparisons between instructors

Instructor	Cohen's *d*
Teacher B, time 1 x Teacher A	.12
Teacher B time 2 x Teacher C	.63
Teacher B time 1 x Teacher B time 2	.86
Teacher A x Teacher C	.10

Small	**0.1-0.299**
Medium	0.3-0.599
Large	0.6- above

Effect Size

How to calculate it?

Cohen's d:

m1-m2 / sd1 or, even better m1-m2/ pooled sd

Or (even better):

http://www.uccs.edu/~lbecker/

Or

SPSS will compute an eta (η^2)

Small Sample Sizes

It can be done! Non-parametrics! Normality testing!

Generally..

- Small samples (less than 100) are less likely to be robust to violations of normality,
- There may be statistical significance, but there isn't enough power to detect it.

Small Sample Sizes

How?

Parametric	Non-Parametric
Independent Samples T-Test	Mann-Whitney U
Independent ANOVA/ MANOVA	Kruskal-Wallis
Dependent (Matched-Pairs) t-test	Wilcoxon Signed Rank Test
Matched ANOVA (more than 2 groupings of the same sample)	Friedman Rank Sum Test
ANCOVA/MANOVA/ MANCOVA	Generalized Estimating Equation

Small Sample Sizes

Recommendation

Run the parametric and non-parametric and if the p values are similar, report the parametric and use a reference stating:

- You ran parametric / non parametric,
- there is no difference between them, thus
- the data is robust to violations of normality.

Reference:

Smith, M. (2003). *Research Methods in Accounting*. Thousand Oaks, CA: Sage.

(See stats.brookerobertshaw.com for text from the book)

Small Sample Sizes

Example

Sample: 7 students, repeated measures, investigating ability to write a scientific argument.

Data: From grading rubric, points assigned to each piece of a scientific argument, continuous

Analyzed using: Friedman Rank Sum Test

Finding: No difference in ability for students to write scientific argument from pre-post $\chi^2(1)=1.29$, $p=.26$

But!

From: Robertshaw, M.B. & Campbell, T. (2013). Constructing arguments: Investigating pre-service science teachers' argumentation skills in a socio-scientific context. *Science Education International*.

Small Sample Sizes

Effect sizes for each TAP component and overall score. (Vargha and Delaney (2000) set values to be 0.56=small, 0.64=medium and 0.71=large)

Component	Value
Backing	A=0.57
Claim	A=0.64
Data	A=0.71
Qualifier	A=-0.64
Rebuttal	A=0.79
Warrant	A=0.57
Argumentation Total Score	A=0.74

From: Robertshaw, M.B. & Campbell, T. (2013). Constructing arguments: Investigating pre-service science teachers' argumentation skills in a socio-scientific context. *Science Education International*.

Reliability

That would be, test reliability.
What does it test? Whether a test tests different samples of the same population the same.

Statistic:

- Cronbach's alpha
- KR-20 (dichotomous)

Reliability

Interpretation	Value
Unacceptable	0-0.499
Poor	0.5-0.599
Acceptable	0.6-0.699
Good	0.7-0.8999
Excellent	Above 0.9

How?

- In SPSS -> analyze ->scale -> reliability analysis
- In Excel: http://languagetesting.info/statistics/excel.html

**Note: This is the most common form of reliability, there are others*

Questions?