# 26:010:557 / 26:620:557
# Social Science Research Methods


**Dr. Peter R. Gillett**

**Associate Professor**

**Department of Accounting & Information Systems**

**Rutgers Business School – Newark & New Brunswick**

# *Overview*

- Analysis of frequencies
- Analysis of Variance
- Multiple Regression Analysis
- Other Statistical Techniques
- References
- Research proposals

# *Analysis of Frequencies*

- Crosstabs
  - E.g. 2 x 2 tables
- Frequencies v. percentages
- Contingency tables
- $\chi^2$ test
- Levels of significance
- Yates' correction when N small
- Fisher exact test for small N
- Cramer's V (measures strength of association)

# *Analysis of Variance*

- Total variance
- Between group variance
- Within groups variance
- Error variance
- t-tests (Student t)
- One-way ANOVA
- Two-way ANOVA

# *Analysis of Variance*

- F tests
- Degrees of freedom
- Strength of relations (effects) – $\eta$ and $\eta^2$, $\omega^2$
- Post Hoc comparison
  - Many alternatives – e.g., Scheffé
- Planned comparisons
  - Orthogonal contrasts
  - Family-wise / experiment-wise alpha risks
  - Bonferroni adjustment (Dunn procedure)
  - Dunn / Sidak adjustment
  - Tukey WSD, Dunnett, Fisher LSD, Duncan, Newman-Keuls, . . .

# *Analysis of Variance*

- Contrasts
  - Deviation
    - Compares the mean of each level (except a reference category) to the mean of all of the levels (grand mean). The levels of the factor can be in any order.
  - Simple
    - Compares the mean of each level to the mean of a specified level. This type of contrast is useful when there is a control group. You can choose the first or last category as the reference.
  - Difference
    - Compares the mean of each level (except the first) to the mean of previous levels. (Sometimes called reverse Helmert contrasts.)
  - Helmert
    - Compares the mean of each level of the factor (except the last) to the mean of subsequent levels.
  - Repeated
    - Compares the mean of each level (except the last) to the mean of the subsequent level.

# *Analysis of Variance*

- Factorial Analysis of Variance (Chapter 14)
- Interactions
- Main effects
- Hard to interpret main effects when interactions are significant
- Simple effects
- Correlated groups ANOVA (Chapter 15)
  - Randomized blocks
  - Within subjects designs
  - Repeated measures designs

# *Contrasts*

- Suppose the means of k factor levels being sampled are $m_1, m_2, \ldots, m_k$
- A linear combination C of these means is said to be a contrast if the sum of its coefficients is zero; i.e., $C = \sum_{i=1}^{k} c_i m_i$ where $\sum_{i=1}^{k} c_i = 0$
- Two contrasts are *orthogonal* if $\sum_{i=1}^{k} \frac{c_{1i} c_{2i}}{n_i} = 0$

# *Analysis of Variance*

- Assumptions
  - Independent random samples
  - Random assignment of treatments
  - Equal population variances of groups
  - Equal cell sizes (sample sizes)
  - Scores normally distributed
- Nonparametric tests
  - Kruskal-Wallis (One way ANOVA)
  - Friedman Test (Two way ANOVA)

# *Multiple Regression Analysis*

- Simple regression (Chapter 32)
  - Slope, intercept
- Multiple regression
  - Ordinary Least Squares (OLS)
  - Multiple Correlation Coefficient (and $R^2$)
  - Significance tests
    - Regression Coefficients – t test
    - $R^2$ – F test

# *Multiple Regression Analysis*

- Assumptions
  - Regression model has linear form
    - $Y = X\beta + \varepsilon$
  - X is an n x K matrix with rank K (identification)
  - The error term has expected value zero for every observation
  - Error variances are constant (homoskedasticity) and error covariances are zero (nonautocorrelation) – collectively – sphericity
  - X is known and constant
  - Errors are normally distributed (normality)

# *Multiple Regression Analysis*

- ANOVA as a regression analysis
- Dummy variables
  - Dummy coding
    - k-1 dummies for k categories
  - Effects coding
  - Orthogonal coding
- ANCOVA

# *Multiple Regression Analysis*

- Discriminant Analysis
  - Estimates the linear combination of (independent) variables that best discriminates between two groups
    - Do the independent variables discriminate
    - If so, which group should each subject belong to

# *Multiple Regression Analysis*

- Canonical Correlation
  - A multivariate technique
  - Given two sets of variables, estimates the linear combinations of variables in each group that have the highest correlation with each other

# *Multiple Regression Analysis*

- MANOVA
  - A multivariate technique
  - Multivariate equivalent of ANOVA
  - Examines how groups differ on linear combinations of a set of dependent variables
  - Important because groups may not differ significantly on any single variable, but may still differ significantly on linear combinations of variables
  - Assumes multivariate normality

# *Multiple Regression Analysis*

- Path Analysis
  - Repeated use of regression
- Ridge Regression
  - Attempts to solve problems with OLS arising from multicollinearity
    - Coefficients too large
    - Coefficients have wrong sign
    - Coefficients unstable
    - Regression weights over- or under-estimate
  - Estimates biased

# *Multiple Regression Analysis*

- Logistic Regression
  - Applicable when criterion variable (dependent variable) distributed binomially instead of normally
  - E.g., when criterion is dichotomous
  - Essentially, applies a transformation then OLS
  - Coefficients, when exponentiated, show how odds of criterion are multiplied
  - Also, polychotomous logistic regression

# *Multiple Regression Analysis*

- Probit is an alternative approach to logit in many instances
- Tobit is used when dependent variables are truncated at zero

# *Multiple Regression Analysis*

- Log-linear Analysis
  - When all variables are categorical
  - Multiway contingency tables
  - Saturated and unsaturated models

# *Multiple Regression Analysis*

- Does it matter whether you are testing the *right* model?
  - Irrelevant Variables
  - Omitted Variables

# *Irrelevant Variables*

- True Model: $Y = \alpha + \beta X$
- Estimate: $Y = \alpha + \beta_1 X + \beta_2 Z + \varepsilon$
- Problems:
  - Estimates unbiased and consistent
  - But generally inefficient (i.e. variances too large)
  - Extra degrees of freedom consumed
  - Introduces additional multicollinearity
  - Reduces precision
  - May lead to wrong conclusions as to significance

# *Omitted Variables*

- True Model: $Y = \alpha + \beta_1 X + \beta_2 Z$
- Estimate: $Y = \alpha + \beta X + \varepsilon$
- Problems:
  - Z and X uncorrelated
    - Estimates of intercept biased but estimates of beta coefficients unbiased
    - Estimates inconsistent
  - Z and X correlated
    - Estimates biased and inconsistent
    - Variances biased (could be too large or too small)
    - Even signs of beta coefficients are not reliable!
    - Worse with binary variables or in presence of heteroscedasticity

# *Factor Analysis*

- A factor is a construct – a latent variable
- Data reduction
- A measure is factorially pure if it measures only one factor
- Factor matrix
  - Factor loadings
    - Correlation between measures and factors
- Communalities
  - Common factor variance

# *Factor Analysis*

- How many factors?
- Communality
- Principal factor method (principal axis)
- Rotation
  - Orthogonal
  - Oblique
- Second order factors
- Factor scores

# *Factor Analysis*

- Exploratory factor analysis (EFA)
- Confirmatory factor analysis (CFA)
- Reification
  - Naming factors does not make them real
- Construct validity and constitutive definitions

# *Other Statistical Techniques*

- General Linear Model
  - Generalizes the techniques we have discussed
- Structural Equation Modeling
  - Analysis of covariance structures
  - Combines factor analysis with a form of path analysis

# References


- Planned Contrasts and Post-Hoc Test for One-Way ANOVA
  A. Karpinski
  http://astro.temple.edu/~andykarp/psych522/06_contrasts2.pdf
- Strengths and limitations of Ipsative Measurement. Journal of Occupational and Organizational Psychology, 69, 49-56
  Helen Baron
- Pitfalls of Using Proxy Variables in Studies of Guns and Crime
  C.E. Moody & T.B. Marvell
  http://papers.ssrn.com/sol3/papers.cfm?abstract_id=473661
- How Variables Uncorrelated with the Independent Variable Can Actually Make Excellent Predictors: The Important Suppressor Variable Case
  K.K. Woolley
- Effect Size
  http://web.uccs.edu/lbecker/Psy590/es.htm
- What is an “Effect Size”
  Robert Coe
  http://www.cemcentre.org/ebeuk/research/effectsize/ESguide.htm
- Measures of Effect Size (Strength of Association)
  http://web.uccs.edu/lbecker/SPSS/glm_effectsize.htm

# *Research Proposals*

- Outline research proposals are due next week
  - By the start of class
  - Via the Digital Drop box on Blackboard, as usual
- Your goal is to demonstrate mastery of the material we have studied together in this class, not mastery of the literature or the practice of research in your own field (and so need not even be in your own field of expertise)
- You will not be required to conduct the research, so you should not allow your proposal to be limited by actual resource constraints – although it should still be research that is potentially doable
- It should be the best *proposal* you can imagine, based on what we have studied together