

Analysis of Data

Claudia J. Stanny
PSY 6217 – Research Design



Organizing Data Files in SPSS

☞ All data for one subject entered on the same line

- ◆ Identification data
- ◆ Between-subjects manipulations: variable to identify group level for that subject
- ◆ Measures for Within-subjects manipulations: each level of a within subject variable appears as a new variable in the data file

ID	SEX	AGE	HAND	WD1	WD2	WD3	PG1	PG2	PG3	RTW1	RTW2	RTW3	KTP1	KTP2	KTP3
1	1	2	2	12	13	11	14	16	15	617	725	710	550	513	490
2	1	1	2	11	15	10	16	20	19	982	880	798	682	773	825
3	2	1	1												
4	2	2	1												
5	1	2	1												
6	2	1	2												

Descriptive Statistics

☞ Biased and Unbiased Estimates of Population parameters

☞ Measures of Central Tendency

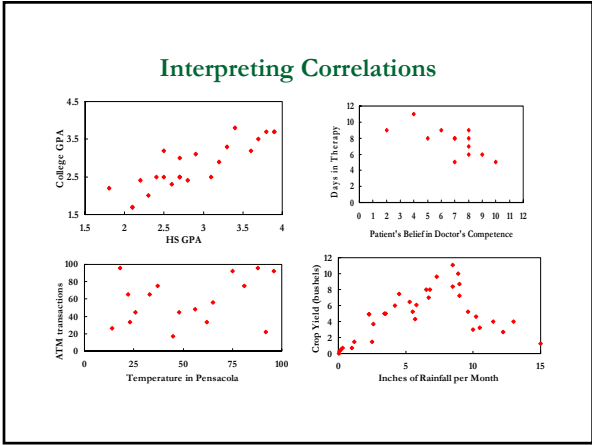
- ◆ Operational definitions of a "typical" score
- ◆ Mean, median, mode

☞ Measures of Variability

- ◆ Range, interquartile range
- ◆ Variance
- ◆ Standard deviation

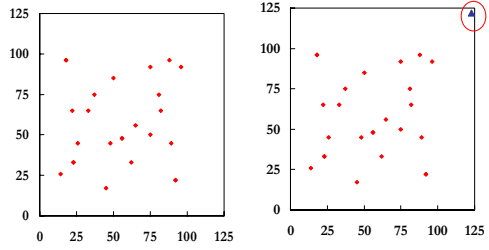
Measures of Association

- ∞ Pearson product-moment correlation (r)
- ∞ Point-biserial correlation
 - ♦ *Equivalent to a between-subjects t-test*
- ∞ Spearman rank order correlation (r_s)
- ∞ Phi coefficient

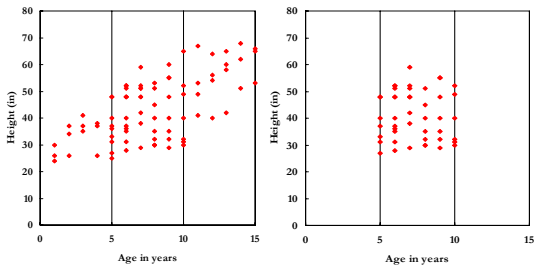




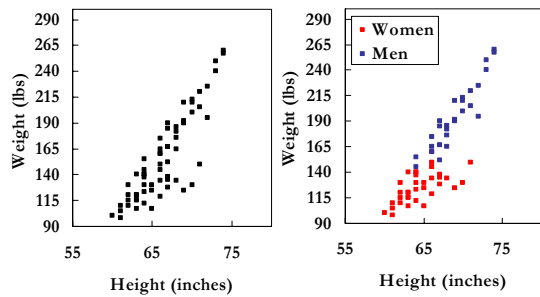
Effects of Outliers



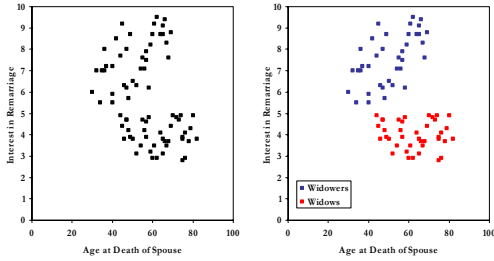
Restricted Range



Combining Data From Different Groups



Data From Different Groups

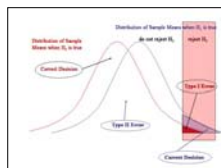
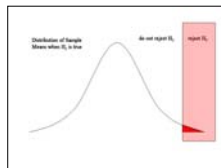


Inferential Statistics

- ☞ Logic of Inferential Statistical Tests
- ☞ Assume that all participants are selected from the same population
 - ♦ $Score (DV) = \mu + \text{random error}$
 - ♦ $H_0: \mu_1 = \mu_2$
- ☞ Assume that the experimental conditions produce systematic changes in the scores of participants on the dependent measure
 - ♦ $Score (DV) = \mu + \text{treatment effect} + \text{random error}$
 - ♦ $H_1: \mu_1 \neq \mu_2$

Types of Statistical Errors

- ☞ Type I Error
 - ♦ Probability = alpha
 - ♦ Rejection of H_0 when H_0 is true
- ☞ Type II Error
 - ♦ Probability = beta
 - ♦ Failure to reject H_0 when H_1 is true
 - ♦ Failure to detect a true treatment effect



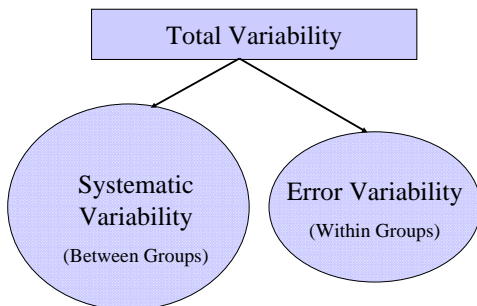
Parametric Statistics

- ∞ Random selection and/or random assignment of scores from the population of interest
- ∞ *Sampling distribution* is normal
 - ♦ *Central Limit Theorem: sampling distribution will become normal as sample size increases, even when raw data are not distributed normally*
- ∞ Homogeneity of variances within groups
 - ♦ *Problems can arise when observations in one condition are more variable than observations in other conditions*
- ∞ Testing these assumptions statistically

Analysis of Variance

- ∞ One-way designs
 - ♦ *One independent variable (factor)*
 - ♦ *two levels of a factor – a t-test will give equivalent results*
 - ♦ *three or more levels of a factor requires use of ANOVA*
- ∞ Partitioning Variance
 - ♦ *Between-groups variability*
 - ♦ *Within-groups variability*

Partitioning Variance in ANOVA



Creating an F Ratio for ANOVA

- ☞ $SS_{total} = SS_{treatment} + SS_{error}$
- ☞ Mean Square = SS / df
- ☞ Expected $MS_{BG} = Var_{TX} + Var_{error}$
- ☞ Expected $MS_{WG} = Var_{error}$
 - F Ratio = MS_{BG} / MS_{WG}
 - = $(Var_{TX} + Var_{error}) / Var_{error}$
- ☞ Value of F ratio when H_0 is true?
- ☞ Value of F ratio when H_0 is false?

Interpreting a Significant F Ratio

- ☞ Need to evaluate pair-wise comparisons of means following a significant F ratio
- ☞ Planned comparisons
 - ♦ *A priori comparisons*
 - ♦ *Comparisons based on specific predictions hypothesized before data collection*
- ☞ Unplanned comparisons
 - ♦ *Post hoc tests*
 - ♦ *Need to control Experiment-wide Error Rate (family-wise error rate)*
 - ♦ *Probability of at least one Type I Error increases as the number of comparisons increases*

Planned Comparisons

- ☞ Contrasts should be orthogonal
 - ♦ *Outcome of each contrast should be independent*
 - ♦ *Knowing the result for one comparison does not constrain the decision for other comparisons*
- ☞ Example: Planned comparisons for a three group design: Drug A, Drug B, Control
 - ♦ *Contrast 1: Average of Drug versus No Drug*
 - *Combines mean for drug A and drug B*
 - ♦ *Contrast 2: Drug A versus Drug B*

Orthogonal Contrasts Produce Independent Outcomes

Assume that the comparison of the Drugs versus Control produces no difference

Mean for Drugs = 5 Mean for Control = 5

3 outcomes possible for the comparison of Drug A with Drug B

- ♦ *No Difference* Drug A = 5 Drug B = 5
- ♦ *Drug A > Drug B* Drug A = 7 Drug B = 3
- ♦ *Drug A < Drug B* Drug A = 2 Drug B = 8

Knowing the decision about the first contrast does not provide information about the outcome of the second contrast

Independence of Contrast Outcomes (continued)

Assume that the comparison of the Drugs versus Control produces a significant difference

Mean for Drugs = 4 Mean for Control = 8

3 outcomes possible for the comparison of Drug A with Drug B

- ♦ *No Difference* Drug A = 4 Drug B = 4
- ♦ *Drug A > Drug B* Drug A = 6 Drug B = 2
- ♦ *Drug A < Drug B* Drug A = 1 Drug B = 7

Knowing the decision about the first contrast does not provide information about the outcome of the second contrast

Post Hoc Tests

Scheffe

- ♦ *Corrects for all possible comparisons, whether or not these are examined*

Bonferroni procedure

- ♦ *Divide alpha by the number of comparisons to be made*
- ♦ *Use the critical value for the smaller alpha value for all tests*

Dunnett

- ♦ *Specifically designed to compare each experimental group with a single control group*

Post Hoc Tests

☞ Tukey HSD (default post hoc in SPSS)

- ♦ Controls *EER* over a set of pair-wise comparisons (adjusts critical value for significance)
- ♦ Specialized procedures available for calculation for means based on unequal sample sizes

☞ Newman-Keuls

- ♦ Less conservative than Tukey HSD
- ♦ Means in a comparison set are rank ordered
- ♦ Critical value for a pair-wise comparison varies with the number of "steps" between means in the ordered set

☞ Duncan

ANOVA – Two Factor Designs

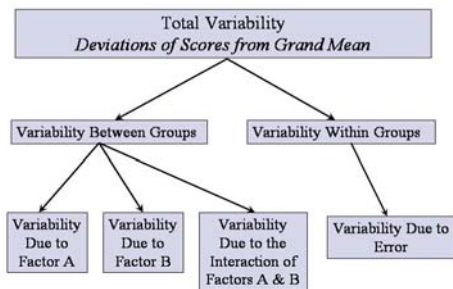
☞ Additional factors require partitioning variance in different ways

☞ Nature of the partitioning will change depending on the type of design

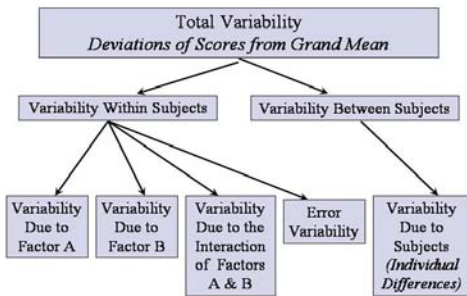
- ♦ Both factors manipulated between-subjects
- ♦ Both factors manipulated within-subjects
- ♦ Mixed designs

☞ Logic of the test remains the same: Isolate the systematic effect in the numerator of the F ratio

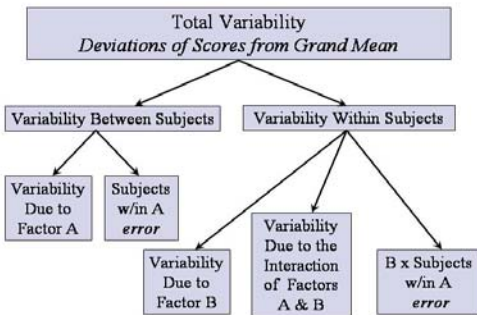
Partitioning Variance in the Between-Subjects Design



Partitioning Variance in the Within-Subjects Design



Partitioning in a Mixed Design



Fixed Effects & Random Effects

Fixed Effects

- ♦ All levels of a factor are in the design
 - men & women
- ♦ All relevant levels of a factor are in the design
 - zero stress, moderate stress, high stress

Random Effects

- ♦ Levels of a factor represent a sample of the possible levels in the design
 - Individual subjects as levels of a factor
 - Unique stimulus materials (e.g., photos, sentences, etc. used to create a manipulation)
- ♦ Analysis problems when 2 or more factors are random effects

Interaction Effects

☞ Additive Model (no interaction)

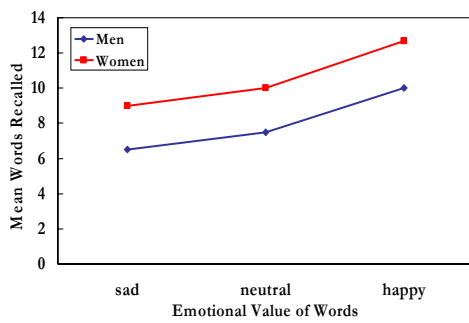
$$\text{Score} = \mu + \text{TXa} + \text{TXb} + \text{error}$$

☞ Non-Additive Model (interaction)

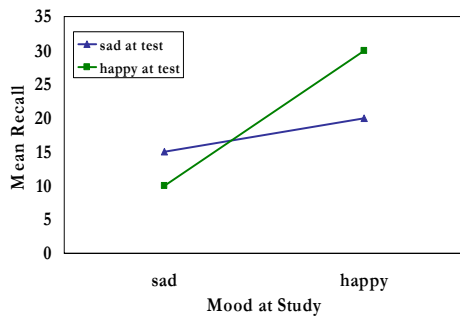
$$\text{Score} = \mu + \text{TXa} + \text{TXb} + \text{TXab} + \text{error}$$

☞ Meaning of interactions: *effect of a treatment varies depending on conditions created by manipulation of a second variable*

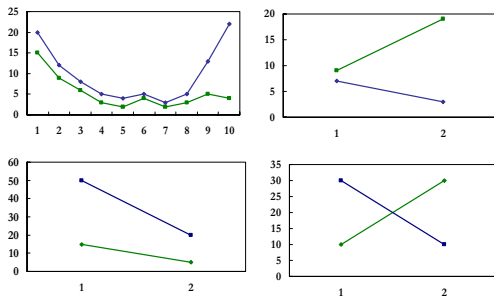
2 x 3 Factorial with No Interaction



2 x 2 Factorial with Significant Interaction



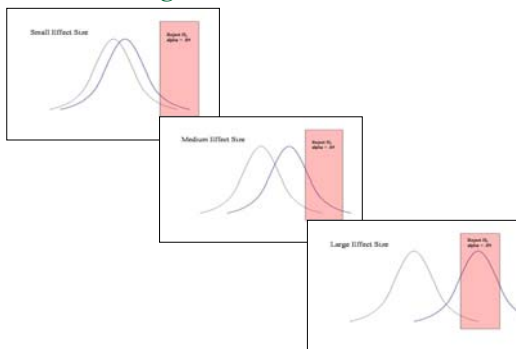
Other Forms of Interactions



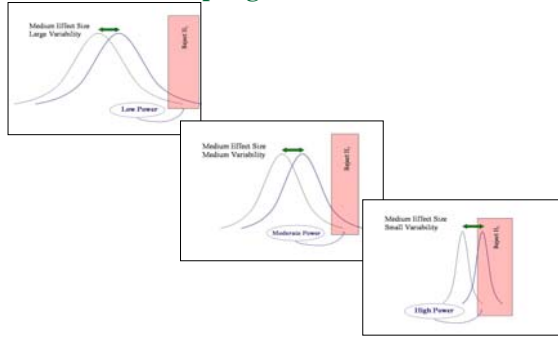
Factors that Affect Statistical Power

- ∞ Selection of alpha level
 - ◆ Comparing alpha = .05 to alpha = .01
 - ◆ Comparing one tailed tests to two-tailed tests
- ∞ Size of the treatment effect produced
- ∞ Variability of the *sampling distribution* for the statistical test
 - ◆ Effects of changes in sample size
 - ◆ Effects of changes in control of extraneous variables

Power Changes with Size of Effect Created



Power Changes with Variability in the Sampling Distribution



Measuring Effect Sizes (Abelson)

☞ Raw effect size

- ♦ Differences created measured in the same units of measurement used for the dependent measure
- ♦ Difference in IQ, number of words recalled, etc.
- ♦ Difficulty of knowing what counts as a “large” effect

☞ Standardized effect size

- ♦ Effect sizes converted to z-scores
- ♦ Creates equivalent scales for comparing different dependent measures
- ♦ Useful for meta-analysis
- ♦ Cohen’s d, variance explained (r^2), omega squared

Interpreting Statistical Significance

☞ Large sample sizes can provide powerful tests in which small numeric differences are statistically reliable

☞ How should these be interpreted?

☞ Numerically small effects can be impressive

- ♦ Created with minimal manipulation of the IV
- ♦ Created when the DV is difficult to change at all
- ♦ Effects of any size have an important theoretical interpretation

Statistical Significance vs Practical Significance

☞ What are the practical implications of an effect?

- ♦ *An educational program reliably produces a 2 point increase in the IQ of 10-year-old children*
- ♦ *A medical intervention reduces the risk of heart attack from 2 per 10,000 ($p(\text{heart attack}) = .0002$) by half (to $p(\text{heart attack}) = .0002$)*
 - In a population of 10 million people, this will translate into a change from 1,000 heart attacks to 500 heart attacks

Data Transformations

☞ Linear transformations

- ♦ *Add a constant value*
- ♦ *Shifts location of distribution without changing the variance of the distribution*
- ♦ *Can make data analysis easier (1K vs 1000)*

☞ Other transformations change both the mean and the variance of the distribution

- ♦ *Transforming data can produce data that meet the distribution assumptions for statistical tests*

Non-Linear Data Transformations

☞ Square Root transformation

- ♦ *Used when larger means correlate with larger variances (common with count data)*
- ♦ *Makes variances more homogeneous*
- ♦ *Reduces skew of distributions*

☞ Arcsine transformation

- ♦ *Used when data consist of proportions (e.g., percent correct data)*

Non-Linear Data Transformations

☞ Log transformation

- ♦ *Reduces skew in distributions*
- ♦ *Common with sensory data – gives manageable ranges for plotting results*

☞ Reciprocal transformation

- ♦ $X = 1/X$
- ♦ *Reciprocal data are less variable than raw data*
- ♦ *Transforms alter the interpretation of the measure*
 - *Reciprocal of a time measure is a speed measure*
- ♦ *New measure is sometimes easier to interpret*
 - *GSR gets smaller as stress increases*
 - *EDR (1/GSR) increases with increased stress*

Non-Parametric Tests

- ☞ Do not make assumptions about the underlying distributions to compute probabilities
- ☞ Useful when data do not meet assumptions for normality, etc.
- ☞ Can be more powerful tests than parametric tests when assumptions have been violated
- ☞ Non-parametric analogs exist for nearly every parametric test
