Multivariate Statistics

Psy 524

Andrew Ainsworth

Stat Review 1

IV vs. DV

- Independent Variable (IV)
  - Controlled by the experimenter
  - **and/or** hypothesized influence
  - **and/or** represent different groups

IV vs. DV

- Dependent variables
  - the response or outcome variable
- IV and DV - “input/output”, “stimulus/response”, etc.

IV vs. DV

- Usually represent sides of an equation

$$x \rightarrow y$$

$$x \rightarrow y \rightarrow z$$

$$\beta x + \alpha = y$$

$$x = \beta y + \alpha$$

Extraneous vs. Confounding Variables

- Extraneous
  - left out (intentionally or forgotten)
  - Important (e.g. regression)

- Confounding –
  - Extraneous variables that offer alternative explanation
  - Another variable that changes along with IV

Univariate, Bivariate, Multivariate

- Univariate
  - only one DV, can have multiple IVs
- Bivariate
  - two variables no specification as to IV or DV (r or $\chi 2$)
- Multivariate
  - multiple DVs, regardless of number of IVs

Experimental vs. Non-Experimental

- Experimental
  - high level of researcher control, direct manipulation of IV, true IV to DV causal flow
- Non-experimental
  - low or no researcher control, pre-existing groups (gender, etc.), IV and DV ambiguous
- Experiments = internal validity
- Non-experiments = external validity

Why multivariate statistics?

Why multivariate statistics?

- Reality
  - Univariate stats only go so far when applicable
  - “Real” data usually contains more than one DV
  - Multivariate analyses are much more realistic and feasible

Why multivariate?

- "Minimal" Increase in Complexity
- More control and less restrictive assumptions
- Using the right tool at the right time
- Remember
  - Fancy stats do not make up for poor planning
  - Design is more important than analysis

When is MV analysis not useful

- Hypothesis is univariate use a univariate statistic
  - Test individual hypotheses univariately first and use MV stats to explore
  - The Simpler the analyses the more powerful

Stat Review 2

Continuous, Discrete and Dichotomous data

- Continuous data
  - smooth transition no steps
  - any value in a given range
  - the number of given values restricted only by instrument precision

Continuous, Discrete and Dichotomous data

- Discrete
  - Categorical
  - Limited amount of values and always whole values
- Dichotomous
  - discrete variable with only two categories
  - Binomial distribution

Continuous, Discrete and Dichotomous data

- Continuous to discrete
  - Dichotomizing, Trichotomizing, etc.
  - ANOVA obsession or limited to one analyses
  - Power reduction and limited interpretation
  - Reinforce use of the appropriate stat at the right time

Continuous, Discrete and Dichotomous data

x1	x2
11	9
10	7
11	10
14	12
14	11
10	8
12	10
10	9
11	8
10	11
…	…

x1di	x2di
1	0
1	0
1	1
1	1
1	1
1	0
1	1
1	0
1	0
1	1
…	…

X1 dichotomized at median >=11 and x2 at median >=10

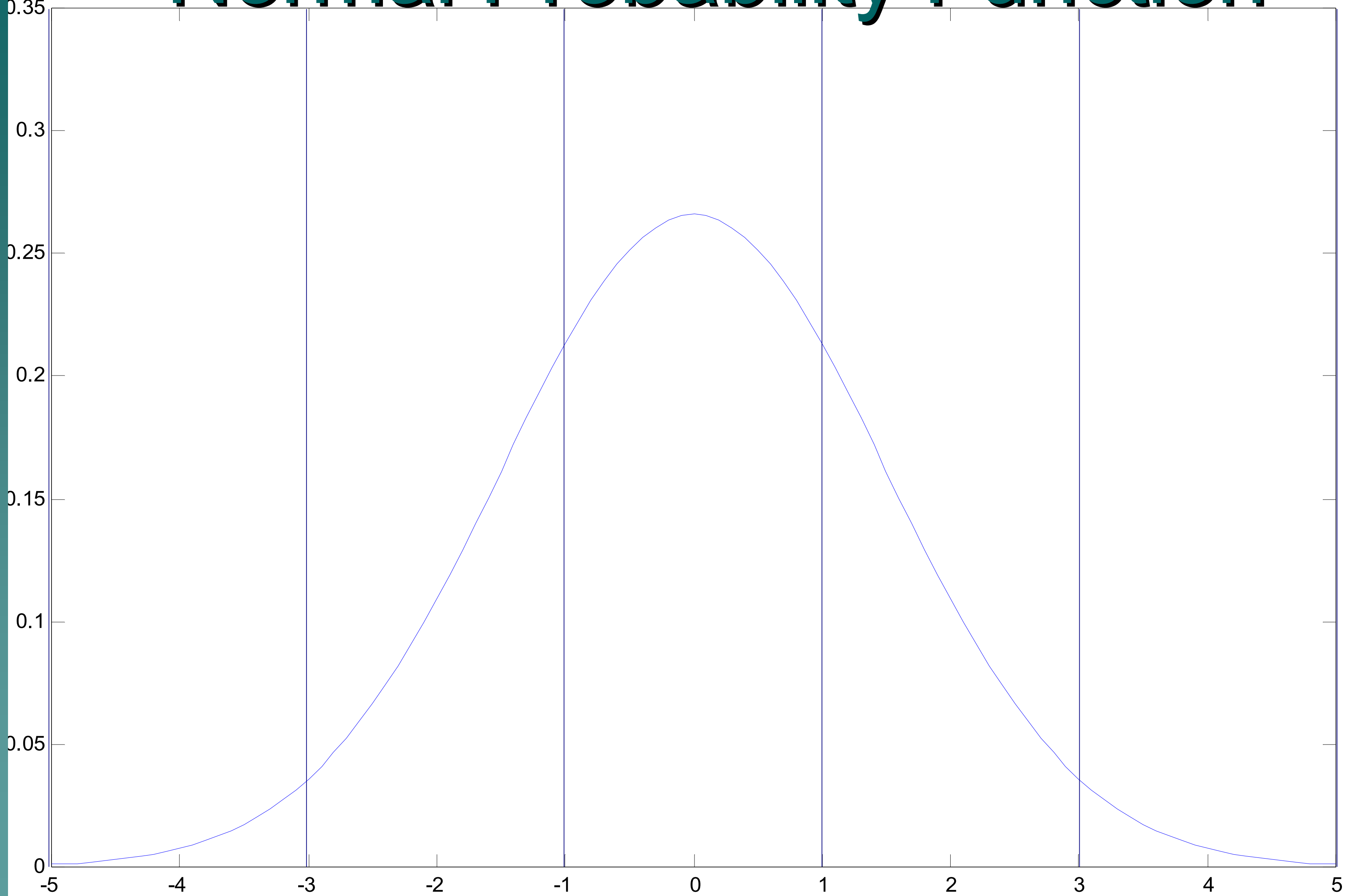
Continuous, Discrete and Dichotomous data

- Correlation of X1 and X2 = .922
- Correlation of X1di and X2di = .570

Continuous, Discrete and Dichotomous data

- Discrete to continuous
  - cannot be done literally (not enough info in discrete variables)
  - often dichotomous data treated as having underlying continuous scale

Normal Probability Function



Continuous, Discrete and Dichotomous data

- Correlation of X1 and X2 when continuous scale assumed = .895
- (called Tetrachoric correlation)
- Not perfect, but closer to real correlation

Continuous, Discrete and Dichotomous data

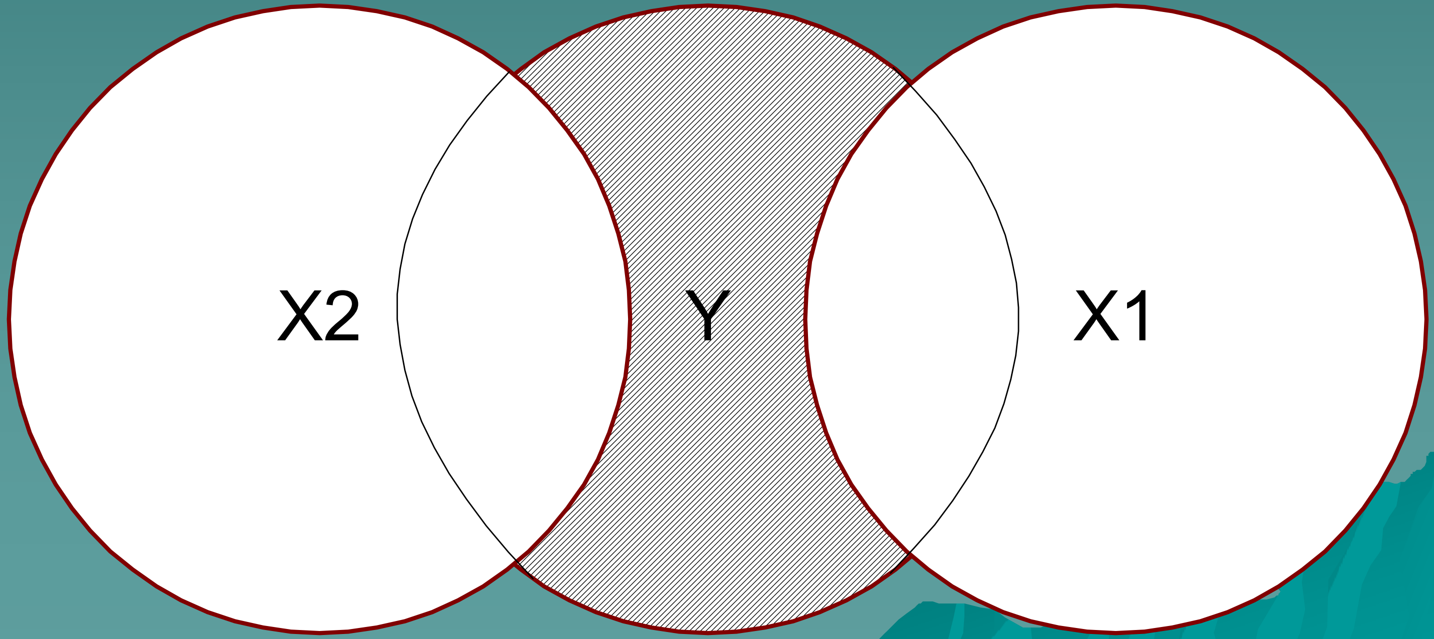
- Levels of Measurement
  - Nominal – Categorical
  - Ordinal – rank order
  - Interval – ordered and evenly spaced
  - Ratio – has absolute 0

Orthogonality

- Complete Non-relationship
- Opposite of correlation
- Attractive property when dealing with MV stats (really any stats)

Orthogonality

- Predict y with two Xs; both Xs related to y; orthogonal to each other; each x predicts additively (sum of x_i/y correlations equal multiple correlation)

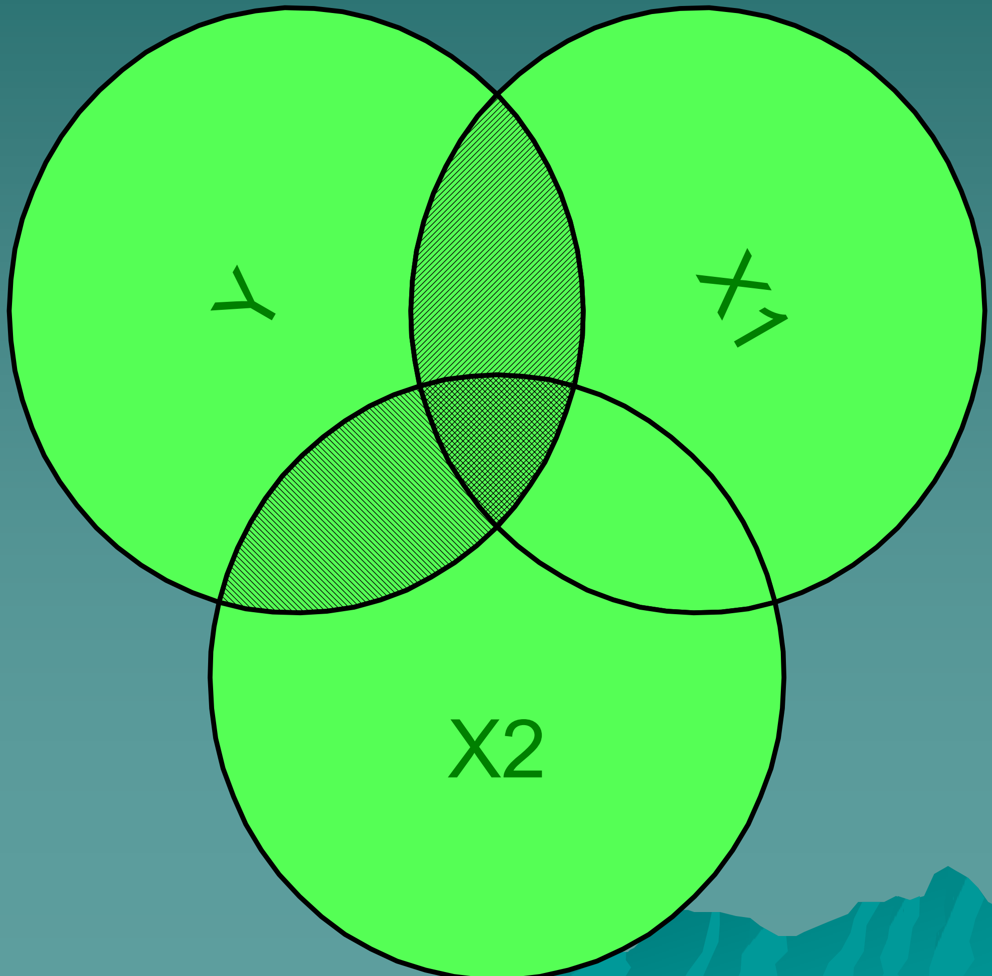


Orthogonality

- Designs are orthogonal also
- With multiple DV’s orthogonality is also advantages

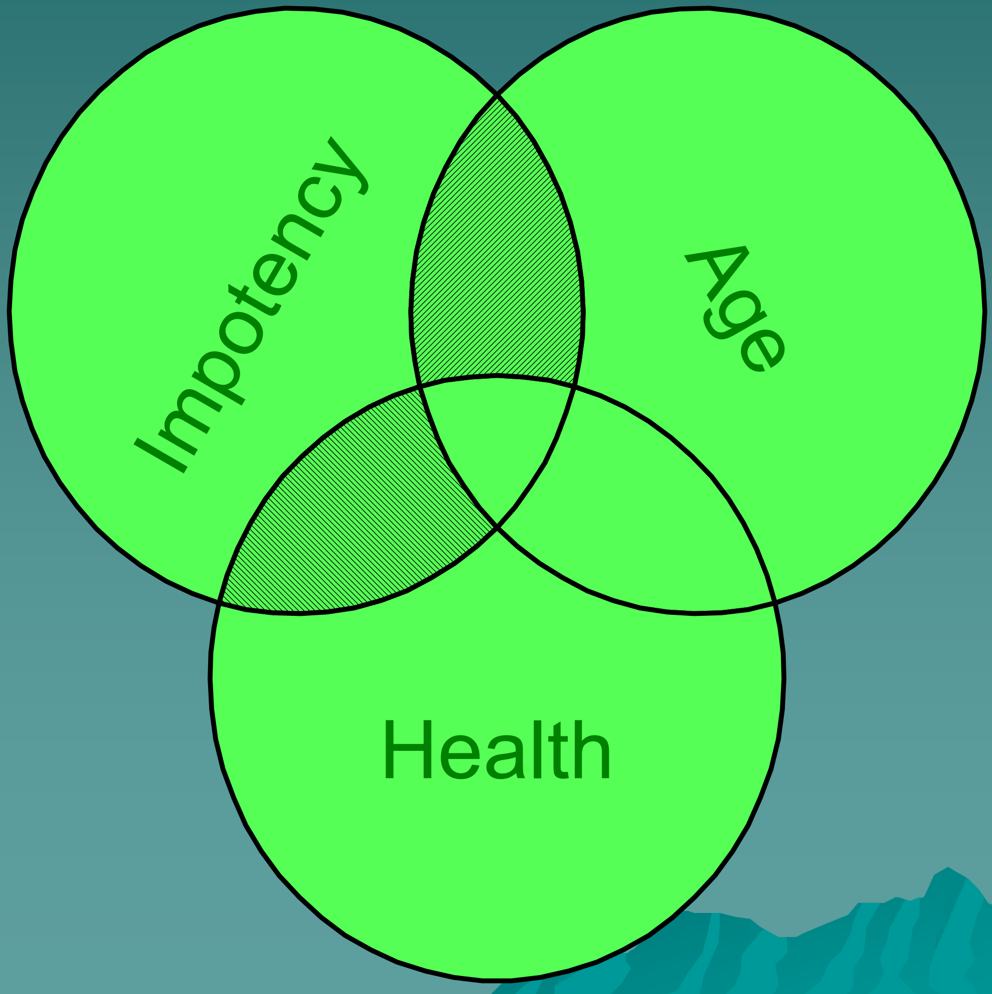
Standard vs. Sequential Analyses

- Choice depends on handling common predictor variance



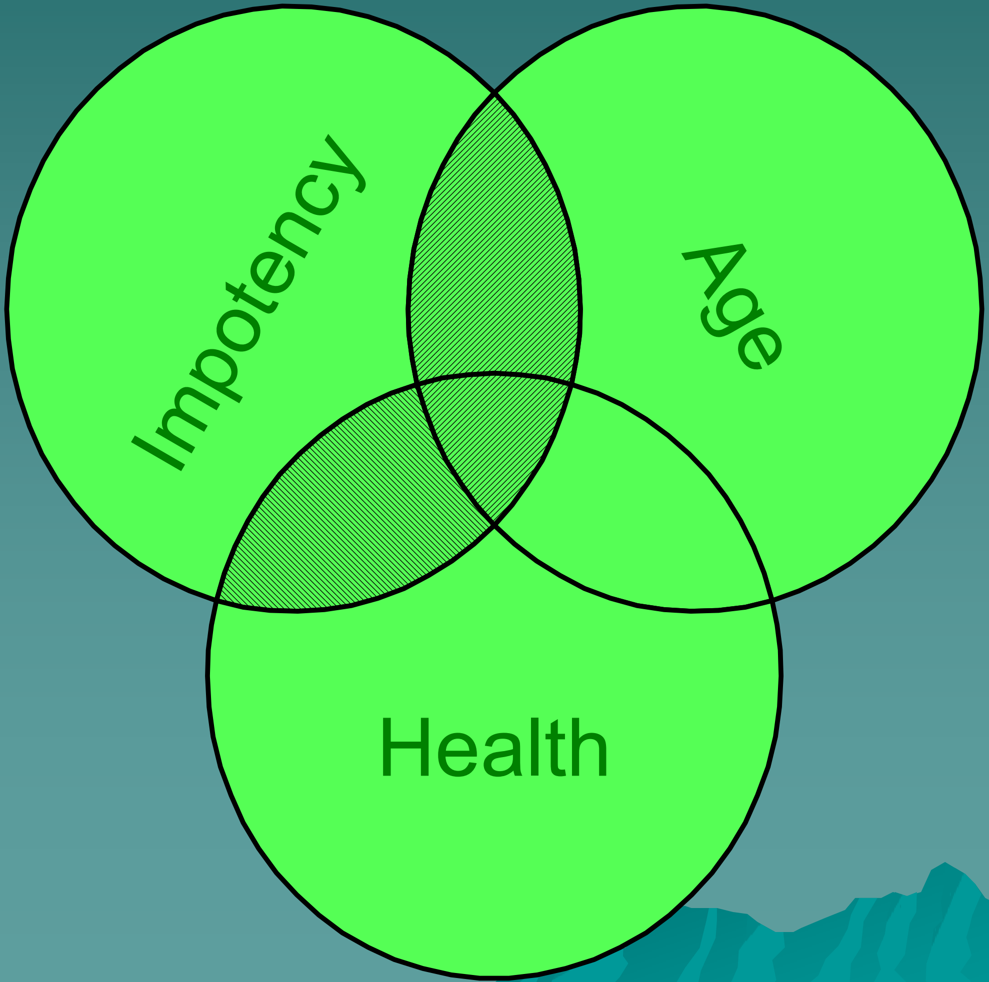
Standard vs. Sequential Analyses

- Standard analysis – neither IV gets credit



Standard vs. Sequential Analyses

- Sequential – IV entered first gets credit for shared variance



Matrices

- Data Matrix

GRE	GPA	GENDER
500	3.2	1
420	2.5	2
650	3.9	1
550	3.5	2
480	3.3	1
600	3.25	2

For gender women are coded 1

Matrices

- Correlation or R matrix

	GRE	GPA	GENDER
GRE	1.00	0.85	-0.13
GPA	0.85	1.00	-0.46
GENDER	-0.13	-0.46	1.00

Matrices

- Variance/Covariance or Sigma matrix

	GRE	GPA	GENDER
GRE	7026.67	32.80	-6.00
GPA	32.80	0.21	-0.12
GENDER	-6.00	-0.12	0.30

Matrices

- Sums of Squares and Cross-products matrix (SSCP) or S matrix

	GRE	GPA	GENDER
GRE	35133.33	164.00	-30.00
GPA	164.00	1.05	-0.58
GENDER	-30.00	-0.58	1.50

Matrices

- Sums of Squares and Cross-products matrix (SSCP) or S matrix

$$SS(X_i) = \sum_{i=1}^{N} (X_{ij} - \bar{X}_j)^2$$

$$SP(X_j X_k) = \sum_{i=1}^{N} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$