Multivariate Statistics

Definition

A collection of procedures which involve observation and analysis of more than one statistical variable at a time

Types of Analyses

- Univariate
  - Only interested in variation in an outcome (dependent variable)
  - t-test, ANOVA
- Bivariate
  - Interested in the variation between two variables (shared variance)
- Multivariate
  - Interested in the variation among several variables (distributed variance)

Variability as the basis of research

- Purpose of research is to explain:
  - situational variability
  - individual differences
  - temporal variability
- All analyses in some way take variability into account

Two causes of variability

- Systematic = What you have identified as a cause of variability
  - e.g., provocation causes aggression
- Error = What you have not accounted for
  - e.g., past behavior, situational cues, personality

Measuring Variability

- Behavior must be quantified in some way
- Measures of behavior must be able to "pick up" variation

Variance

The degree to which scores in a distribution are spread out or clustered together

Why Is It Important?

- The basis of most statistical analyses
  - How much the variation in scores of one variable can be explained by variation of scores of other variables
  - The degree to which two variables covary

Deviation Score

- The distance of a given score from the mean
- Formula: $X - \mu$
- Pro: Accounts for all scores
- Cons:
  - Involves several steps
  - Not a summary score of variability

Sum of Squares (*SS*)

- The sum of the ***squared*** deviation scores
- Remember: a negative number squared becomes a positive number (e.g., $-3^2 = 9$)
- Pro: Summarizes deviation scores
- Con: Too large (will resolve this later)

Definitional Formula

- $SS = \Sigma(X - \mu)^2$
- Steps in calculation:
  1. Find the mean (μ)
  2. Find all deviation scores ($X - \mu$)
  3. Square all deviation scores $(X - \mu)^2$
  4. Add the squared deviation scores $\Sigma(X - \mu)^2$
- Preferred method for getting *SS*

Variance

- Average value of the squared deviation scores
  - Typical distance of scores from the mean
- How is an average obtained?
  - Sum of scores ÷ Number of scores
- Variance (σ^2)= *SS* ÷ *N*
  - Used to obtain ***Population Variance***

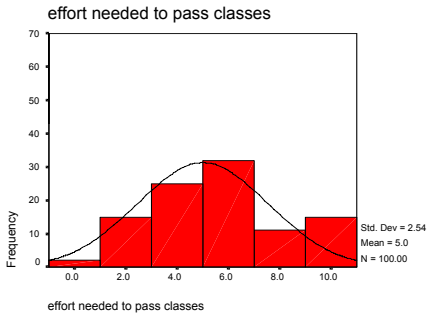
Problem

- By squaring the deviation scores, we are exaggerating the distance from the mean
- How do we correct it?
  - How do you reduce a squared value?
  - Square root ($\sqrt{\ }$)
- Standard Deviation (σ) = $\sqrt{\text{Variance}}$
  - Conversely, Variance = SD^2

Sample Estimates

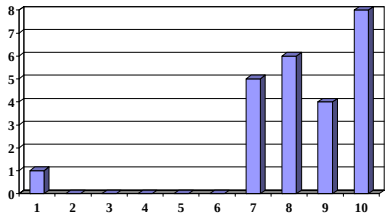
- Problem: Samples are almost always less variable in scores than the population
- Thus, our variance statistic is always underestimating the population parameter
  - Need to make a correction in the formula
- Sample Variance (s^2)= $SS \div (n - 1)$
  - Sometimes shortened as $SS \div df$
- Sample Standard Deviation (s)
  - Still $\sqrt{\text{variance}}$, but use the sample variance

Factors That Affect Variability



Extreme Scores (Outliers)

- Exclude cases if you have a good reason
  - These could otherwise REALLY mess you up
- Interpret Variance and SD carefully



Sample Size (*n*)

- Each additional score can potentially be the highest or lowest