Collecting & Making Sense of Quantitative Data

Deborah Eldredge, PhD, RN
Director, Quality, Research & Magnet Recognition
Oregon Health & Science University

Margo A. Halm, RN, PhD, ACNS-BC, FAHA
Director, Nursing Research & Quality
Salem Hospital

Objectives

1. Identify principles of measurement as they relate to research design

2. Describe strategies for data collection and data management

3. Describe the differences between statistical and clinical significance

Measurement

- The process of assigning numbers to objects in accord with some rule
- Goals for Measurement
  - Capture intended phenomenon = Validity
  - Measure perform consistently = Reliability
  - Control error – Data collection & processing

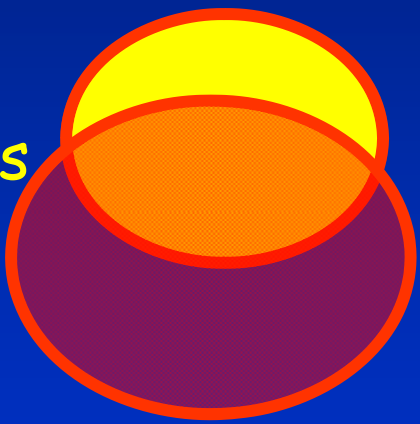
Measurement Error

Observed score = true score + error

- Systematic Error:

    Influences direction of the mean

    - Variation in administration of scales
    - Data processing errors
    - Situational factors



Random Error:

Influences variation around the mean

Levels of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

Level of Measurement: Nominal

- Name or category only
- Can not be ordered or compared
- Mutually exclusive categories
- Numbers assigned as labels only
  - Ethnicity, religion, marital status
  - Gender: males = 1 ; females = 2
  - Experimental group, control group

Level of Measurement: Ordinal

- Attribute that can be ranked
- Interval not equal
- Intensity of pain, ability to provide self-care
- Daily exercise:
  - 0 = no exercise
  - 1 = moderate, no sweating
  - 2 = sweating, altered breathing
  - 3 = strenuous, heavy breathing

Level of Measurement: Interval

- Rank ordering with equal interval
- Continuum of values
- No absolute zero
  - Temperature:
  - 40^{o} versus 70^{o} same as 100^{o} to 130^{o}

Level of Measurement: Ratio

- Absolute zero, and all other constraints
  - Weight, length, and volume
- Most commonly, interval and ratio level data are treated the same way

Types of Measures

- Direct vs. Indirect
- Physiological vs. Psychological/ Attitudinal
- Objective vs. Subjective

Physiological Measures

- Physiology has greater precision than attitudes
  - devices include thermometers, sphygmomanometers, stethoscopes, electrocardiograms, IC U monitors
- Calibration (reliability, validation)
- Inter-rater consistency
  - identify error range

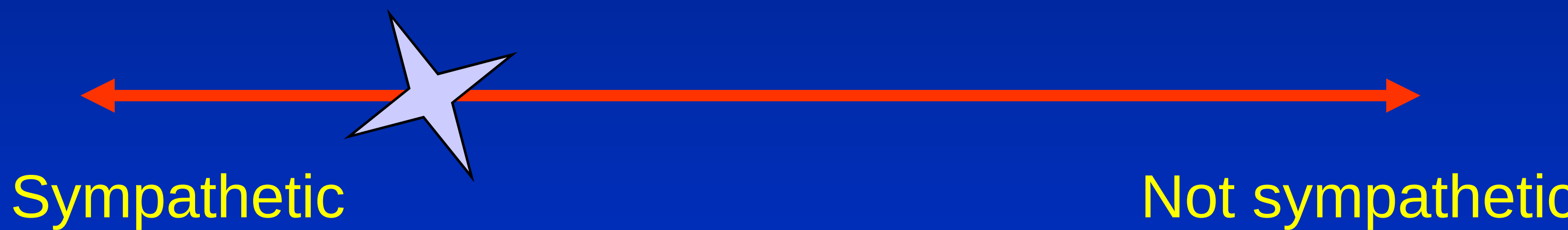
Likert-type Scales

Continuum between 'agree and disagree'
Typical range is 5-7 per item

Do you do this type of help for your family member?	No	If Yes, circle how **hard** it is for you to do that.						How **frequently** do you do this activity?			
			Very Hard	Pretty Hard	Some-what Hard	Not too Hard	Easy	**2-3x wk**	**4-5x wk**	**1x daily**	**2-3x daily**
	No	Yes	5	4	3	2	1	**1**	**2**	**3**	**4**
1. Do you check in on your family member to make sure he or she is OK?	No	Yes	5	4	3	2	1	**1**	**2**	**3**	**4**
2. Do you monitor the number of people who come to see him or her?	No	Yes	5	4	3	2	1	**1**	**2**	**3**	**4**

Semantic Differential Scales

Continuum between two adjectives
(e.g. 'friendly and unfriendly')

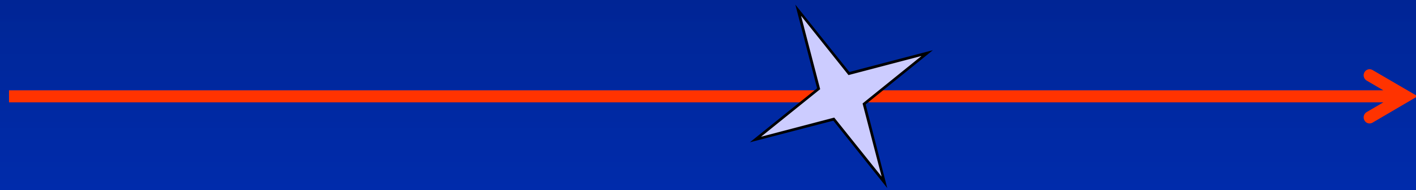


Visual Analog Scales

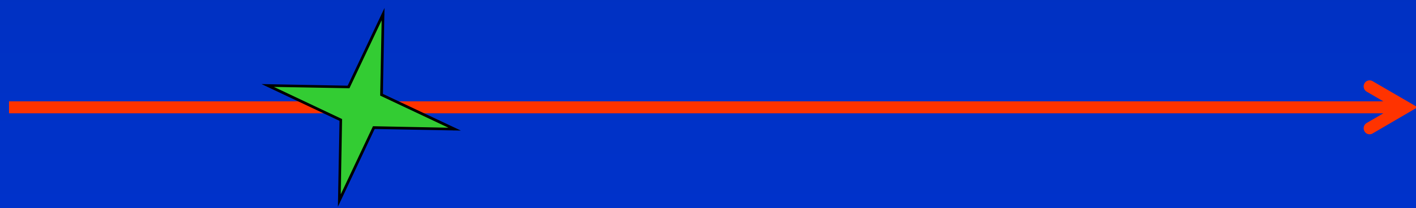
10 millimeter line

Asked to mark how much they feel a certain way

Mark along the line below how SAD you feel at this time



Mark along the line below how ANXIOUS you feel at this time



Questionnaires

- Typically self-reported data
  - Demographic
  - Open-ended vs. closed-ended
  - Contingency and filler
- If mailed, send instructions re: completion

Interviews

- Types of questions
  - Open ended vs. closed ended
  - Order of questions
  - Timing and setting
- Interviewer training / guidelines
- Influence of interviewer on respondents

Observational Methods

- What is being observed?
  - Structured vs. unstructured observations
  - Event sampling and time sampling
- Training for data collection
- Relationship between observer and subjects (Hawthorne effect)
- Role of nurse vs. researcher

Plan for data collection

- What methods will be used to collect data?
- How will the data be collected?
- Who will collect the data?
- Where will data be collected?
- When will data be collected?

Variables

- Dependent variable: characteristic or outcome that researcher is interested in understanding, explaining, predicting, or affecting

- Independent variable: presumed cause of, or antecedent to, or the influence on the dependent variable

Selecting Instruments

- What information is available
  - Reliability
  - Validity
  - Sample for development? For testing?

- Scoring instructions

- Feasibility issues

Reliability

- Consistency of a measure, equivalence
- Does the instrument perform the way you expect it to across items, over time, between persons, different settings?
- Same scale used by two data collectors or in two settings would yield the same results

Internal consistency

- Correlation among items on long scale
- Expressed as a correlation coefficient (Cronbach's alpha) with a score ranging from 0 to 1.0.
- Typically should be above .80+, new instruments acceptable at .70+
- Determined from a specific sample; may not hold true with other samples.

Stability

- Test-retest reliability, often with 2 or more measurements.
- STATE vs. TRAIT characteristics
- Consider: Did the phenomenon being measured change
  - pain or anxiety
  - optimism or attachment style

Equivalence

- Compares two versions of the same instrument
- Interrater reliability
  - 2 data collectors
  - # of agreements/# of possible agreements
- Alternate forms, parallel forms

Validity

- The extent to which the instrument actually reflects the concept being measured.
- Are you measuring what you intended to measure (or is there another concept that you might have captured)?

Content Validity

- Examines the extent to which the measure captures ALL of the relevant elements
- Literature, clinical experts, and lay experts used to generate items
- Content Validity Index
  - Relevance, comprehensiveness, readability/clarity, and face validity

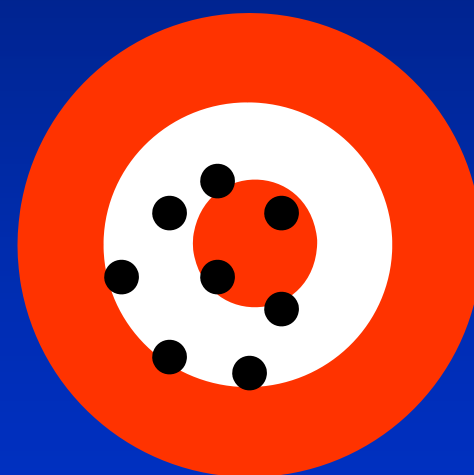
Construct Validity

- Determines validity of measure by exploring a set of relationships that SHOULD map out as expected
  - Factor analysis
  - Structural equation modeling

Which is most important?

Reliability

Validity



Feasibility for Data Collection

- How difficult is instrument to complete?
- Reading level, language choices
- Is phenomena under study sensitive – would data collection change the experience?
- Cost for data collection or processing

Making Sense of Data
(statistics)

What are statistics really about?

- Describe what is going on with data
- Is this group different from another group?
- How BIG is this difference? Relationship?
- Is this difference due to chance?
- What else accounts for the difference?
- Will this difference / relationship be important to patients?

Descriptive Statistics

- Used to describe major elements of the sample:
- Demographic characteristics
  - (Did randomization work?)
- Predictors
- Outcomes

Distribution

- How do the data look?
- Frequency distribution or counts
- Graphed
  - pie charts
  - bar charts
  - histograms

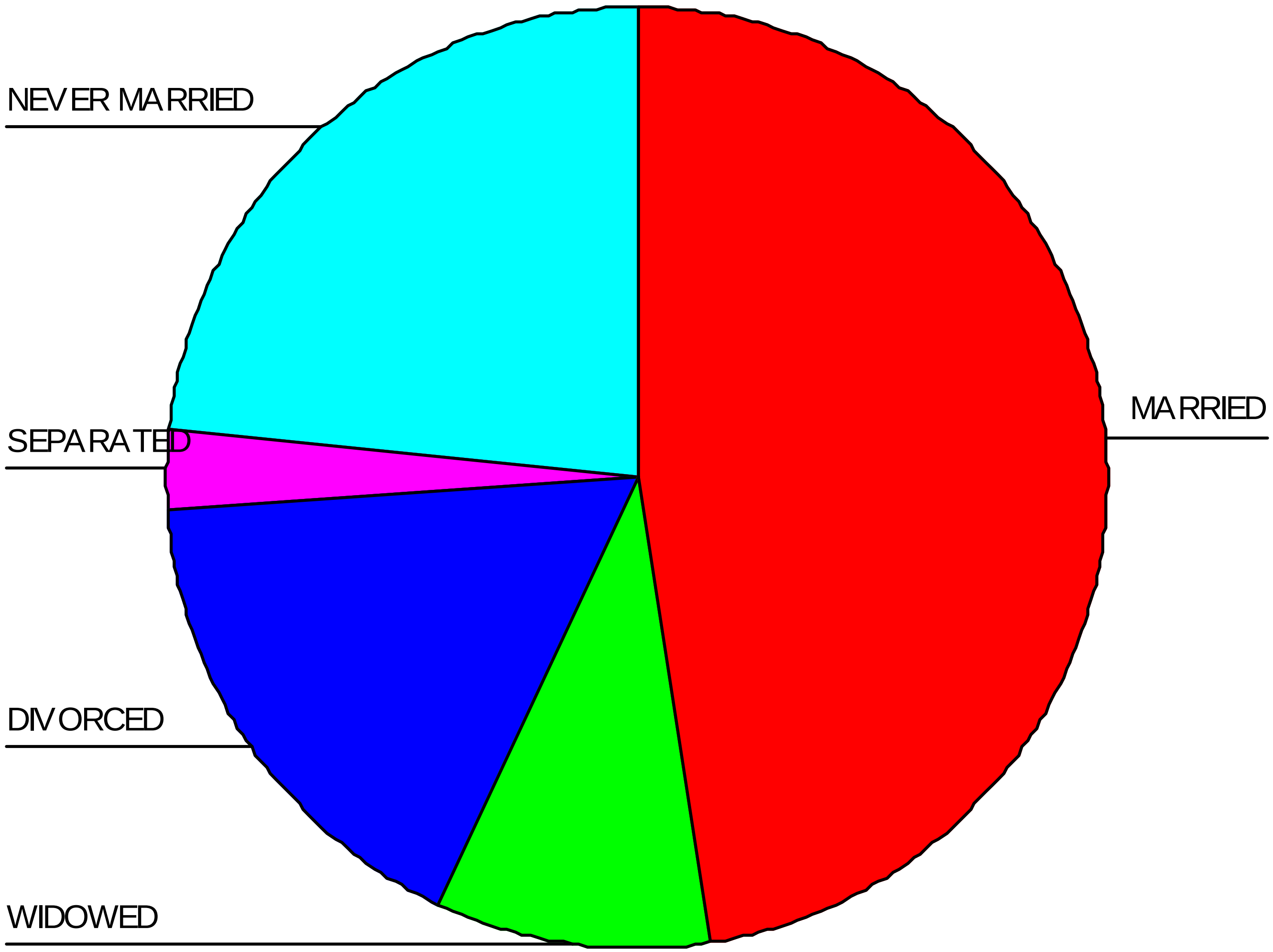
Central Tendency

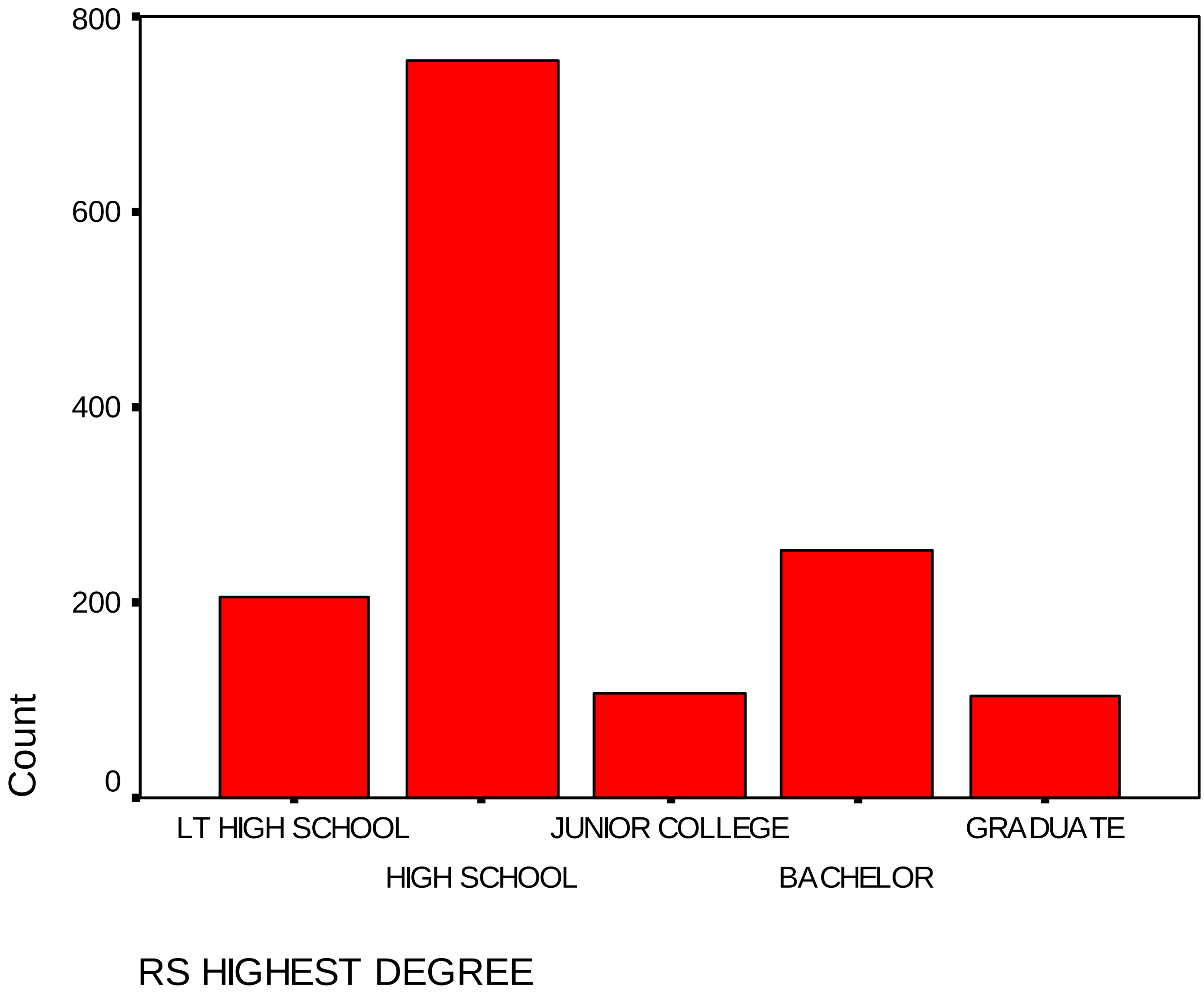
- How alike are the group members?
- Mode (nominal)
  - Score of greatest frequency
- Median (ordinal)
  - Score at center of distribution; 50%ile
- Mean (interval/ratio)
  - sum of scores / number of scores

Dispersion

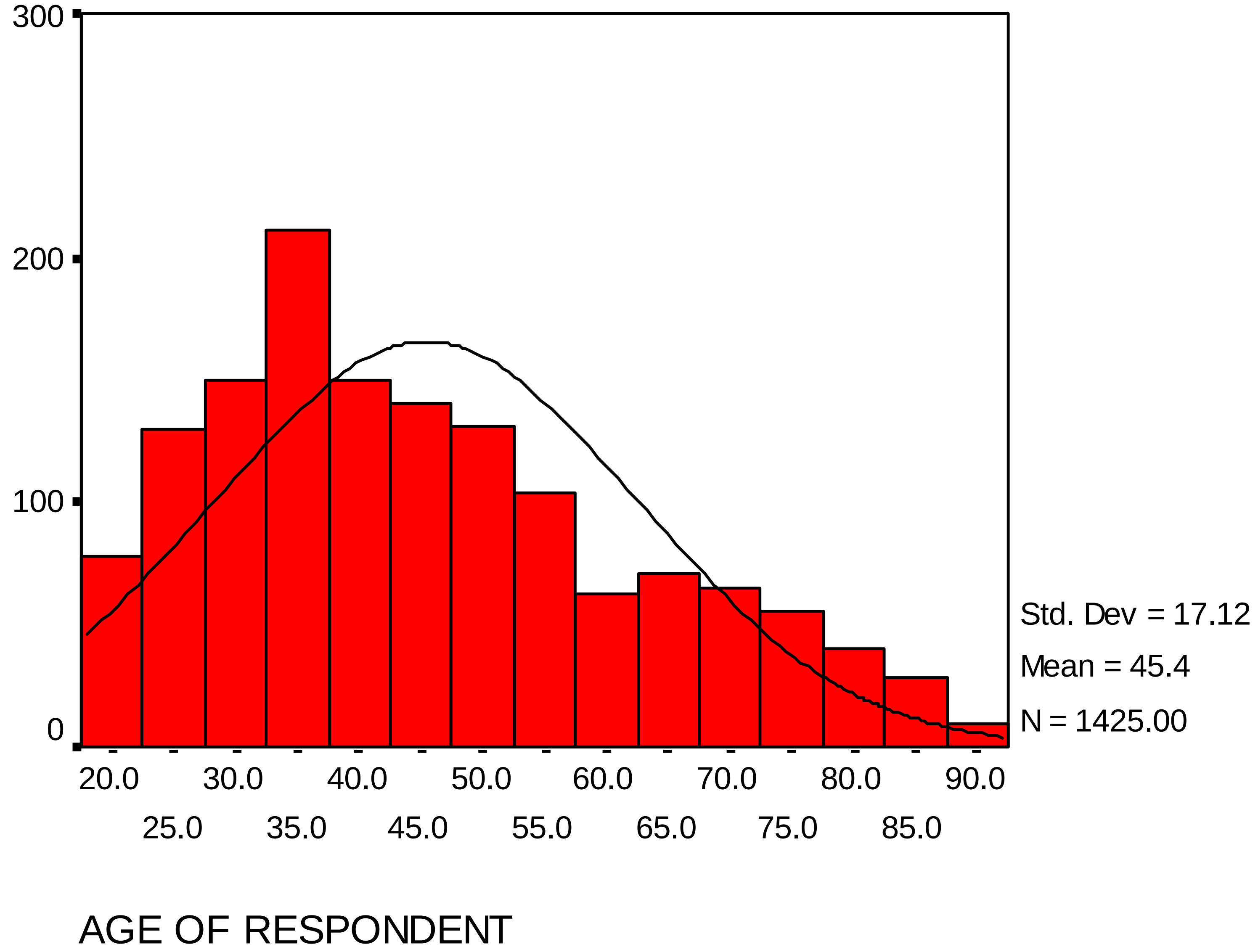
- How different are the group members?
- Range
  - High to low score
- Difference scores
  - Score minus the mean
- Variance / Standard Deviation
  - Average deviation score
  - How different is mean from any individual score

Marital Status, GSS 1998; n = 1427



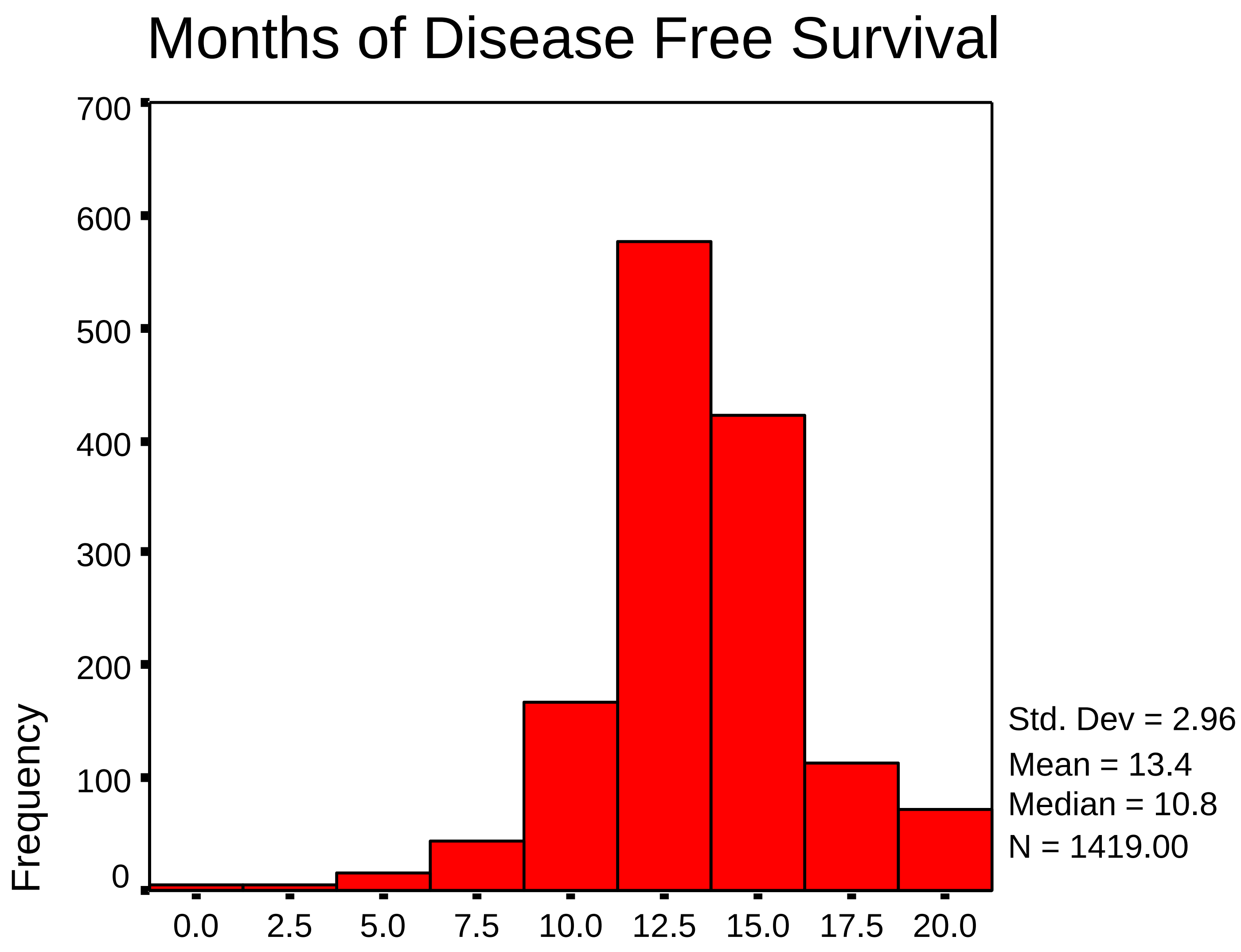

800
600
400
200
0
Count
LT HIGH SCHOOL
HIGH SCHOOL
JUNIOR COLLEGE
BACHELOR
GRADUATE
RS HIGHEST DEGREE

Age, GSS 1998; N = 1425



Skewed Data

- Extreme scores (either high or low) will skew data, distribution will be shifted to right or left
- Non-normal data
- Determine how outliers have been handled
- When data are skewed, important to consider both MEAN and MEDIAN



Months of Disease Free Survival
Frequency
700
600
500
400
300
200
100
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Std. Dev = 2.96
Mean = 13.4
Median = 10.8
N = 1419.00

Inferential Statistics

- Use inductive reasoning to infer from specific case to general truth
- The process of estimating that what is true in sample is true in population
- Use sample descriptive statistics to generalize to population

Inferences about…

- Tests of means (differences between groups)
  - T-test and ANOVA
- Magnitude and direction of relationship
  - Correlation and regression

Tests of Means

- Are these groups drawn from the same population?
  - Ratio of difference to variability

- T-test formula:

$$\frac{M_1 - M_2}{\text{(pooled standard dev)}}$$

Types of T-Tests

Independent Groups

- Boys vs. Girls
- Drug A vs. Drug B
- Experimental vs. control groups

Dependent Group or Scores

- Pre-test, post-test design
- Partners' scores

Analysis of Variance (ANOVA)

- Simultaneously analyzes the differences between several means at one time
  - 3 or more groups
  - The same group 3 or more times
- Examines the ratio of the differences between groups and the differences within groups
  - *F*-distribution, *F*-value

ANOVA is omnibus

- ANOVA tells you THAT differences exist
- Need to do comparisons to determine WHERE exact differences are, called post-hoc tests

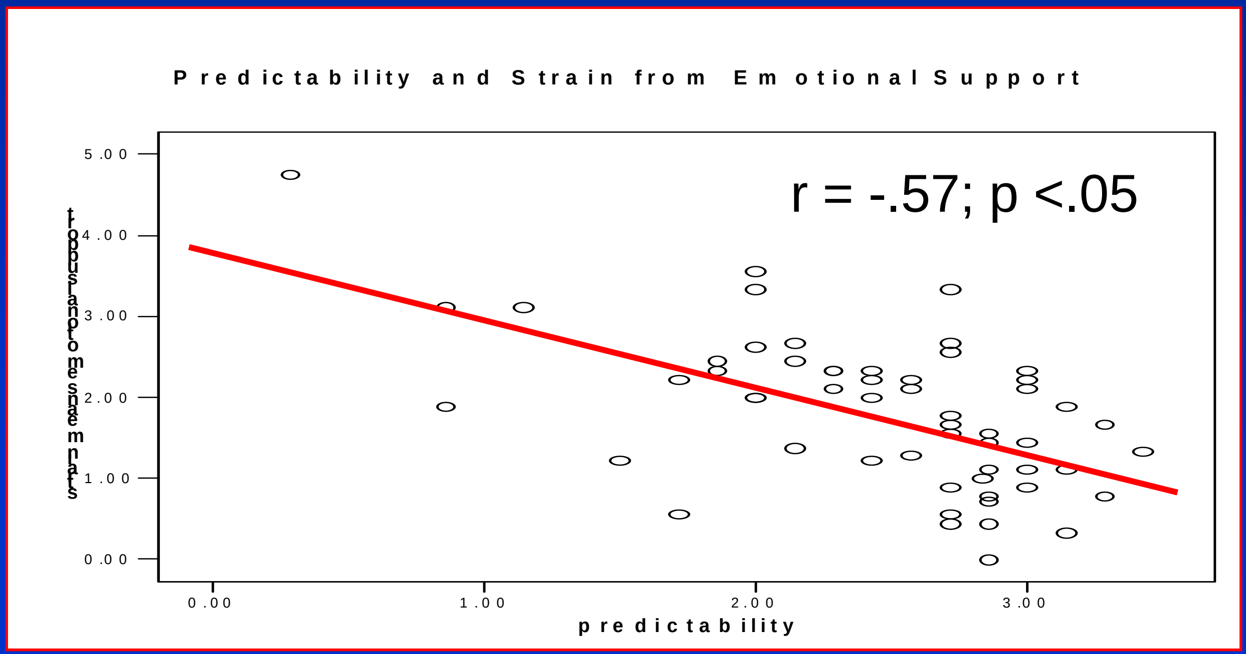
Tests of Association

- Examines degree to which the values of one variable (X) are related to the values of another variable (Y)

- Correlation: Extent of linear relationship between two variables
  - Pearson product-moment correlation (r) - Interval or ratio data
  - Spearman's rho (r_s, r_{rho}) - Ordinal data
  - Contingency coefficient (C) & Chi-square (X^2) – Nominal data

Correlation Coefficients

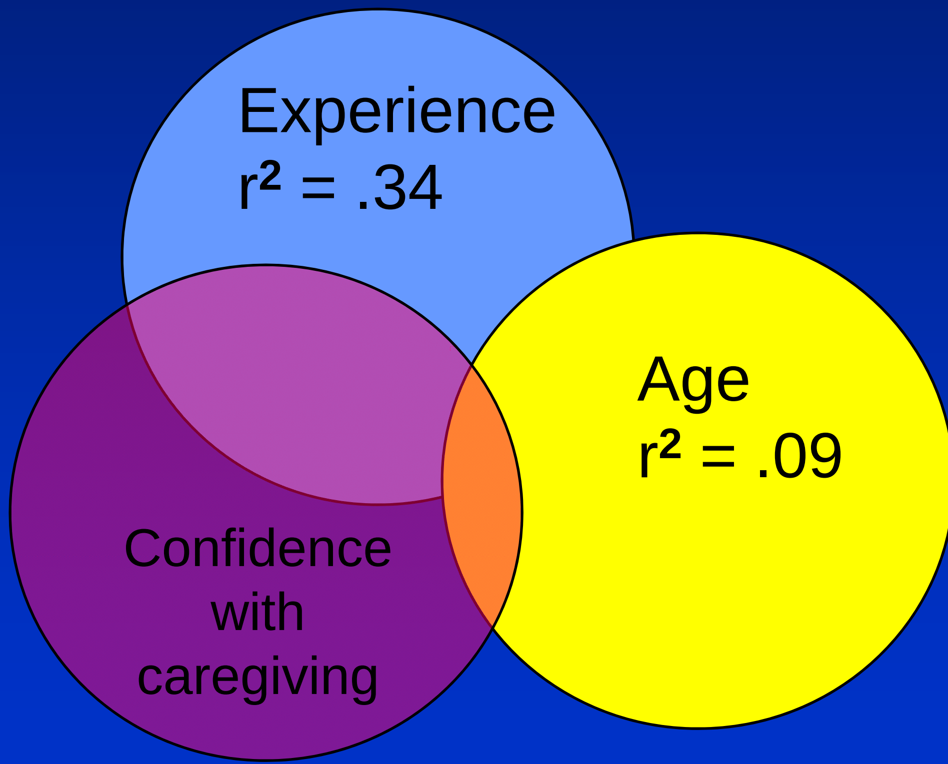
- Direction & magnitude of relationship
- Range from -1.00 to 0 to +1.00
- Graphed as a scatterplot



Regression Analysis

- Any technique for modeling and analyzing several variables
  - focus is on the relationship between a dependent (outcome) variable and one or more independent (predictor) variables
- helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Age and Experience on Confidence with Caregiving



Multiple Regression

- Way to estimate the value of the dependent variable based on a set of independent variables.
- Can also determine unique contribution to variance
- Predict or explain as much variance as possible in dependent variable
- $y = a + b_1x + b_2X + b_3x$

Statistical & Clinical Significance

- Statistical significance is focused on whether groups different than what would happen by chance alone
  - p level .01 or .0001 does not reflect MAGNITUDE of difference
- Clinical significance is focused on whether that difference or association matters to patients

Clinical Significance

- Are the results big enough to be clinically or practically important?
- Would it make a difference to my population?
- Size of the benefit
- Depends on clinical expertise
- Rough estimate for clinical significance
  - Half a standard deviation

Some additional tidbits

- Probability / normal distribution
- Power
- Hypothesis testing

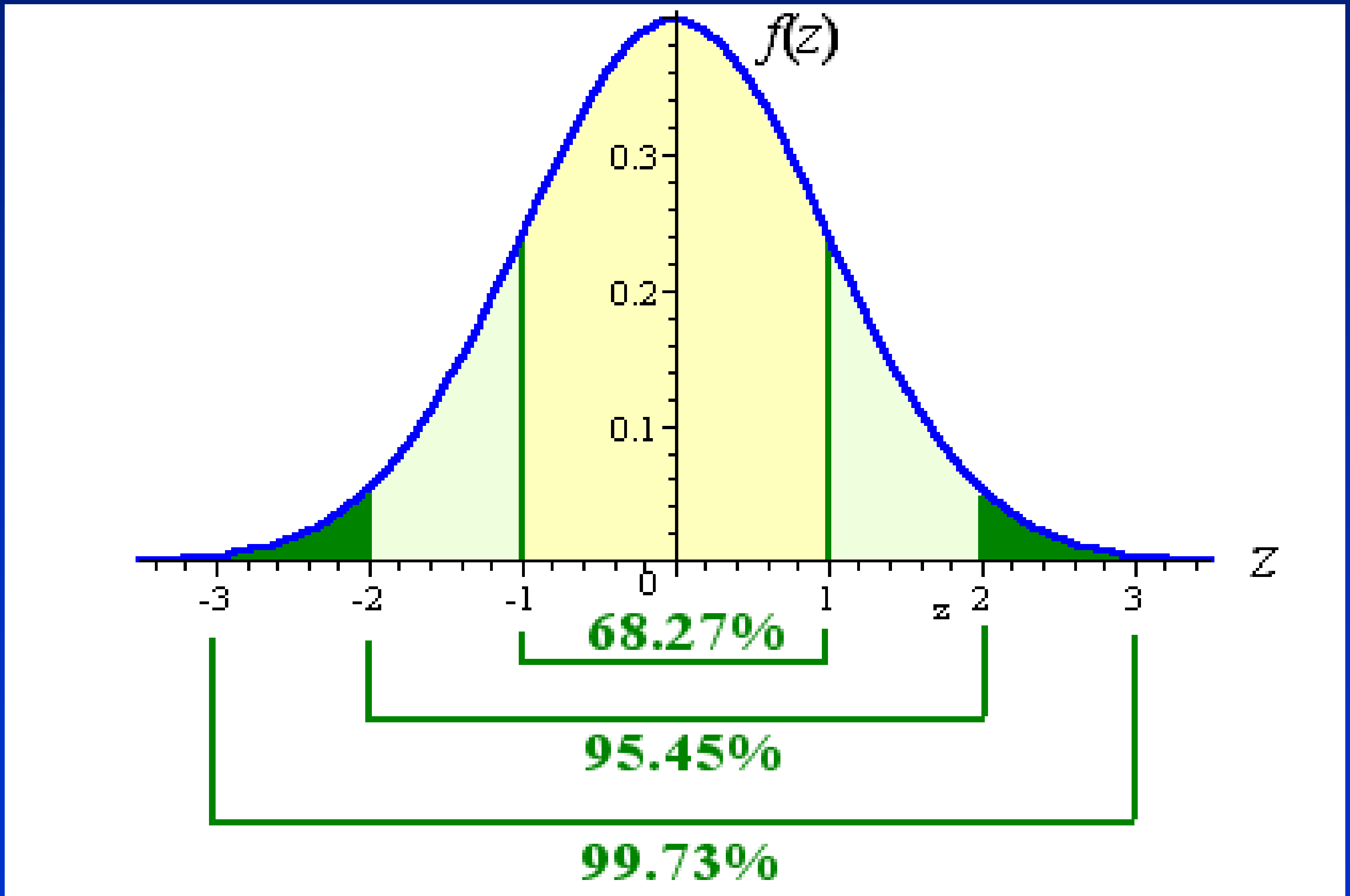
Probability Theory

- Probability of accurately predicting an event or extent of a relationship
- Chances are…
- Expressed as percentage or decimal
- Level of significance set by researcher
  - Alpha before analysis: a
  - Probability after analysis: *p*

Normal Curve

- Theoretical frequency distribution of all possible scores
- The larger the sample, the more certain we are the distribution is normal.
- Extreme scores at tails
  - One tail (directional hypothesis)
  - Two tail (test of any difference)

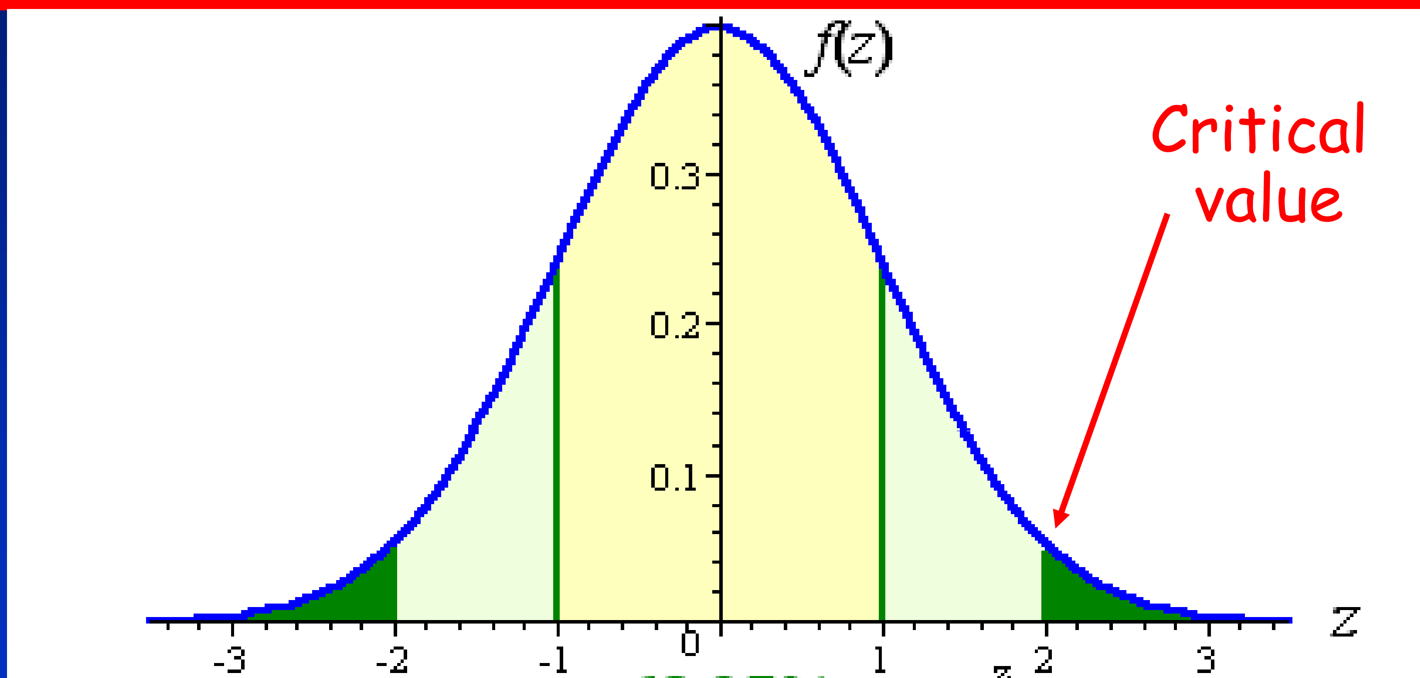
Normal Distribution



Significance Level: a and *p*

- Cutoff point used to determine whether samples being tested are different from the population
- Willingness to reject hypothesis when it should be retained
- Determined by researcher
  - Typically .05 (nursing) or .01 (pharmacologic)
- Corrected with multiple research questions

Critical Value



- Scores more extreme than critical value are 'significantly different'.
- 95% of the time, the difference would hold up in the population

Power

- The probability that a statistical test will detect a difference that exists
- How much POWER do you have to find a true difference?
- Effected by sample size, effect size, and level of significance
- .80 ideal level of power

Hypothesis Testing Outcomes

True State of the World

Decision

	Ho is True	Ho is False
Reject Ho	False Rejection Type I error Alpha Significance level	Correct Power 1-beta
Do not reject Ho	Correct 1-alpha	Miss Type II error beta

Correct Decisions

- Power (microscope)
  - Reject the null hypothesis when it is false
  - Conclude that the intervention does increase positive mood when fact it does
  - Area under the H_A curve
- Correctly not rejected
  - Do not reject the null hypothesis when it true
  - Conclude that the intervention does not increase positive mood when in fact it doesn't
  - Area under the Ho curve

Errors in Hypothesis Testing

- Type I: false rejection
  - reject null hypothesis when in fact it is true
  - Area under the Ho curve
  - conclude that the intervention works when it does not (adopt faulty intervention)

- Type II: missed result
  - do not reject null hypothesis when in fact it is false
  - Area under the H_A curve
  - conclude that the intervention does not work when it does (reject successful intervention)

Trade-off of Errors

- Need to weigh the importance of the different errors in the context of your study
  - Type I error: accept drug when it is not effective
    - What if it has lots of side effects?
  - Type II error: miss a drug that is effective
    - Stop a line of inquiry that may be productive

Reporting statistical findings

- Report statistic
  - t, F, r, R^2
- Report degrees of freedom
- Report probability
  - p value
- Okay to report selected findings
  - Most interesting findings
  - Don't forget non-significant findings