

# Methods and Statistics in I/O Psychology



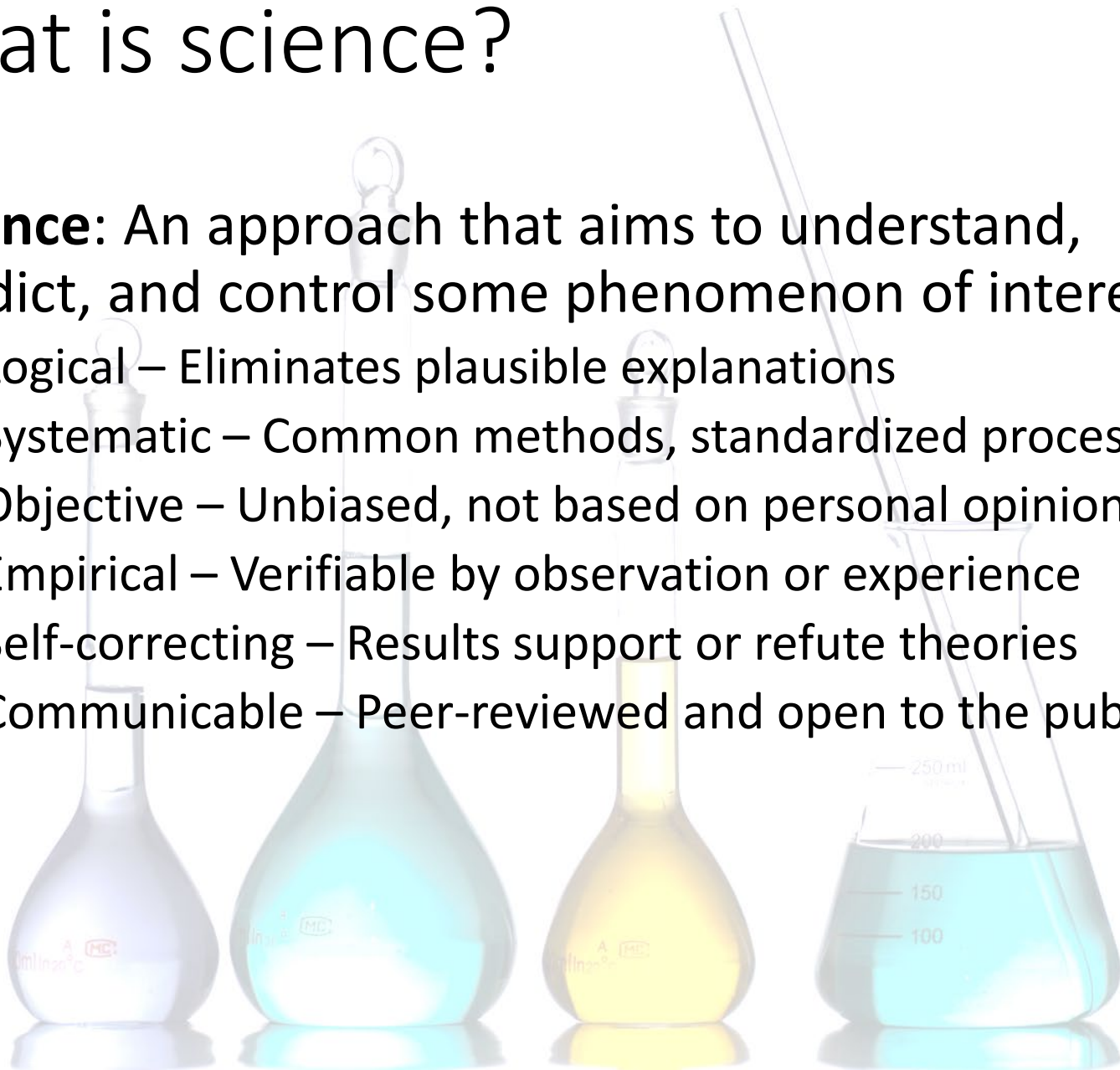
# Agenda

- What is science? Why do I/O psychologists rely on science? Why are generalizability and control important to science?
- What research designs do I/O psychologists use? What research methods do I/O psychologists use?
- What is an experiment? A quasi-experiment? A non-experimental design? What is correlational research?
- What are descriptive and inferential statistics?
- What is reliability? Test-retest reliability? Equivalent forms reliability? Internal consistency? Inter-rater reliability?
- What is validity? Test validation? Criterion validity? Content validity? Construct validity?

What is science? Why do I/O psychologists rely on science? Why are generalizability and control important to science?

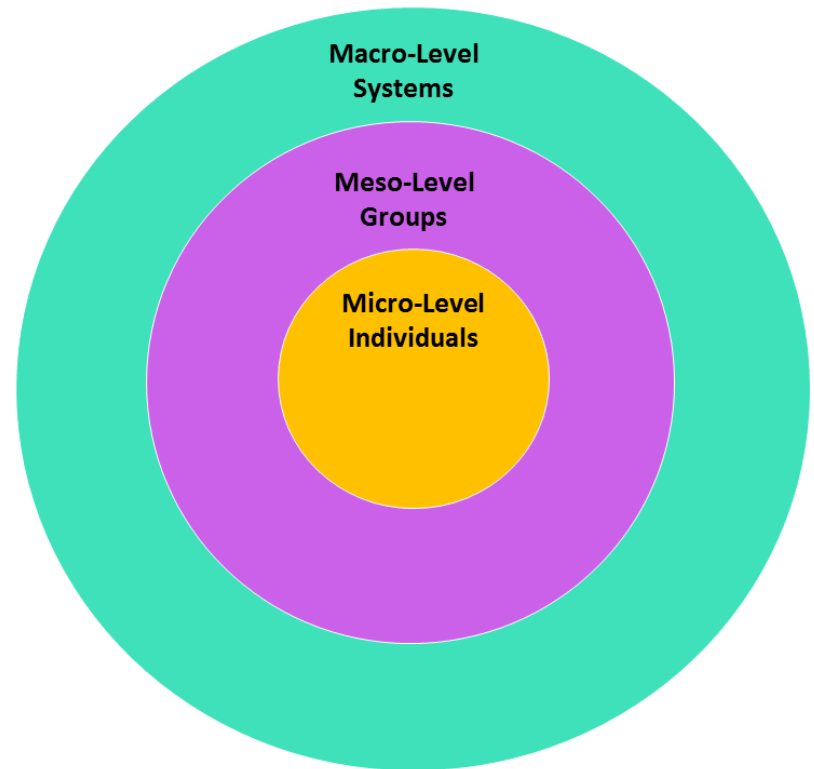
# What is science?

- **Science:** An approach that aims to understand, predict, and control some phenomenon of interest
  - Logical – Eliminates plausible explanations
  - Systematic – Common methods, standardized process
  - Objective – Unbiased, not based on personal opinion
  - Empirical – Verifiable by observation or experience
  - Self-correcting – Results support or refute theories
  - Communicable – Peer-reviewed and open to the public



# What is science?

- 3 levels of analysis
  - **Micro-level research:**  
The study of individual behavior
  - **Meso-level research:**  
The study of the interaction of individual and collective behavior
  - **Macro-level research:**  
The study of collective behavior

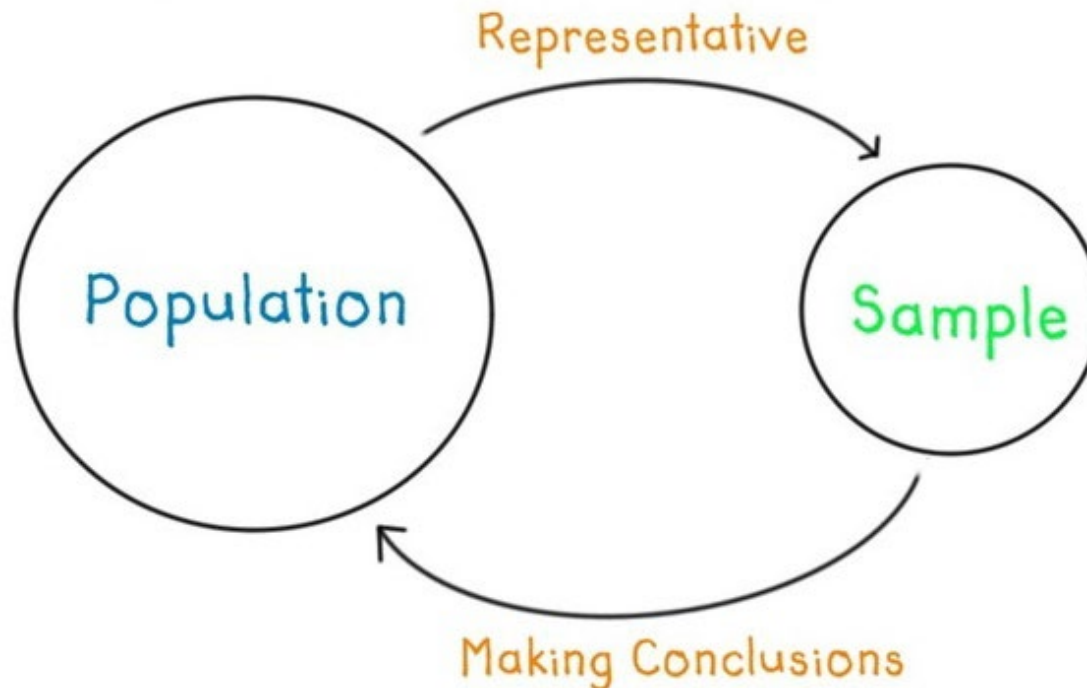


# Why do I/O psychologists rely on science?



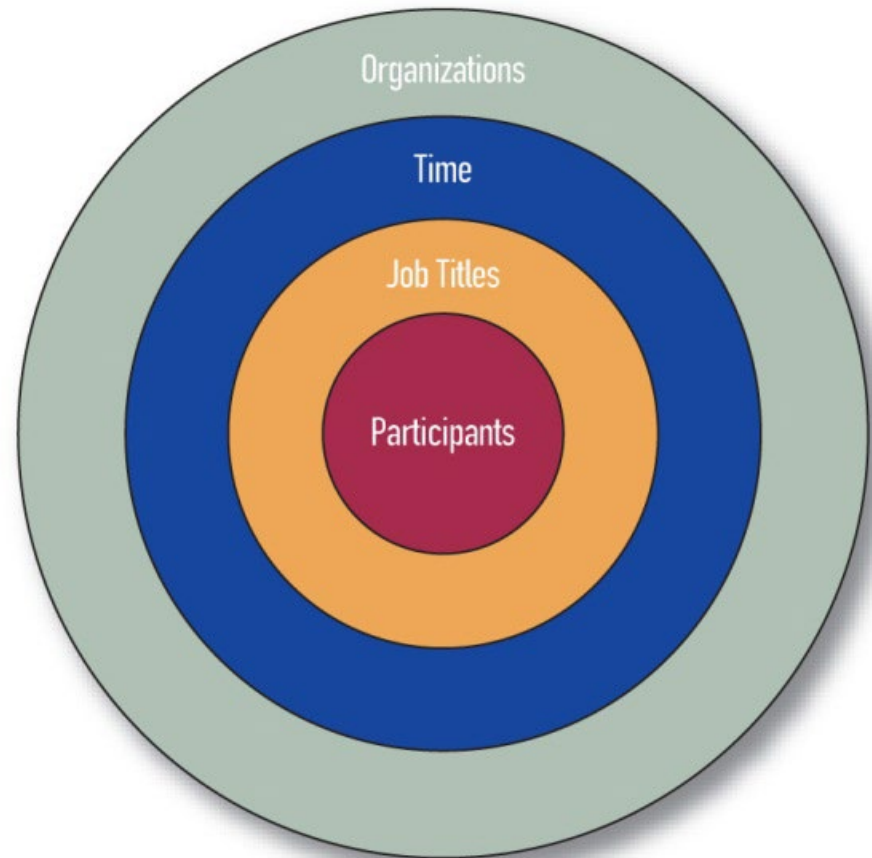
# Why is generalizability important to science?

- **Generalizability:** The extent to which the results from one sample can be applied to other people



# Why is generalizability important to science?

- Increase generalizability by:
  - Recruiting a large representative sample
  - Including multiple sampling domains





# Why is control important to science?

- **Experimental control:**

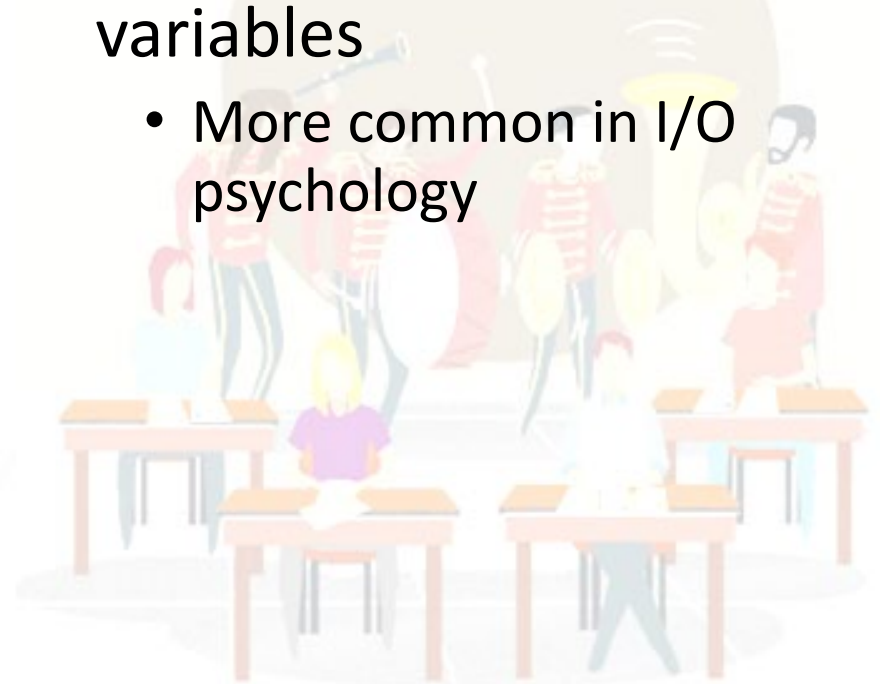
Techniques used to minimize the influence of extraneous variables

- Can reduce the practical value of results

- **Statistical control:**

Statistical techniques used to minimize the influence of subject variables

- More common in I/O psychology



# Why is control important to science?

**CONTROL  
GROUP**



**EXPERIMENTAL  
GROUP**



What research designs do I/O psychologists use? What research methods do I/O psychologists use?

# What research designs do I/O psychologists use?

- **Research design:** Provides the overall structure for the research study
  - Contingent on the goals of the research study
  - 3 basic types:



**Experimental  
research**

**Quasi-  
experimental  
research**

**Non-  
experimental  
research**

# What research methods do I/O psychologists use?

## **Quantitative Methods**

- Rely on tests, rating scales, questionnaires, and physiological measures
- Yield numerical data

## **Qualitative Methods**

- Rely on observation, interviews, case studies, and analyses of records
- Yield descriptive data

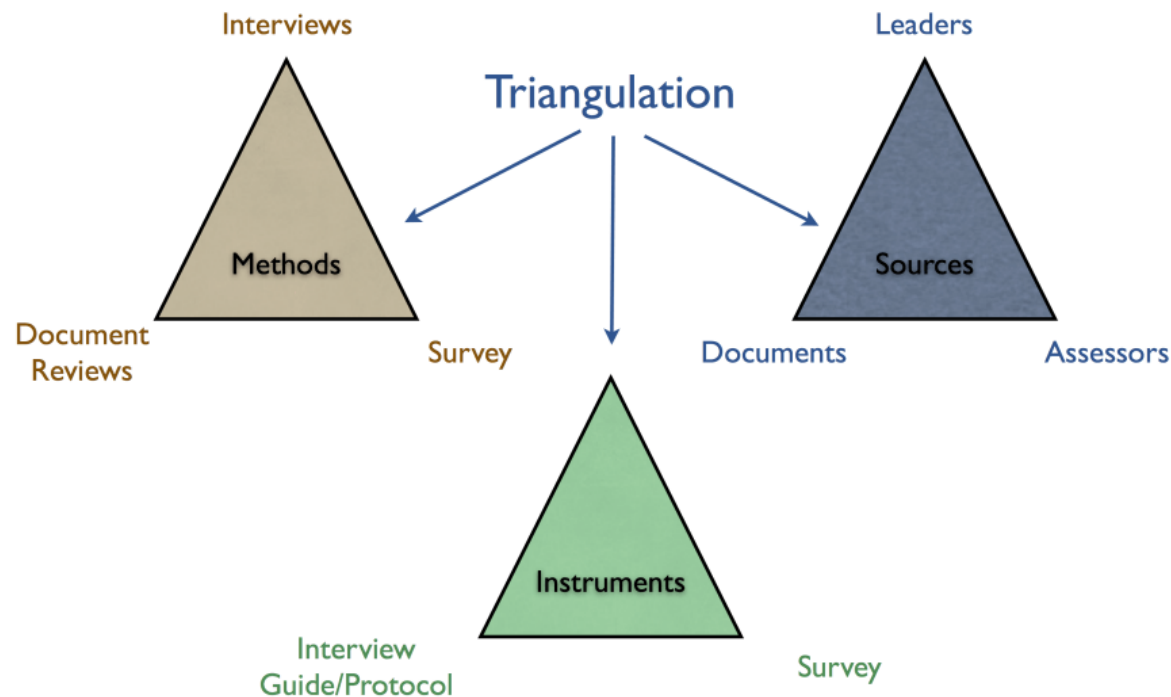


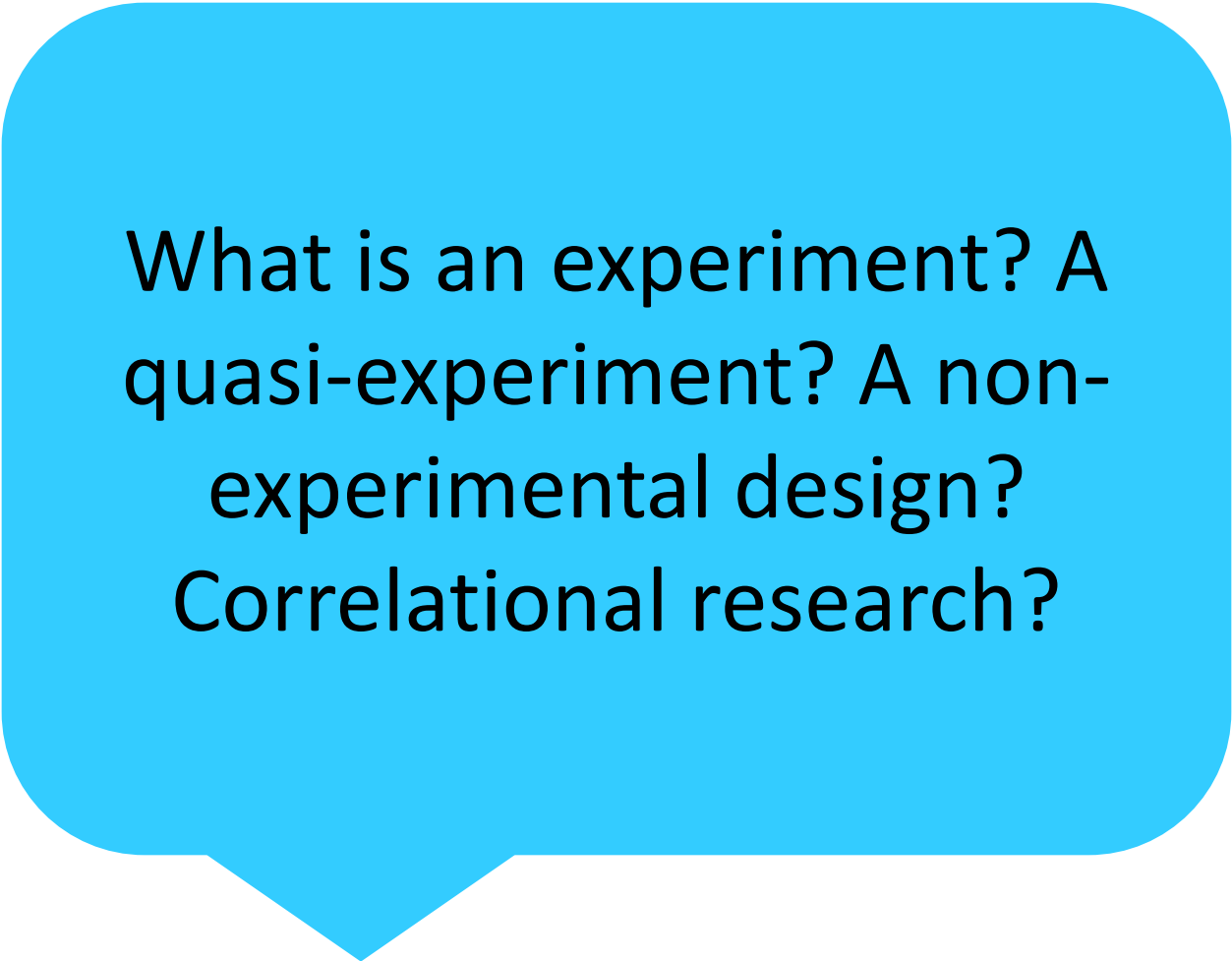
# What research methods do I/O psychologists use?

	Quantitative	Qualitative
<b>Advantages</b>	<ul style="list-style-type: none"><li>• Answers how much or how many</li><li>• More statistically robust</li><li>• Preferred method</li><li>• More legally defensible</li></ul>	<ul style="list-style-type: none"><li>• Answers why and how to fix</li><li>• Requires fewer participants</li><li>• Procedures more flexible</li></ul>
<b>Disadvantages</b>	<ul style="list-style-type: none"><li>• Procedures fixed</li><li>• Requires more participants</li></ul>	<ul style="list-style-type: none"><li>• Less statistically robust</li><li>• Least preferred method</li><li>• Less legally defensible</li><li>• Small sample size means less generalizability</li></ul>

# What research methods do I/O psychologists use?

- **Triangulation:** Approach that aims to seek converging information from different methods, sources, and instruments





What is an experiment? A  
quasi-experiment? A non-  
experimental design?  
Correlational research?

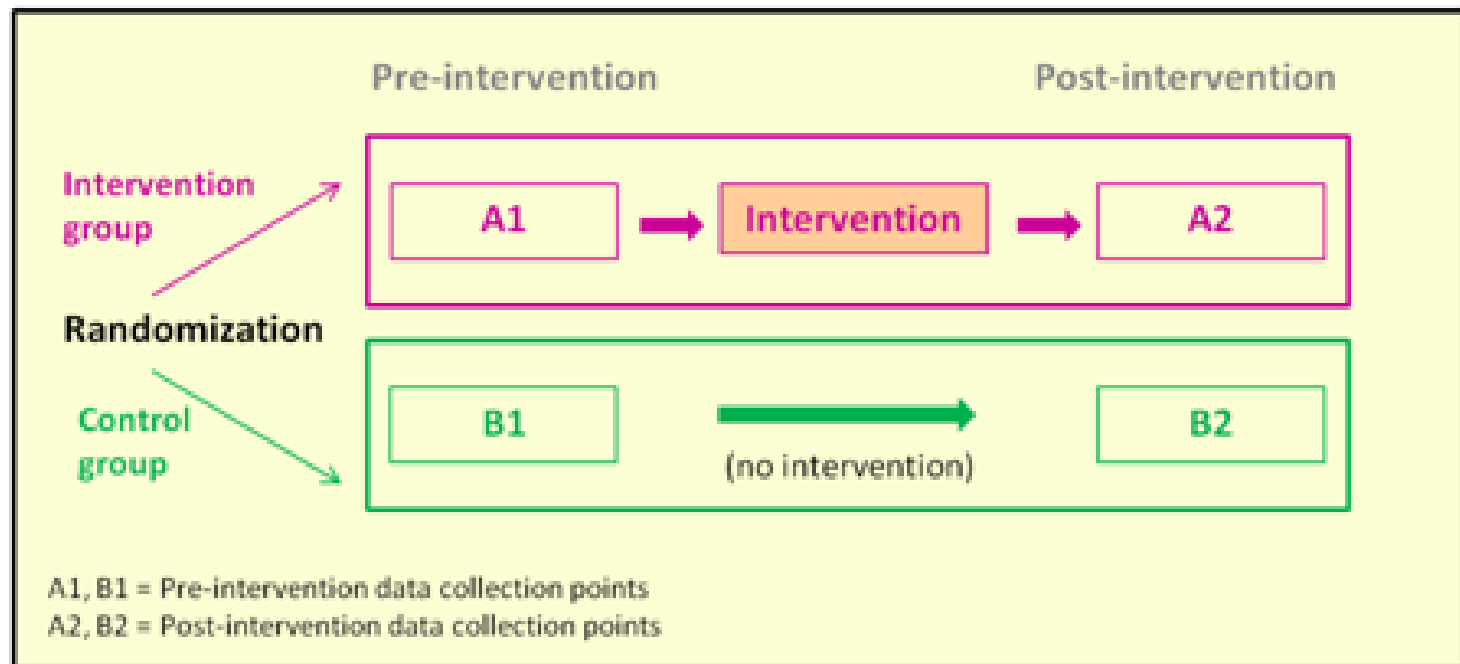


# What is an experiment?

- **Experimental research:** Examines cause-and-effect relationships
  - Random assignment to conditions + manipulation of IV + measurement of DV
  - Independent variable = Predictor
    - E.g., job satisfaction, organizational commitment
  - Dependent variable = Criterion
    - E.g., job performance, turnover



# What is an experiment?



# What is an experiment?

- Statistics determine whether differences between conditions are statistically significant

$p < 0.05$  = Differences significant

- Probability of replication by chance less than 5%

$p > 0.05$  = Differences not significant

- Probability of replication by chance more than 5%



# What is an experiment?

## Advantages

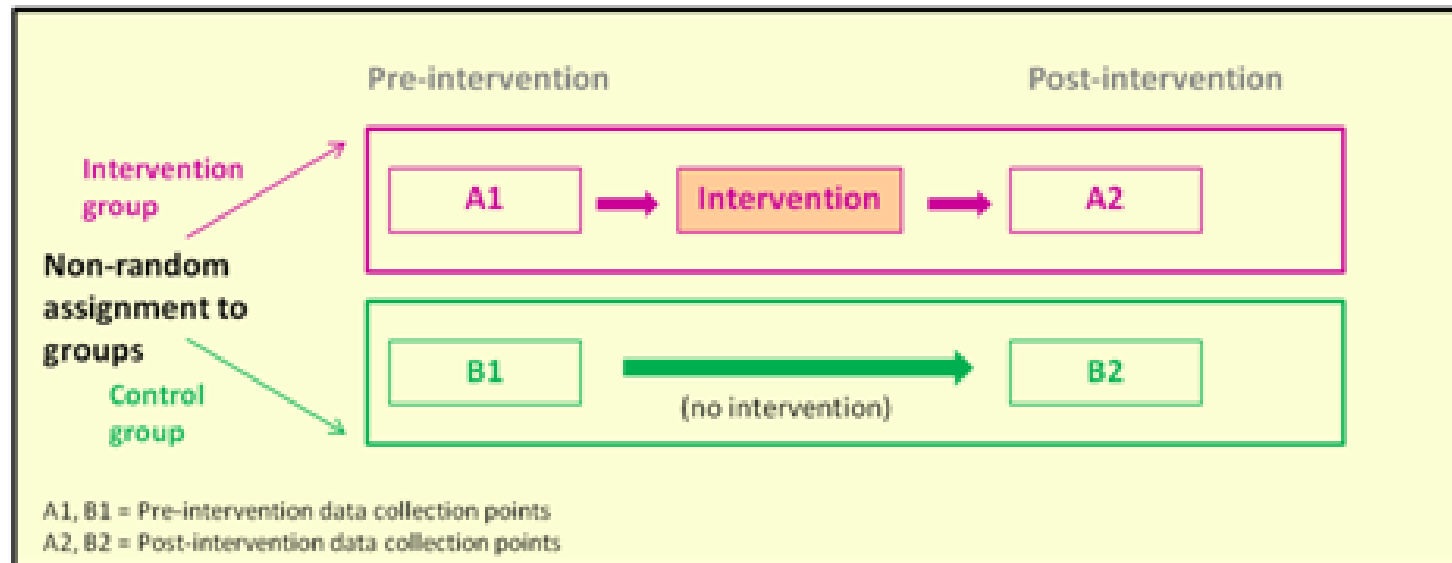
- Lab experiments
  - More control
  - More likely to lead to causal explanations
- Field experiments
  - Better able to simulate behaviors
  - Rely on actual employees performing their jobs

## Disadvantages

- Lab experiments
  - Limited ability to simulate behaviors
  - Rely on convenience samples
- Field experiments
  - Less control
  - Less likely to lead to causal explanations
  - Not appropriate when workers cannot be randomly assigned or when investigating sensitive topics

# What is a quasi-experiment?

- **Quasi-experimental research:** Non-random assignment to conditions + manipulation of IV + measurement of DV

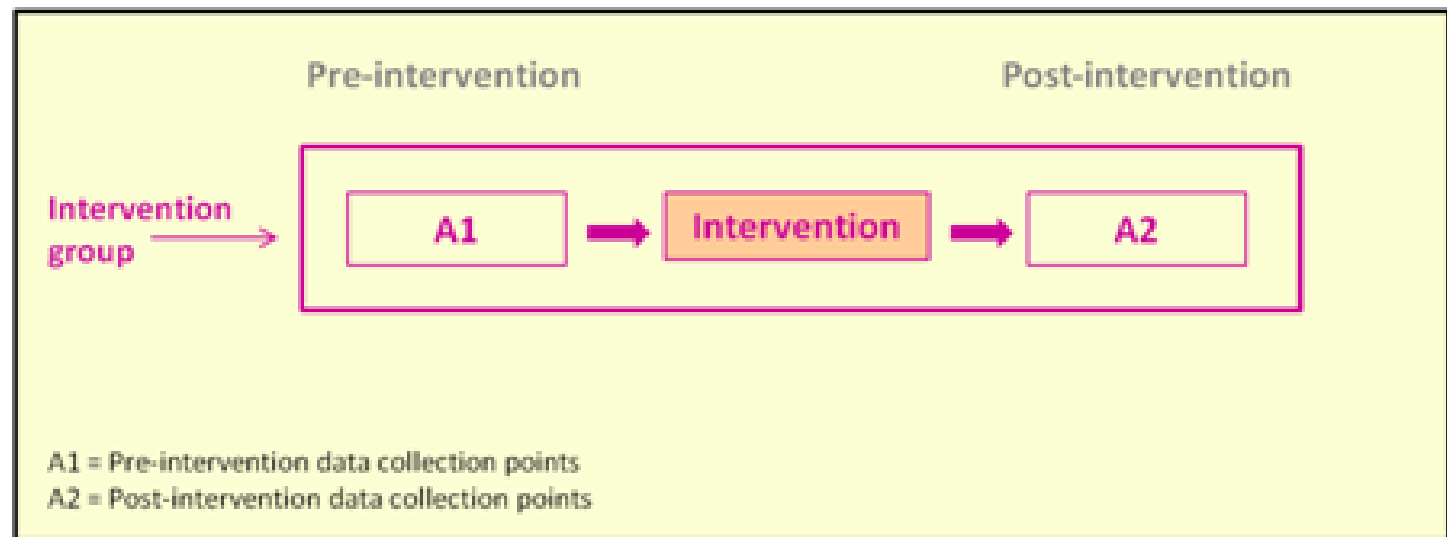


# What is non-experimental research?

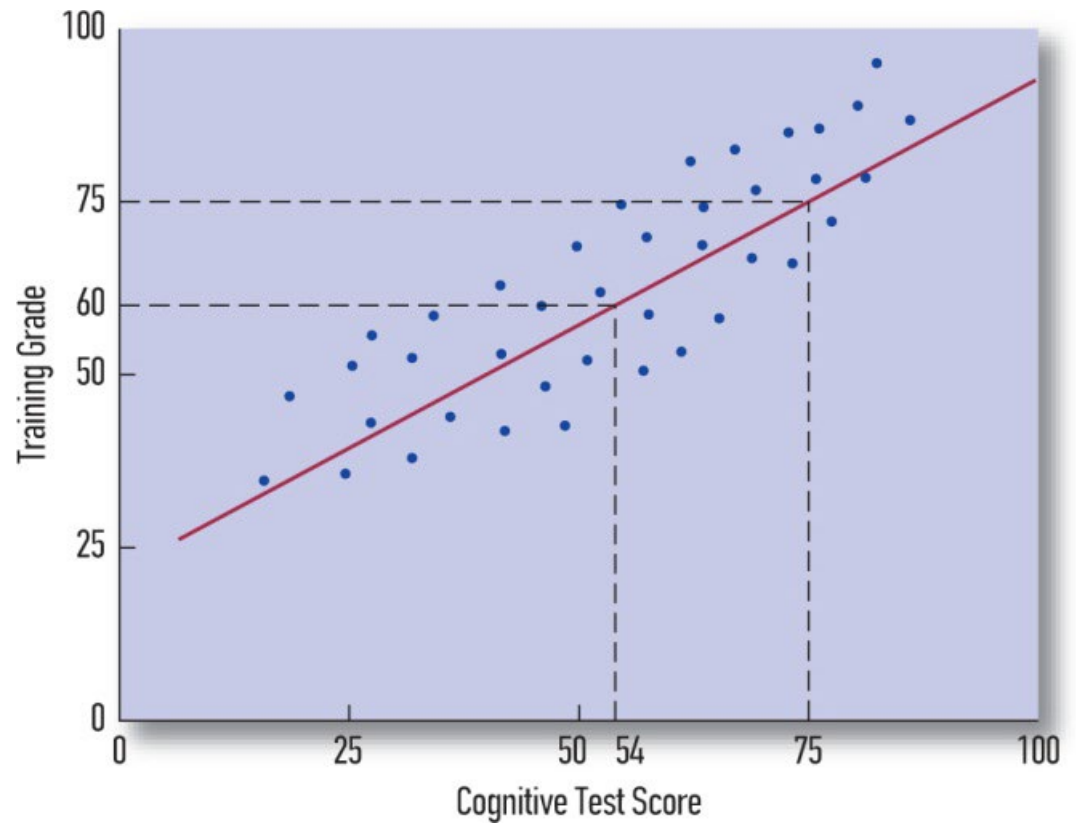
- **Non-experimental research:** Measurement of two or more variables
  - **Descriptive research:** Describes variables
    - Observational studies
    - Survey studies
    - Archival studies
  - **Correlational research:** Measures the relationship between two or more variables



# What is non-experimental research?



What is  
correlational  
research?





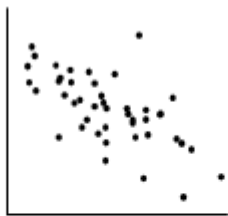
# What is correlational research?

- **Correlation coefficient:** A statistic that describes the strength and direction of the relationship
  - Ranges from -1.00 to +1.00
  - # closer to 1 = stronger relationship
  - # closer to 0 = weaker relationship
  - Positive # = change in the same direction
  - Negative # = change in different directions

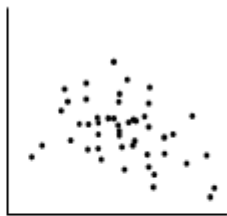


$r = -1$

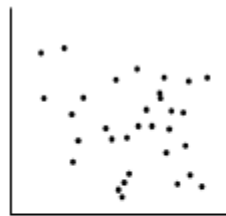
Points fall exactly  
on a straight line



$r = -0.7$

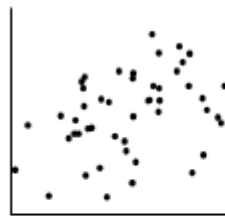


$r = -0.4$

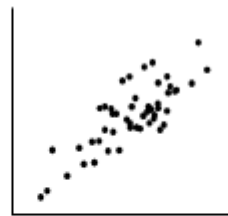


$r = 0$

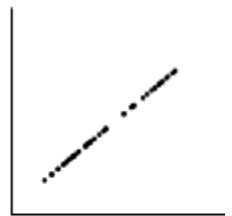
No linear  
relationship



$r = 0.3$



$r = 0.8$

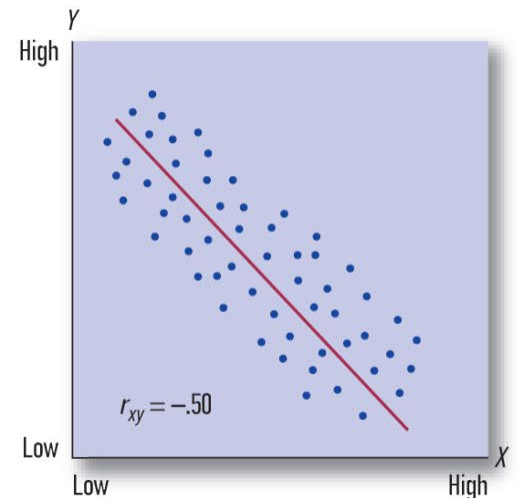
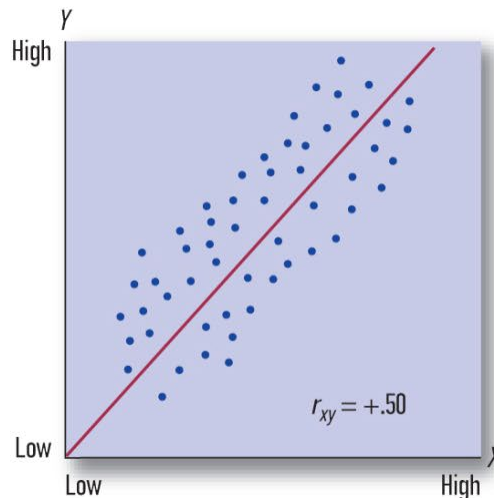
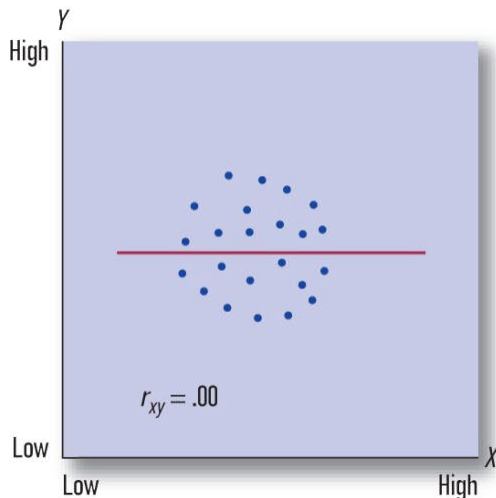


$r = 1$

Points fall exactly  
on a straight line

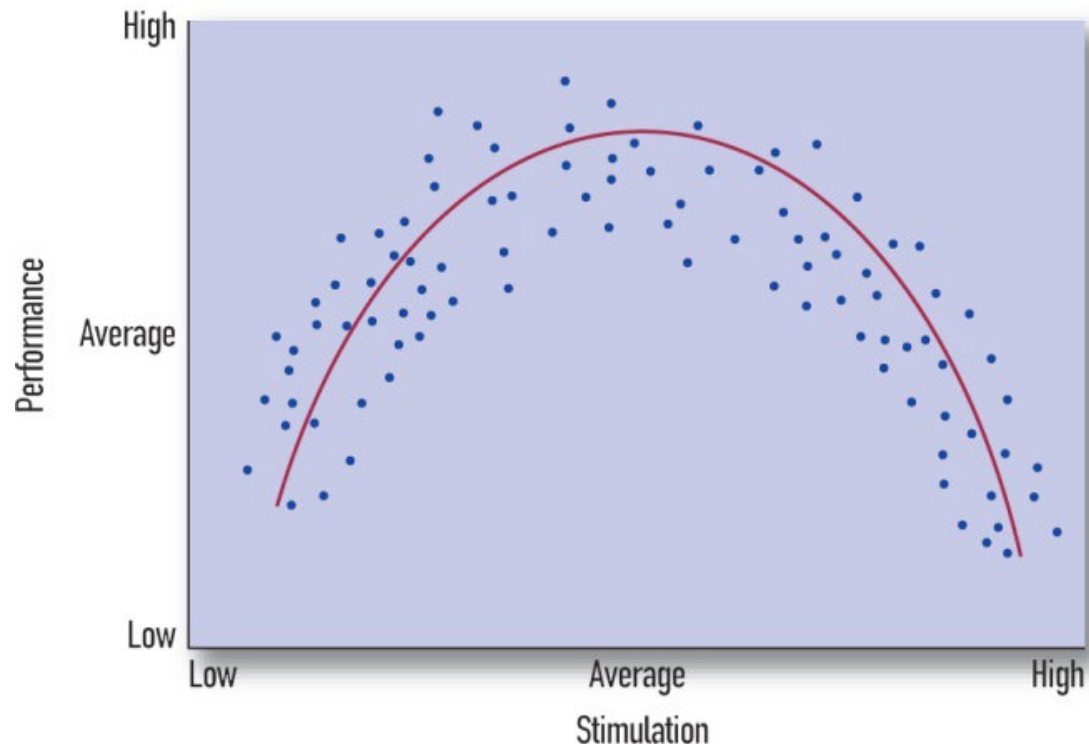
# What is correlational research?

- **Scatterplot:** Graphical display of the correlation between two variables
  - **Regression line:** Straight line that best “fits” the scatterplot and represents the relationship



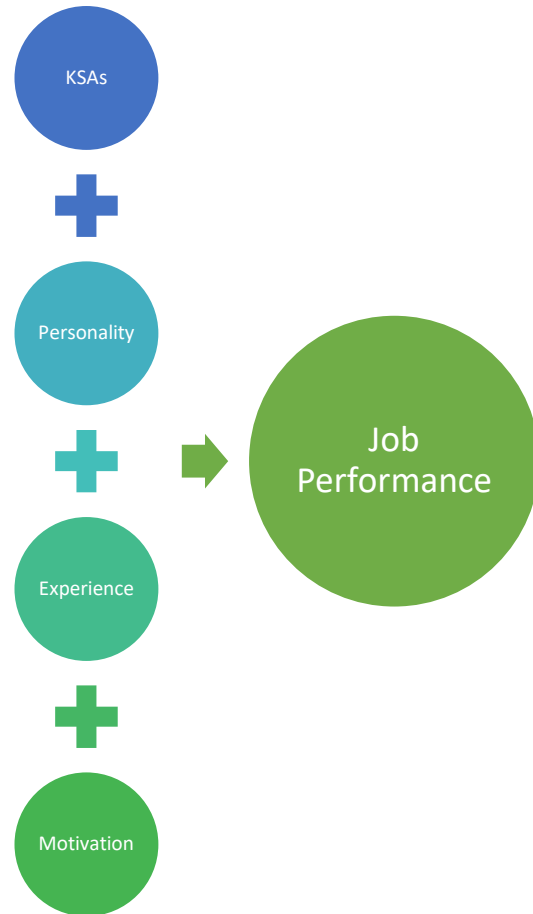
# What is correlational research?

- Relationship may be curvilinear
  - E.g., stimulation and performance relationship



# What is correlational research?

- **Multiple correlation coefficient:** Statistic that represents overall linear association between several variables and a single outcome variable



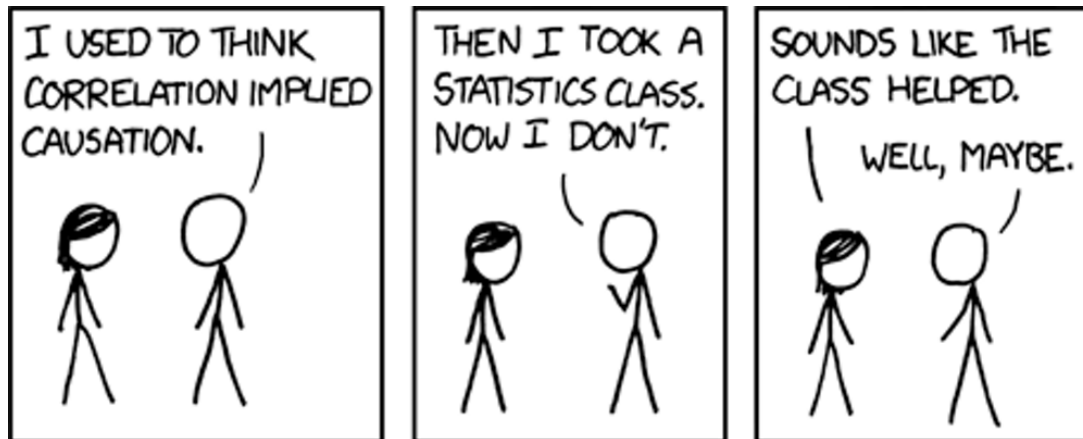
# What is correlational research?

## Advantages

- Enables the study of variables that cannot be manipulated
- Laboratory or field

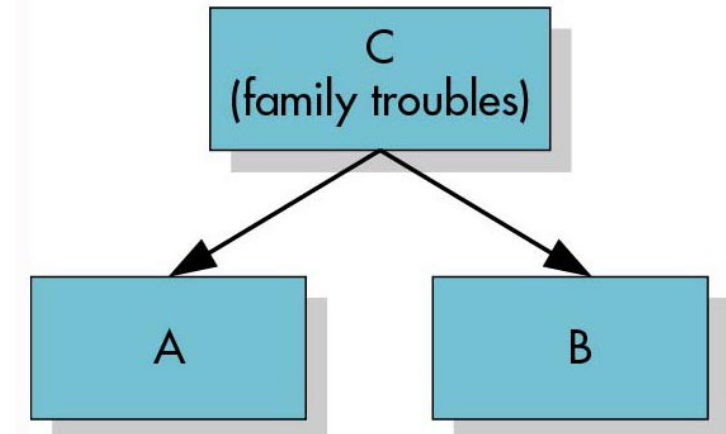
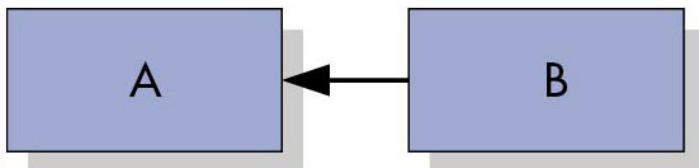
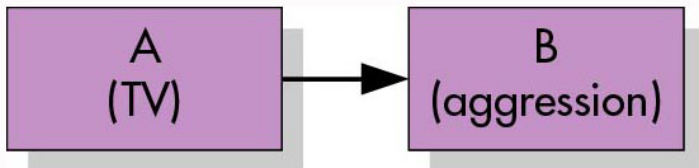
## Disadvantages

- Correlation  $\neq$  causation!
- Cannot rule out extraneous variables



# What is correlational research?

- Other explanations for the relationship between variables:





What are descriptive and  
inferential statistics?

# Statistics

```
graph TD; A[Statistics] --> B[Descriptive Statistics]; A --> C[Inferential Statistics];
```

## Descriptive Statistics

Presenting, organizing  
and summarizing data

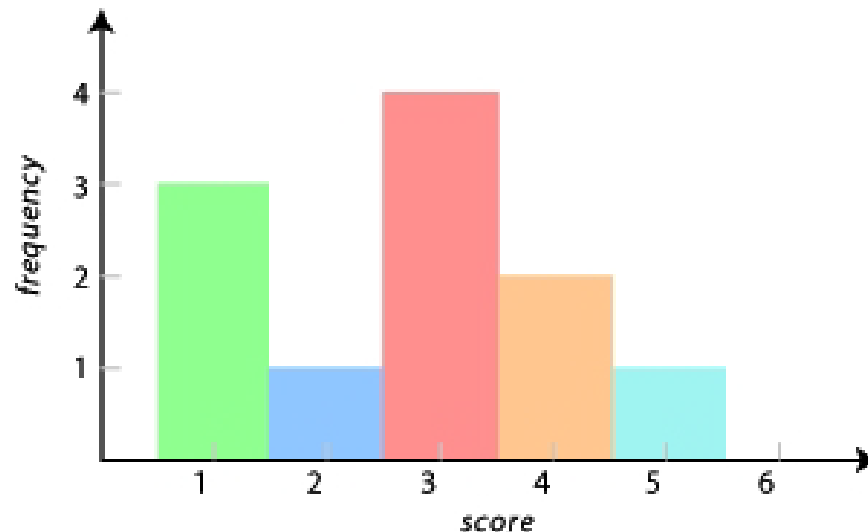
## Inferential Statistics

Drawing conclusions  
about a population based  
on data observed in a  
sample



# What are descriptive statistics?

- **Descriptive statistics:** Used to summarize, organize, and describe a sample of data
- Used to create visual representations of the data (**frequency distribution tables and histograms**)



# What are descriptive statistics?

- 3 characteristics of frequency distributions
  - **Measures of central tendency:** Statistics that identify the center of a distribution
    - **Mean:** The average score
    - **Mode:** The most frequently occurring score
    - **Median:** The score that splits the distribution in half

## Mean

Add all the numbers then divide by the amount of numbers

9, 3, 1, 8, 3, 6

$$9 + 3 + 1 + 8 + 3 + 6 = 30$$

$$30 \div 6 = 5$$

The mean is 5

## Median

Order the set of numbers, the median is the middle number

9, 3, 1, 8, 3, 6

1, 3, 3, 6, 8, 9

The median is 4.5

## Mode

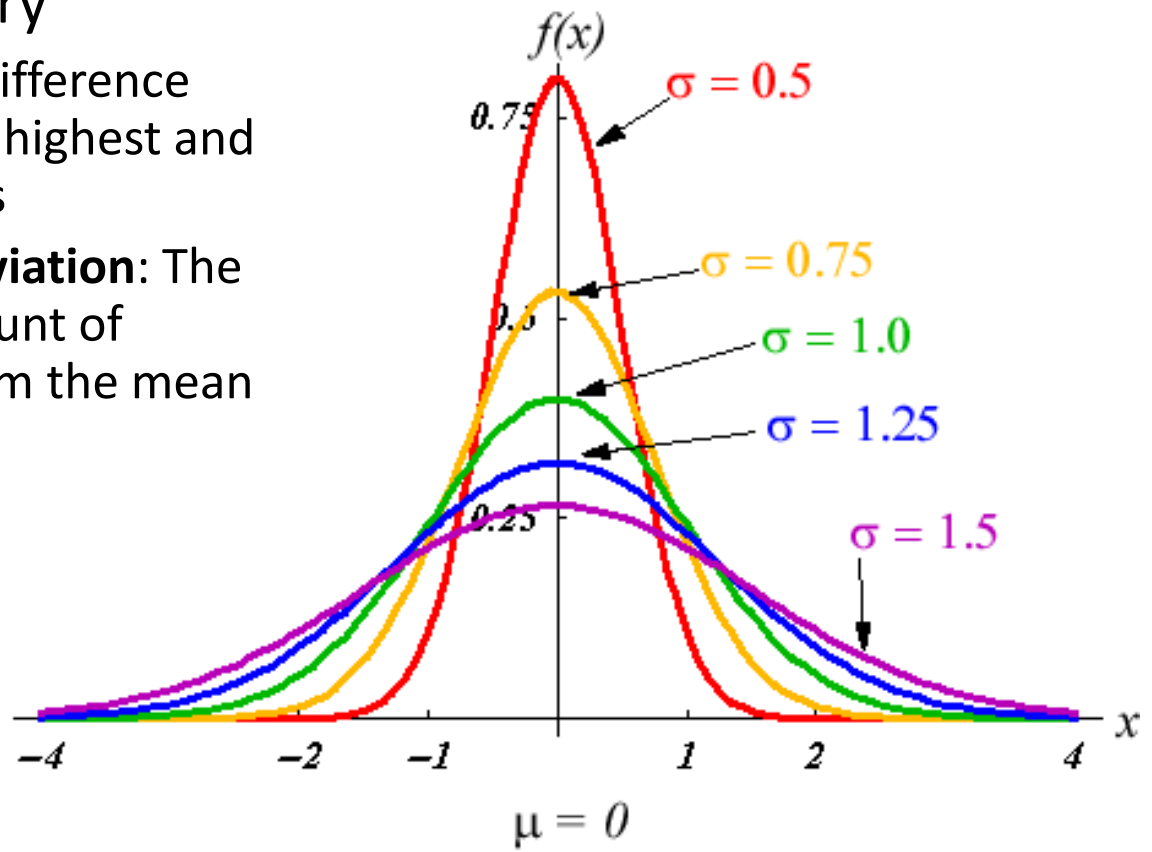
The most common number

9, 3, 1, 8, 3, 6

The mode is 3

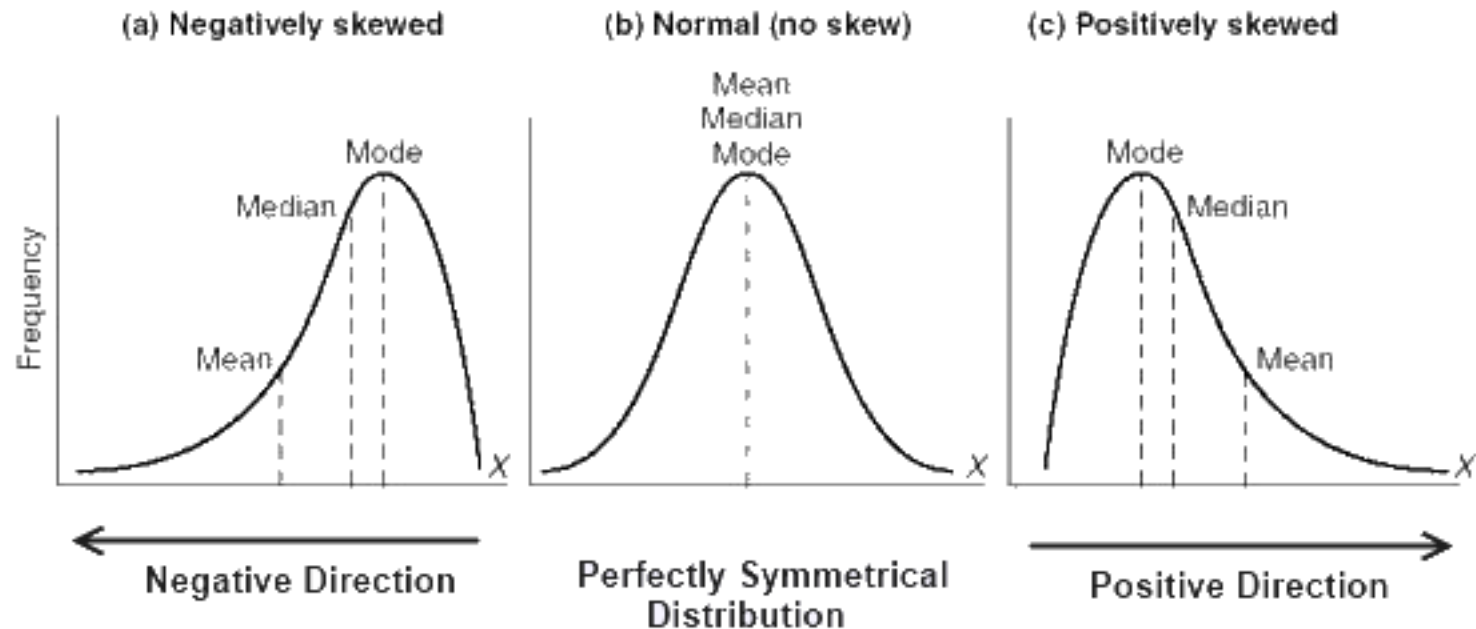
# What are descriptive statistics?

- 3 characteristics of frequency distributions
  - **Variability:** The extent to which scores in the distribution vary
    - **Range:** The difference between the highest and lowest scores
    - **Standard deviation:** The average amount of deviation from the mean



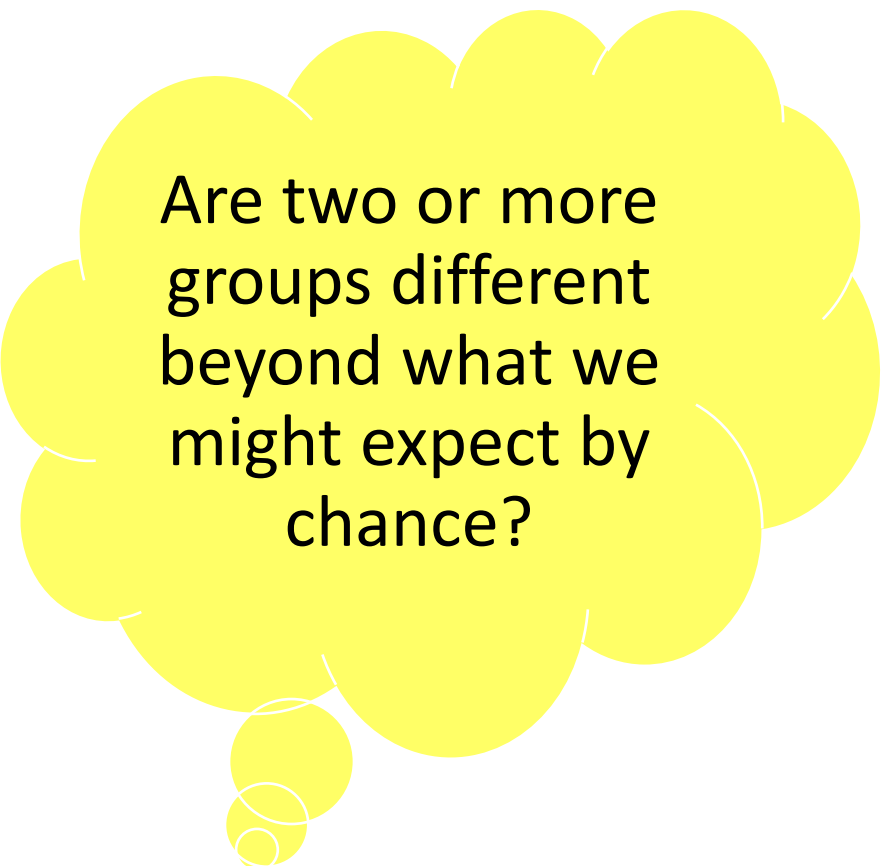
# What are descriptive statistics?

- 3 characteristics of frequency distributions
  - **Skewness:** The extent to which scores tend to fall on the left or right of the distribution



# What are inferential statistics?

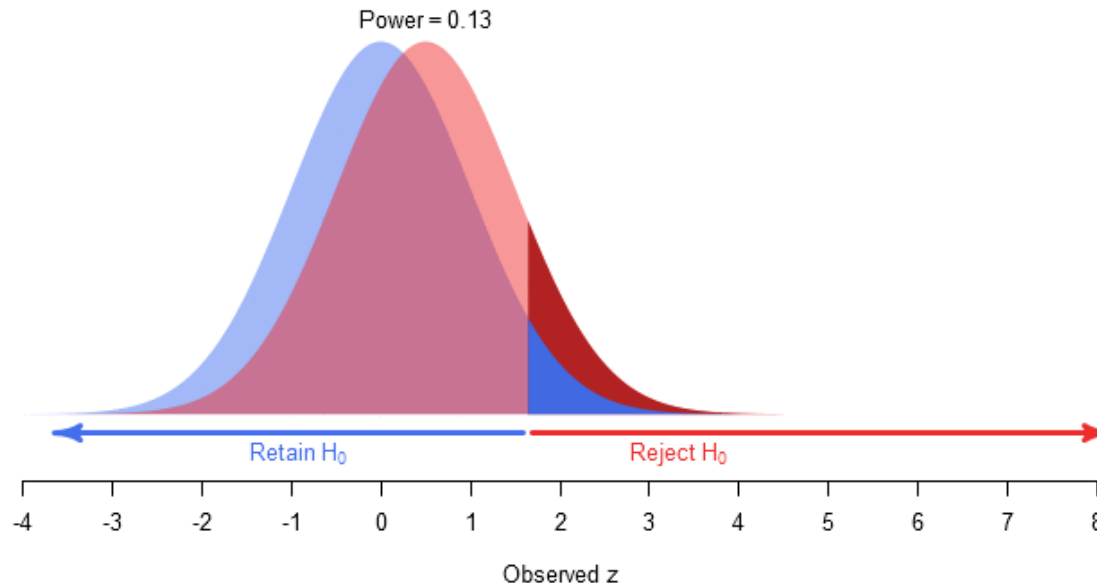
- **Inferential statistics:**  
Used to test hypotheses, draw conclusions about relationships between variables, and make inferences about the population of interest
- E.g., t-test, F-test, chi-square test



Are two or more groups different beyond what we might expect by chance?

# What are inferential statistics?

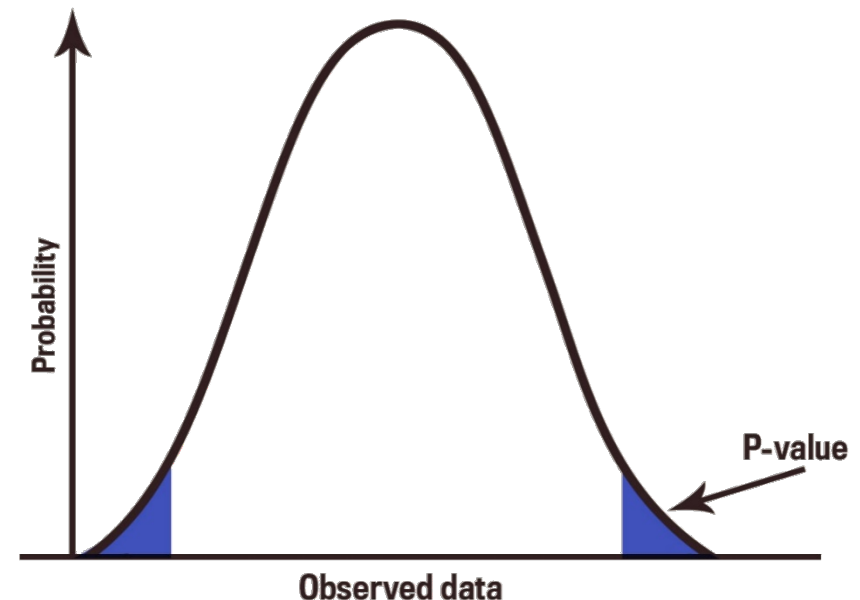
- **Statistical power:** The likelihood of finding a statistically significant difference when a true difference exists
  - Calculated before data collection to determine sample size
  - The smaller the sample size, the lower the power



# What are inferential statistics?

- **Statistical significance:** Indicates the probability that the results are due to random chance
  - Does not indicate the strength of a relationship or the importance of results

$p < \alpha$	Results statistically significant (not likely due to random chance)
$p > \alpha$	Results not statistically significant (likely due to random chance)

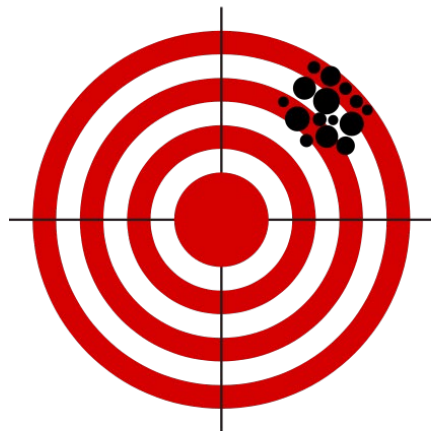


What is reliability? Test-retest reliability? Equivalent forms reliability? Internal consistency? Inter-rater reliability?

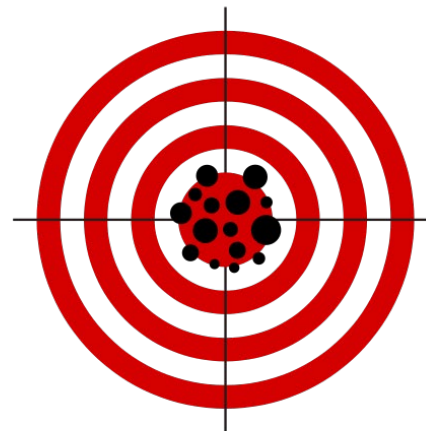


# What is reliability?

- **Reliability:** A test's consistency or stability
- High reliability ( $\pm 0.8-1.0$ ) = Test produces consistent results across time, forms, raters, and/or items
- Perfect correlations ( $\pm 1.0$ ) rare due to errors



Reliable, Not Valid



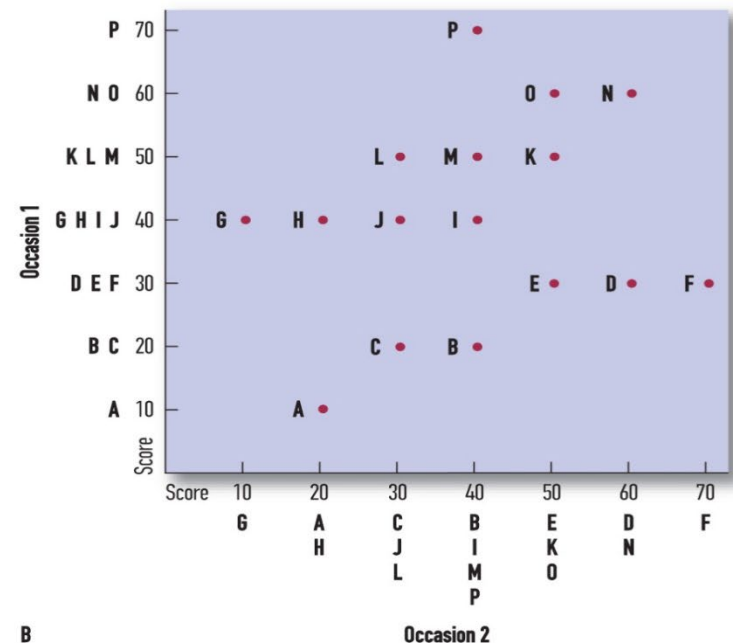
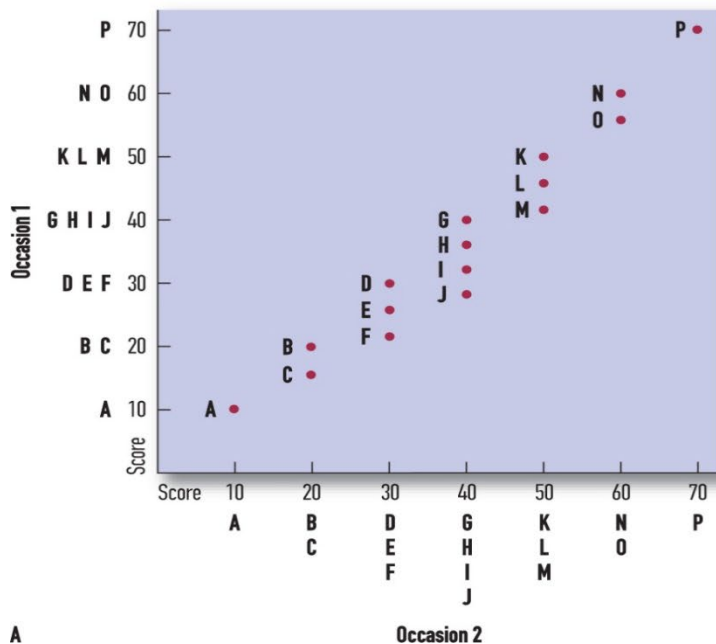
Both Reliable & Valid

# What is reliability?

Type of Reliability	What It Is	How You Do It	What the Reliability Coefficient Looks Like
<b>Test-Retest</b>	A measure of stability	Administer the same test/measure at two different times to the same group of participants	$r_{\text{test1} \cdot \text{test1}}$
<b>Parallel Forms</b>	A measure of equivalence	Administer two different forms of the same test to the same group of participants	$r_{\text{form1} \cdot \text{form2}}$
<b>Inter-Rater</b>	A measure of agreement	Have two raters rate behaviors and then determine the amount of agreement between them	Percentage of agreements
<b>Internal Consistency</b>	A measure of how consistently each item measures the same underlying construct	Correlate performance on each item with overall performance across participants	Cronbach's alpha Kuder-Richardson

# What is test-retest reliability?

- **Test-retest reliability:** Reflects a test's temporal stability across Time 1 and Time 2 measurements in the same group



# What is equivalent forms reliability?

- **Equivalent (AKA parallel or alternate) forms reliability:**  
Reflects a test's equivalence across two forms administered to the same group
- E.g., SAT, ACT, GRE

SAT Alternate-Form Reliability  
by Skill

<i>Skill</i>	<i>Alternate-Form Reliability</i>
Reading_Sk1	.65
Reading_Sk2	.77
Reading_Sk3	.79
Reading_Sk4	.60
Math_Sk1	.78
Math_Sk2	.78
Math_Sk3	.71
Math_Sk4	.72
Writing_Sk1	.68
Writing_Sk2	.44
Writing_Sk3	.56

# What is internal consistency?

- **Internal consistency:**  
Reflects a test's consistency across even-numbered and odd-numbered items

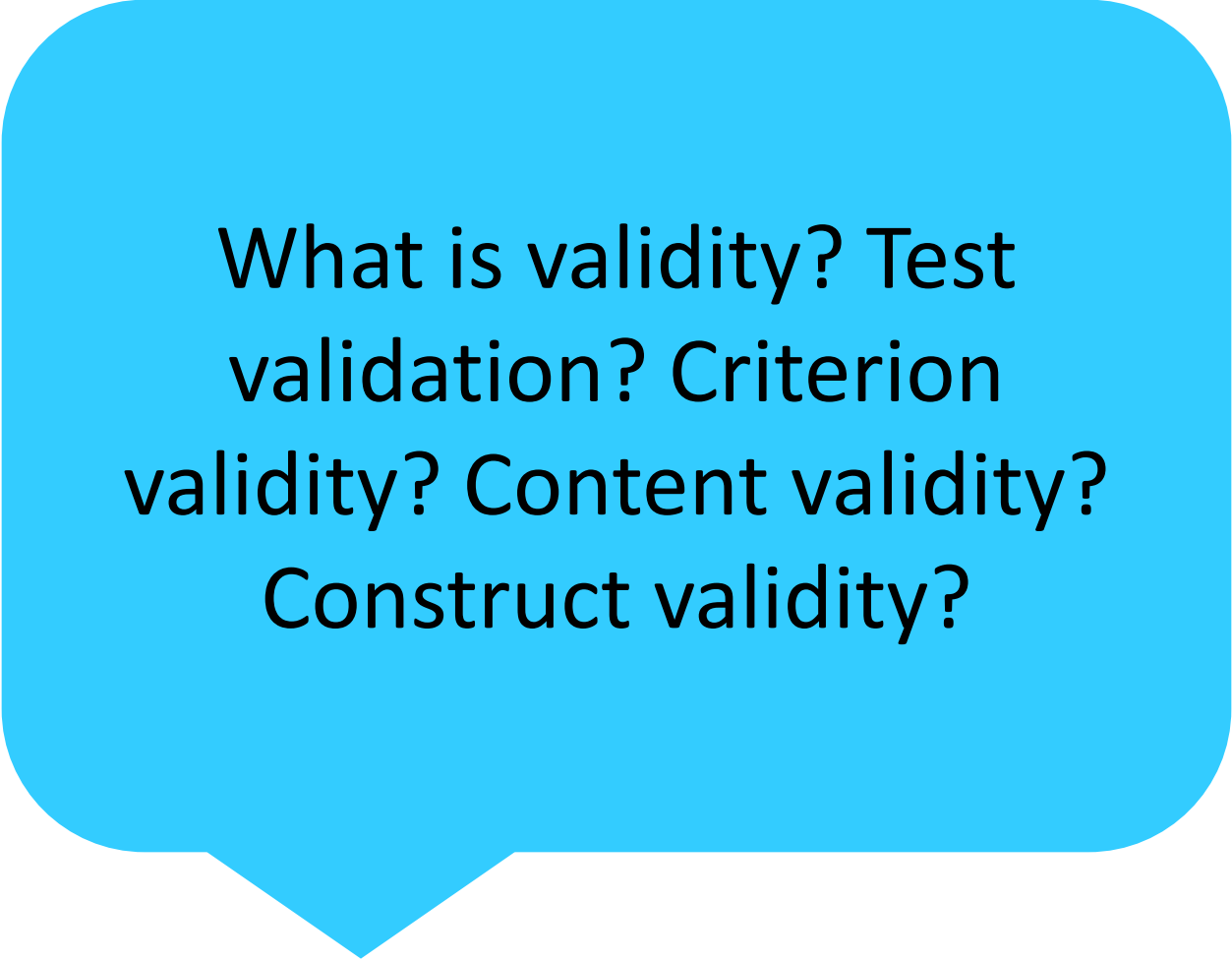
SAT Internal Consistency by Skill

<i>Skill</i>	<i>Internal Consistency Form 1</i>	<i>Internal Consistency Form 2</i>
Reading_Sk1	.69	.60
Reading_Sk2	.84	.82
Reading_Sk3	.82	.87
Reading_Sk4	.69	.68
Math_Sk1	.80	.82
Math_Sk2	.77	.79
Math_Sk3	.68	.80
Math_Sk4	.81	.64
Writing_Sk1	.64	.68
Writing_Sk2	.40	.45
Writing_Sk3	.67	.56

# What is inter-rater reliability?

- **Inter-rater reliability:** Reflects the level of agreement among raters
- Assumed to reduce human error and bias





What is validity? Test  
validation? Criterion  
validity? Content validity?  
Construct validity?

# What is validity?

- **Validity:** A test's accuracy and comprehensiveness
- **Validity coefficient:** The correlation between a predictor (test) and a criterion (outcome)
- **Test validation:** The process of determining whether a test's results accurately and completely represent what was hoped to be measured





# What is test validation?

- Step 1: Job Analysis
  - Identify job demands and job-related attributes
- Step 2: Define Performance
  - **Criterion:** An outcome variable that defines performance in terms of meeting job demands

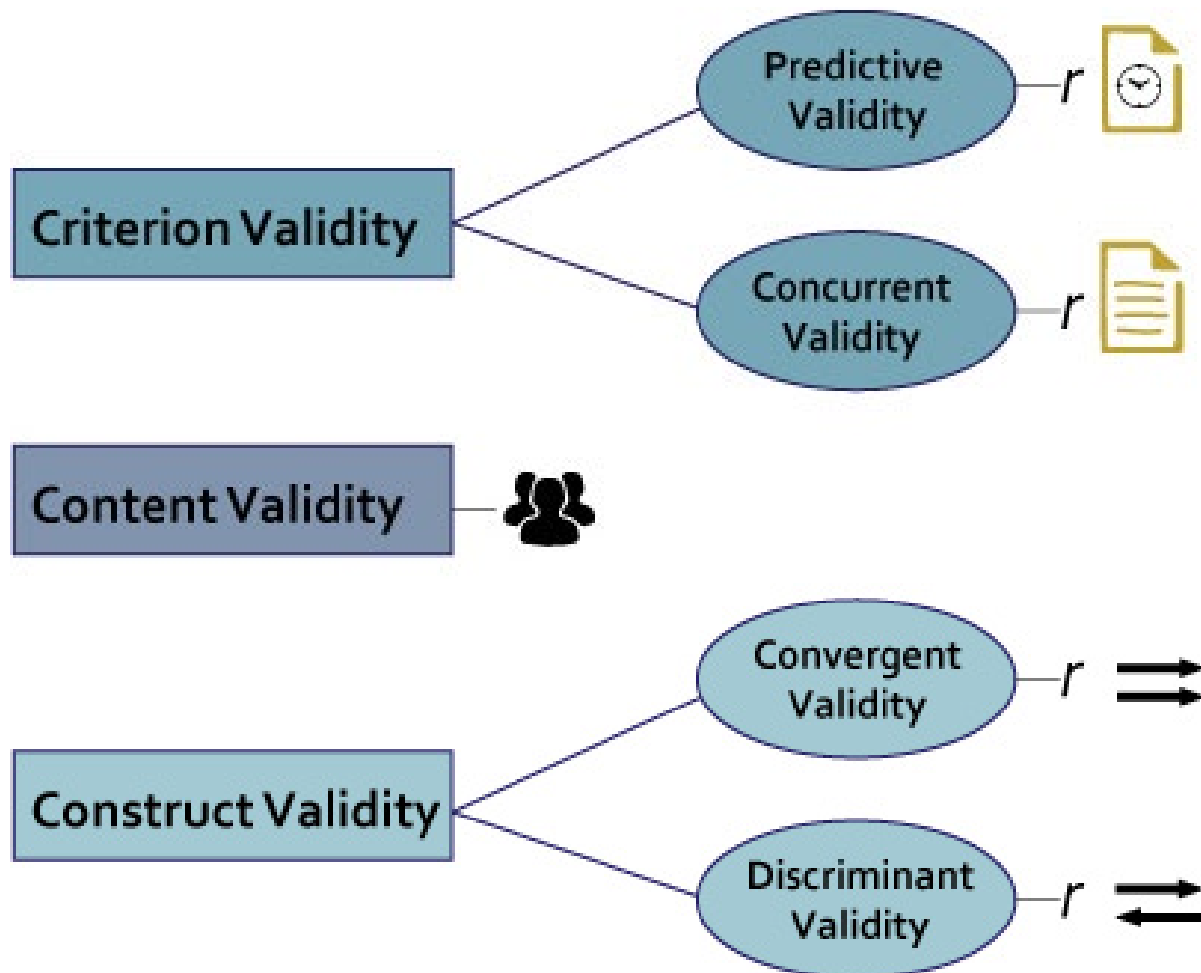


# What is test validation?

- Step 3: Test Development
  - **Predictor:** A test that assesses job-related attributes
- Step 4: Test Validation
  - Test hypothesis that people who have high predictor scores will also have high criterion scores

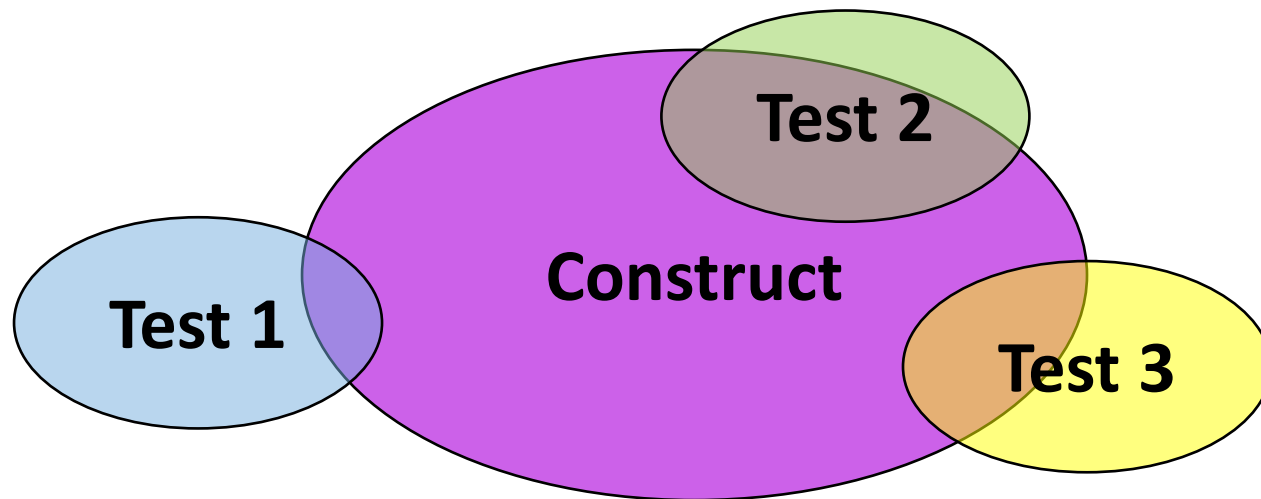


# What is test validation?



# What is construct validity?

- **Construct validity:** The extent to which a predictor accurately and comprehensively measures a construct
- **Construct:** A psychological concept indirectly measured by a predictor
  - E.g., intelligence, creativity, integrity, strength



# What is construct validity?

## **Convergent Validity**

- Are two predictors of the same construct correlated (as they should be)?

**Amy's  
Intelligence  
Test**

**Weschler  
Adult  
Intelligence  
Test**

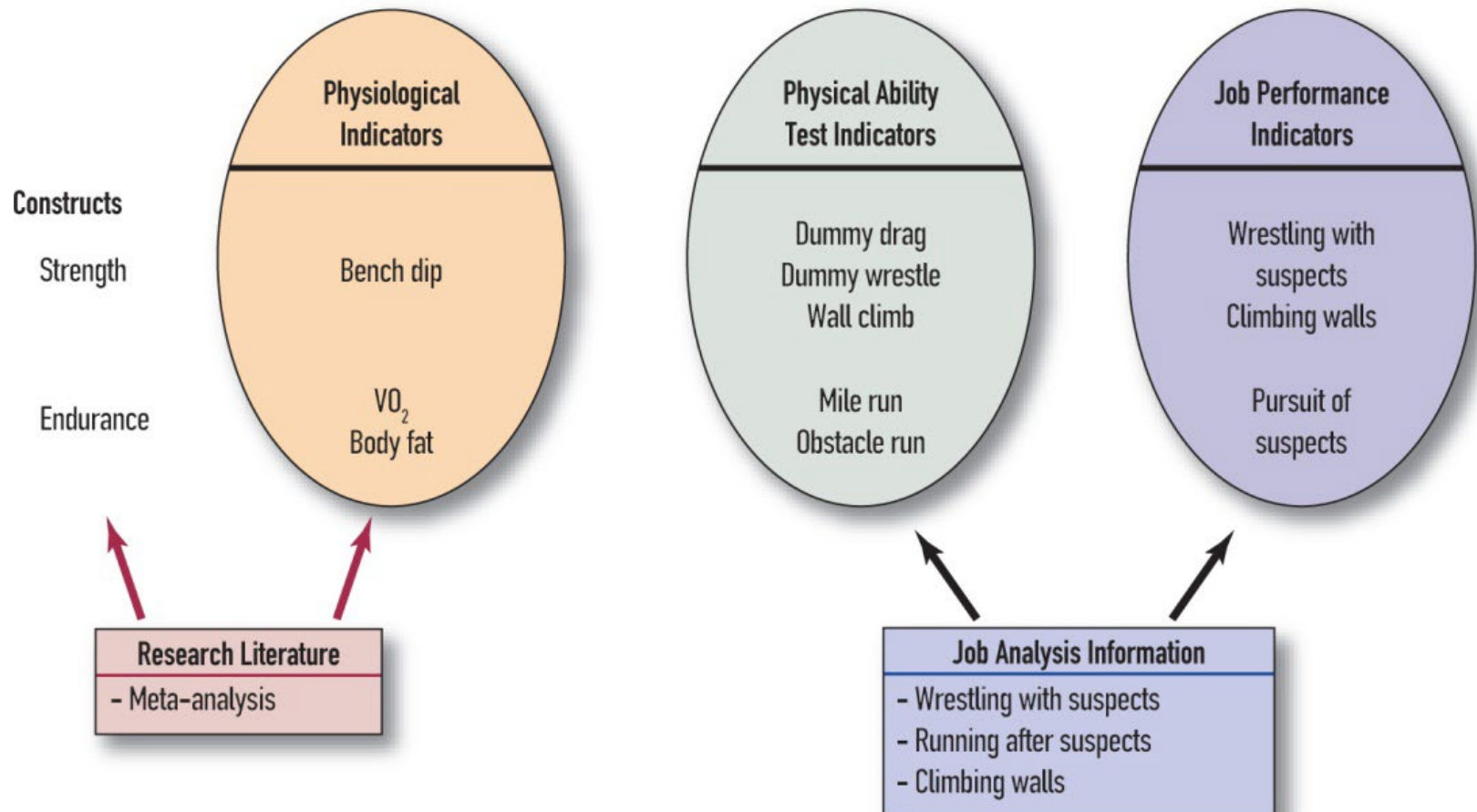
## **Discriminant Validity**

- Are two predictors of the same construct correlated more so than one of the predictors and an unrelated predictor?

**Amy's  
Intelligence  
Test**

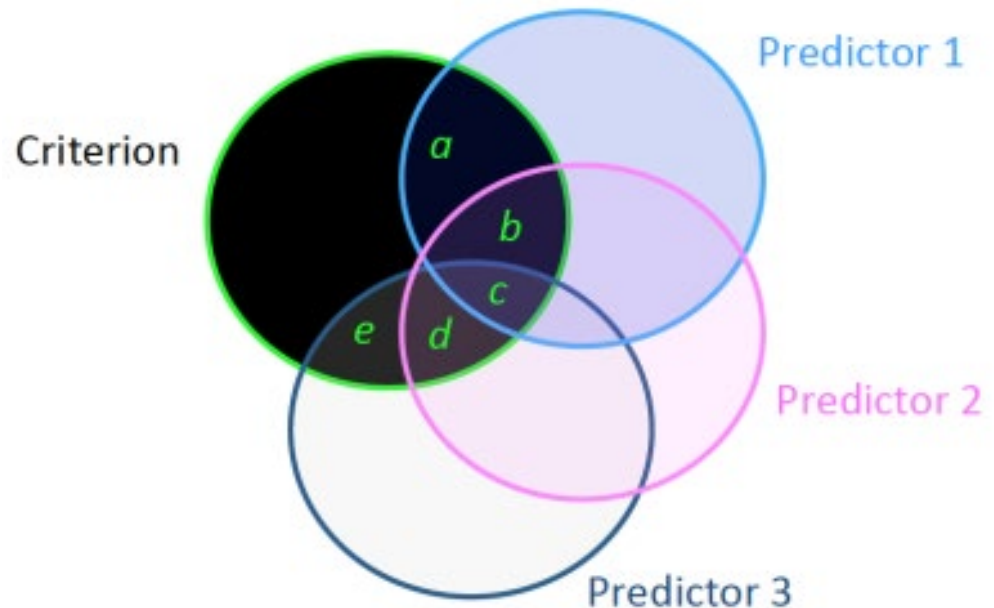
**Physical  
Abilities  
Test**

# What is construct validity?



# What is criterion-related validity?

- **Criterion-related validity:** The extent to which a predictor score correlates with a criterion score
- Evidence enhances confidence that the test can be used to predict the outcome
- 2 types:
  - Predictive validity
  - Concurrent validity



# What is criterion-related validity?

- **Predictive validity:** The extent to which a predictor score taken at Time 1 correlates with a criterion score taken at Time 2
  - Time 1 – Administer test to applicants (but do not use the results to hire)
  - Time 2 – Measure the performance of applicants who were hired





# What is criterion-related validity?

- **Concurrent validity:** The extent to which a predictor score correlates with a criterion score measured simultaneously
  - Time 1 – Administer test to current employees and measure their current performance

# What is content-related validity?

- **Content-related validity:** The extent to which a test represents an adequate sample of work behaviors or employee KSAOs as defined by job analysis
- Job incumbents and SMEs compare the test and the job (no criterion data or validity coefficient)

