► Chapter 10: USING BETWEEN-SUBJECTS AND WITHIN-SUBJECTS EXPERIMENTAL DESIGNS ◄

An experimental design is used when you want to establish a causal relationship and you can manipulate a variable (IV). Manipulate a variable; set value to two different values (levels); observe performance.

"Three aspects of the experiment play the biggest part in determining the design:
(1) the number of independent variables
(2) the number of treatment conditions
(3) are the same or different subjects used in each of the treatment conditions.

TYPES OF EXPERIMENTAL DESIGN
Three types of experimental designs

A. BETWEEN-SUBJECTS DESIGN:
- Different groups of subjects are randomly assigned to the levels of the independent variable.

B. WITHIN-SUBJECTS DESIGNS:
- Only one group of subjects; subjects receive all levels of the independent variable at different times.

C. SINGLE-SUBJECT DESIGNS:
- Use the same method of varying the level of the independent variable used by the with-in subjects designs.
- Focus on change of individual subjects under the different treatment conditions.

THE PROBLEM OF ERROR VARIANCE IN BETWEEN-SUBJECTS AND WITHIN-SUBJECTS DESIGNS

Error Variance: is the statistical variability of scores caused by the influences of variables other than your independent variable.

If we simply look at the means, there could be two explanations:

1)Reflects nothing more than ______________________________

____

2)Independent variable ____________________________________

BETWEEN-SUBJECTS DESIGNS

(Family of Designs)

SINGLE-FACTOR RANDOMIZED GROUPS DESIGNS

Randomized Group Design: randomly assign subjects to level of independent; variable to form "groups" of subjects

1) THE RANDOMIZED TWO-GROUP DESIGN:

+ Randomly assign subjects to two groups
+ Expose each to different levels of the independent variable.

ADVANTAGES: simple to carry out

1. Economical in terms of time and material.
2. No Pre-testing of subjects in necessary
3.

DISADVANTAGES:

1. Provides limited amount of information about effects of independent variable.
2.
3. Sensitivity:

2) THE RANDOMIZED MULTIGROUP DESIGN

1. Single-Factor Parametric Design: assess independent variable at more than two levels.

2. Single-Factor Nonparametric Design:
   Nonparametric Design - nominal scale

3. Multiple Control Group Design:
   - Include multiple control groups when a single control group is not adequate to rule out alternative explanations of your results.

WITHIN-SUBJECTS DESIGNS (repeated measures)

The comparison of treatment effects involves looking at changes in performance within each subject across treatments.
- Expose a single group of subject to all the treatments

TYPES OF WITHIN-SUBJECTS DESIGNS

(family of designs)

1. **SINGLE-FACTOR TWO-LEVEL DESIGN**: 2 levels of a single independent variable. All subjects receive both levels of the variable.
   - If dependent variable not strongly affect by subject-related variable design will be less effective.

2. **SINGLE-FACTOR MULTILEVEL DESIGNS**:
   - More than two levels of the independent variable
   - Single group of subject is exposed to three or

more levels of a single independent variable.

3. **MULTIFACTOR WITHIN-SUBJECTS DESIGNS**: Includes two or more independent variables
Factorial Designs: Each subject is exposed to every combination of levels of all the factors (independent variable)
- Have main effects and interactions.

4. **OTHER WITHIN-SUBJECTS DESIGNS**:
Nonfactorial design

5. **MULTIVARIATE WITHIN-SUBJECTS DESIGNS**:
Use of more than one dependent variable

WHEN TO USE A WITHIN-SUBJECTS DESIGN:

1. **SUBJECT VARIABLES CORRELATED WITH THE DEPENDENT VARIABLE**:
   - Use when subject differences contribute heavily to variation in the dependent variable.

2. **ECONOMIZING ON SUBJECTS**:
   - Use when # of subjects is limited and carryover absent or minimized.

3. **ASSESSING THE EFFECTS OF INCREASED EXPOSURE ON BEHAVIOR**:
   - Measured a number of trials, passage of time, etc.
   - Looking at changes as a function of earlier exposure

A. ADVANTAGES:
- Close to matched groups. Within provides the ultimate in matching of characteristics.
- Tends to be more powerful than equivalent between-subject design.
- Increase power may allow for use of less subjects.

B. DISADVANTAGES:
- Amount of time in the experiment
- Carryover effects

C. SOURCES OF CARRYOVER EFFECTS: Potential sources for carryover effects
1. <u>Learning</u>: if learn a task; second performance is likely to be better if similar.
2. <u>Fatigue</u>: may lead to deterioration in later performance.
3. <u>Habituation</u>: reduction in responsiveness for repeated exposure.
4. <u>Sensitization</u>: exposure to one stimulus can cause subjects to respond more strongly to another stimulus.

5. Contrast: exposure to one condition may alter responses in other conditions

6. Adaptation (adjustments): may lead to different results due to adaptive changes (e.g., drug tolerance)

D. DEALING WITH CARRYOVER EFFECTS: Three ways to deal with carryover effects.

1. COUNTERBALANCING: assign various treatments of the experiment in a different order for different subjects.

a) COMPLETE COUNTERBALANCING: every possible ordering of treatments assign at least 1 subject to each ordering.

1) Every treatment follows every other treatment equally often.

2) Every treatment appears equally often in each position.
   - Minimum number of subjects is equal to the number of different ordering of the treatments.
   - k treatments have k! (k factorial).
     3 x 2 x 1 = 6;
     For four treatments:
     4 x 3 x 2 x 1 = 24

b) PARTIAL COUNTERBALANCING: include only some possible treatment orders. Orders retained are chosen randomly from total set.

Latin Square Design: partially-counterbalanced design
- Number of treatment order equivalent to number of treatments.
- Each item appears exactly once in each column and row.

2. TAKING STEPS TO MINIMIZE CARRYOVER: if minimize the error variance – increases power
- Pre-train subject
- Allow breaks

3. MAKING TREATMENT ORDER AN INDEPENDENT VARIABLE:
Order of treatment as a second independent variable

MATCHED GROUPS DESIGNS

Some subject characteristics correlate significantly with dependent variable

Matched Group Design: matched set of subjects are distributed at random.

Assess (a head of time)

1) LOGIC OF THE MATCHED GROUPS DESIGNS:

- Effect of characteristic is distributed across treatments.

2) ADVANTAGES AND DISADVANTAGES:

<u>Advantages</u>:

- Control over variable that may otherwise obscure the effects of independent variable
- matching increases experiments sensitivity

<u>Disadvantages</u>:

- If matched characteristics does NOT have an effect then statistics some what less powerful

3) MATCHED PAIRS DESIGN:

4) MATCHED MULTIGROUP DESIGNS:

Same approach used involving multiple levels of a single factor a multiple factor

WITHIN-SUBJECTS VERSUS MATCHED GROUPS DESIGNS:

- Both attempt to reduce error variance by directly comparing the response of similar participants.
- Both take advantage of correlation of scores across treatments.
- Both will be less powerful than equivalent randomized group design if correlation is low or absent.

Advantages of Within over matched:

- Do not have to measure variable on which subjects are matched.
- Tend to reduce variance better than matched.
- Requires fewer subjects.

ERROR VARIANCE for the various Designs

A. SOURCES OF ERROR VARIANCE:
- Each design deals with error variance in a different way.
- Impossible to hold constant all extraneous variables that could influence your dependent variable
- Different in subjects on number of levels

B. HANDLING ERROR VARIANCE in BETWEEN-SUBJECTS DESIGNS:

Ways to increase detecting effect of independent variable (if there is one).

1. Reducing Error Variance:
- Hold extraneous variables constant
- Treat each subject as similar as possible

2. Increasing the Effectiveness of Your Independent Variable:
- Look at literature
- Problem of a weak manipulation

3. Randomizing Error Variance Across Groups: randomize error across groups (for between subjects designs).
- Random Assignment:

4. Statistical Analysis:
- Given results is not likely to have occurred by chance - result is said to be statistically significant (probably reliable)

THE PROBLEM OF ERROR VARIANCE IN WITHIN-SUBJECTS DESIGNS

Error Variance: differences in subjects' scores that have nothing to do with the effect of the independent variable.

Between: randomly assign subjects to groups and then expose each group to a single experimental treatments.

Two differ in the way you expose subjects to the experimental treatment, but are the same in the following:
1. Expose subjects to different experimental treatments
2. Average score within each treatment
3. Compare means to determine the influences of the Independent Variable.
4. Statistical analysis (formula may differ)

Table for Error Variance

	Groups y/n	Error Variance	Advantage/Disadvantage
Between-Subjects			
Within-Subjects			
Matched-Subjects			

Between-Subjects

Control	Exp
S1: 74	S5: 79
S2: 90	S6: 97
S3: 87	S7: 92
S4: 94	S8: 98
$\bar{X}$ =	$\bar{X}$ =

Within-Subjects

	Control	Exp.	Diff
S1	74	79	5
S2	90	97	7
S3	87	92	5
S4	94	98	4
			$\bar{X}$ =

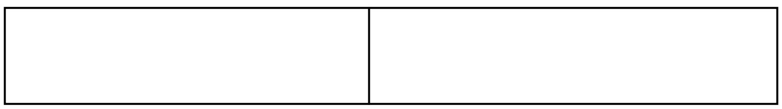
DESIGNS WITH TWO OR MORE INDEPENDENT VARIABLES

1) Conduct separate experiments for each independent variable
   - OR -
2) Experiment with more than one independent variable. Can get more information at less expense by using a design that incorporates two or more independent variables.
   - some time avoid confounding two variables.

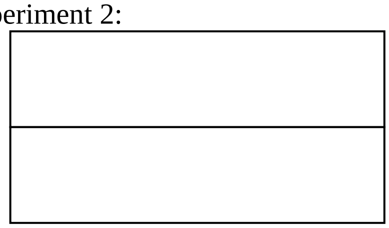
Example

Two Simple Experiments:

Experiment 1:



Experiment 2:



Three of the four cells of a 2x2 Experiment

A. FACTORIAL DESIGNS:

* Include a separate group for each possible combination of the levels of the independent variable.

1) MAIN EFFECTS: Separate effects of each independent variable. Compare column means or row means.
- A reliable difference in column indicates effect independent of other variable.

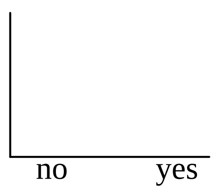
2) INTERACTIONS:
- Effect of one independent variable changes as the levels of another independent variable
- In general if lines are not parallel an interaction exists (take into account random variability)

Study on the effects of Expectancies on Memory
(2x2 between subjects)
Double Blind test of subliminal self-help audiotapes
(Greenwald, et al.)
Independent Variable:
Independent Variable:
Dependent Variable:

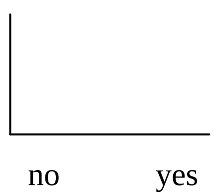
	Actual Tape: S-E	Actual Tape: Mem	
Label S-E	.20	.14	.17
Label Mem	.46	.45	.46
	.33	.30	

It Depends (looking at interactions)

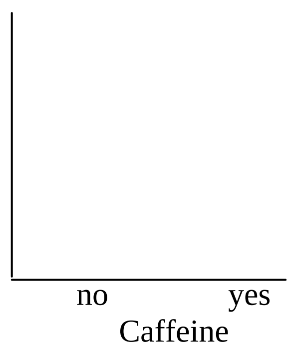
Nicotine



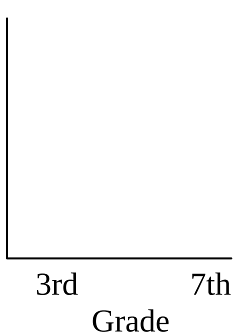
Caffeine

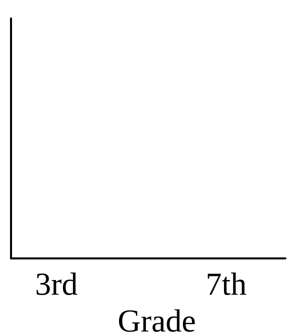


Nicotine & Caffeine



Nicotine:
no nicotine
nicotine





Subtraction Method for identifying interactions

Glass, D.C., & Singer, J.E. (1972) . Urban stress: Experiments on noise and social stressors

- Need to state a more complex rule about the effects of noise

- Some qualifiers: "depending on," "but only if," or "however, that holds only under certain conditions."

3) FACTORIAL WITHIN-SUBJECTS DESIGNS:
Each subject is exposed to every combination of levels of all the factors (independent variables).

4) HIGHER ORDER FACTORIAL DESIGNS:
- Can include any number of levels of a given factor
- In general you want a minimum of 5 subjects per group for a reasonable ability to detect the effect of the independent variable.

Issue Subjects (minimum):

(2 x 2)	4 x 5 =	20 subjects
(2 x 2 x 2)	8 x 5 =	40 subjects
(3 x 3 x 3)	27 x 5 =	135 subjects

Number and Complexity of interactions with three factors:
- Main Effects: A, B, C
- Three TWO-WAY Interactions: (A & B; B & C; A & C)
- Three-Way Interaction: (Usually limit to a three factors)

B. OTHER GROUP BASED DESIGNS:
- Choose designs that best address the question.
- Some designs can not do combinations of every level of independent variables (e.g., certain drug combinations)

DESIGNS WITH TWO OR MORE DEPENDENT VARIABLES

A. MULTIVARIATE EXPERIMENTAL DESIGNS:

Univariate: single dependent variable

Multivariate: multiple dependent variables (besides experimental can also be correlational)

Example:

Independent Variables
1)
2)
3)
4)

Dependent Variables:
1)
2)
3)

B. ADVANTAGES AND DISADVANTAGES OF MULTIVARIATE DESIGNS:

<u>Advantages</u>:
1. Economize subjects' time and effort with single experiment
2. Increase chances that at least one of your dependent variables will be sensitive to effects of you independent variable.
3. Can construct a composite dependent variable by combining values according to a mathematical formula.

<u>Disadvantages</u>:
1. Value of dependent variables tend to be correlated;
2. Statistics maybe more complicated and difficult to interpret.

CONFOUNDING AND EXPERIMENTAL DESIGN

Confounding Variable: one that varies along with your independent variable (affects internal validity of experiment; may not establish causal relationship).

Random Assignment: of subject to condition

Experimenter Bias: assignment of subjects to certain groups based on probable performance.

Counterbalance:

► Chapter 14: USING INFERENTIAL STATISTICS ◄

INFERENTIAL STATISTICS: you infer the characteristics of a population from the characteristics of the samples comprising your data.

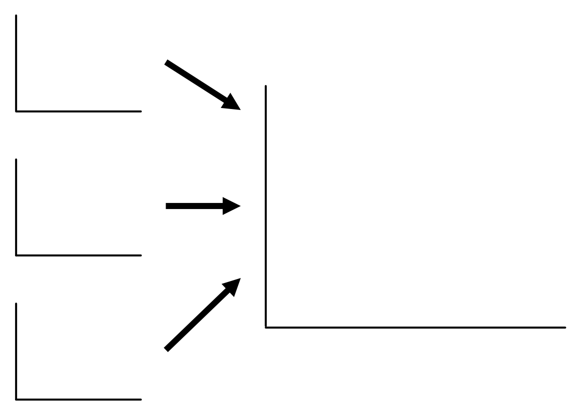
INFERENTIAL STATISTICS: BASIC CONCEPTS

A. SAMPLING DISTRIBUTION: using the sampling distribution for each statistics one can determine the PROBABILITY that a value of a statistic as large or larger than the obtained value could have occurred by chance (called p).

Sampling Distribution of the Mean: Distribution of means

1. Tends to follow normal distribution
2. Mean of sampling distribution is same as mean of population from which samples were drawn from.

Central Limit Theorem: Distribution of means will tend to be normal even if distribution of population scores deviates widely from normality.



B. SAMPLING ERROR: Mean of sample will probably differ from population mean.

Standard Error of the Mean: an estimate of the amount of variability in the expected sample means across a number of such sample $\sigma_{\bar{x}}$

C. DEGREES OF FREEDOM (df): for single sample n-1. Sample of 10; 9 scores are free to vary,

THE LOGIC BEHIND INFERENTIAL STATISTICS

Experiment: different groups (different levels) of your independent variable

- Each mean is assumed to represent the mean of an underlying population.

If we took TWO means from the same population:

We would expect them to differ only because of sampling error.

- Each sample provides independent estimate of population mean.
- Hypothesis stated in terms of the effects of the independent on the dependent variable.

NULL HYPOTHESIS (Ho): means were drawn from same population ($\mu 1 = \mu 2$) [no difference]

- Assumed to be true
- Null is the DEFAULT

REJECT Ho: If evidence show assumption not well founded we Reject Ho and accept alternative.

FAIL TO REJECT Ho: If sufficient evidence to reject null does not exist simply report not enough to reject null and alternative not supported.

ALTERNATIVE HYPOTHESIS: means were drawn from different populations ($\mu 1$ not equal to $\mu 2$)

TWO FORMS:

(1) Directional: specifies an expected direction of difference

(2) Non-directional:

EXAMPLE:

NULL HYPOTHESIS: among teenage males, the mean level of self-esteem is the same for those who have failed at an experimental task as for those who have succeeded.

ALTERNATE HYPOTHESIS: among teenage males, the mean level of self-esteem is lower for those who have failed at an experimental task than the mean level for those who have succeeded.

Two Different Populations means: due to sampling error, the two means might or might not differ, even though a difference exits between the underlying population means.

PRINCIPLE OF FALSIFICATION (Popper, 1963): the process of showing an argument to be false – was absolutely central to understanding how science works.

- One can never prove that a theory is finally and completely correct.
- Quite definitely possible to falsify it.

A. STATISTICAL ERRORS: When comparing two sample means there are two possible outcomes (Ho is true or false) and two decisions (reject or not).

Type I Error (alpha error; false positive, false alarm): independent variable had no effect but you think it did.
- Rejected Ho and say independent variable had an effect.

Type II Error (beta; miss; false negative): independent variable did have an affect but you incorrectly decide not to reject Ho.

Correct Decision: Reject Ho when it is false

Correct decision: Do not reject Ho when Ho is true

True State of Affairs

Decision		Ho True	Ho False
	Reject Ho		
	Do not Reject Ho		

B. STATISTICAL SIGNIFICANCE: used to determine the probability that your observed means come from one or 2 underlying populations.
- Difference is so great that it is unlikely, as some level of probability, to have occurred by chance.

Alpha Level: criterion used to accept or reject the null hypothesis (Type I Error)

Significance Level .05: 5 chances in 100 that sampling error could have produced a difference at least as large as the one observed.

Observed Value: base on sample data

Critical Value: typically value in table
depends on
1. Number of observations per treatment
2. Number of treatments
3. Desired alpha level
4. One tail vs. two tail.

Statistically Significant: exceeds critical level. If the obtained p is less than or equal to alpha, your comparison is statistically significant.

C. ONE-TAILED VERSUS TWO-TAILED TEST:
Critical Region: determined by alpha level
ONE TAIL t TEST: Research assumes a difference & in a particular direction!
- Calculation of t test is the same; only the critical t-value changes
- Specify before experiment

TWO-TAIL t TEST: looking for a difference in either direction.
$\mu 1$ not equal $\mu 2$ (alternative hypothesis)

PARAMETRIC STATISTICS: estimates the value of a population parameter from the characteristics of a sample.
- Use depends on:

A. ASSUMPTIONS: violations may lead to bias (Type I Error more or less)
1. Scores randomly sampled from population
2. Sampling distribution of mean is normal
3. With-in group variances are homogeneous (variance should be similar)

B. INFERENTIAL STATISTICS WITH TWO SAMPLES: Is the observed difference reliable or due to sampling error?

C. THE t TEST: used when you have TWO levels of an independent variable.
1. Independent Samples: two groups - subjects randomly assigned
   Error term:
   - Unpooled:
   - Pooled: error term based on 2 samples

2. T-Test for Correlated Samples: when two samples are not independent. Scores come in pairs

D. ANALYSIS OF VARIANCE (ANOVA): beyond two groups - includes more than 2 groups.

Partition Variation for between subjects:

Any score depends on (sources of variability)

1. Characteristics of subjects at time score was measured
2. Measurement or recording errors (experimental error)
3. Value of independent variable

Two sources of variability:

- Between Groups
- Within Groups

Within Group Variability attributed to: Error, Individual differences, experimental error

The F Ratio: both types of variance constituting the ratio are expressed as variance called MEAN SQUARE.

Difference due to treatment - effect is unique to between-group variables.

$$F = \frac{\text{between group variability}}{\text{within group variability}}$$

IF NULL HYPOTHESIS IS TRUE

$$F = \frac{\text{random error}}{\text{random error}}$$

IF NULL HYPOTHESIS IF FALSE

$$F = \frac{\text{random error + treatment effect}}{\text{random error}}$$

F. ONE-FACTOR BETWEEN-SUBJECTS ANOVA: One factor (with more than 2 levels) and different subjects in each experimental condition.

1. p value: actual probability of making a Type I error.

2. Interpreting Your F Ratio: Significant F ratio tells you that at least some of the differences among means are probably not caused by chance.

3. Planned Comparison: based on pre-experimental hypotheses.
   - Separate Fs or ts are calculated

4. Unplanned Comparison: looking for differences that might exist.

Per-comparison Error: alpha for each comparison between means. If set alpha to .05, then per-comparison is equal to .05.

Familywise Error: "takes into account" the increasing probability of making at least one Type I Error as the number of comparisons increase

$\alpha_{fw} = 1-(1-\alpha)^c$ [c=# of comparisons]

5. Sample Size: adjusted computational formulas are used when groups contain unequal number of subjects.

6. Unweighted Means Analysis: Instead of discarding to make even; use all subjects - each group receives equal weight in analysis.

7. Weighted Means Analysis: Groups weighted to number of subjects in each group.

G. ONE-FACTOR WITHIN-SUBJECTS ANOVA: Treat subjects as factor; contribution of individual differences taken out.
- Reduces amount of error in denominator make F more sensitive to independent variable

Well Rested	Awake 24 hours	Awake 48 hours

H. THE TWO-FACTOR BETWEEN-SUBJECTS ANOVA: Two independent variables; subject are randomly assigned to the conditions
- Determine significance of both main effects and interaction

Main Effects and Interactions: If interaction is significant need to be careful about interpreting main effects (interpret interaction first)

Sample Size: Can complete ANOVA with unequal sample sizes

		Type of Audience	
		Alone	Audience
Self-Esteem	High		
	Low		

I. THE TWO-FACTOR WITHIN-SUBJECTS ANOVA:

Exposed to every possible combination of level of two independent variables

		Fatigue		
		Well Rested	Awake 24 hrs.	Awake 48 hrs.
Type of environment	Noisy			
	Quiet			

J. MIXED DESIGNS: mixing between-subjects and within-subjects components; issue of error term to calculate F ratio.

K. HIGHER ORDER AND SPECIAL CASE ANOVAs:

"ANOVA involves forming a ratio between the variance (mean square) caused by your independent variable and the variance (mean square) caused by experimental error."

NONPARAMETRIC STATISTICS: makes no assumptions about the distribution of scores underlying your sample.

- Use when assumptions of a parametric test are not met.

A. CHI-SQUARE: Significant Chi-Square tells you that your two variables are significantly related.

Chi-Square for Contingency Tables:
- For frequency data
- Compares observed cell frequencies with expected cell frequencies.

Limitations of Chi-Square:
- Problem if expected cell frequencies is less than five. Chi-Square may be artificially inflated.

B. The MANN-WHITNEY U TEST:
- Used when data is scaled on at least ordinal scale.
- Good alternative to t-test when assumptions are violated.

C. PARAMETRIC VERSUS NONPARAMETRIC STATISTICS:

- Appropriate versions are not always available for more complex designs.

SPECIAL TOPICS IN INFERENTIAL STATISTICS

A. POWER OF A STATISTICAL TEST:

1. **Power**: the ability of a test to detect difference in data that are inconsistent with the null hypothesis.

AFFECTED BY THE FOLLOWING:

a) Alpha Level:
   - Reduces alpha level from .05 to .01 reduces the probability of making a type 1 error -> also reduces power.

b) Sample Size: power increases with increase in sample size

c) One-Tailed versus Two-tailed Test:
   - 2-tailed less powerful than one-tail
   - Easier to reject Ho with one-tail

   ONE-TAILED:
   - Specifies an expected direction of difference.
   - Computed value is compared to values at one end of distribution.

   TWO-TAILED (non-directional):
   - The computed value of the test statistic is compared with the values in the regions of rejection at both ends (tails).

d) Effect Size: “The degree to which the manipulation of your independent variable changes the value of the dependent variable.”

2. **Determining Power:** to much power can be bad Conceivably, one could find statistical significance in the most minute or trivial of difference.

B. STATISTICAL VERSUS PRACTICAL SIGNIFICANCE: "To say results are significant (statistically speaking) merely indicates that the observed differences between means are probably reliable, not the results of chance."

Practical Significance:

C. THE MEANING OF THE LEVEL OF SIGNIFICANCE:
Alpha Level: provides a criterion for deciding whether the differences you have obtained are reliable.
Applied Research: may be better evaluated at a less conservative alpha level

D. DATA TRANSFORMATION: Converting original data to a new scale
- Adding or subtracting a constant (adding constant to remove negative numbers)

E. ALTERNATIVES TO INFERENTIAL STATISTICS:
Inferential statistics: tool to help you make a decision about Ho provides way to test reliability of a single experiment.

Replication: one way of testing reliability
Repeat the experiment.
- If reliable, should find similar results.
- Small or 1 subject experiment.

$p < .05$ 1 in 20 difference that occurred by chance

THE NULL HYPOTHESIS AS MODUS TOLLENS

(reasoner denies the consequent)
When we reject the null hypothesis we are saying that the two means are not equal and that they probably were sampled from two different populations.

Ho: $\mu1 = \mu2$

CONDITIONAL PROPOSITION:
IF two means come from the same distribution
THEN they are statistically equal

EXISTING CONDITION:
They are not statistically equal

INFERENCE:
Therefore they did not come from the same distribution