Advanced Quantitative Data Analysis

May 31, 2016
Presenter: Aynslie Hinds
umhinds0@myumanitoba.ca

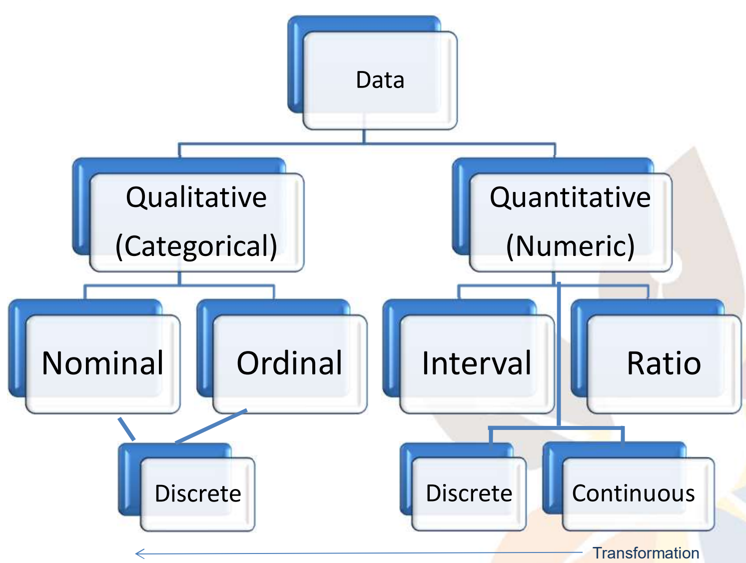
Main Objectives

- To learn about basic inferential analyses and when to perform them
- To consider the value and limitations of various quantitative methods
- To understand how to interpret output

Steps Involving Data

- (Design and test data collection instruments)
- (Collect the data)
- (Data entry)
- Clean/refine the data
- Analyze the data
- Interpret the results

Types of Data



6

Scales of Measurement

Scale	Measurement Scale	Example
Nominal	Do students study? categories = yes and no	Melissa – yes Jeff – Yes Gina - No
Ordinal	Who studies more? Scale adds "more than"	Melissa > Jeff> Gina
Interval	Who studies more? Students rate their studying on a 5-point scale	Melissa – 5 (always) Jeff – 3 (sometimes) Gina – 1 (never)
Ratio	Who studies more? Measure the amount time spent studying	Melissa – 12 hours/day Jeff – 2 hours/day Gina – 0 hours/day

Scale of Measurement

- Understanding the type of data is key to knowing
  - How to create a data file
  - The correct method for analyzing data and presenting the results

Statistics

- Statistical investigation and analyses of data fall into two broad categories:
  - Descriptive statistics
  - Inferential statistics

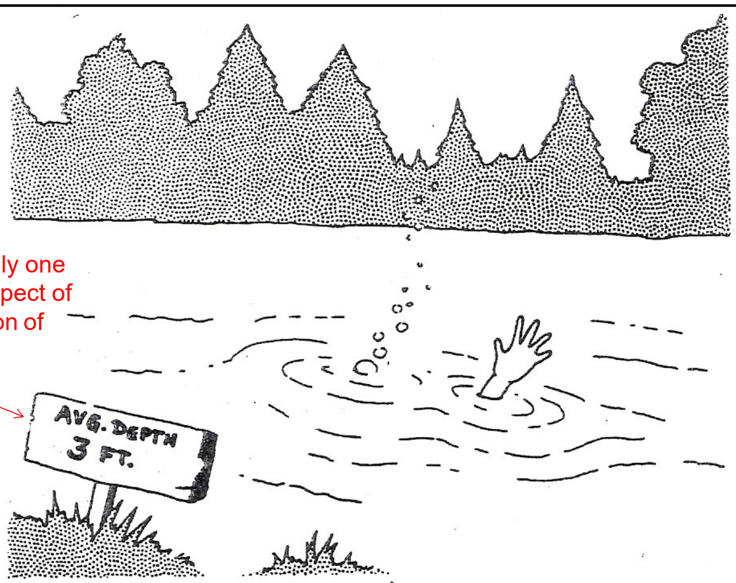
Descriptive Statistics

- Methods for summarizing and presenting data
- May involve:
  - Presenting data in graphical or tabular form
  - Calculating summary statistical measures
- E.g., bar charts, histograms, scatterplots, averages, variance

Typical Value

- Measures of Central Tendency
  - Mean
  - Median
  - Mode

Describes only one important aspect of the distribution of the data



Does the average (mean) tell us all we need to know about a distribution for a set of measurements?

Need to consider the amount of variation or scatter

Example

Data Set 1	Data Set 2
56	30
58	45
60	60
60	60
60	60
66	105

Mean = Median = Mode = 60

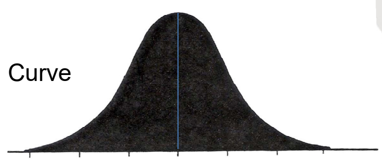
Measures of Dispersion

- Range
- Variance
- Standard deviation

Normal Distribution

- Widely observed in natural and behavioral sciences
- Description:
  - Most results are close to the mean (typical)
  - Few results are atypical
  - The more atypical a result, the less frequent it occurs

Normal Curve

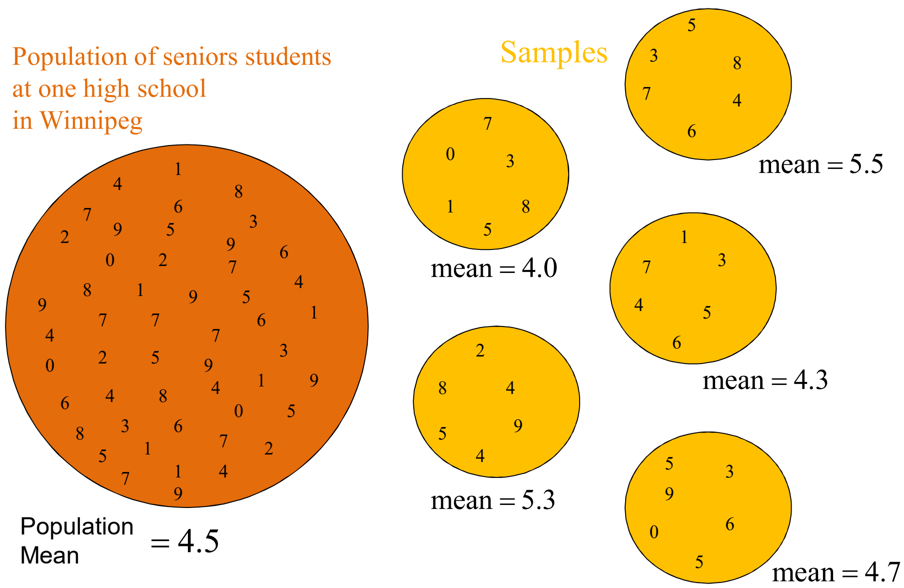


Normal Distribution

- Many statistical tests are based on the assumption of normality
- Parametric VS Non-Parametric

Sampling variability with repeated sampling

Number of cigarettes smoked yesterday



Estimates vs Parameters

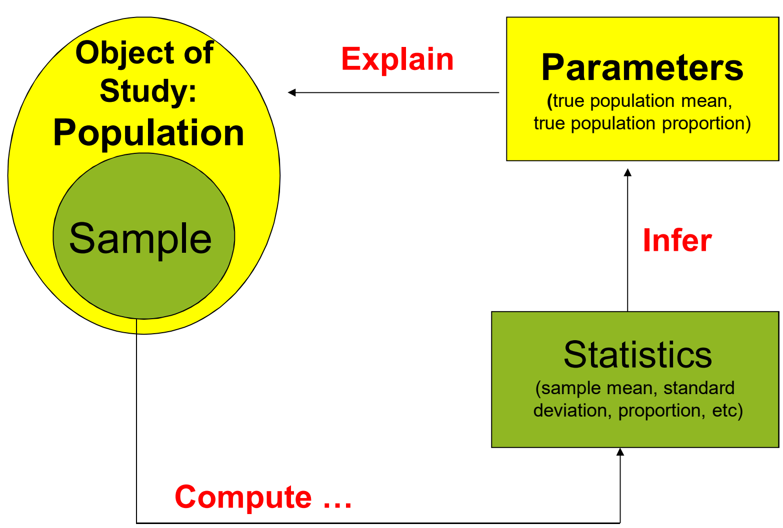
Sample Statistics	Population Parameters
Mean: $\overline{x}$	Mean: μ
Variance: s^2	Variance: σ^2
Standard Deviation: s	Standard Deviation: σ
Proportion: p	Proportion: ρ

- We know or observe $\overline{x}$ from a sample, but we don't know or observe μ
  - Can observe a sample statistic and use it as an estimate of the true, unknown population parameter

Inferential Statistics

- Involves:
  - Using sample information to draw inferences or test hypotheses about a characteristic of a population
  - Making inductive generalizations from the particular (the sample) to the general (the population)
  - Hypothesis Testing & Estimation

Inferential Analysis



Estimation: Asking and Answering Questions

- What is the true proportion of pregnant women who will quit smoking if they undergo a smoking cessation program?
- What is the true mean change in self-esteem scores of individual participating in a skill-based employment training program between pre and post program?
- What is the true mean change in perceptions of safety among community members pre and post program (e.g., improved street lightening, graffiti removal, etc.)?

Estimation

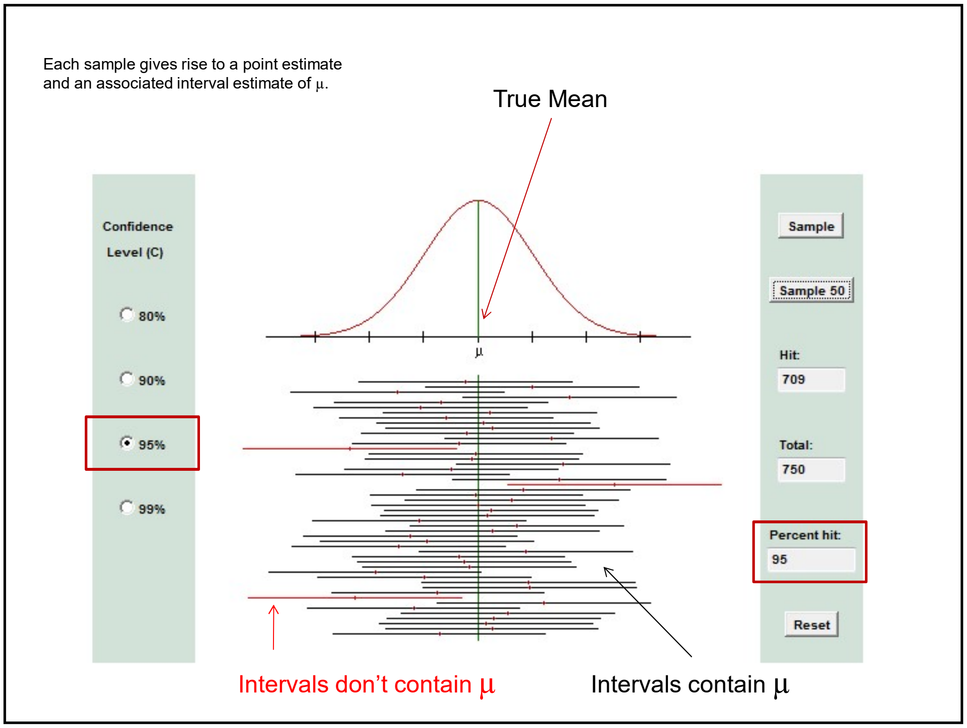
- Process of calculating some statistic that is offered as an approximation (a "guess") to an unknown population parameter from which the sample was drawn
- Two methods for providing an estimate of a parameter…
  - Point estimate
  - Interval estimation (i.e., confidence interval)

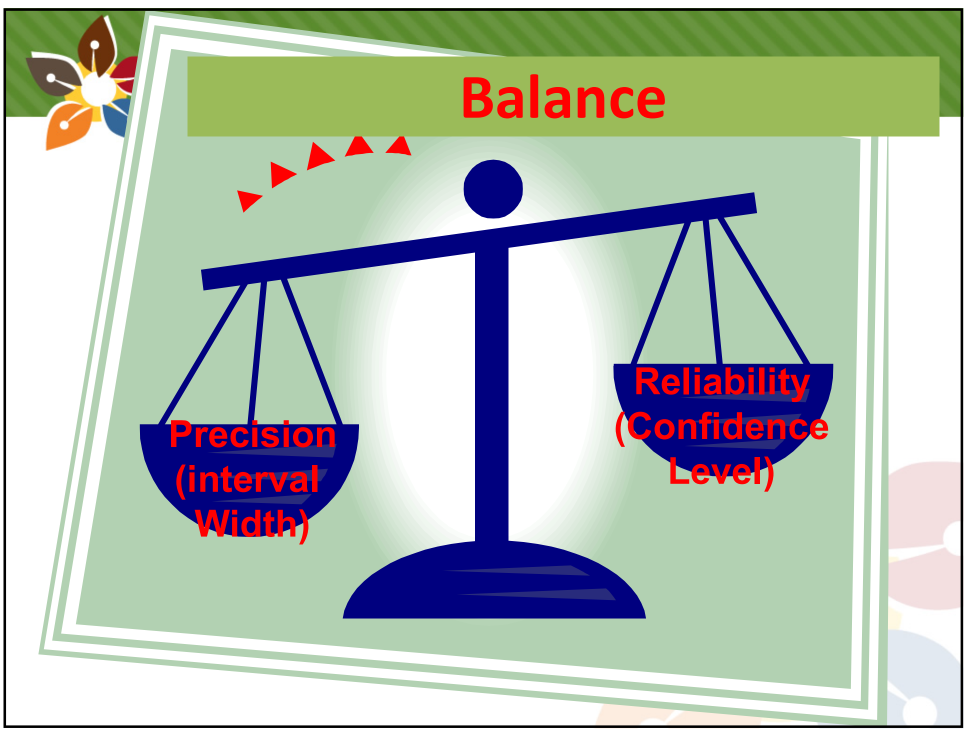
Interval Estimation/ Confidence Interval

- Range of values (interval) that is believed to contain the parameter of interest together with a certain degree of confidence (probabilistic statement) in the assertion that the interval does contain the parameter
- Levels of confidence:
  - 90%, **95%**, 98%, 99%

Confidence Level

- Describes the chance or probability that intervals of this kind "capture" the population value *in the long run*


Each sample gives rise to a point estimate
and an associated interval estimate of μ.
True Mean
Confidence
Level (C)
80%
90%
95%
99%
μ
Sample
Sample 50
Hit:
709
Total:
750
Percent hit:
95
Reset
Intervals don't contain μ
Intervals contain μ


Balance
Precision
(interval
Width)
Reliability
(Confidence
Level)

Hypothesis Testing: Asking and Answering Questions

- Did counseling reduce smoking rates during pregnancy?
- Did the school-based “Just Say No” campaign reduce drug use?
- Did participants’ self-esteem increase as a result of participating in a skills-based employment training program?
- Did having a safety outreach worker in the community increase community members sense of safety?

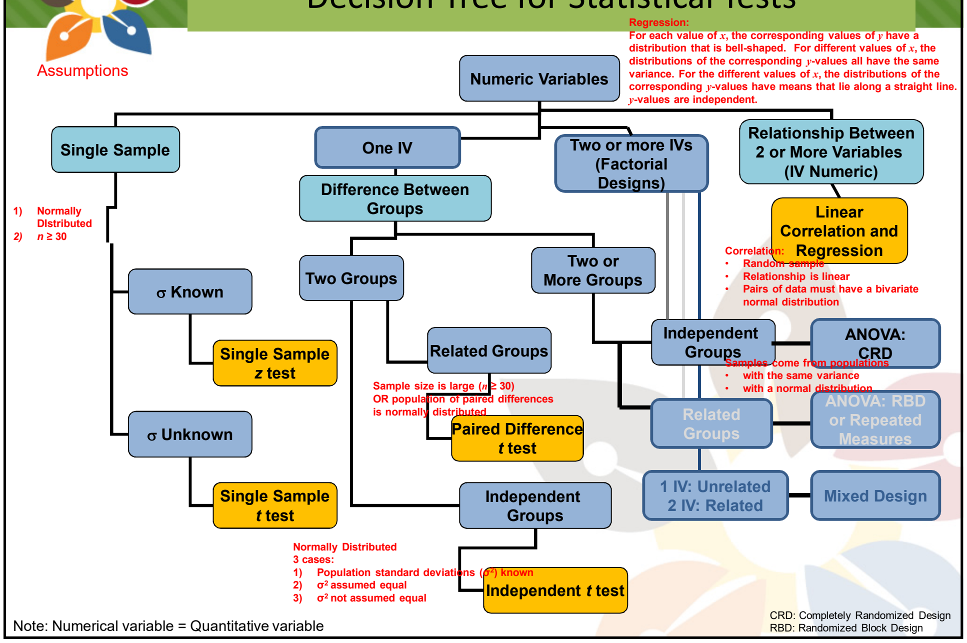
Statistical Tests

- Lots of different statistical tests
- Challenge to know which one to use
- Parametric VS Non-Parametric

Terminology

- IV = independent variable
- DV = dependent variable
- EV = extraneous variables
- True experimental design requires manipulation of IV and control over EVs
- If lack control or have other threats to internal validity, cannot unequivocally establish cause and effect relationships → design is then non-experimental or quasi-experimental

Decision Tree for Statistical Tests



Equivalent Tests

Parametric	Non-Parametric
Paired-difference t-test	Wilcoxon Signed Ranks test
Independent t-test	Wilcoxon Rank-Sum test Mann-Whitney U-test
One-way ANOVA	Kruskal-Wallis test
Linear correlation	Rank correlation

Non-Parametric Tests

Advantages	Disadvantages
Can be applied to wide variety of situations	Waste info because exact numerical data are reduced to qualitative form
Can be applied to categorical data	Not as efficient (as parametric) and therefore need stronger evidence to reject H_0
Usually involves simpler computations (easier to understand and apply)	If outliers aren't errors (result of contaminating factors) then may result in underestimating effect of contaminating factors
Effect of outlier is much less (than parametric)	

Null Hypothesis

- Assumption of no effect
- Hypothesis that is tested

- Examples
  - Counseling had no impact on the proportion of women who smoked during pregnancy
  - School-based "Just Say No" campaign had no effect on drug use among students
  - Skill-based employment program did not increase participants' self-esteem
  - A safety outreach worker had no effect on community members' sense of safety

Alternative Hypothesis

- Complement of the null hypothesis
- Is generally the hypothesis that is believed to be true
- Directional or non-directional

- Examples
  - Counseling had decreased the proportion of women who smoked during pregnancy
  - School-based "Just Say No" campaign had an effect on drug use among students
  - Skill-based employment program did increase participants' self-esteem
  - Having a safety outreach worker increased community members' sense of safety

Basic Idea of Hypothesis Testing

- Assume the null hypothesis is true
- If, under a given assumption, the probability of a particular observed outcome is extremely small, we conclude that the initial assumption probably is not correct

But wait…

- What probability is considered 'extremely small'?
- Usually, 0.05, sometimes 0.01 or 0.10

p-value

- "p" = probability
- Reject the null hypothesis if p-value ≤ **0.05** **(the result is significant)**

  0.01 or 0.10
- The probability of seeing this result (or one that's more extreme), by chance if the initial assumption is correct, is less than or equal to **5**%

p-value > 0.05

- p-value = 0.15, p-value = 0.45, etc.
- No evidence against the null hypothesis (initial assumption), therefore stick with it
- Not a statistically significant result

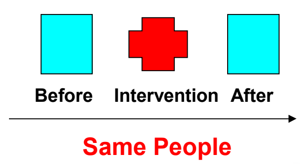
p-value Produced by Statistical Programs

- Statistical programs *usually* provide the p-value for a two-tailed/non-directional alternative hypothesis
- NOTE: If the results are in the direction (+/-) expected, divide the p-value by 2

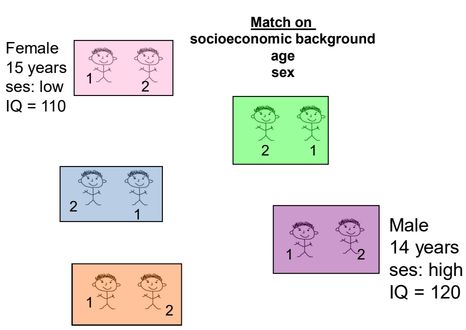
Two Related Groups

1. Pre-Post Study

   Repeated Measures Design



2. Matched on Relevant Characteristics



Test Statistic = Paired Difference t-test

Is there a statistically significant difference in the means between the two conditions?

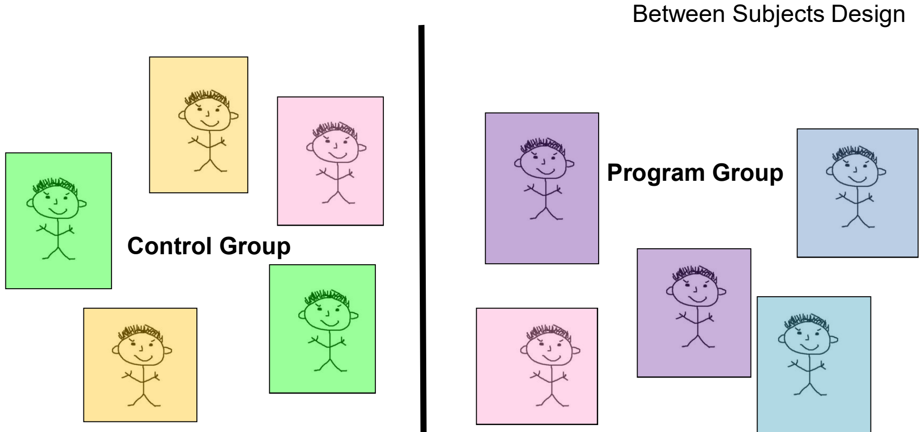
Within-Subjects Design

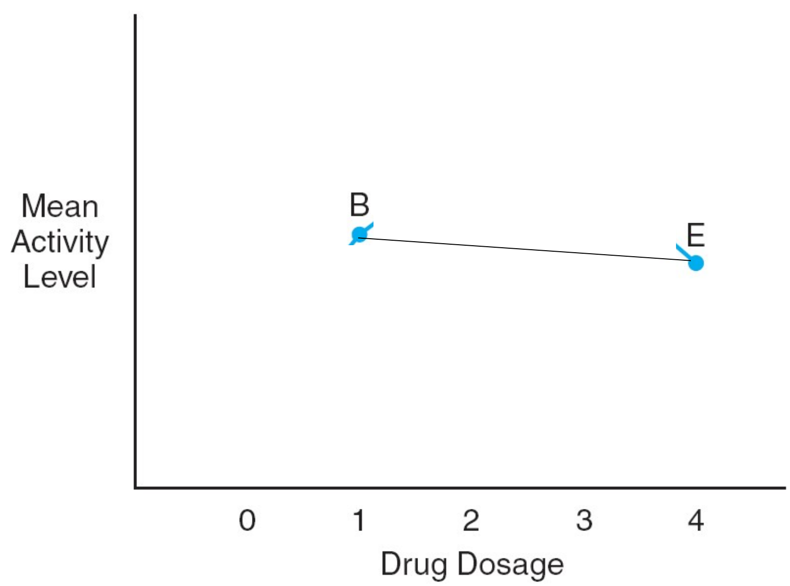
Two Unrelated Groups

Test Statistic = Independent t-test

Is there a statistically significant difference in the means between the two conditions?

Between Subjects Design





Two or More Groups/Conditions

- Analysis of Variance (ANOVA)
  - Extension of an independent t-test (CRD)
  - Extension of a paired-difference t-test (RBD, repeated measures)
  - One-way ANOVA
  - Two-way ANOVA (Factorial Design)
  - Follow-Up Analyses
    - Multiple Comparison Procedures
    - Contrasts

Example

Independent Variable

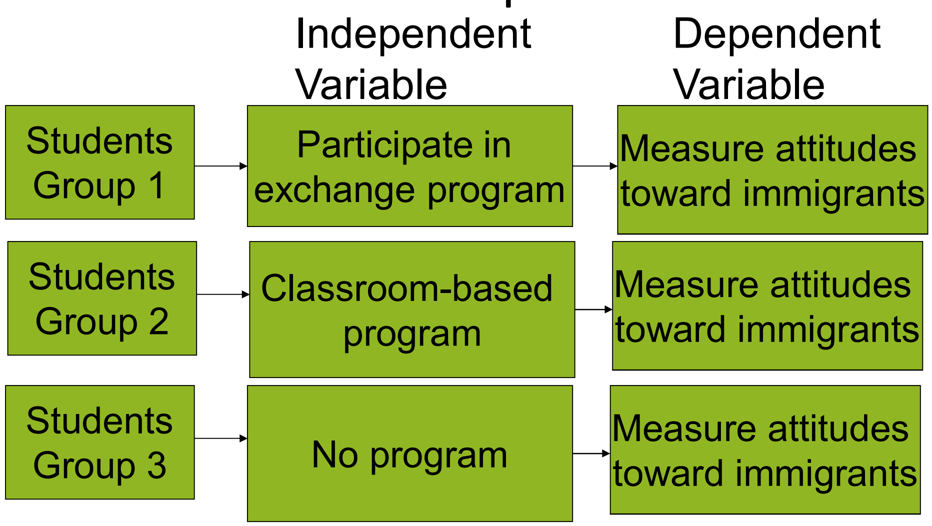
Dependent Variable

Students Group 1 → Participate in exchange program → Measure attitudes toward immigrants

Students Group 2 → No program → Measure attitudes toward immigrants

Posttest-Only Non-Equivalent Control Group Design
Analysis: Independent t-test

Example



Posttest-Only Non-Equivalent Control Group Design
Analysis: Single-factor ANOVA

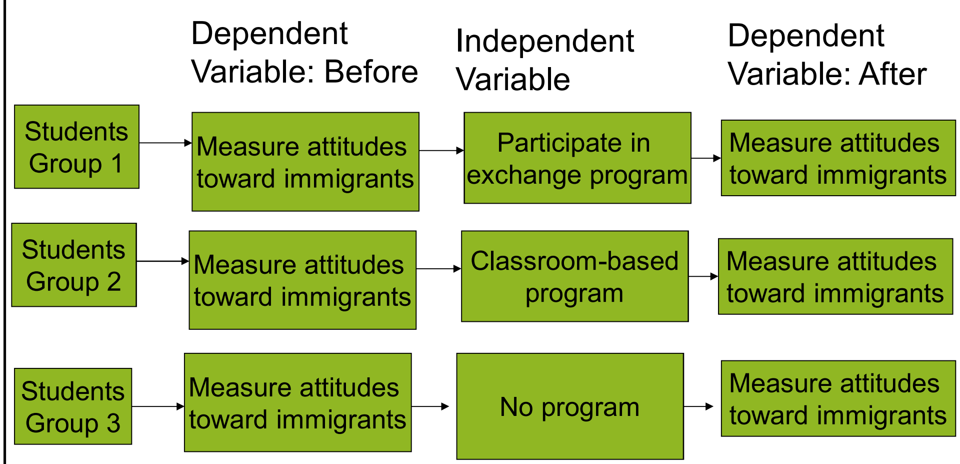
ANOVA

- *F*-test – tells if significant group differences exist but not which group(s) are significantly different from the others
- Planned comparisons, a priori comparison, contrasts
  - Prediction made before the data is collected about which groups differ and in what direction based on theory
- Post hoc comparison, a posteriori comparison
  - *F* is significant
  - If don't have an a priori prediction
    - Bonferroni, Tukey, SNK, LSD, etc.
  - Built in procedures for dealing with problem of Type I Error

Factorial Designs

- Manipulation of 2 or more IVs
  - Could also have quasi-IVs (e.g., age, gender)
- Able to study the individual and interactive effects of the IVs on the DV

Example



Pretest-Posttest Non-Equivalent Control Group Design
Factorial Design (Mixed)

Main Effect

- Mean differences among the levels of one factor
  - Differences between the column means define the main effect for one factor
  - Differences between the row means define the main effect for the second factor
- Reflect the results that would be obtained if each factor was examined in its own experiment (used to describe the overall effect of the IV)

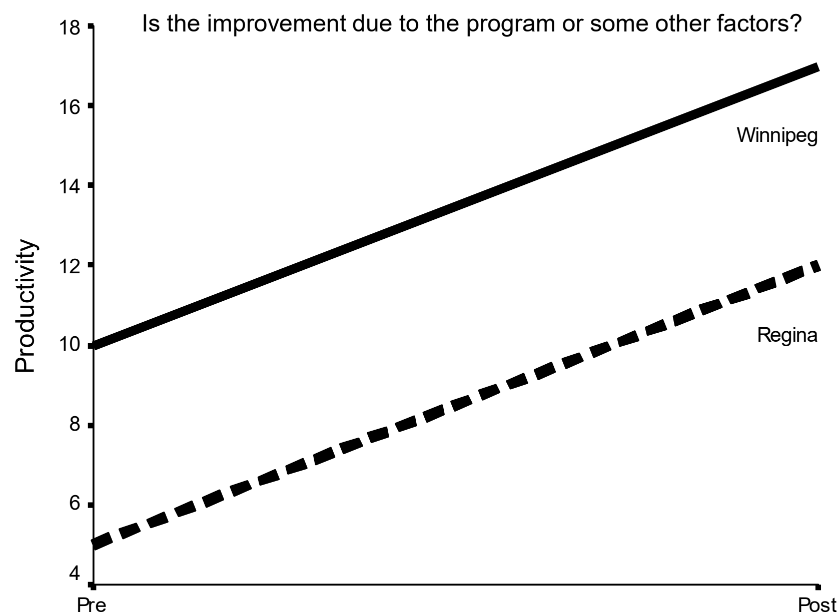
Interaction

- Exists when one factor (IV) depends on the different levels of the second factor (IV)
- Effects of a factor vary depending on levels of the other factor
- Graphical representation
  - Parallel lines = no interaction
  - Non-parallel lines = interaction

Example

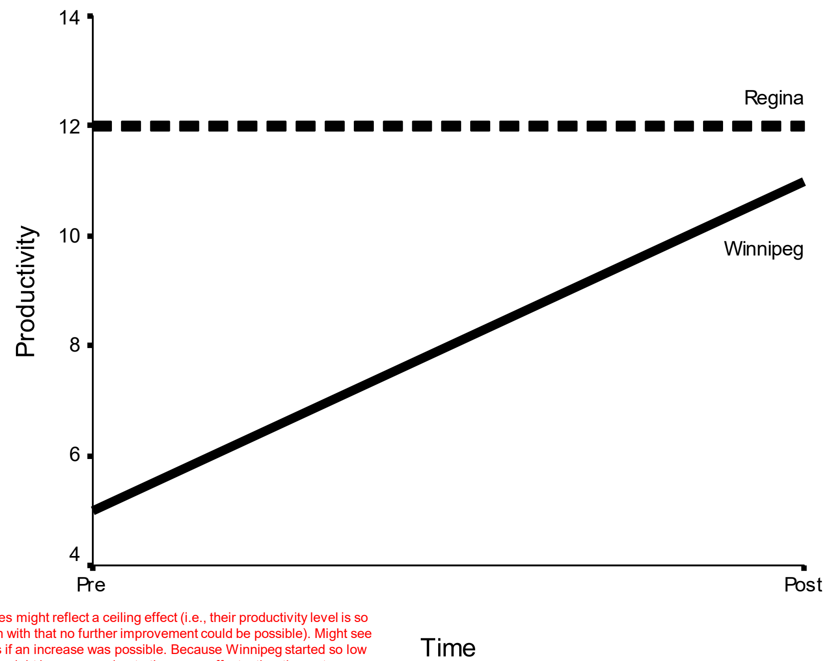
	Pretest	Treatment	Posttest
Winnipeg plant	Average productivity for 1 month prior to instituting flextime	Flextime instituted for 6 months	Average productivity during 6th month of flextime
Regina plant	Average productivity for 1 month prior to instituting flextime in Winnipeg	None	Average productivity during 6th month that flextime is in effect in Winnipeg

Pretest-Posttest Non-Equivalent Control Group Design
Factorial Design (Mixed)

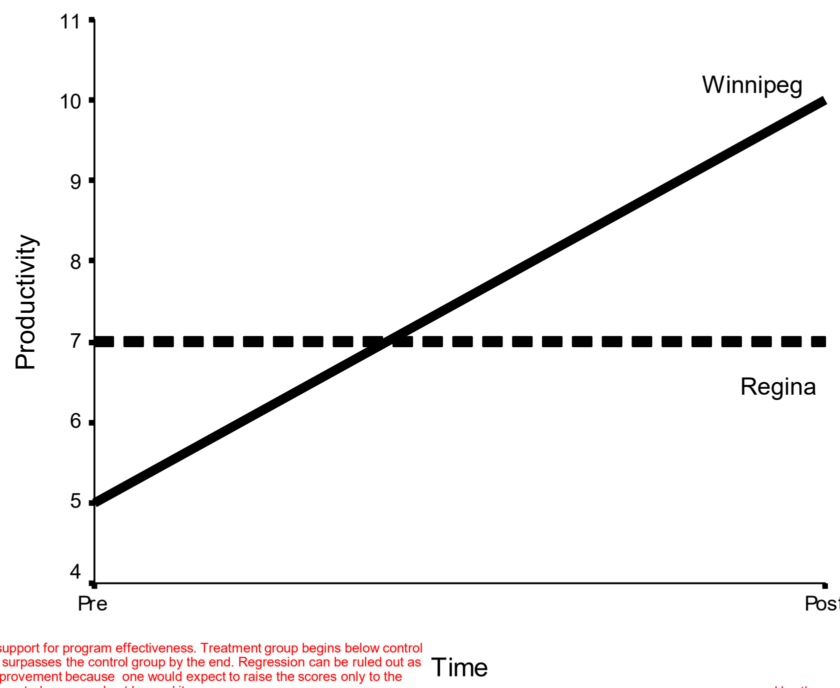


Something other than flextime produced the improvement (e.g., history, maturation) because both plants increased productivity. Example explanations: National election/Olympic victories/Canadian hockey team wins championship between pre and post tests that workers everywhere felt more optimistic leading to increased productivity or improvement due to increased experience.

Time



Regina scores might reflect a ceiling effect (i.e., their productivity level is so high to begin with that no further improvement could be possible). Might see parallel lines if an increase was possible. Because Winnipeg started so low the increase might be a regression to the mean effect rather than a true one.



Strongest support for program effectiveness. Treatment group begins below control group , but surpasses the control group by the end. Regression can be ruled out as causing improvement because one would expect to raise the scores only to the level of the control group and not beyond it.

Hawthorne Effect

Time-Series & Interrupted Time-Series Designs

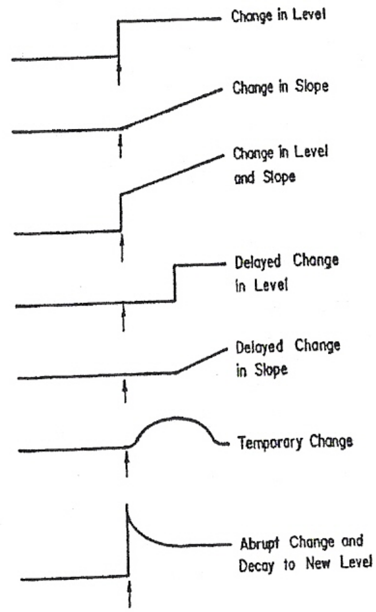
- Involve repeated measurements or observations for each participant over time (before and after treatment or event)

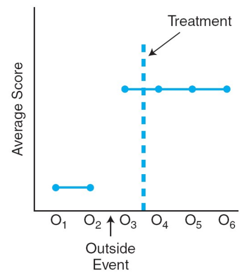
  O O O O O X O O O O O

O = observation
X = treatment or event (the 'interruption' in interrupted time series)

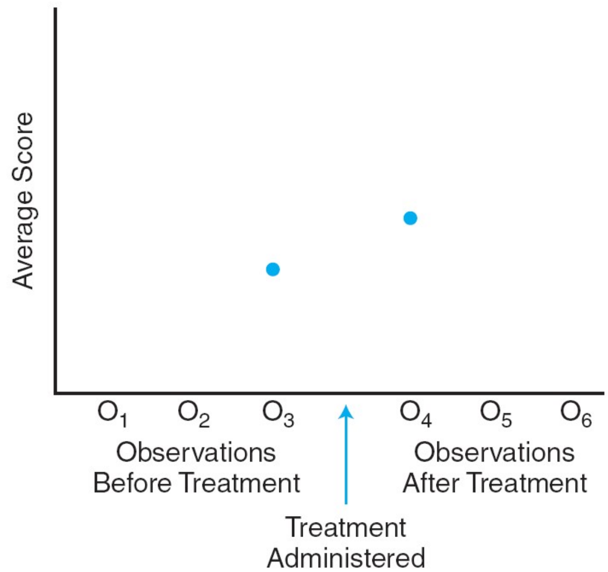
Time-Series & Interrupted Time-Series Designs

- Pre-observations allow for trends to be observed before treatment introduced
  - Trends are indications that scores are influenced by some factor unrelated to treatment
    - Practice, fatigue, instrumentation effects, maturation effects, regression
  - If no trend or fluctuations before treatment then more confident that potential threats to internal validity are not influencing participants
  - Series of observations allow for determination of threats to internal validity



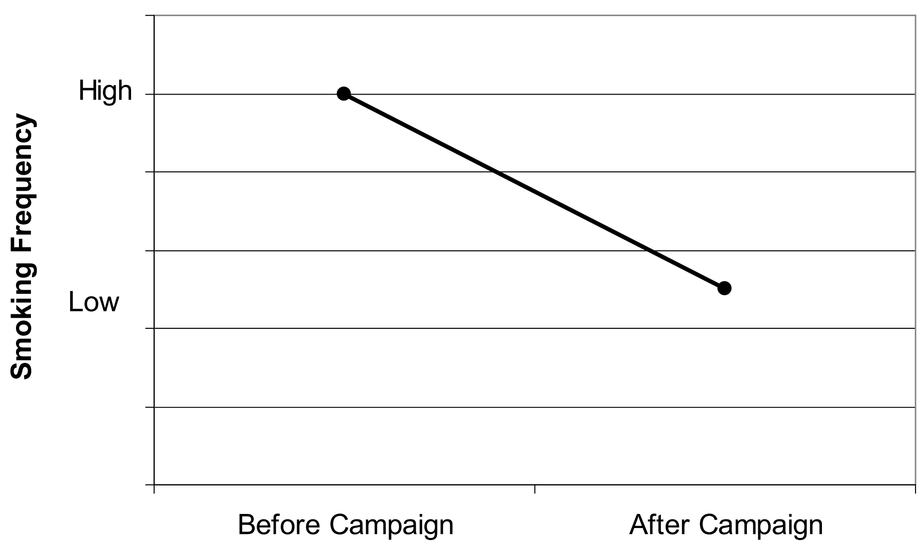


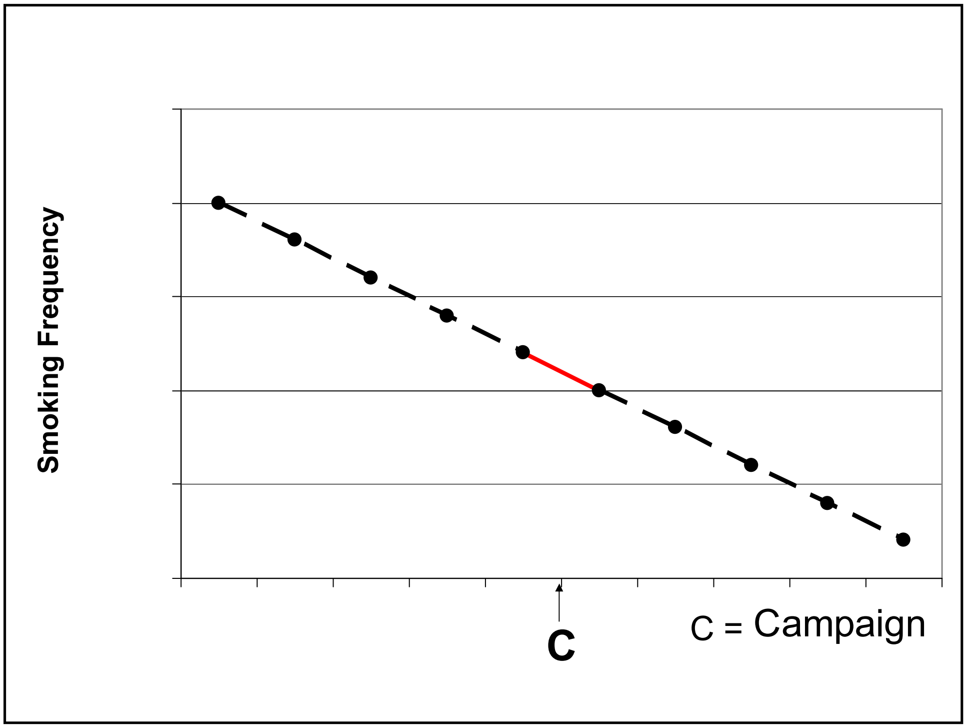
Can't distinguish treatment effect from outside event

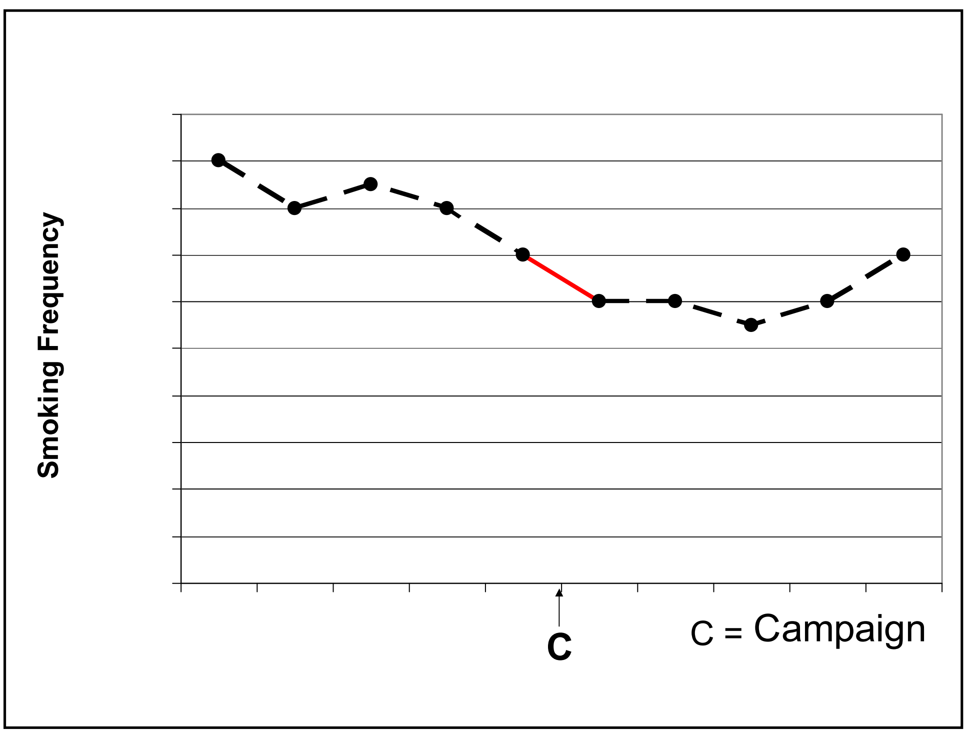


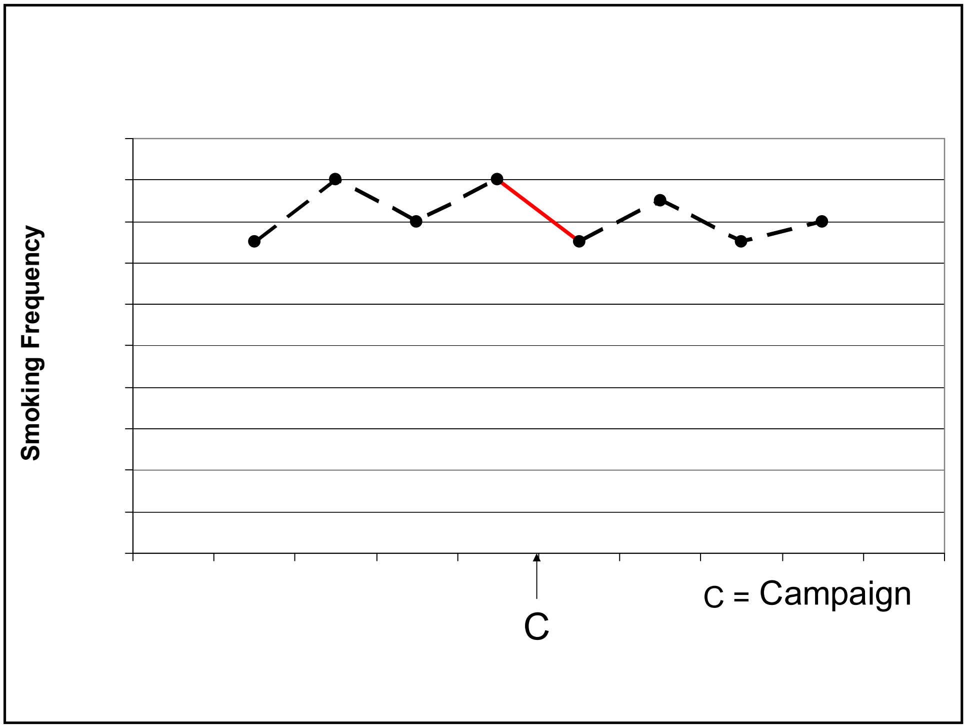
Scores consistently increasing
No evidence that the treatment had any effect

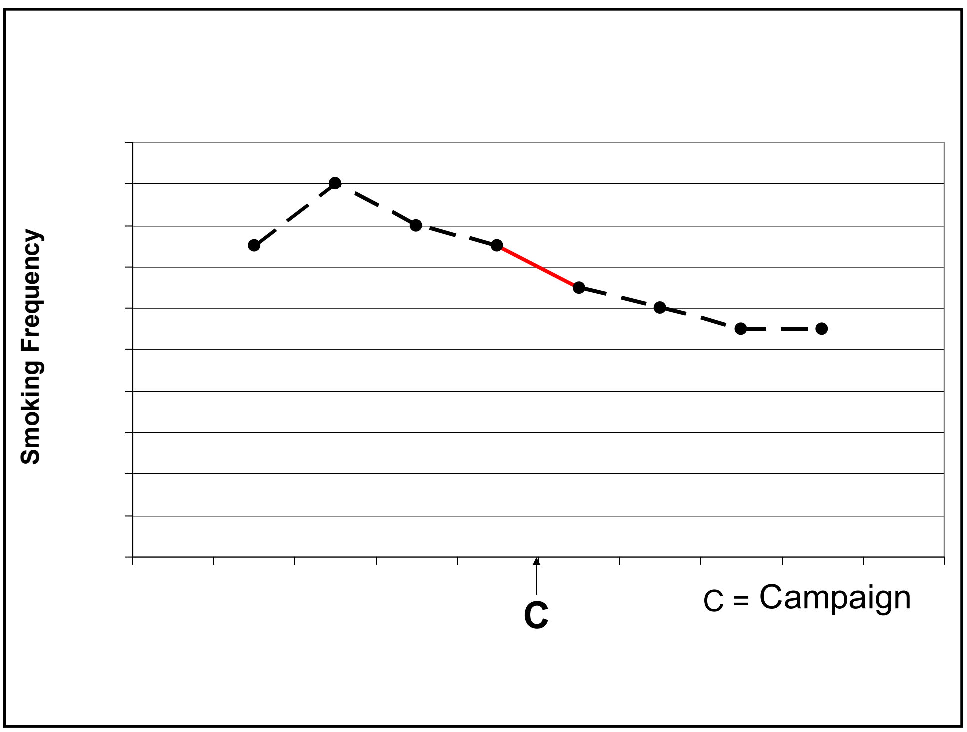
Anti-Smoking Program Example




Smoking Frequency
C
C = Campaign


Smoking Frequency
C
C = Campaign


Smoking Frequency
C
C = Campaign


Smoking Frequency
C
C = Campaign

Strengthening Time Series Designs

- Add a non-equivalent control group

O O O O O X O O O O O

O O O O O   O O O O O

Strengthening Time Series Designs

- Use a second experimental condition (a 'switching replication')
  - treatment is introduced earlier or later

O O O O O X O O O O O

O O O X O O O O O O O

O O O O O   O O O O O

- Helps evaluate history threat
- Enhance external validity

Equivalent Time-Series Design

- Treatment is repeatedly administered and removed during series of observations
- Series of events can be extended as long as evaluator wants

O O O X O O O N O O O X O O O N O O O…

O = observation
X = treatment
N = no treatment (treatment withdrawn)

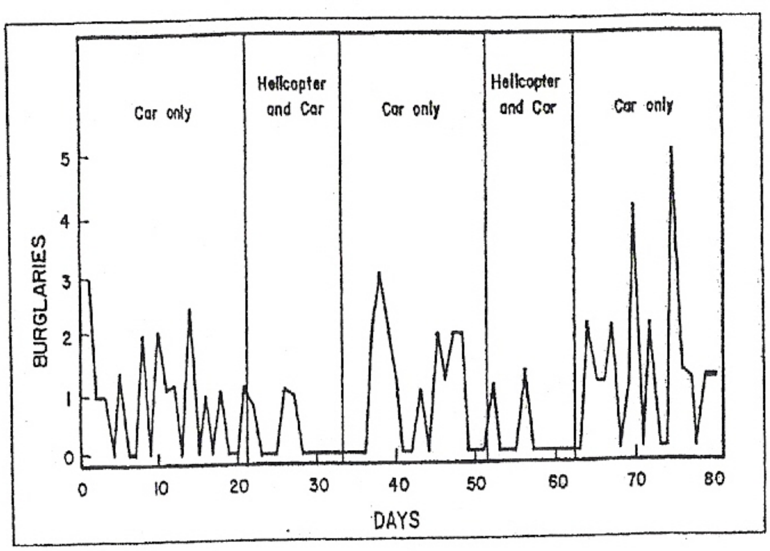
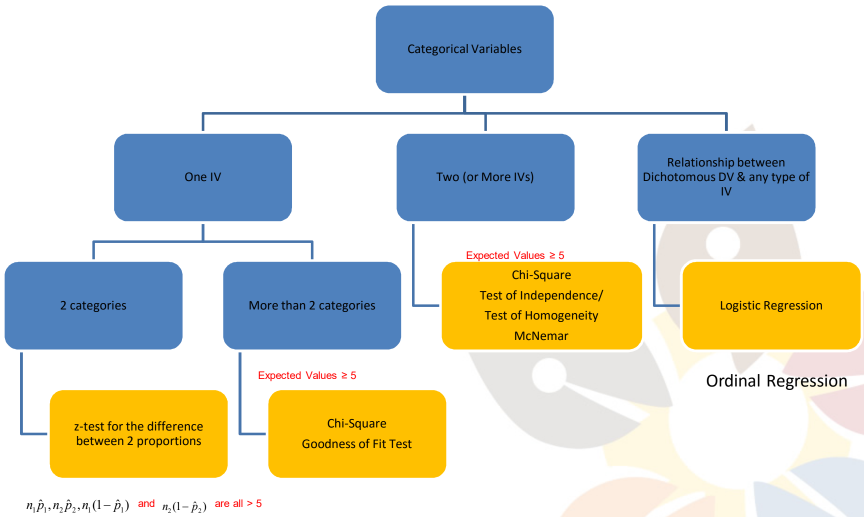


Figure 7.4. The Number of Home Burglaries When Either Police Cars or Both Police Cars and Helicopters Patrolled the Neighborhood
SOURCE: Adapted from Schnelle et al. (1978, p. 15) by permission.

Decision Tree for Statistical Tests



Contingency Table

- Cross tabulation
- Two-way table
- Enumeration or **count data** classified according to **two criteria**

2 x 2 Contingency Table

	Completed the Program		
Sex	Yes	No	Total
Male Female	95 65	40 50	135 115
Total	160	90	250

Chi-Square Test of Independence

- Goal:
  - To determine whether two attributes (categorical variables) are independent

- Expected Value ≥ 5

McNemar Test

- Paired categorical data
  - each observation in the first group has a corresponding observation in the second group
  - observations are counts (1 if attribute is present, 0 if not)

Summary Table

	Variable 2	
Variable 1	Yes	No
Yes	a	b
No	c	d

concordant pairs = (# yes/yes pairs) + (# no/no pairs) = a + d
discordant pairs = (# yes/no pairs) + (# no/yes pairs) = b + c

McNemar Example

Abdominal pain after treatment?	Abdominal pain before treatment?	
	Yes	No
Yes	11	1
No	14	3

Requirements of McNemar

- Sample data are randomly selected
- Sample data consist of matched pairs of frequency counts
- Nominal level data
- Each observation can be classified in 2 ways
- $b + c \geq 10$

Correlation, r

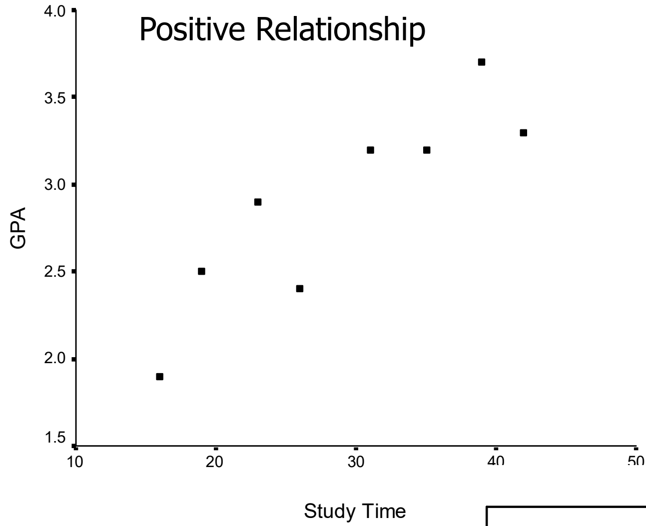
- Are the two continuous variables measured on the same people related?
- Assess the strength and direction (of linear relationships)

- Example
  - Is there a relationship between the number of sessions participants attended a nutrition program and their confidence rating in cooking healthy meals?

Properties of the Correlation Coefficient

1. Positive r (r>0) indicates a positive linear or direct association
   - As x increases, y increases (best fit line slopes up)
2. Negative r (r<0) indicated a negative linear or indirect association
   - As x increases, y decreases (best fit line slopes down)
3. r always between -1 and +1 $(-1 \leq r \leq 1)$
   - Values close to +1 or -1 show strong linear associations (points are scattered closely around a line)
   - r = +1 or -1 a perfect relationship (all the points fall on a line)
   - Values near 0 show no/weak *linear* associations

Positive Relationship



Correlation Matrix
(produced in SPSS)

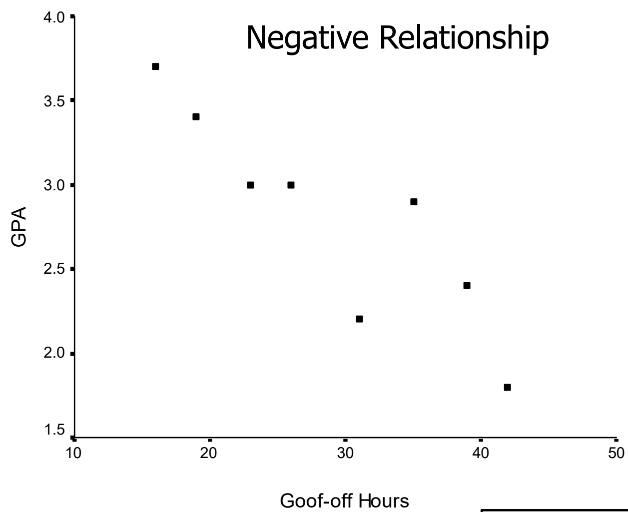
r

Correlations

		STUDYTIM	GPA
STUDYTIM	Pearson Correlation	1	.884**
	Sig. (2-tailed)	.	.004
	N	8	8
GPA	Pearson Correlation	.884**	1
	Sig. (2-tailed)	.004	.
	N	8	8

**. Correlation is significant at the 0.01 level (2-tailed).

Negative Relationship



r

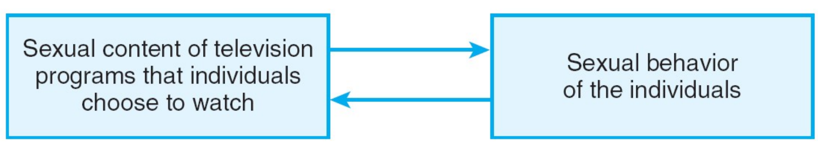
Correlations

		GPA	GOOFOFF
GPA	Pearson Correlation	1	-.892**
	Sig. (2-tailed)	.	.003
	N	8	8
GOOFOFF	Pearson Correlation	-.892**	1
	Sig. (2-tailed)	.003	.
	N	8	8

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation & Causality

- Correlational research
  - Lack of control makes it impossible to conclude anything about cause and effect
  - Directionality problem
  - Third (lurking) variable problem



Regression Models

- Linear, logistic, ordinal, proportional hazards, etc.
  - analysis depends on outcome variable
- Takes into account all sorts of explanations to explain an outcome; able to tease apart the individual contributions of the explanatory variables
- Which of the following variables best predict/explain level of confidence – number of sessions attended, sex, age, …

Errors in Hypothesis Testing

- Process of hypothesis testing is not perfect
- Never certain of decision
- Decision is reached on the basis of evidence presented through data, two kinds of errors may occur

Errors in Hypothesis Testing

H_0: Innocent

Jury Trial

	The Truth	
Verdict	Innocent	Guilty
Innocent	**Correct**	**Error**
Guilty	**Error**	**Correct**

Hypothesis Test

	The Truth	
Decision	**H_0** True	**H_0** False
Do Not Reject **H_0** (no difference)	**Correct Decision**	**Type II Error**
Reject **H_0** (difference)	**Type I Error**	**Correct Decision (Power)**

Type I Error

- p-value = probability of making a Type I Error
- p-value < 0.05

Large Group Discussion

- What are the strengths/opportunities of using quantitative methods in evaluation?
- What are the challenges?
- What type of inferential analysis might you use for your case studies and why?