Advanced Quantitative Methods

June 5, 2018

Presenter: Aynslie Hinds
umhinds0@myumanitoba.ca
204-250-8325

Main Objectives

- To review descriptive statistics
- 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

Question

- If I wanted to determine how “good” your statistics knowledge is, what could I do?

How's your statistics knowledge?

- Please rate your level of knowledge about statistics using the following scale:

1 2 3 4 5 6 7

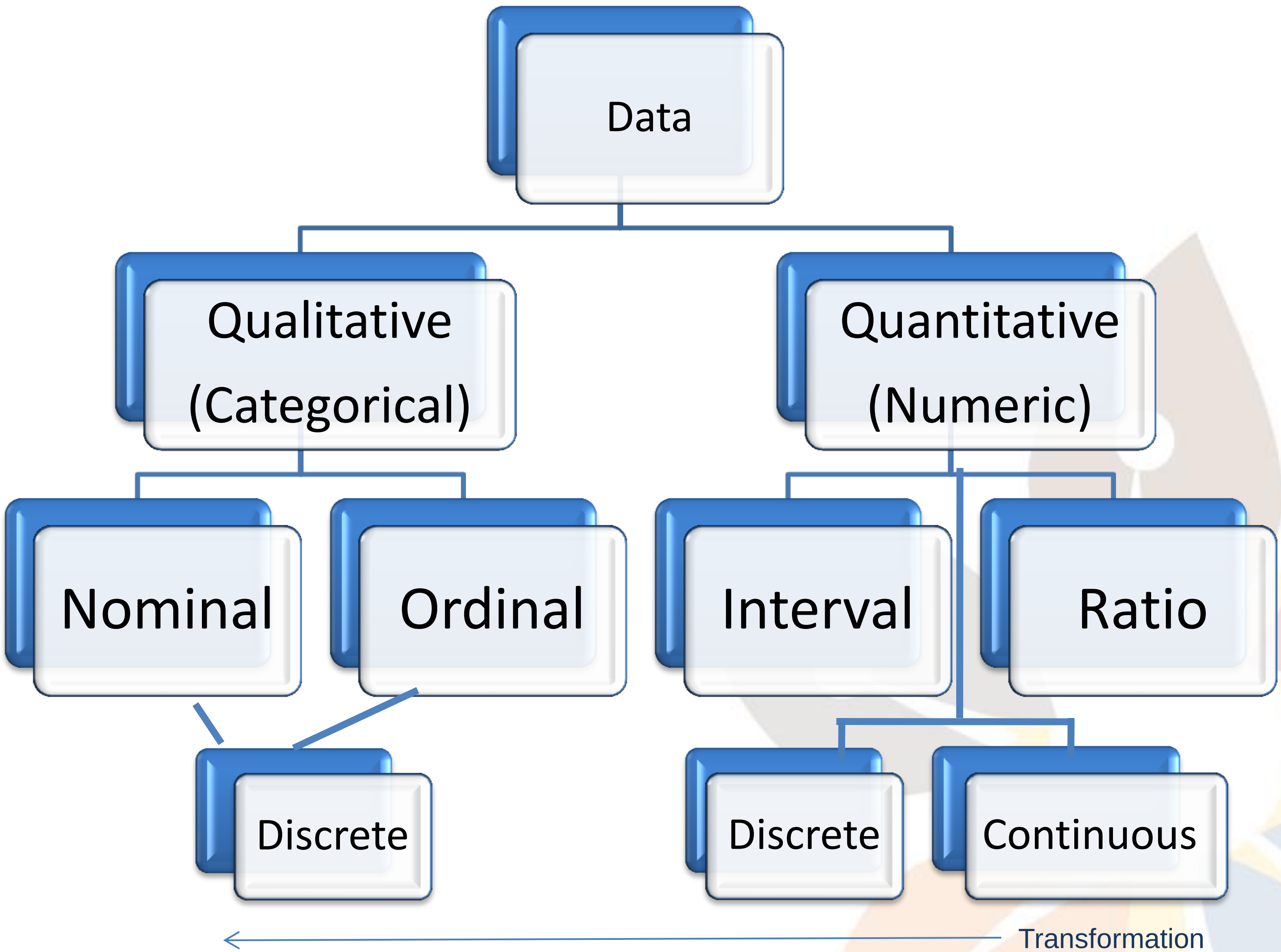
1 – I know nothing about statistics

3 – I have a basic understanding of statistics

5 – I have an intermediate level of understanding about statistics

7 – I'm a statistics guru as I have a PhD in statistics and work in the field

Types of Data



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/Average
  - Median
  - Mode

Mean or Average

$$\text{mean} = \frac{\text{sum of all values}}{\text{number of values}}$$

Data: 2, 8, 4, 10

$$\text{mean} = \frac{2+8+4+10}{4} = \frac{24}{4} = 6.0$$

- Can use when data is numeric (interval or ratio)
- Sensitive to extreme values

Median

- Middle value when the values are arranged in <u>order</u> of magnitude from smallest to largest

Examples	Odd # of Values	Even # of Values
Data set: Arrange in order:	1, 0, 3, 2, 4 0, 1, 2, 3, 4	0, 1, 3, 7, 2, 4 0, 1, 2, 3, 4, 7

Median = 2

$$\text{Median}=\frac{2+3}{2}=2.5$$

- Can be used with ordinal, interval, and ratio data
- Less sensitive to extreme values/outliers
- Use when the data is skewed

Mode

- Value which occurs most frequently

0, 0, 1, 2, 3 → Mode = 0 (unimodal)

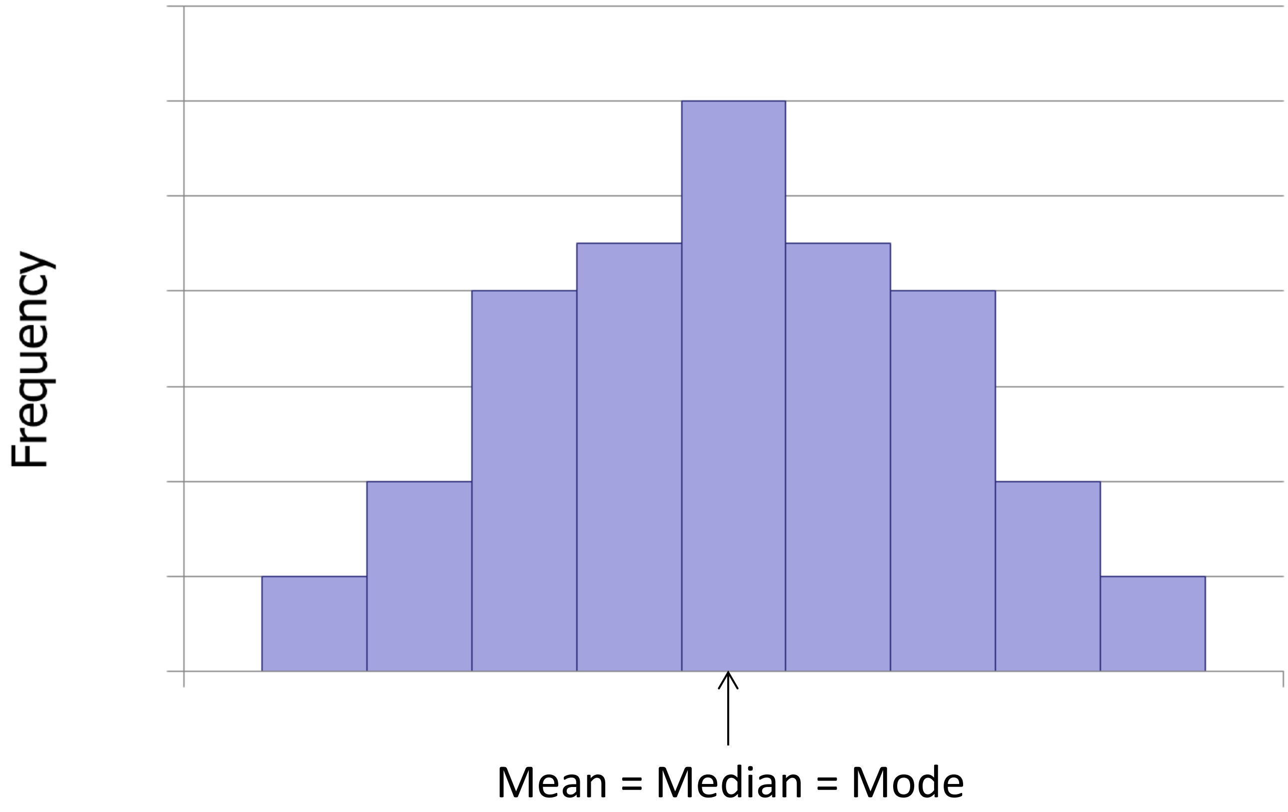
0, 0, 1, 2, 3, 3 → Modes = 0,3 (bimodal)

0, 1, 2, 3 → No mode (all values occur equally often)

- Can be used with all types of data
- Not sensitive to extreme values

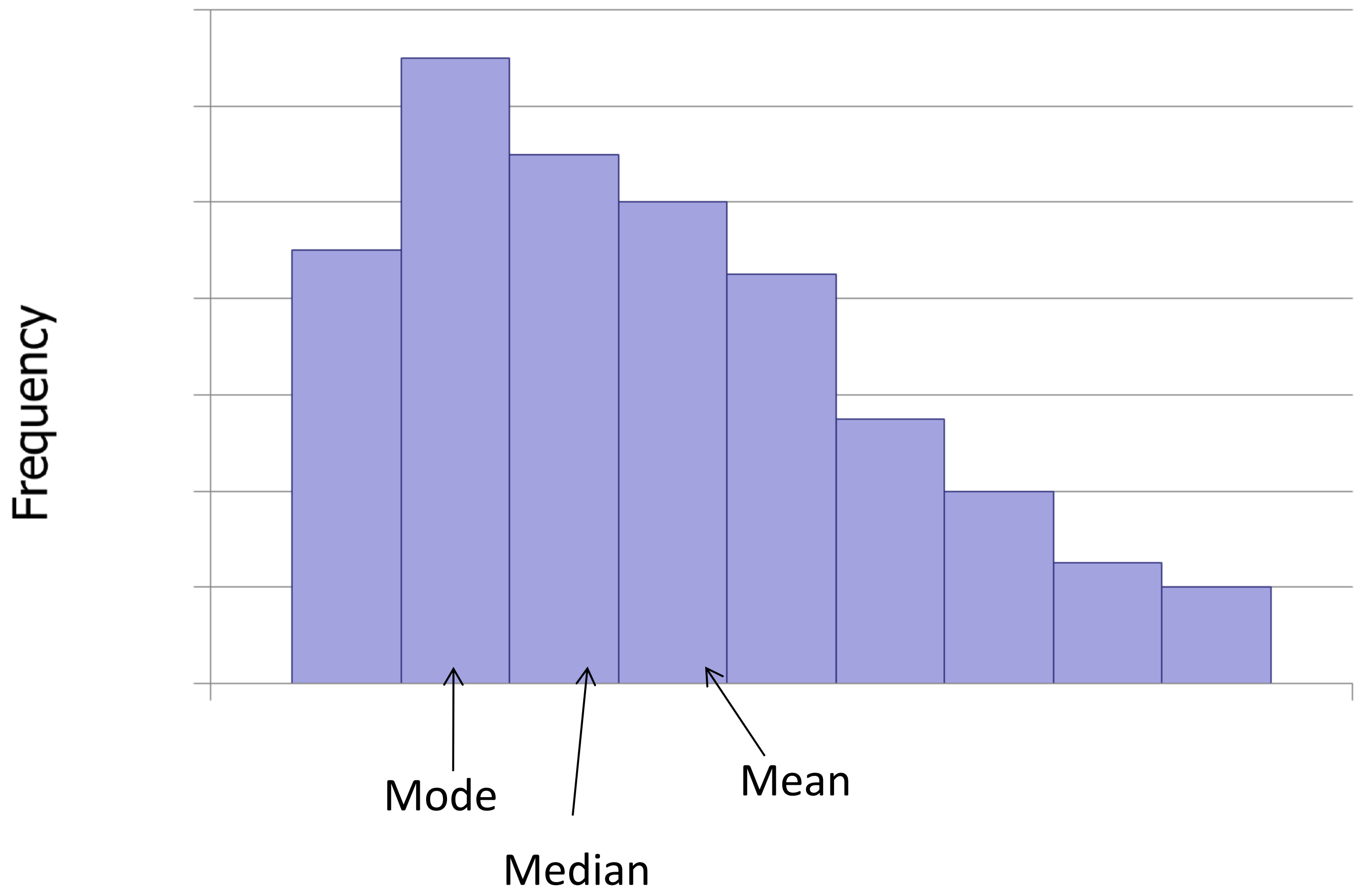
Comparison of Typical Values

symmetric



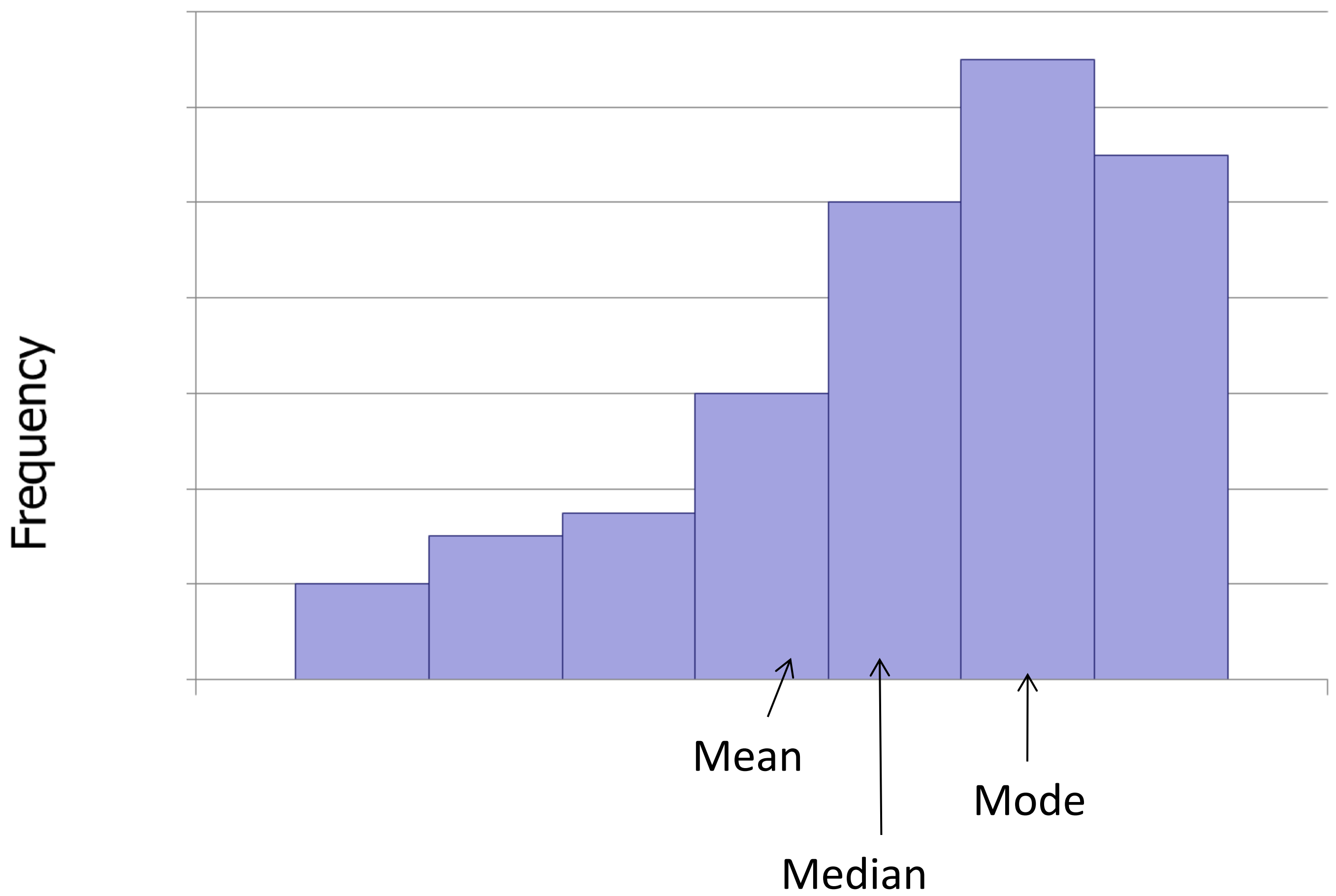
Comparison of Typical Values

Skewed

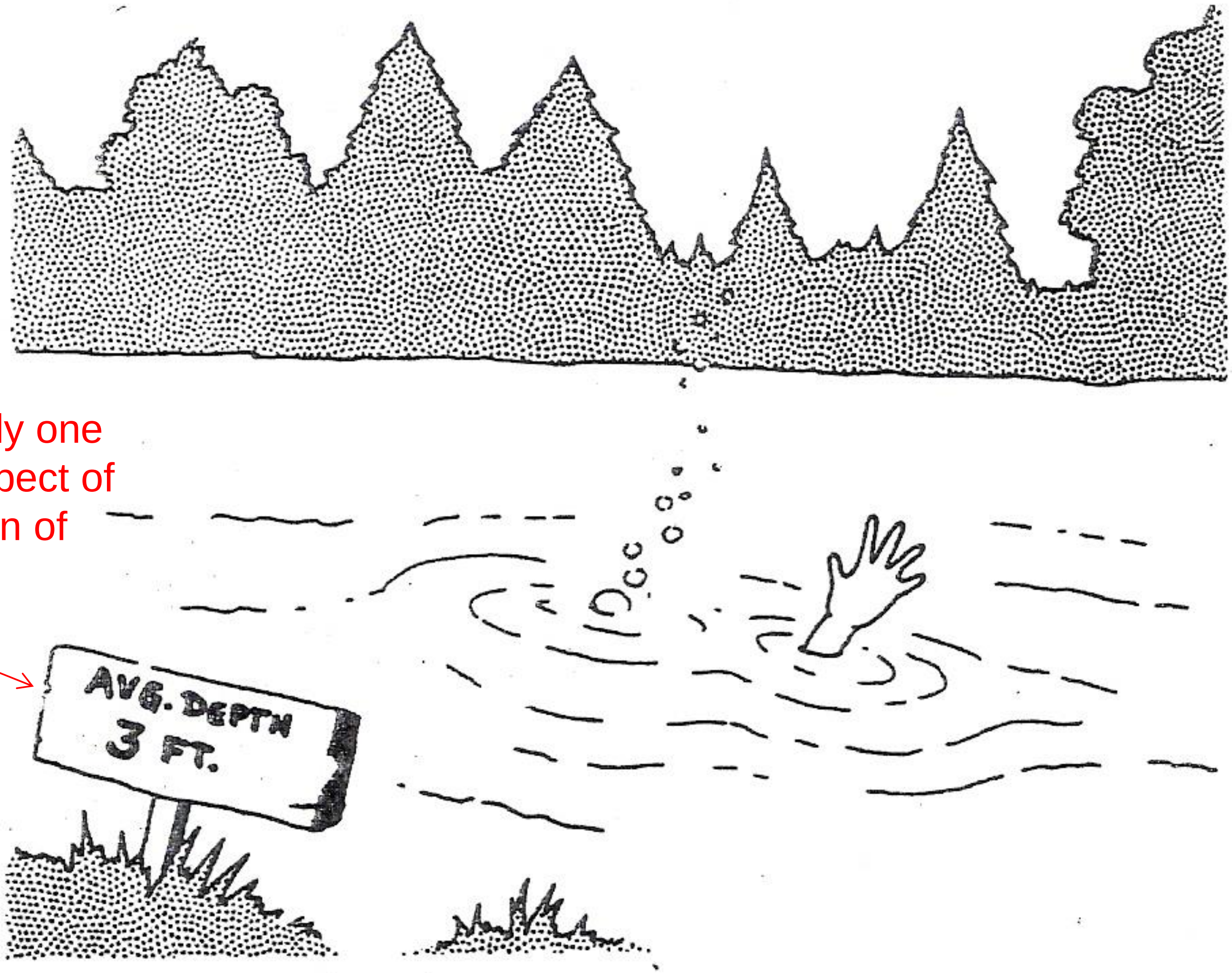


Comparison of Typical Values

Skewed



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
- Interquartile Range

Use when reporting the mean (Variance, Standard deviation)

Use when reporting the median (Interquartile Range)

Range

Range = largest value – smallest value
maximum minimum

- Data set: 5, 6, 8, 9, 10, 12, 14, 15, 17, 18

Range = 18 – 5 = 13

- Data set: 5, 5, 5, 5, 5, 5, 5, 5, 6, 18

Range = 18 – 5 = 13

Interpreting the Standard Deviation

- What is the significance of standard deviation in explaining the spread of the data about the mean?
  - A very small value will indicate that the data values are concentrated around the mean
  - A very large value will indicate that the data values are widely dispersed about the mean
  - If the measurements are all the same (i.e., no variability), then the standard deviation and variance are 0

$$\text{standard deviation} = \sqrt{\text{variance}}$$

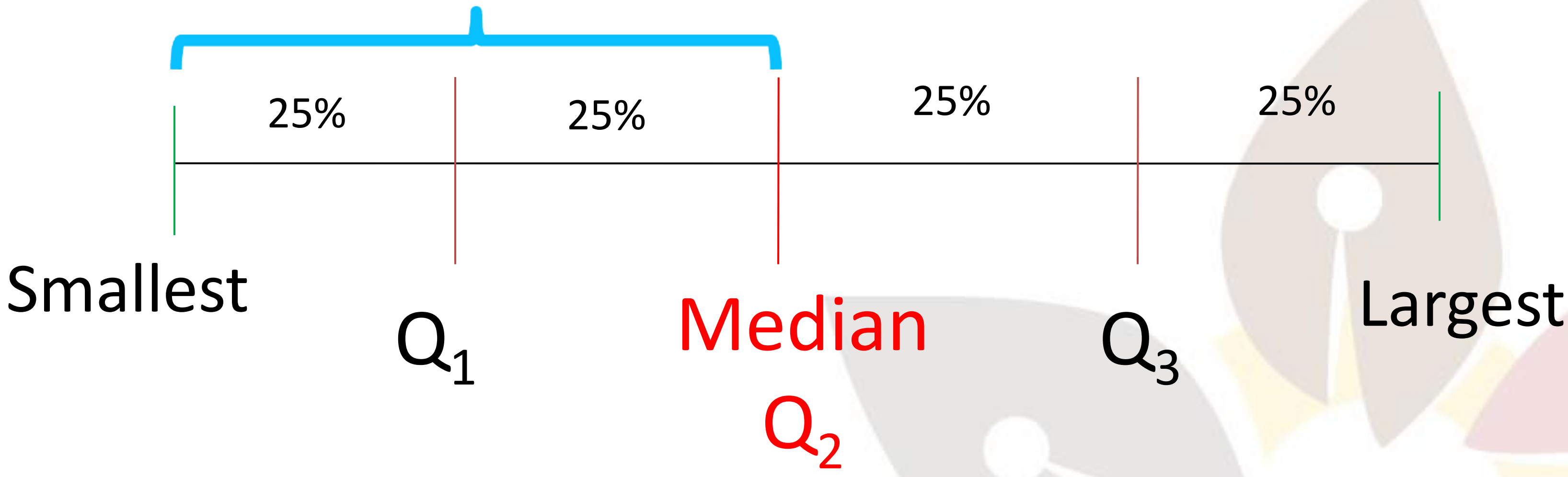
Original units Squared units

Quartiles

- Three summary measures that divide a ranked/ordered data set into 4 (nearly) equal parts
- Is represented on a box plot

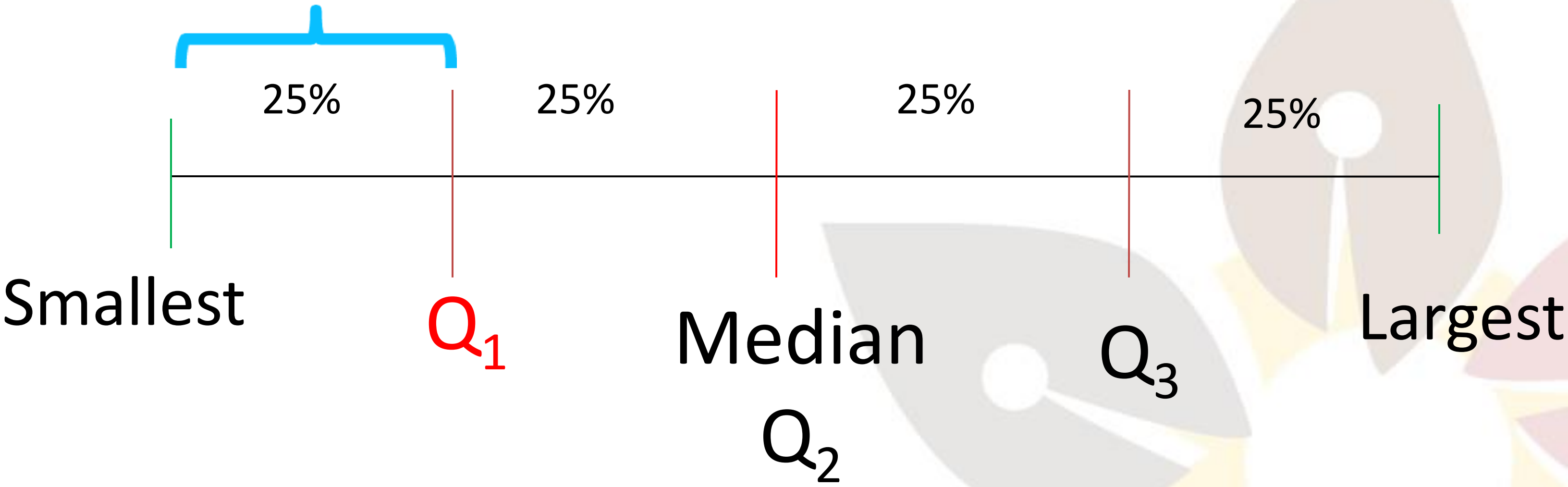
Second Quartile (Q_2)

- Median
- 50th percentile (P_{50})



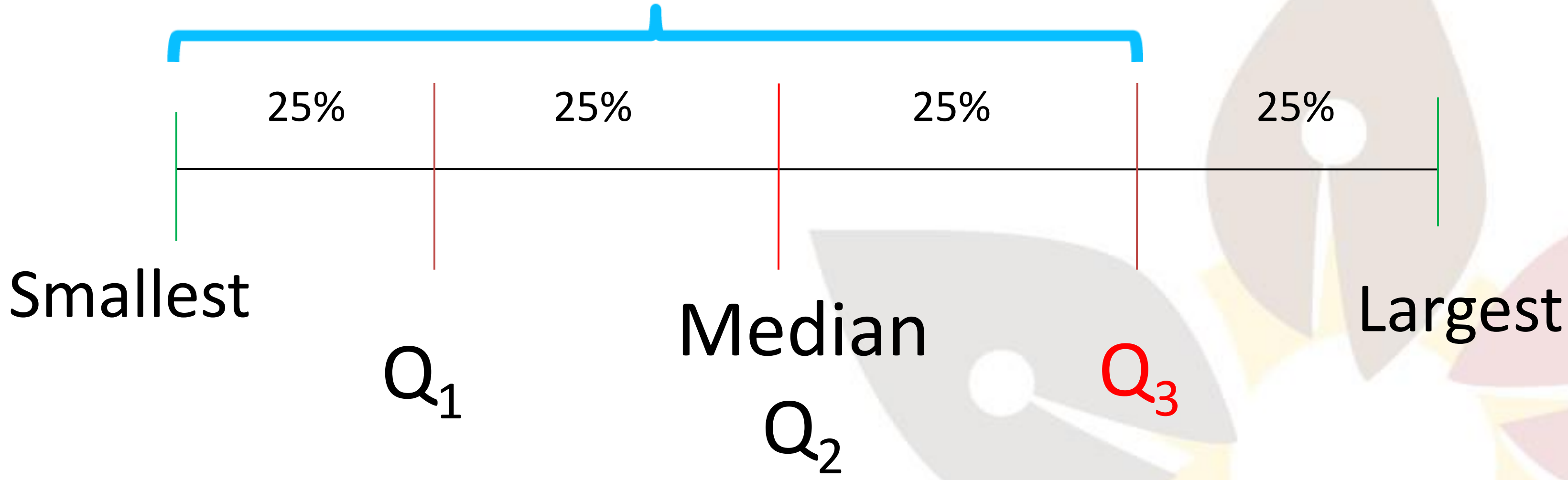
First Quartile (Q_1)

- Median of lower half of dataset
- 25th percentile



Third Quartile (Q_3)

- Median of upper half of data set
- 75th percentile



Inter-Quartile Range (IQR)

$$IQR = Q_3 - Q_1$$

Range of the middle 50% of vales

Example

2 3 6 7 10 12 16 22 35 40

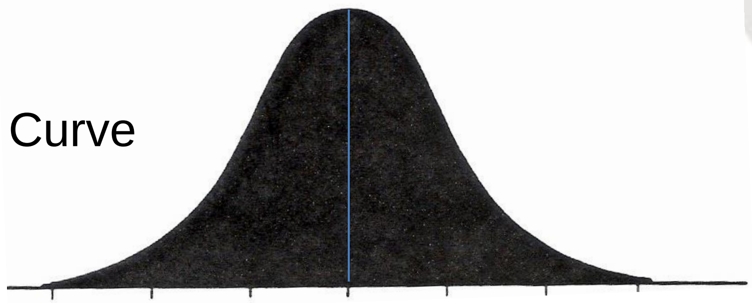
$Q_1 = 6$ $Q_2 = 11$ $Q_3 = 22$

$$IQR = 22 - 6 = 16$$

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

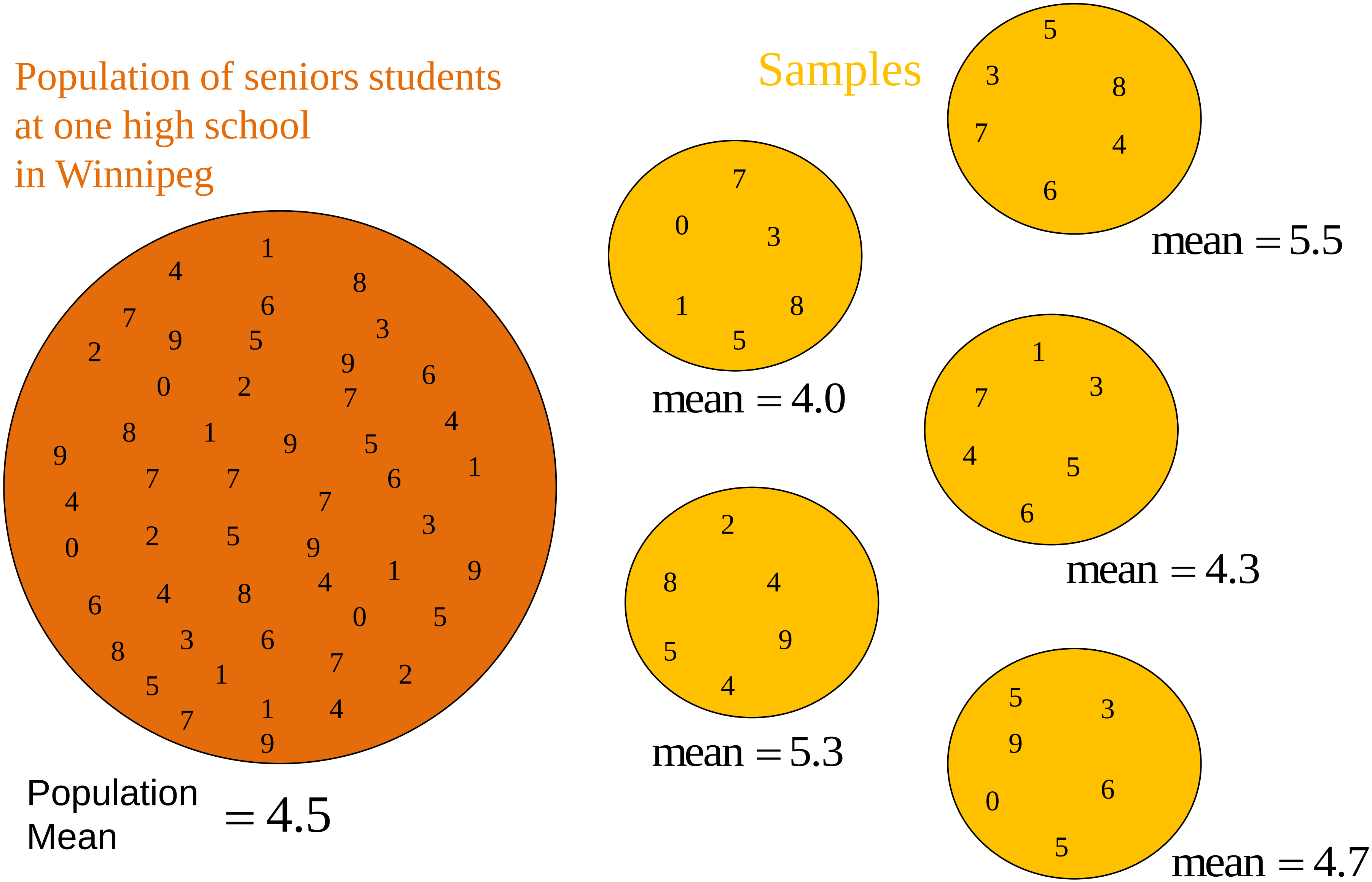


Normal Distribution

- 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

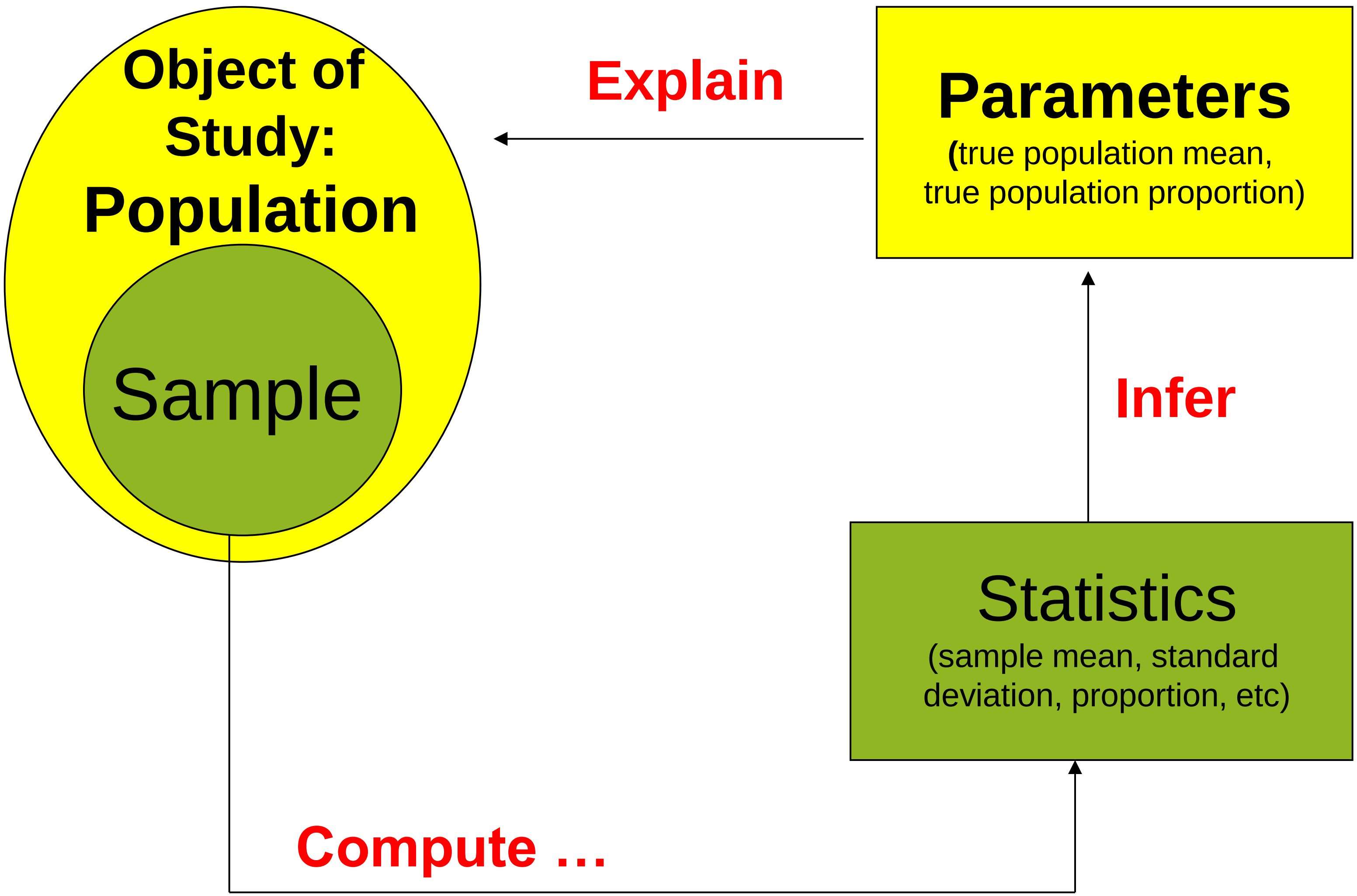
Sample Statistics (measurable)	Population Parameters (unknown)
Mean: $\bar{x}$	Mean: μ
Variance: s^2	Variance: σ^2
Standard Deviation: s	Standard Deviation: σ
Proportion: p	Proportion: ρ

use to estimate

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

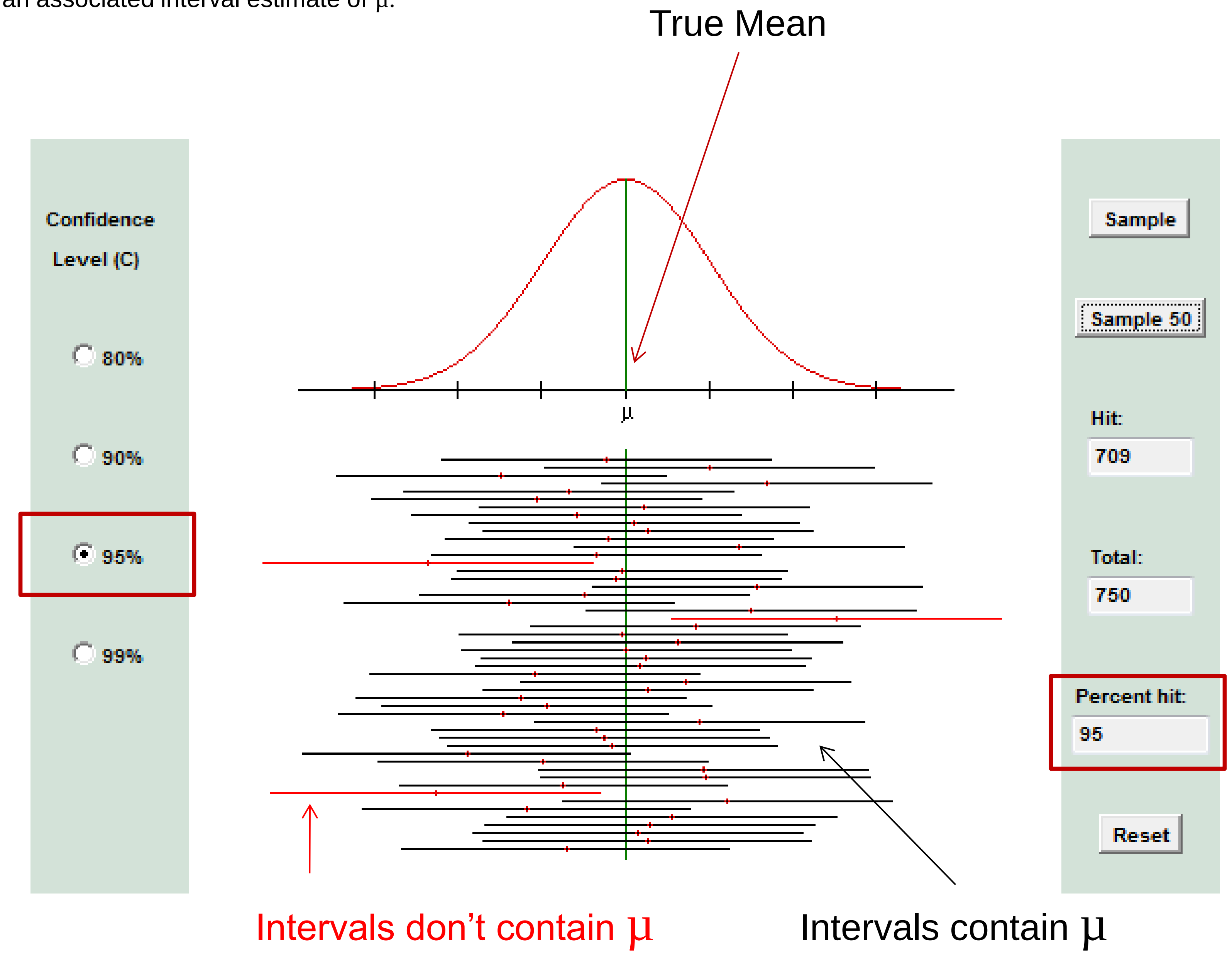
- Two methods for providing an estimate of a parameter…
  - Point estimate
  - Interval estimation (i.e., confidence interval)

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%

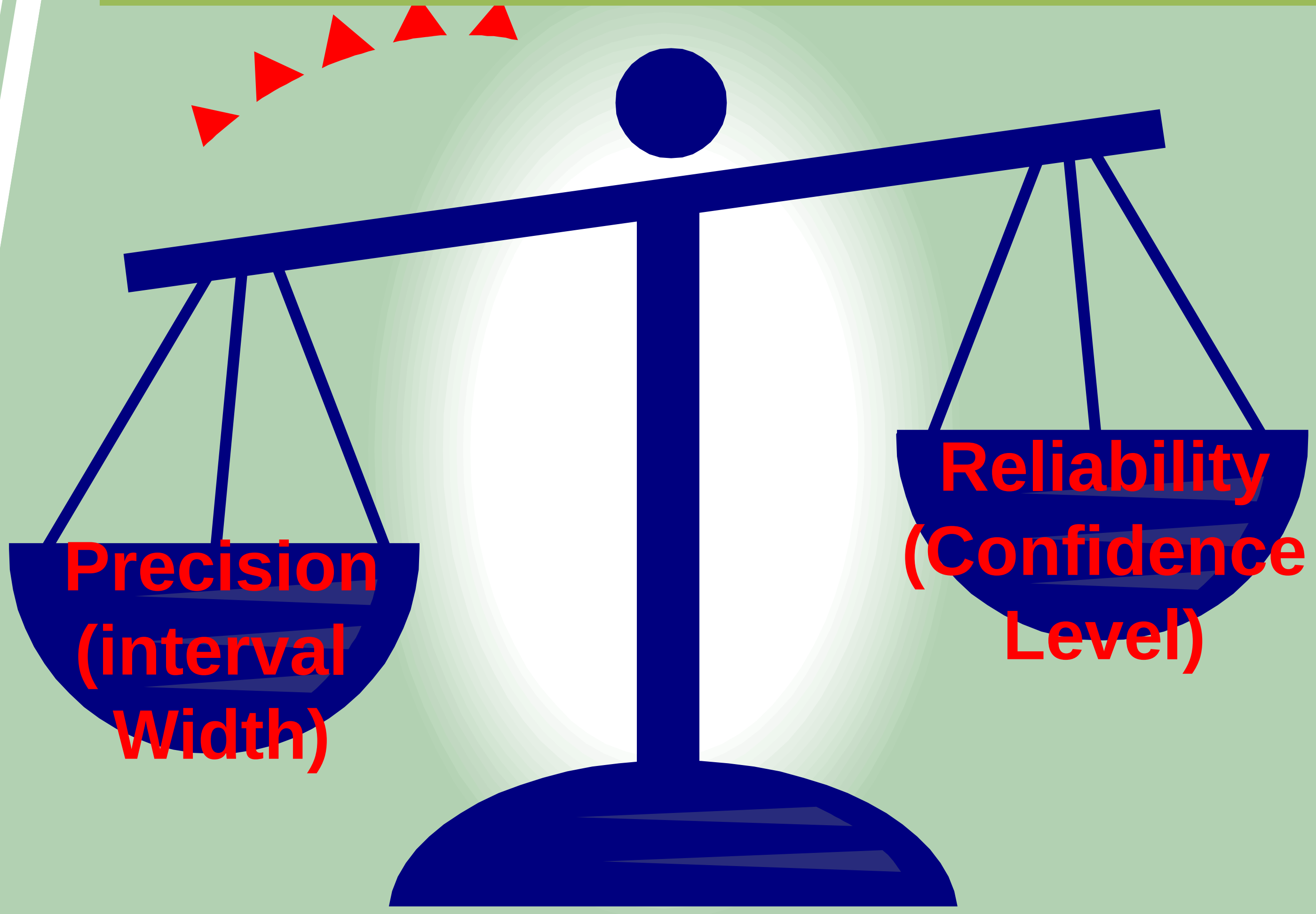
Each sample gives rise to a point estimate and an associated interval estimate of μ.



Example

- We are 95% confident that the true proportion of pregnant women who will quit smoking if they undergo a smoking cessation program is (0.09, 0.12).

Balance



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/outcome 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

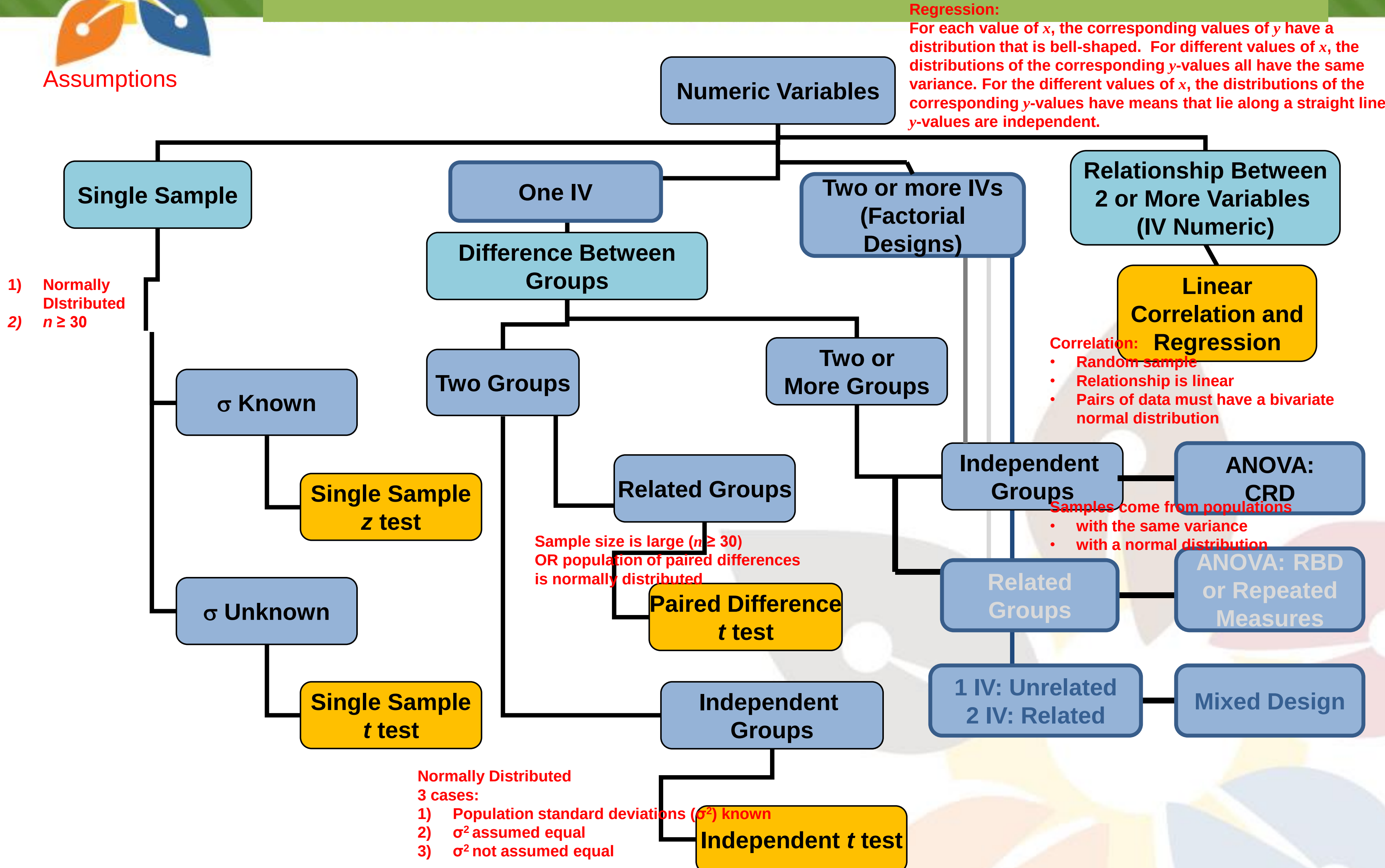
Threats to Internal Validity

- **History effect**
  - Events that occur besides the treatment (events in the environment)
- **Maturation**
  - Physical or psychological changes in the participants
- **Testing**
  - Effect of experience with the pretest - - become test wise
- **Instrumentation**
  - Changes observed from pre to post-test due to changes (instrument, administrators, method of administration)

Threats to Internal Validity

- **Selection**
  - Effect of treatment confounded with other factors because of selection of participants, problem in non random sample
- **Statistical regression**
  - Tendency for participants whose scores fall at either extreme on a variable to score nearer the mean when measured a second time
- **Attrition**
  - Participants lost from the study

Decision Tree for Statistical Tests



Note: Numerical variable = Quantitative variable

CRD: Completely Randomized Design
RBD: Randomized Block Design

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)	

p-value

0.01 or 0.10

- "p" = probability
- If the p-value ≤ **0.05 (the result is significant; there is a difference)**

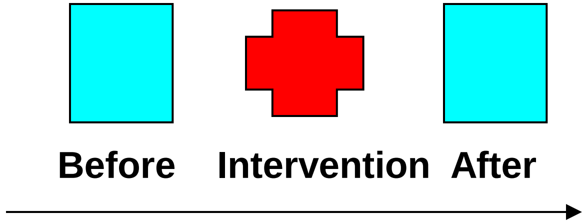
p-value > 0.05

- p-value = 0.15, p-value = 0.45, etc.
- Not a statistically significant result; no evidence of a difference

Two Related Groups

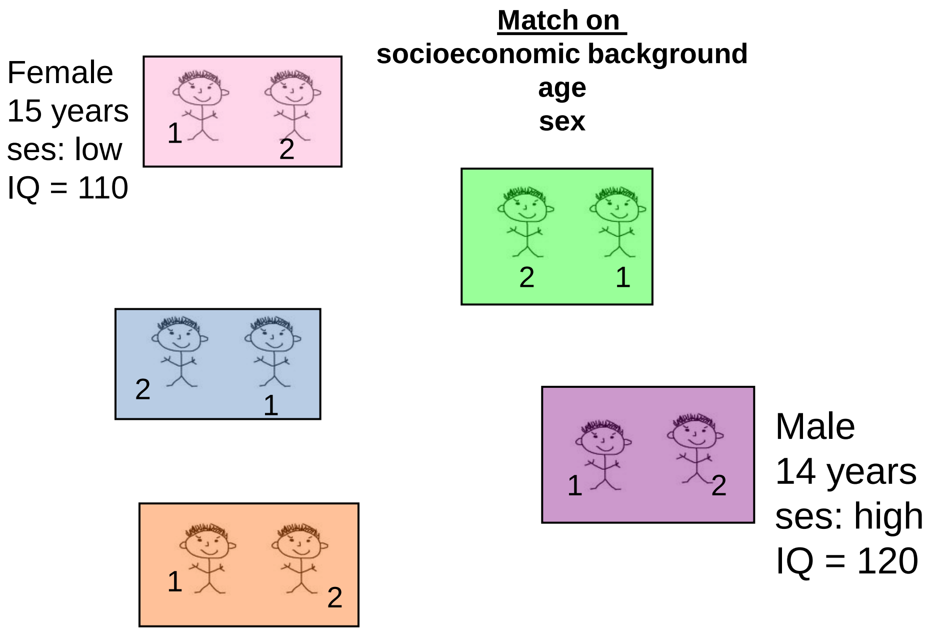
1. Pre-Post Study

Repeated Measures Design



Same People

2. Matched on Relevant Characteristics



Test Statistic =
Paired Difference t-test

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

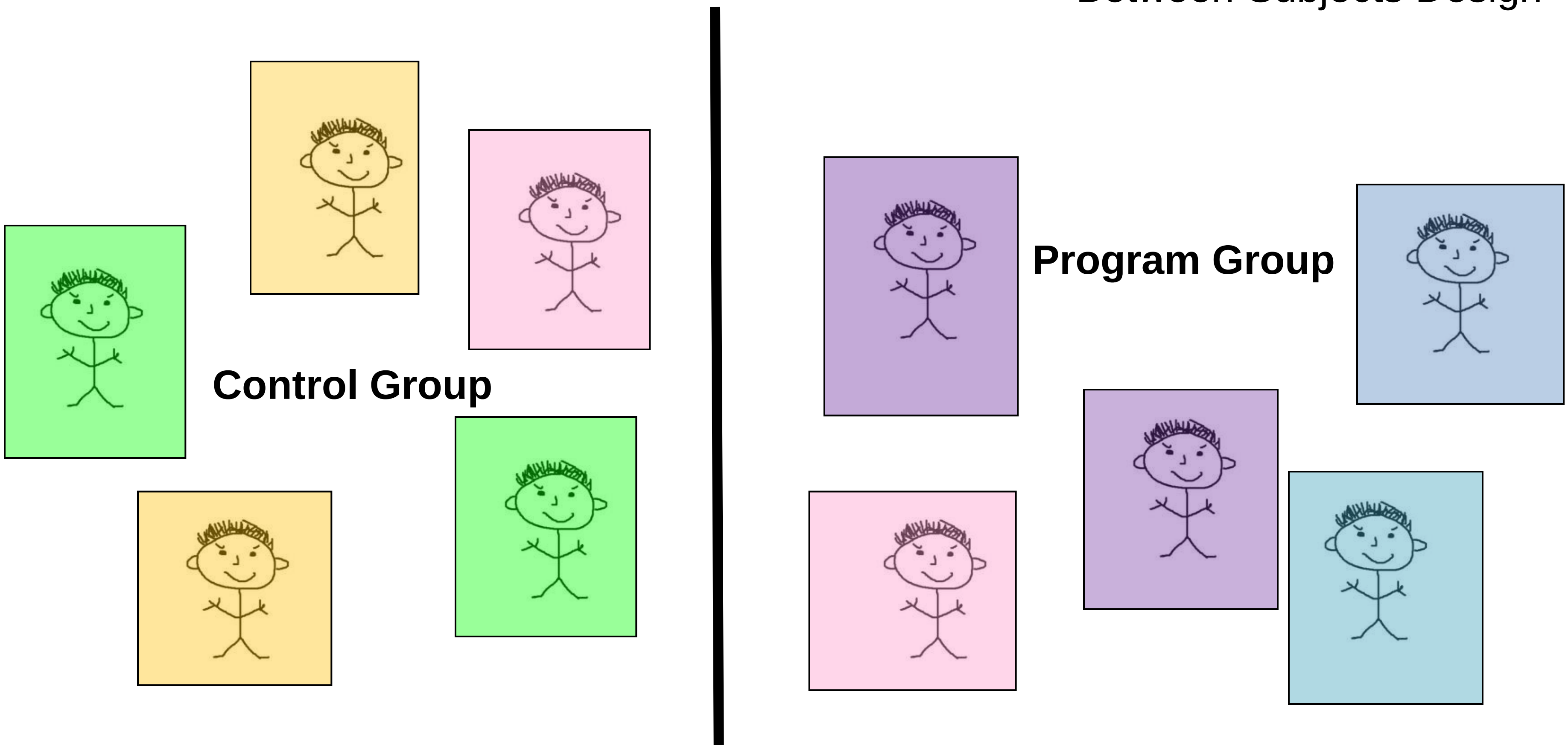
Within-Subjects Design

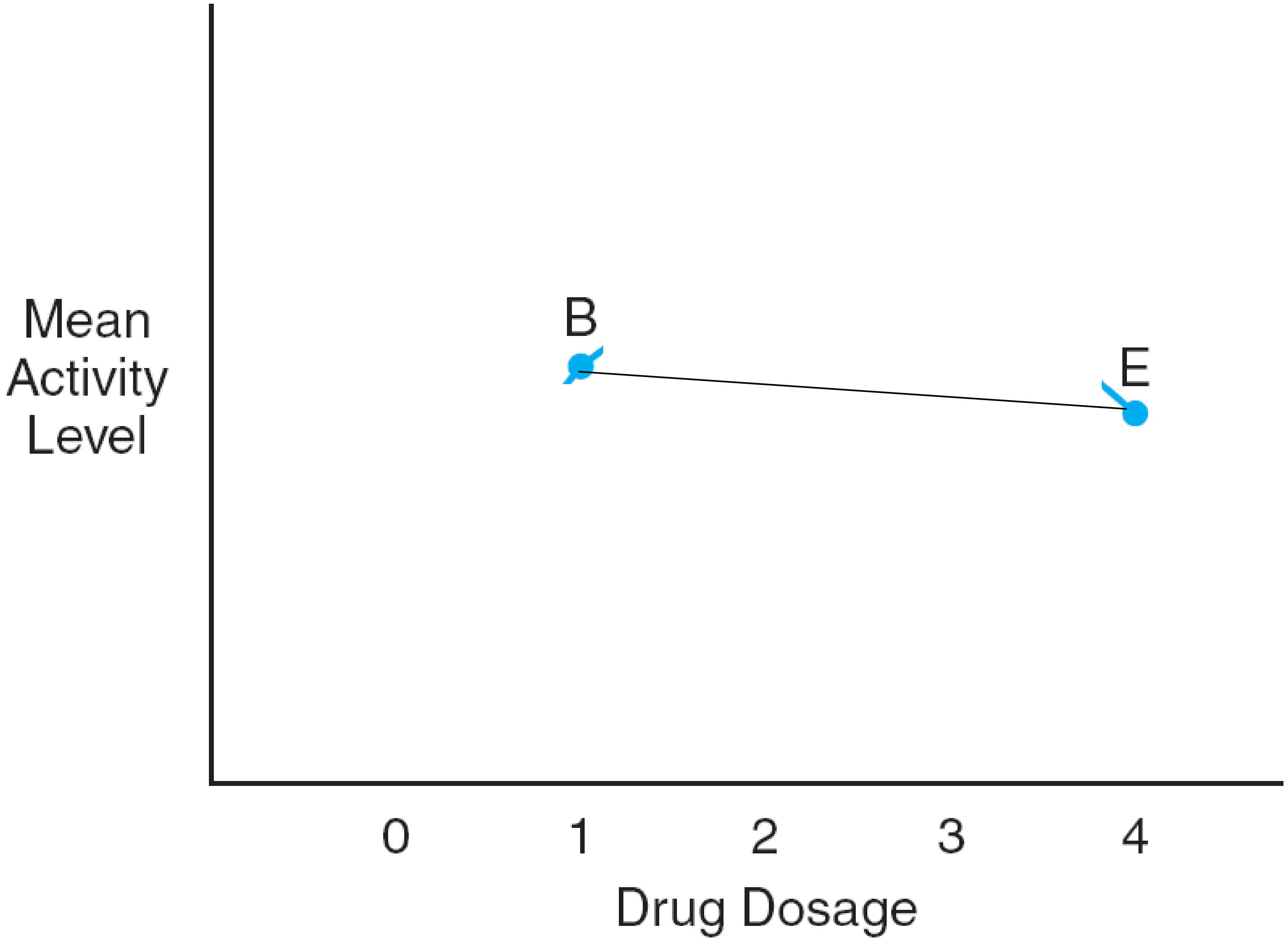
Two Unrelated Groups

Test Statistic =
Independent t-test

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

Between Subjects Design



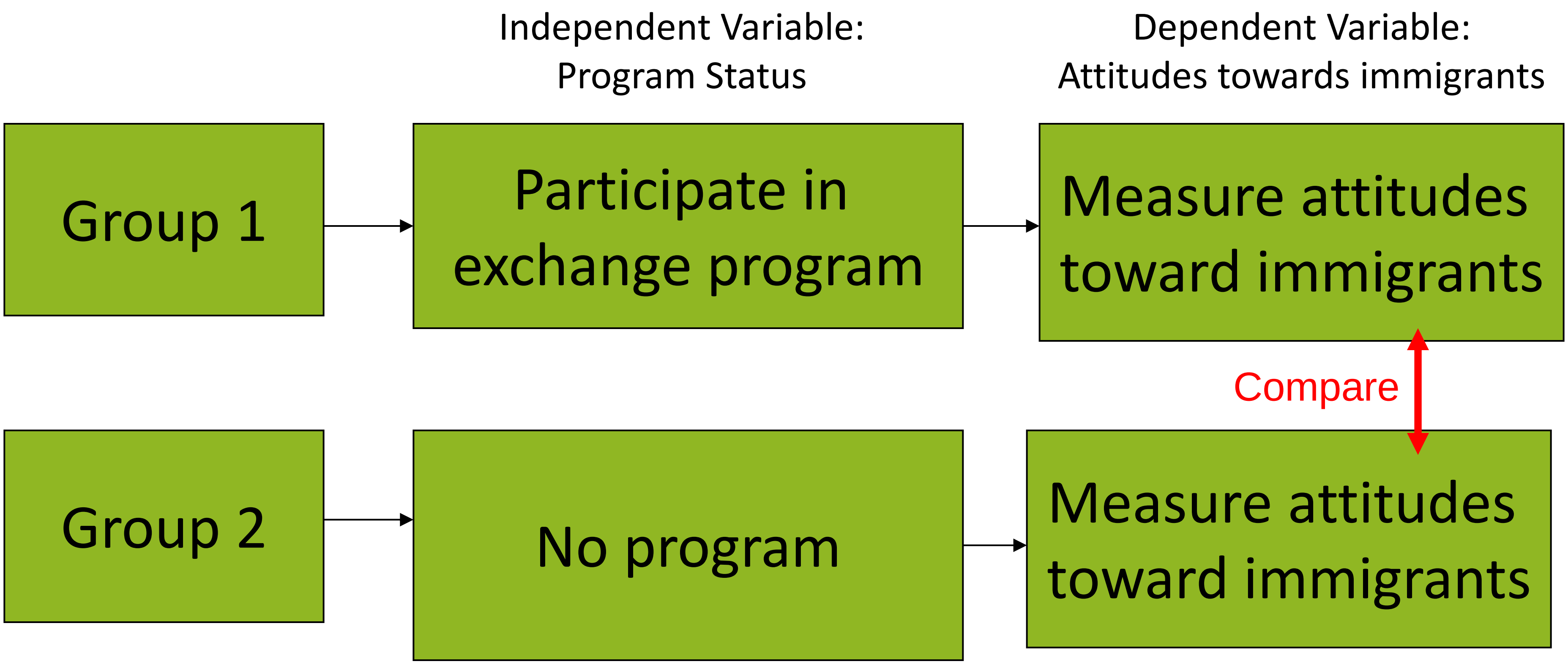

Mean
Activity
Level
B
E
0
1
2
3
4
Drug Dosage

Two or More Groups/Conditions

- Analysis of Variance (ANOVA)
  - F statistic
  - Extension of an independent t-test
  - Extension of a paired-difference t-test
  - One-way ANOVA
  - Two-way ANOVA (Factorial Design)

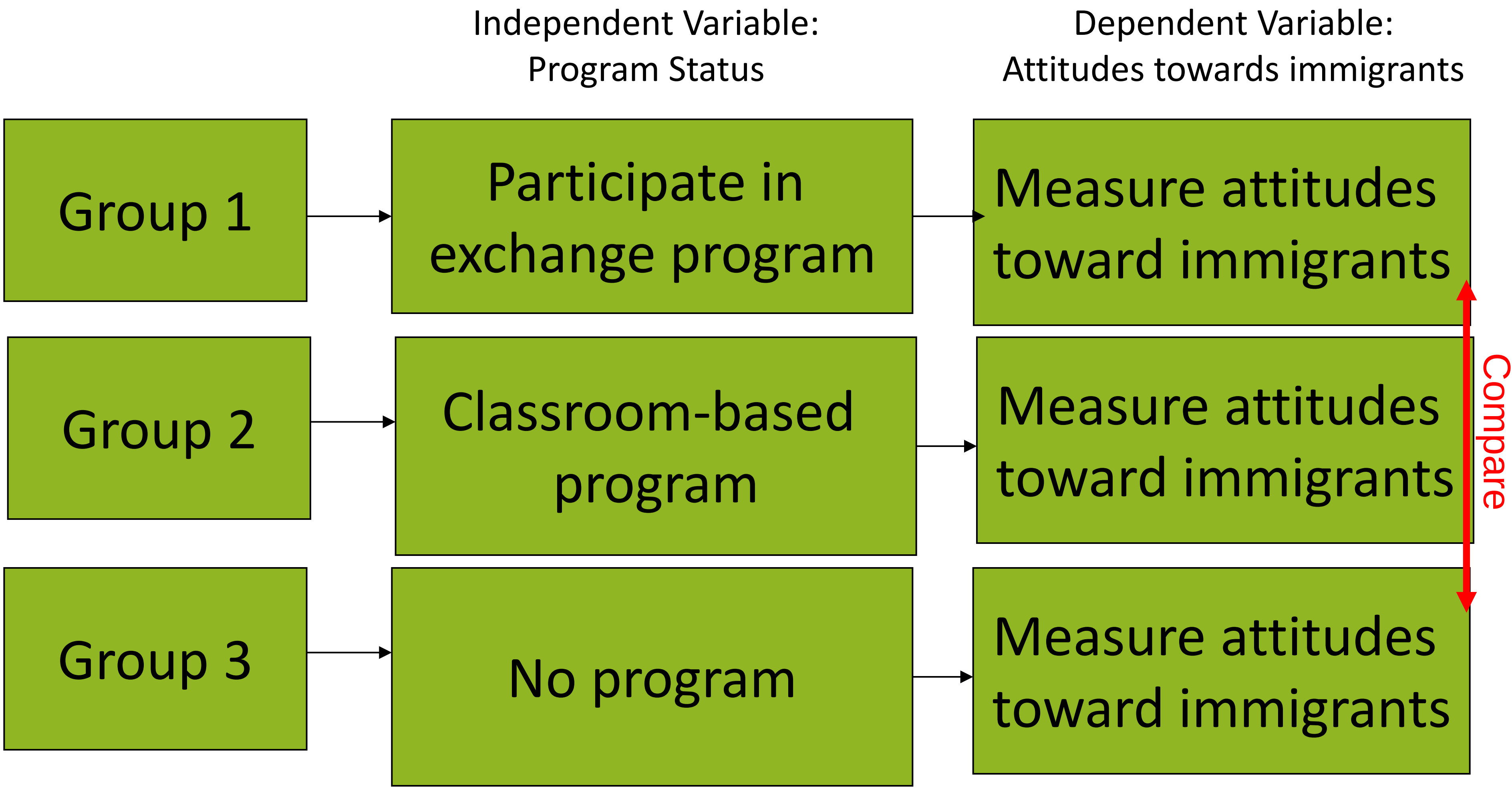
  - Follow-Up Analyses
    - Multiple Comparison Procedures
    - Contrasts

Example



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

Example



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

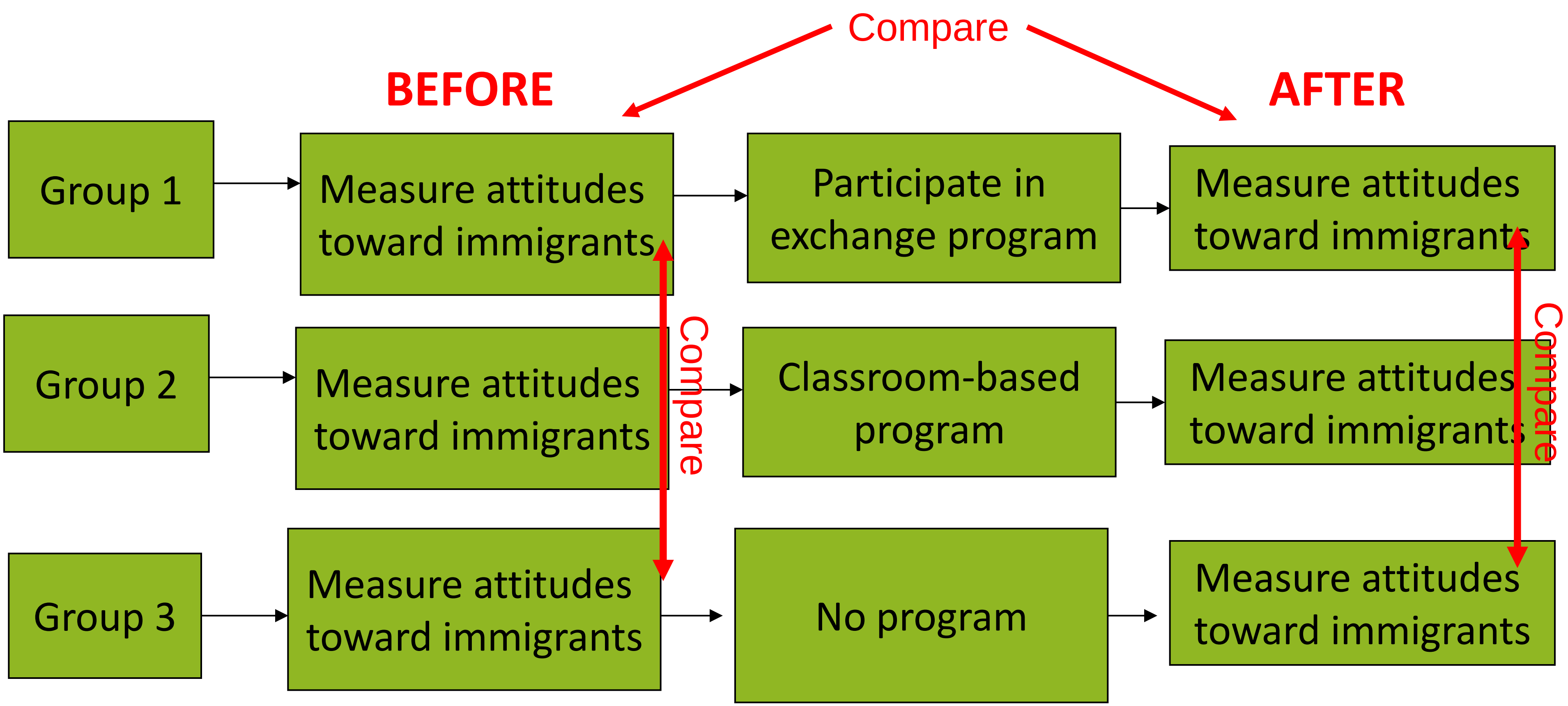
ANOVA

- *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 (factors)
  - Could also have quasi-IVs (e.g., age, gender)
- Able to study the individual (main effects) and interactive effects of the IVs on the DV

Example



Pretest-Posttest Non-Equivalent Control Group Design
Factorial Design (Mixed): Time (same people), Group (different people)

Main Effect

- Mean differences among the levels of one factor/IV (ignoring the other factor/IV)
- Example
  - Do the groups differ? (i.e., collapse over time; do groups 1, 2, and 3 differ?)
  - Do attitudes towards immigrants change over time? (i.e., collapse over groups; do the before values differ from the after values?)

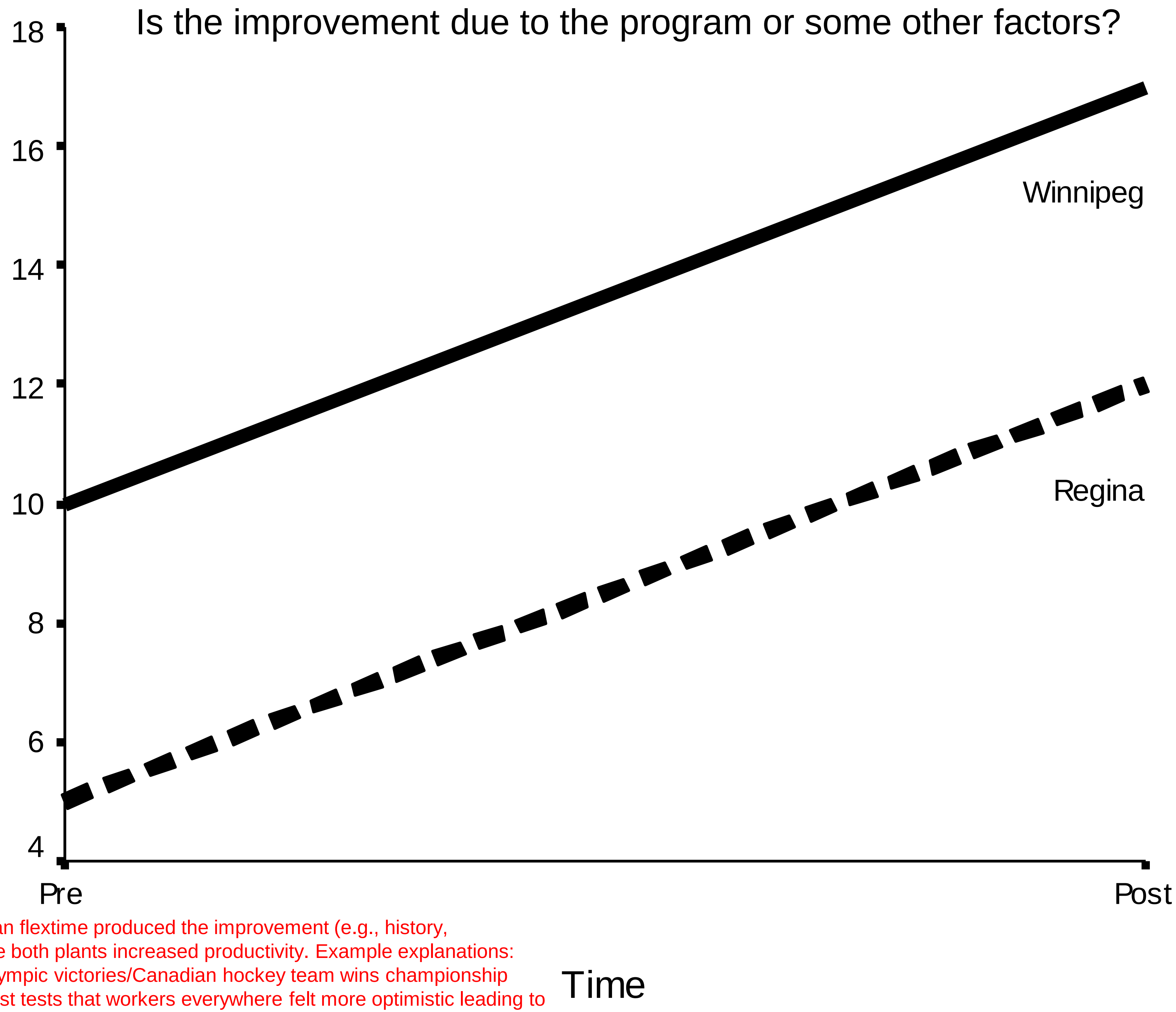
Interaction

- Effects of one factor/IV vary depending on levels of the other factor/IV
- Example
  - No change over time in the “no program group”
  - Greater increase over time in the “exchange group” than the “class-room based program” group
- Graphical representation
  - Parallel lines = no interaction
  - Non-parallel lines = interaction

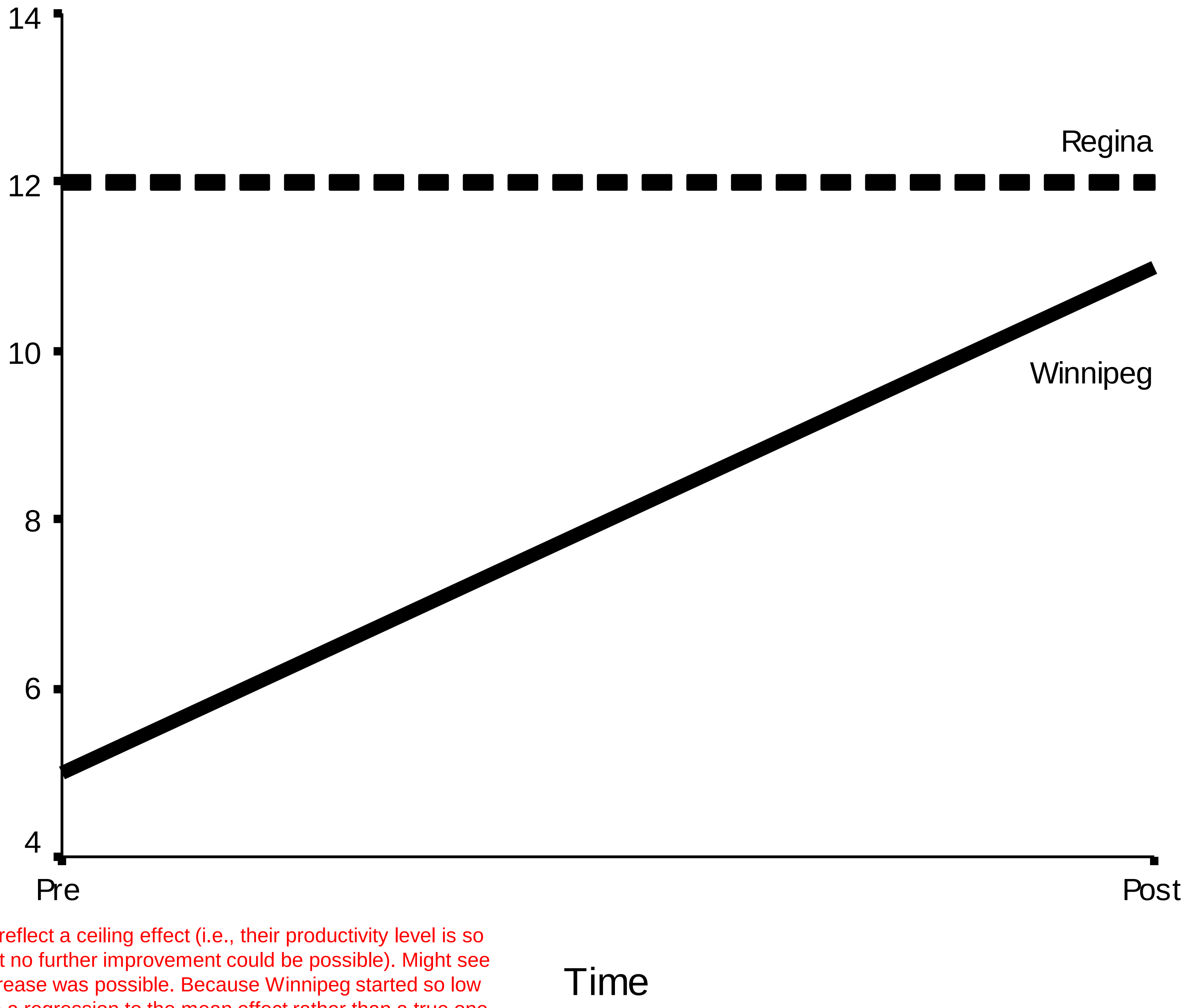
Example

	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

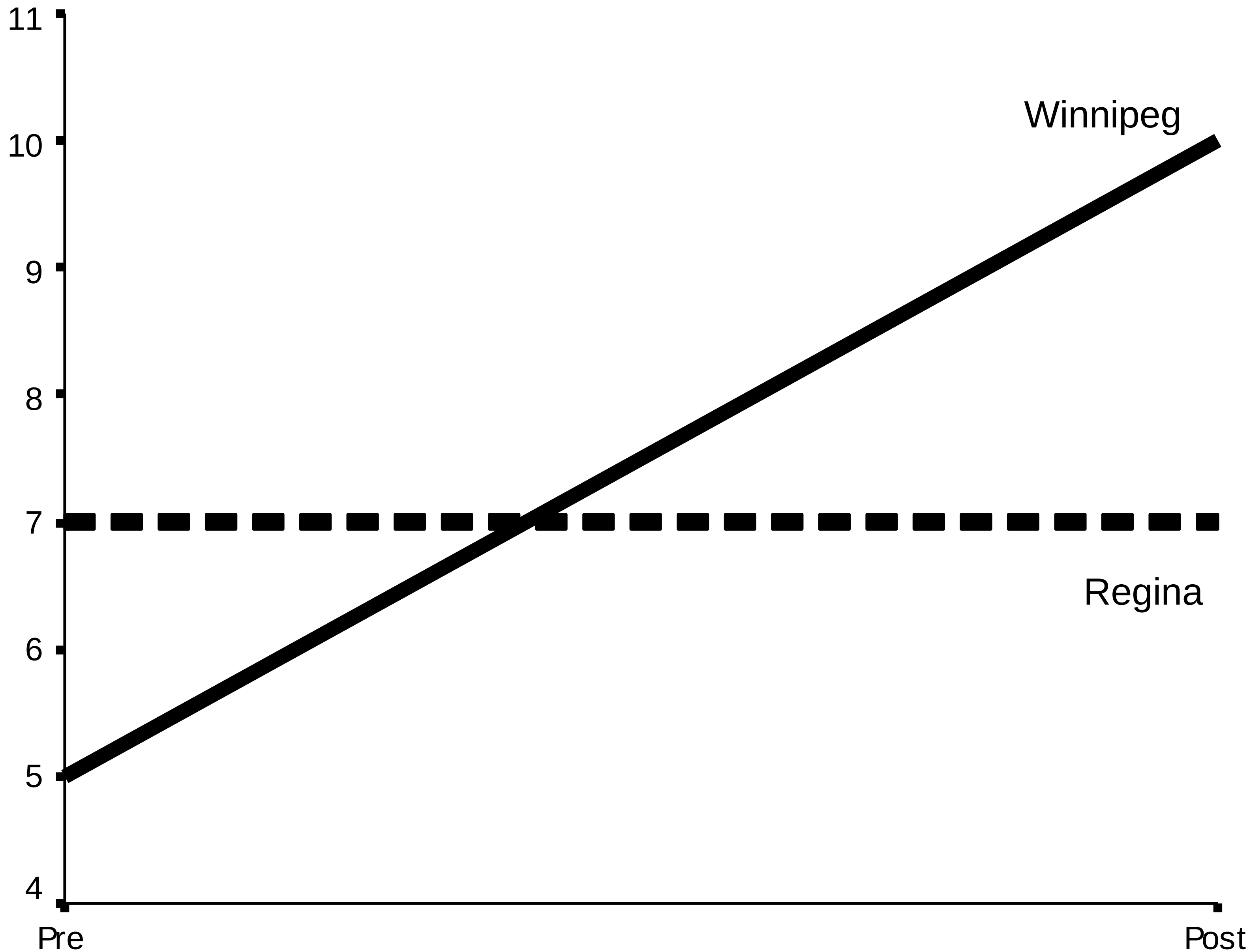
Pretest-Posttest Non-Equivalent Control Group Design
Factorial Design (Mixed): Time (same people), City (different people)



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.



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

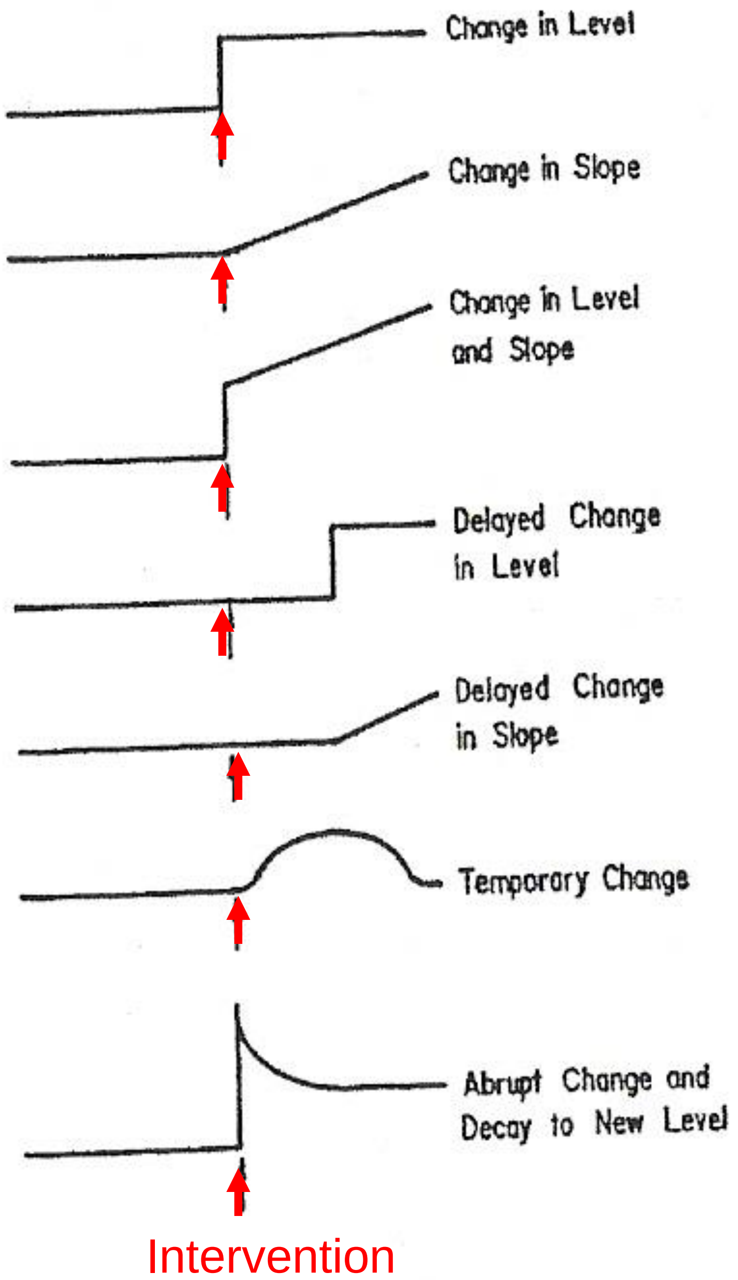
- Involves repeated measurements or observations for each participant over time (before and after treatment or event)

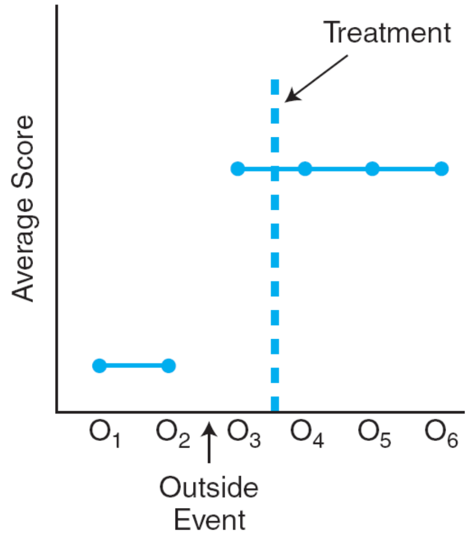
O O O O O X O O O O O

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

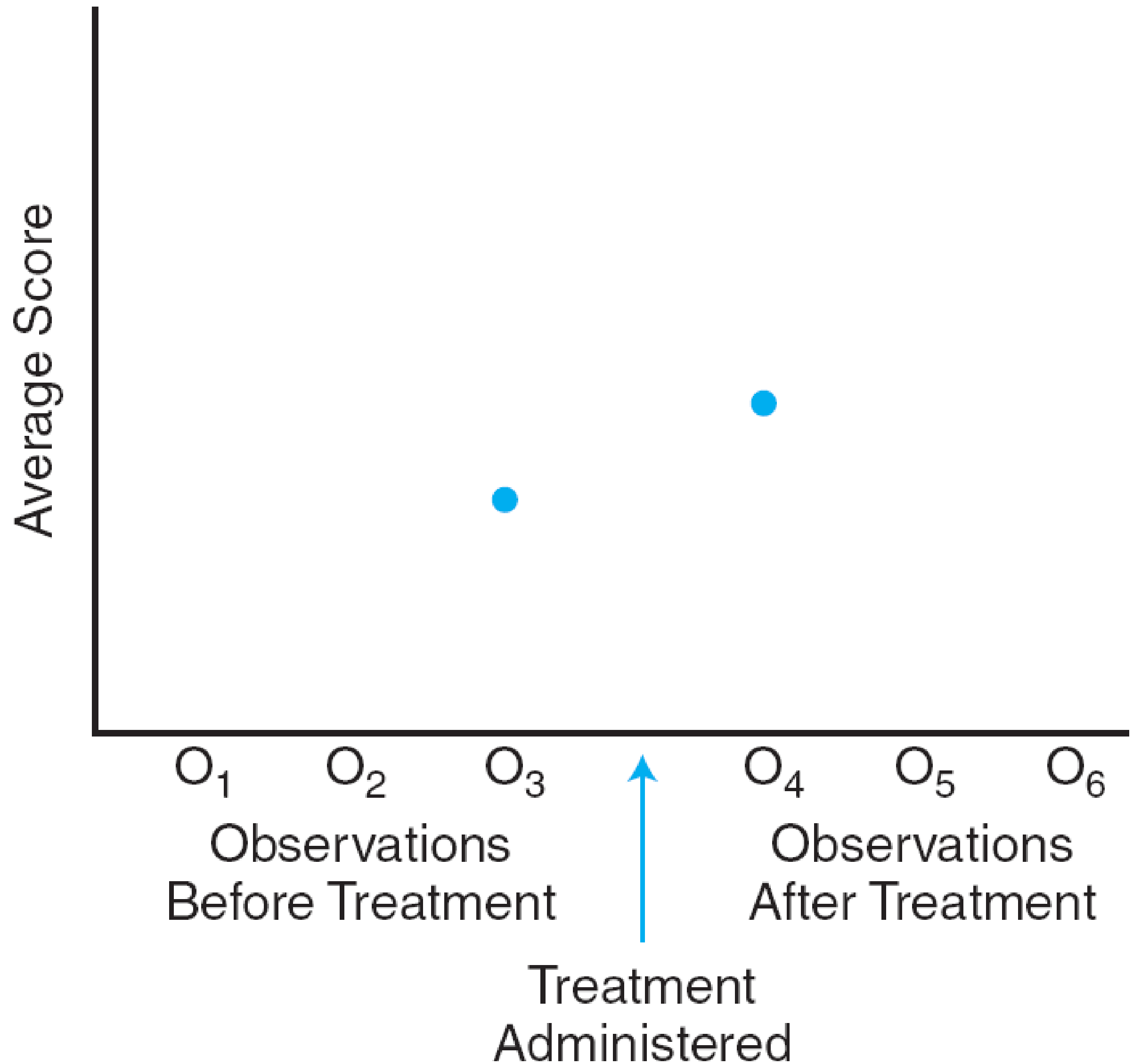
Time-Series & Interrupted Time-Series Designs

- Pre-observations allow for <u>trends</u> 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


Change in Level
Change in Slope
Change in Level and Slope
Delayed Change in Level
Delayed Change in Slope
Temporary Change
Abrupt Change and Decay to New Level
Intervention

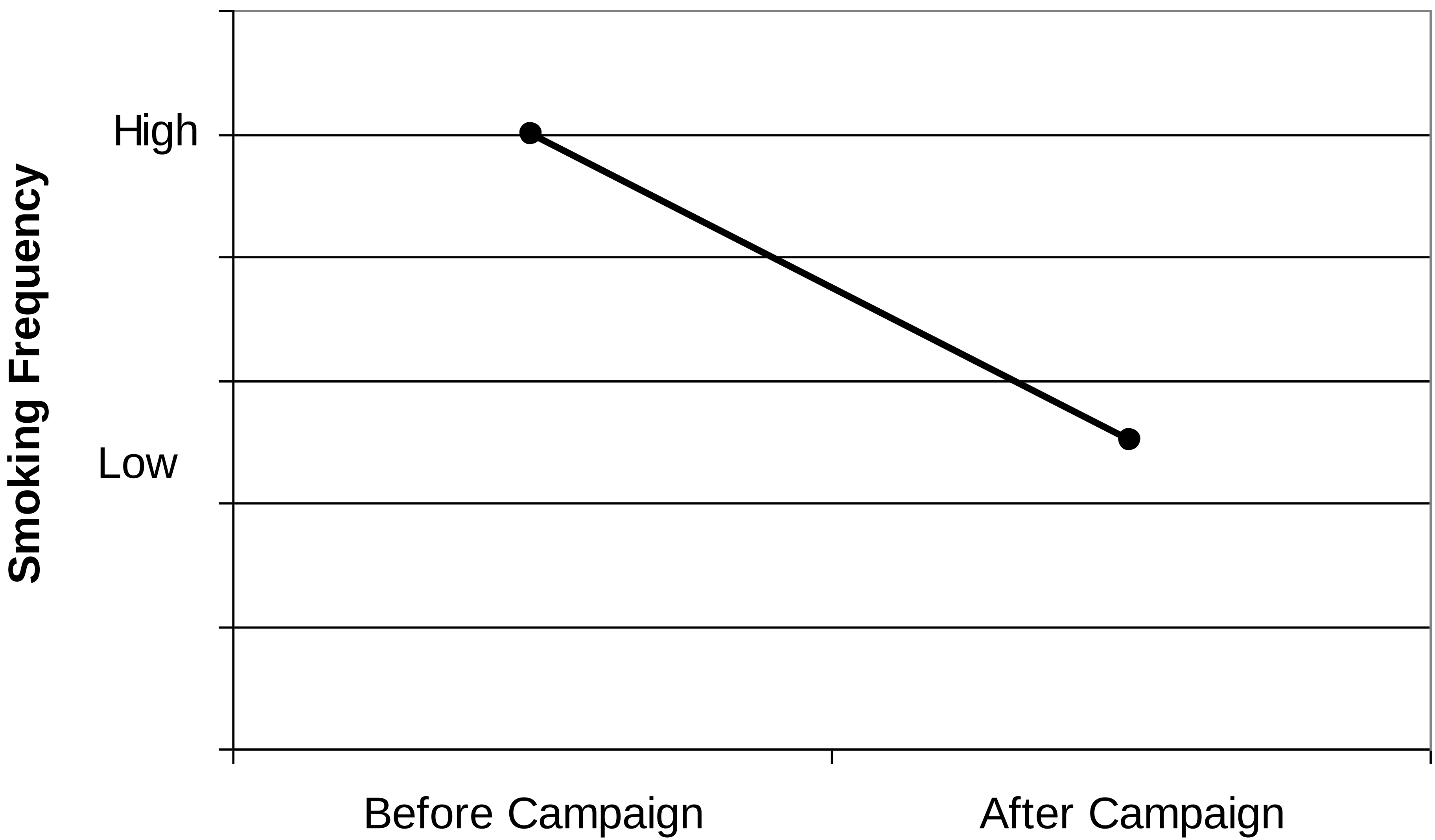


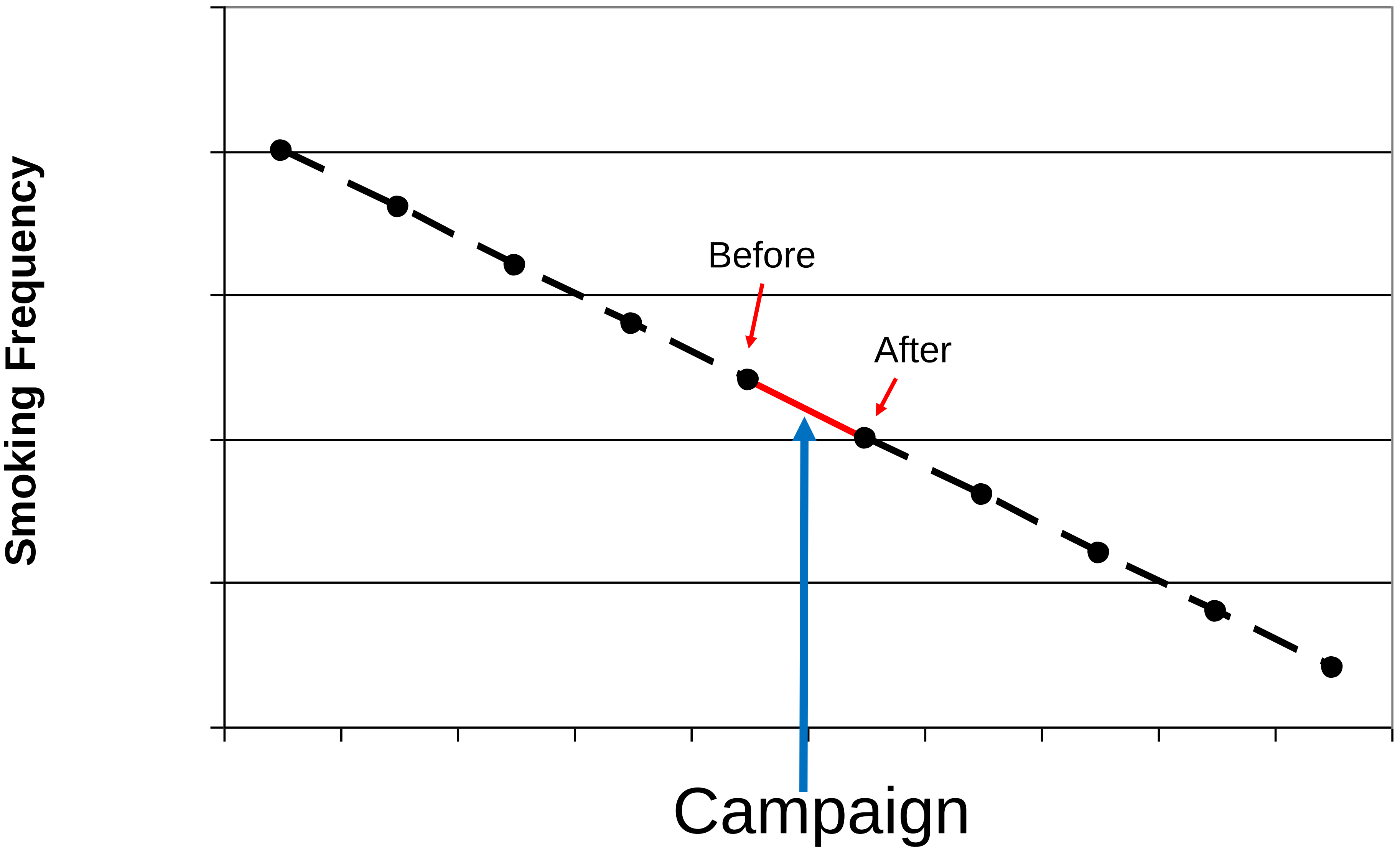
Can’t distinguish treatment effect from outside event



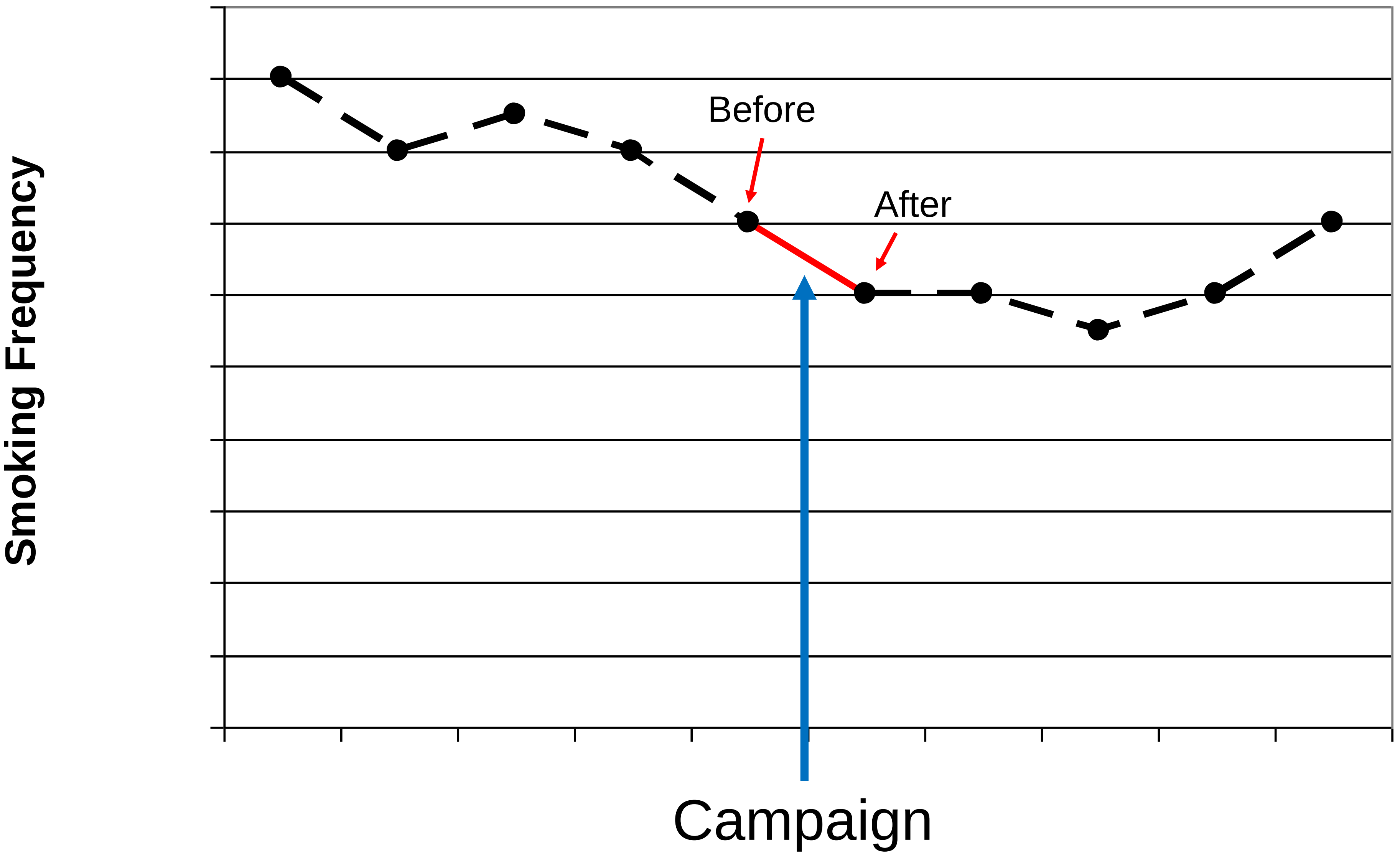
Scores consistently increasing
No evidence that the treatment had any effect

Anti-Smoking Program Example

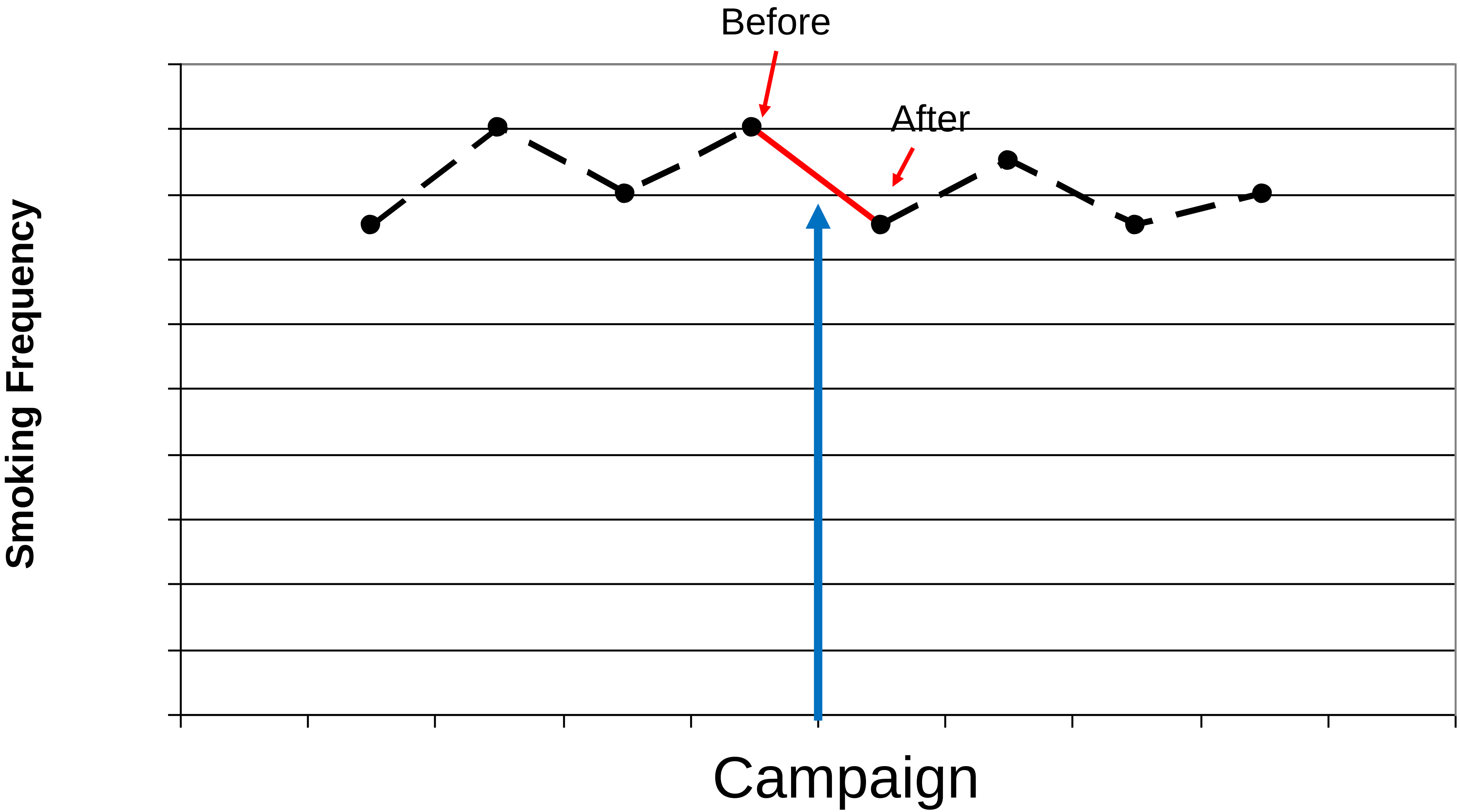




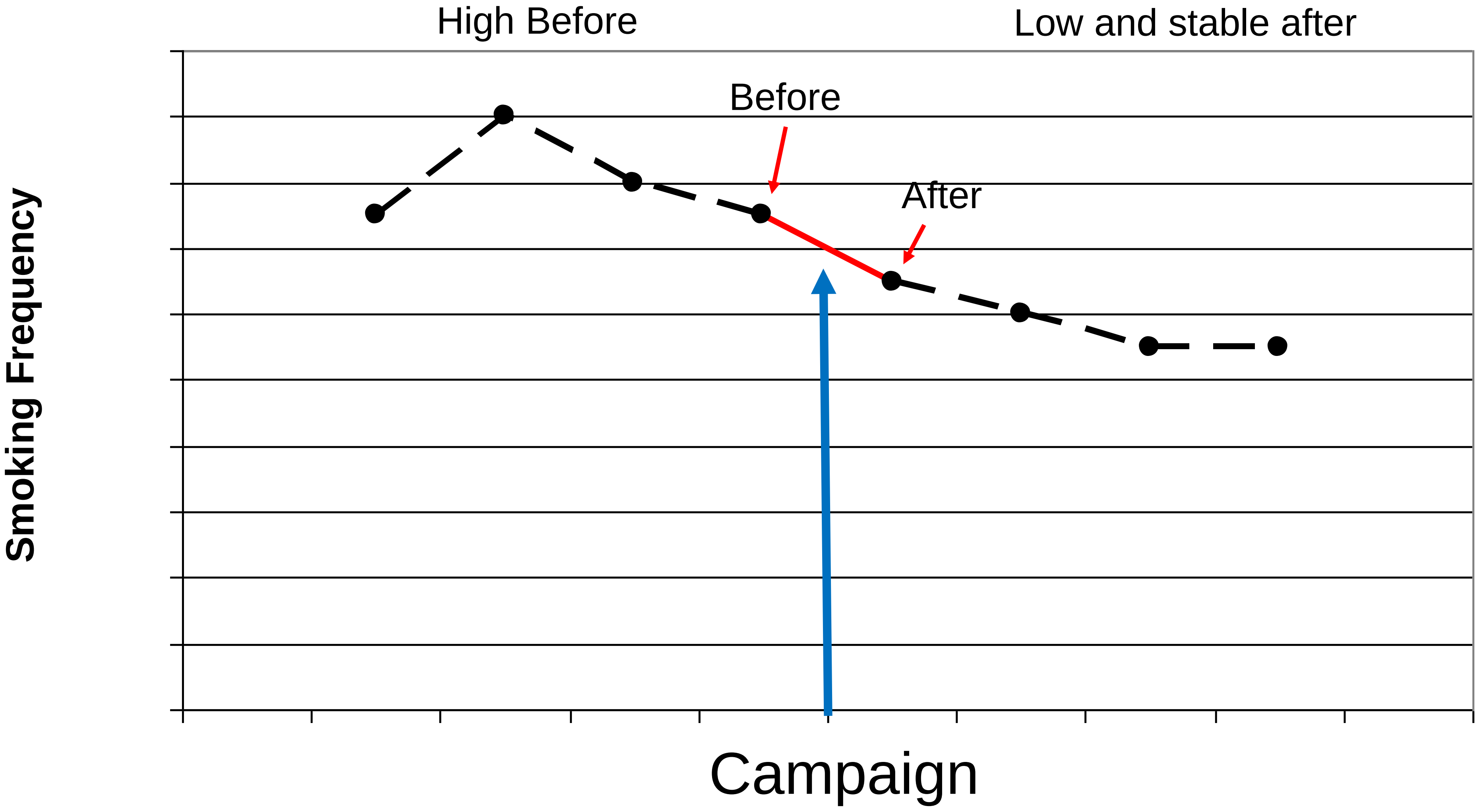
General trend toward reduced smoking



If the antismoking program had any effect it was short lived



General trend; periods of fluctuations



Ideal outcome

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)

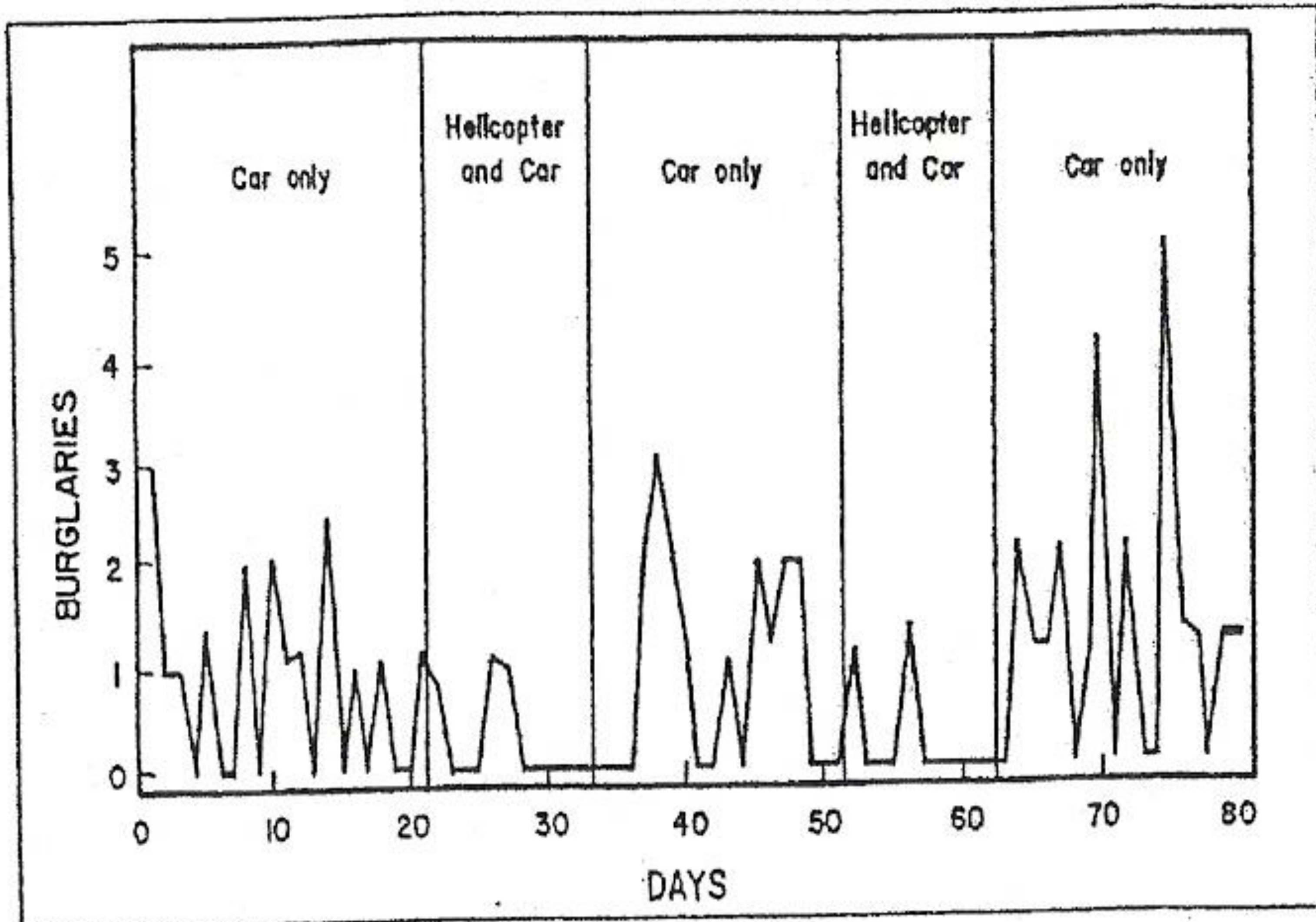
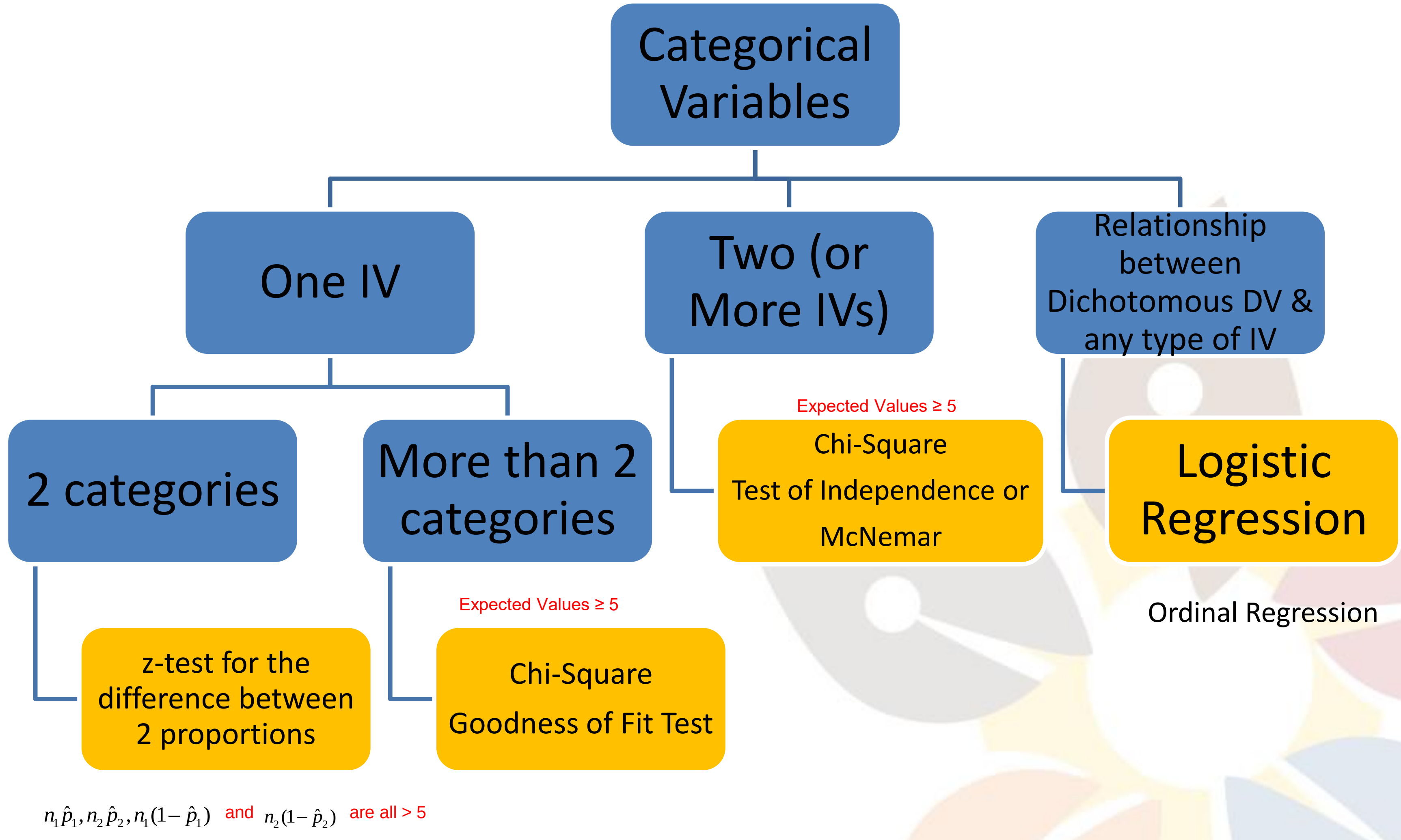


Figure 7.4. The Number of Home Burglaries When Either Police Cars or Both Police Cars and Helicopters Patrolled the Neighborhood

SOURCE: Adapted from Schnelle et al. (1978, p. 15) by permission.

Decision Tree for Statistical Tests



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 <u>independent</u>

- 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)
  - $b + c \geq 10$

McNemar Example

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

Correlation, *r*

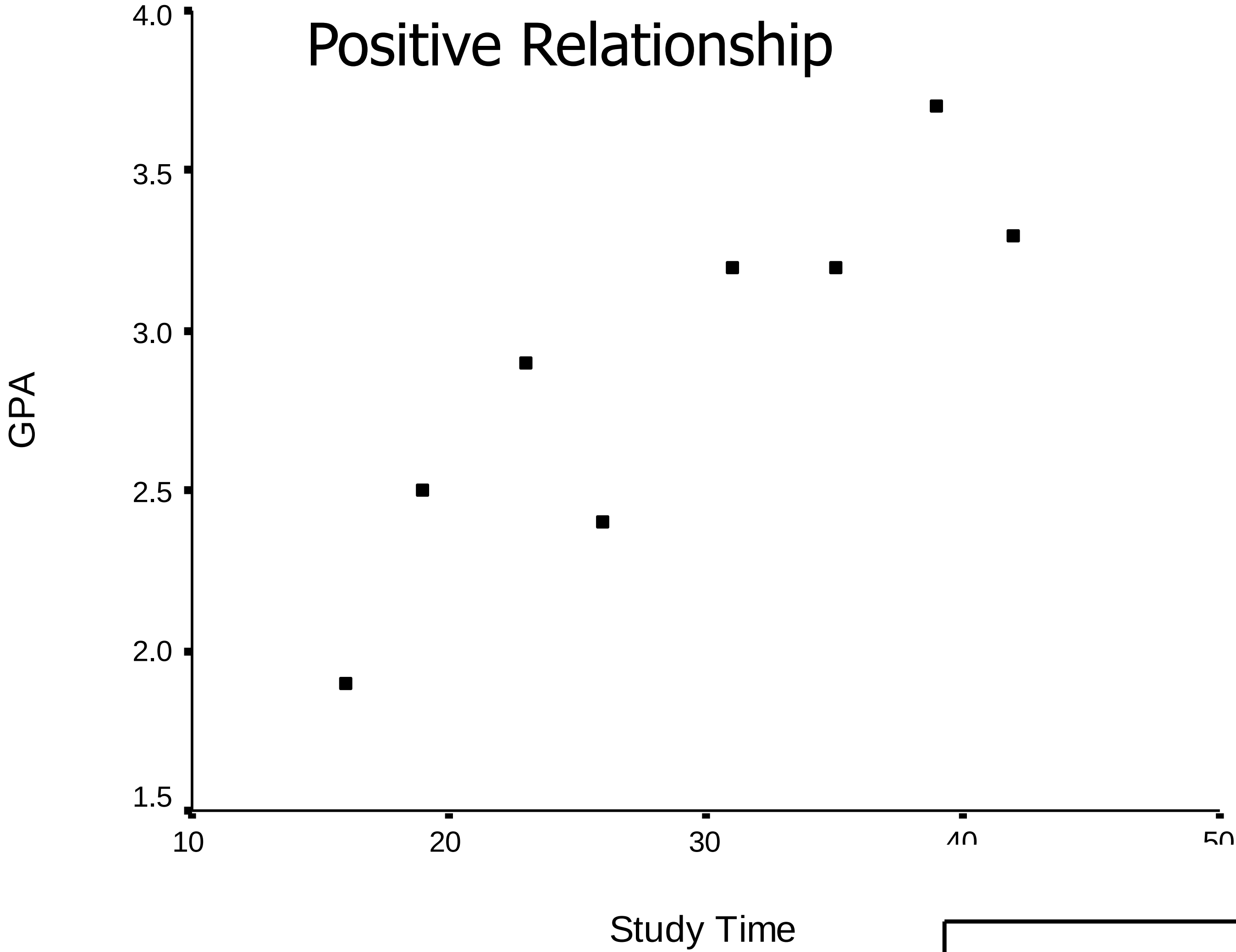
- Are the two continuous variables measured on the same people related?
- Assess the strength and direction (of linear relationships)

- 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)

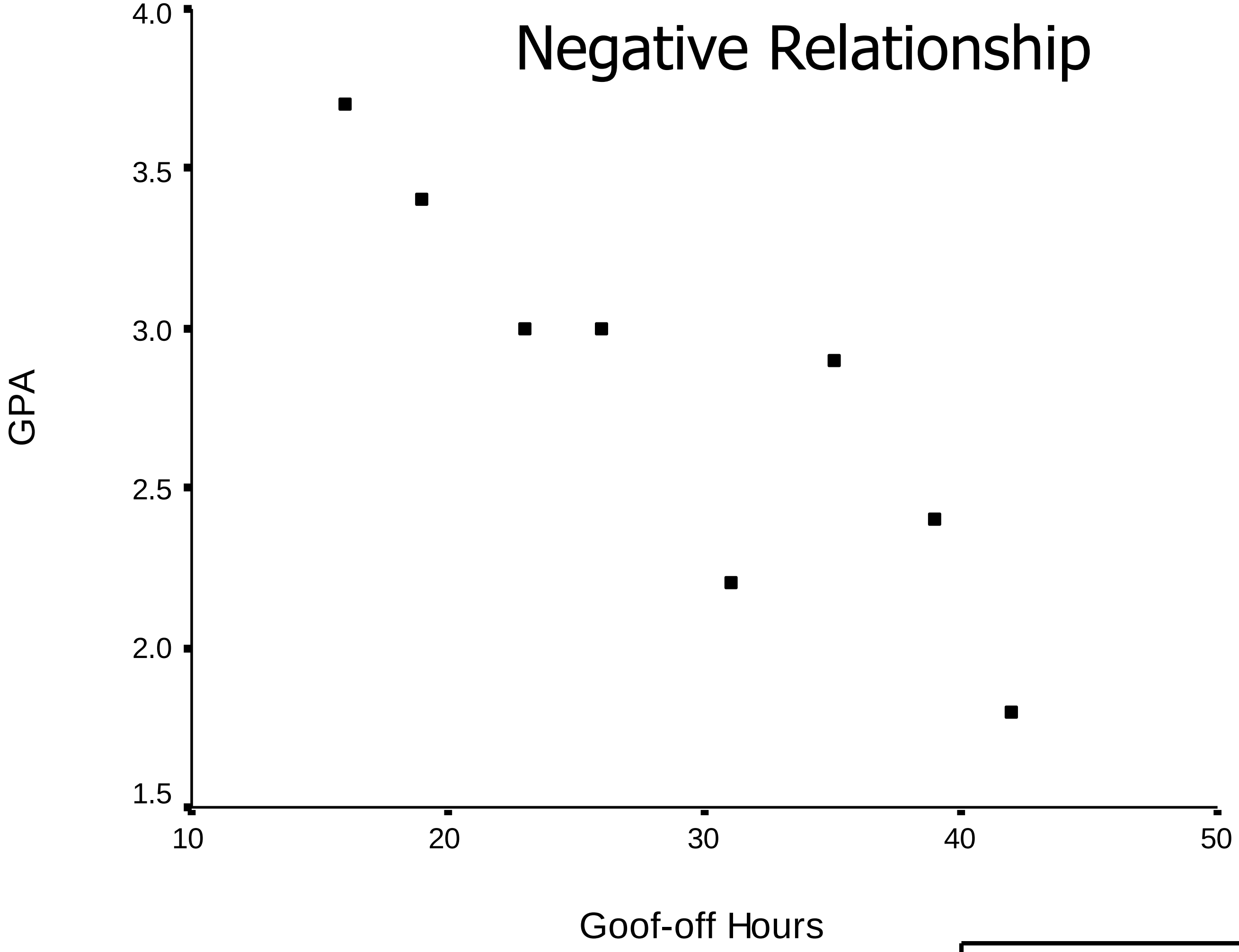
r

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



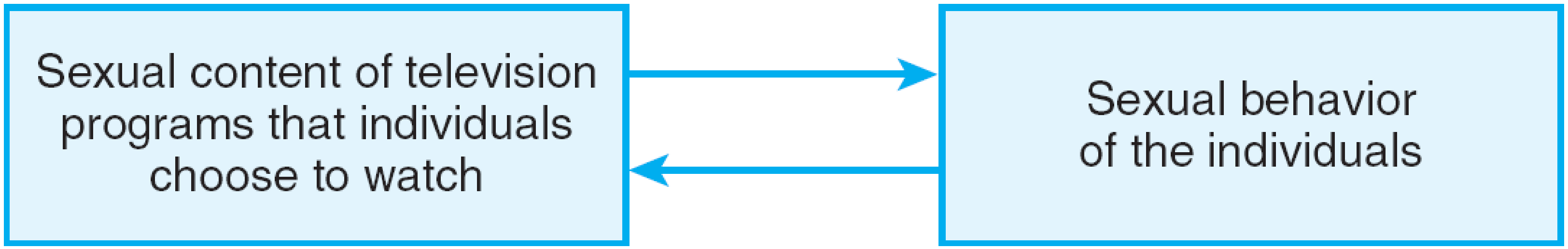
Correlations

		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


Sexual content of television programs that individuals choose to watch
Sexual behavior of the individuals


Variable A
Decision to participate
in fitness training program
Variable B
Days absent due
to illness

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, …

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?