Biostatistics: All About the Basics

Fern Jureidini Webb, Ph. D.

Associate Professor

University of Florida College of Medicine

Department of Community Health and Family Medicine

Department of Epidemiology

April 21 2016

Presented at NIH NIDDK NMRI's 14th Annual Workshop

Speaker's Disclosure

I, **Fern Jureidini Webb** have no conflict of interest to disclose.

Speaker's Instruction for Interaction

Go to: http://www.fernjwebb.participoll.com/

Today's Presentation

- Definition of Epidemiology
- Epidemiologic Research Cycle
- Study Factors/Variables
- Types of Data (Variables)
- Analysis and Evaluation
  - Measures of Frequency
    - Descriptive Analysis
  - Measures of Association
    - Statistical Analysis
    - Inferential Analysis
- Take Home Message
- Questions

http://www.fernjwebb.participoll.com/

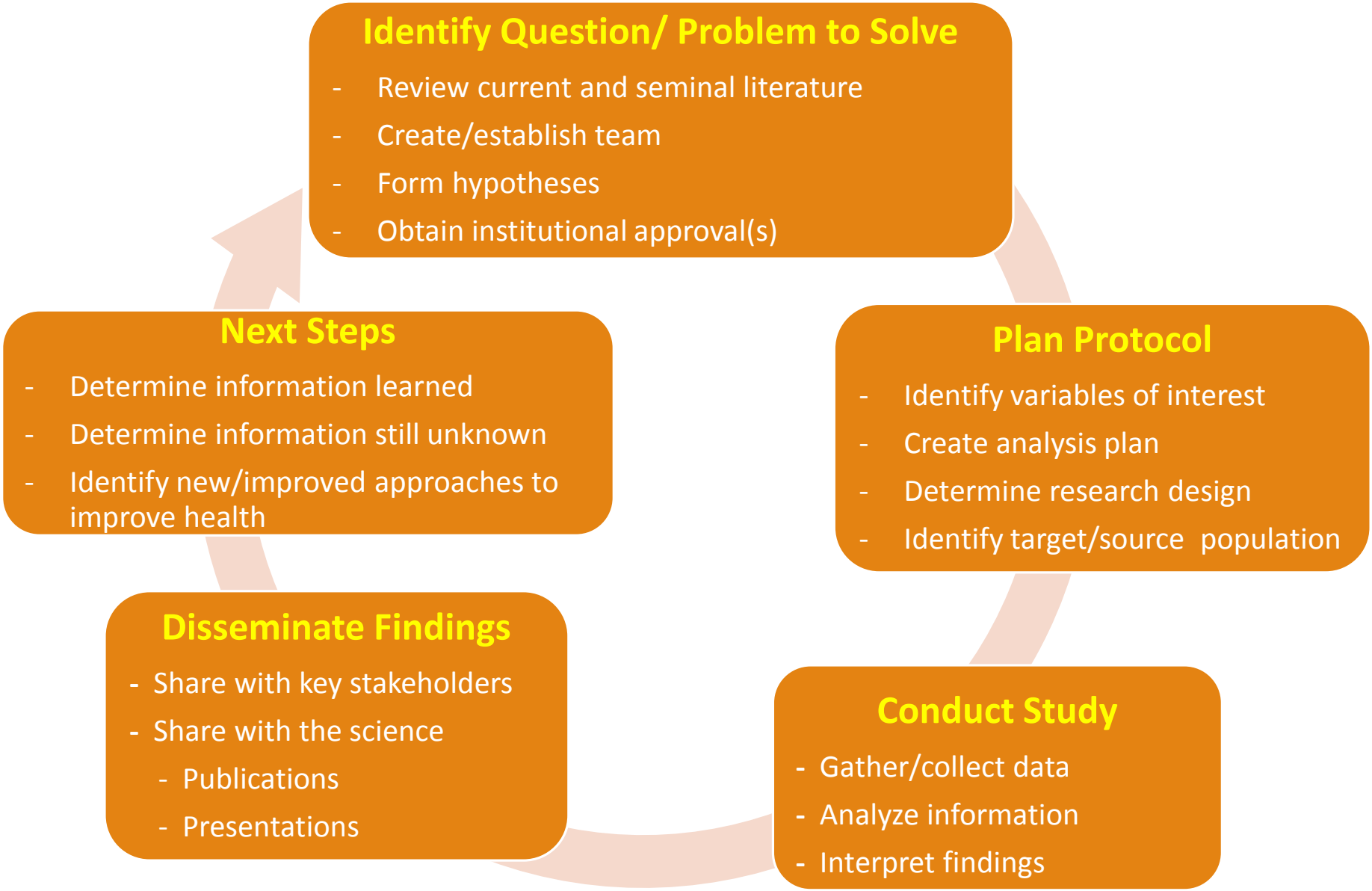
Definitions of Epidemiology

- A branch of medical sciences involving the analysis of the incidence, distribution and control of disease and/or health in a population
  - *Merriam-Webster online dictionary, 2015: http://www.merriam-webster.com/dictionary/epidemiology*

- The study of the distribution and determinants of disease frequency [and health in the population with the premise that disease and/or health are not random]
  - *Gordis L. Epidemiology, 2nd Edition. ISBN: 0-7216-8338-X*
  - *Hennekens C, Buring J. Epidemiology in Medicine, 1st Edition. ISBN: 0-316-35636-0*

Underlying assumption: disease or health distributions are not random events.

Epidemiologic Research Cycle



http://www.fernjwebb.participoll.com/

Epidemiologic Research Cycle

Plan Protocol

- **Identify variables of interest**
- Create analysis plan
- Determine research design
- Identify target/source population

http://www.fernjwebb.participoll.com/

Study Factors/Variables

What is (are) the **exposure(s)** of prime interest?

- How is (are) the **exposure(s)** defined?
- How is (are) the **exposure(s)** measured?

What is (are) the **outcome(s)** of prime interest?

- How is (are) the **outcome(s)** defined?
- How is (are) the **outcome(s)** measured?

http://www.fernjwebb.participoll.com/

Study Factors/Variables

We term the…

exposure(s) of interest:	**outcome(s)** of interest:
◦ Exposure	◦ Outcome
◦ Treatment	◦ Condition
◦ Independent	◦ Dependent
◦ Antecedent	◦ Consequent
◦ Predictor	◦ Response (or “Criterion”)

Types of Data (Variables)

- **Nominal data**
  - Unordered categories (i.e. ethnicity, gender, blood type)
  - No group/category is better/worse than the other

- **Ordinal data**
  - Ordered categories although distance between levels not exactly defined (i.e. excellent, very good, good, fair, poor)

- **Interval data**
  - Ordered and difference between points comparable
  - No ‘true’ zero (i.e. temperature)

- **Ratio data**
  - True zero point (i.e. cost, heart rate, blood pressure)
  - Defined difference/unit between values
  - Also called continuous

http://www.fernjwebb.participoll.com/

Let's Practice ~ Which are nominal data?

A. Blood pressure, weight, age, income

B. Gender, race, hair color, religion/faith

C. Pain measures, education level, satisfaction

D. Temperature, money trends/stock market

http://www.fernjwebb.participoll.com/

Let's Practice ~ Which are ordinal data?

A. Pain measures, education level, satisfaction

B. Blood pressure, weight, age, income

C. Gender, race, hair color, religion/faith

D. Temperature, money trends/stock market

http://www.fernjwebb.participoll.com/

Let's Practice ~ Which are interval data?

A. Blood pressure, weight, age, income

B. Gender, race, hair color, religion/faith

C. Temperature, money trends/stock market

D. Pain measures, education level, satisfaction

http://www.fernjwebb.participoll.com/

Let's Practice ~ Which are ratio data?

A. Gender, race, hair color, religion/faith

B. Blood pressure, weight, age, income

C. Pain measures, education level, satisfaction

D. Temperature, money trends/stock market

http://www.fernjwebb.participoll.com/

Epidemiologic Research Cycle

http://www.fernjwebb.participoll.com/

Analysis and Evaluation

Measures of Frequency

- **Descriptive Analysis**

Measures of Association

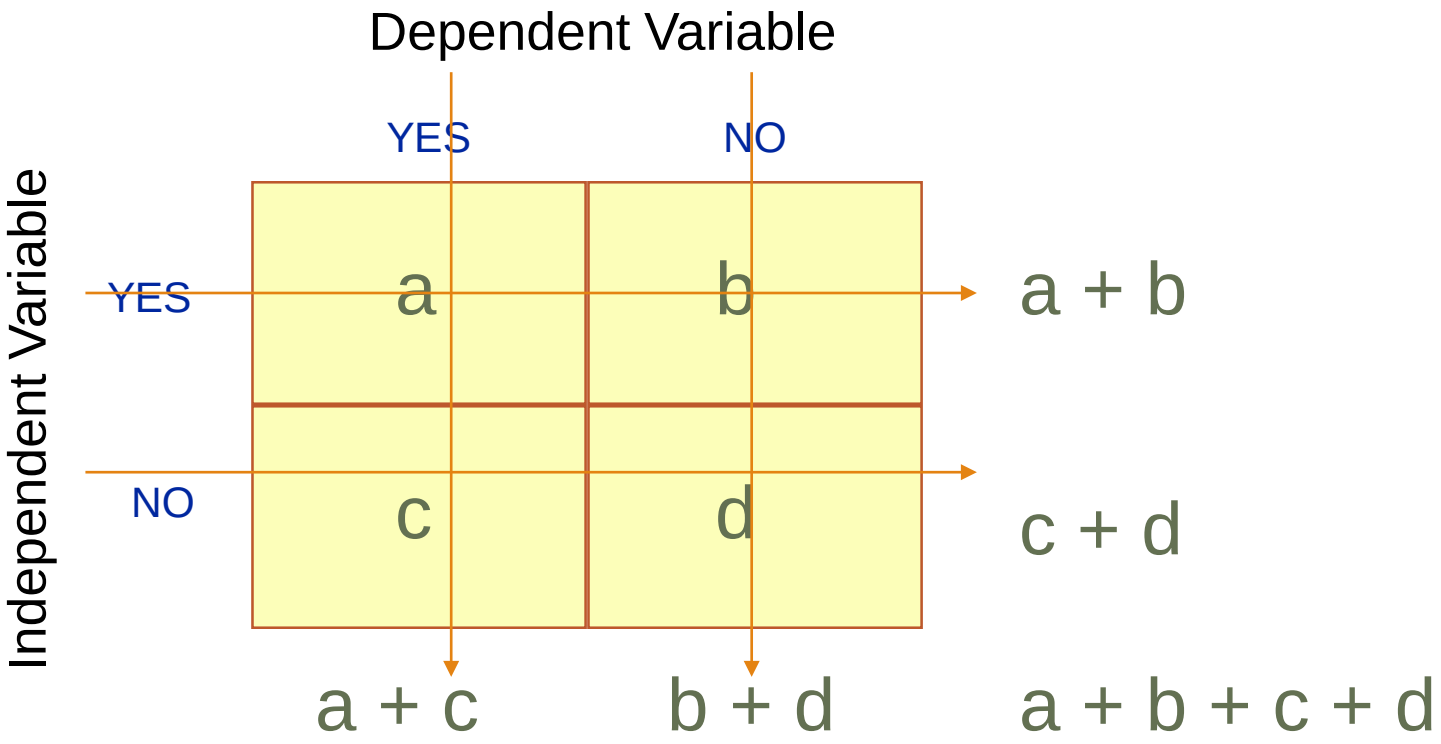
- **Statistical Analysis**
- **Inferential Analysis**

Measures of Frequency

Review of Epidemiologic Measures

http://www.fernjwebb.participoll.com/

The “2 x 2” table



What are some of the uses of the 2x2 table?

A. Measures of Frequency

B. Measures of Association

C. Measures of Screening

D. Hypothesis Testing

E. All of the Above

http://www.fernjwebb.participoll.com/

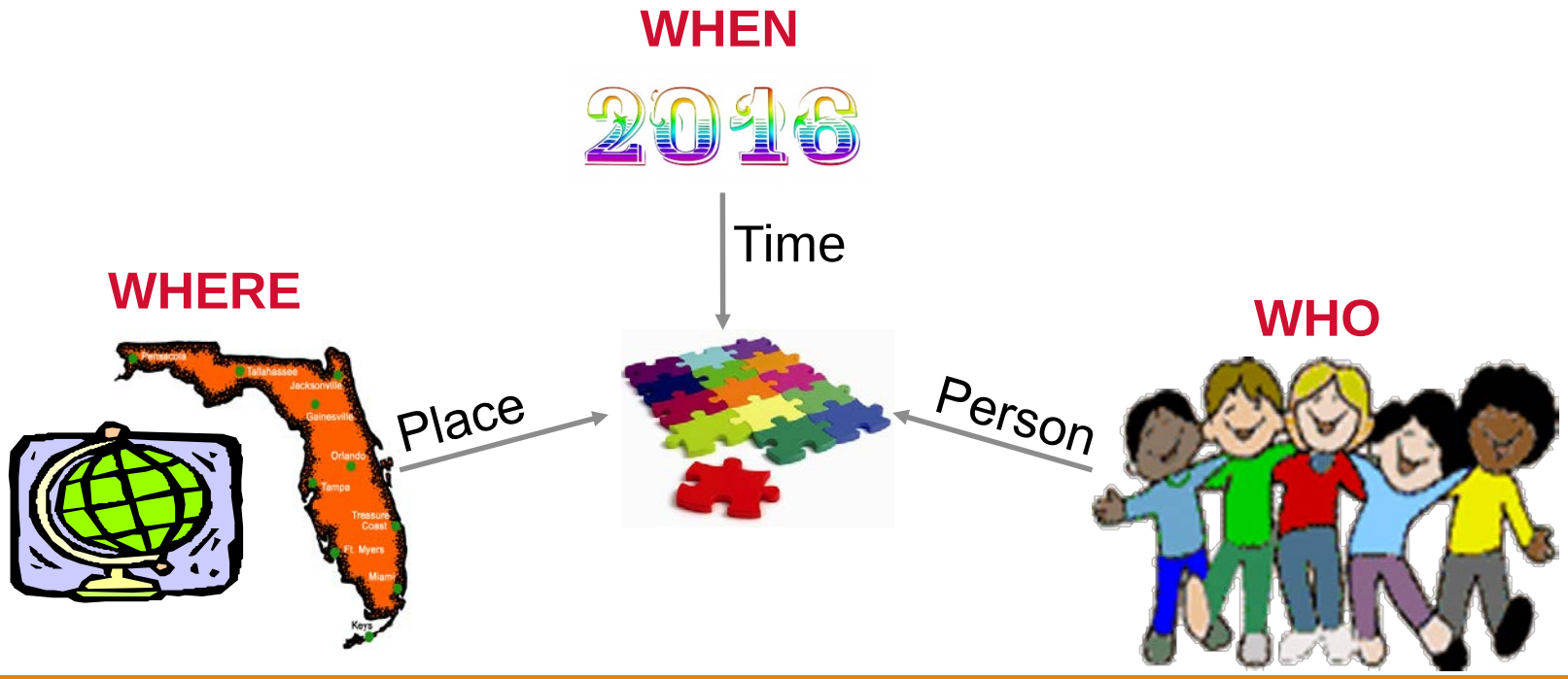
Basic Measures of Frequency

- Counts - n
- Proportions – a/(a+b) (no time period) - i.e., percent
- Rates - a/(a+b) per 1,000, 10,000, 100,000 over a specified period of time
- Ratios - a/b (numerator and denominator are mutually exclusive)

Analysis

Descriptive Analysis

Used to describe information (measured by "variables") or characteristics of those participating in our study



http://www.fernjwebb.participoll.com/

Selected sociodemographic characteristics of participants (*n*=292)*

	N	Percentage
Education		
None	2	< 1
GED/HS diploma	91	32
Some college/Associate's Degree	40	14
Bachelor's degree	95	33
Master's degree	49	17
Doctorate degree	11	4
Marital Status		
Married	114	40
Other (including single, divorced, widowed & separated)	173	60
	Mean	**Range**
Age	35	18-73
Household income	$30,000-$49,000	< $10K - ≥ $130K

* Webb F, Khubchandani J, Doldren M, Balls-Berry J, Blanchard S, et al. African-American Womens' Eating Habits and Intention to Change: A Pilot Study. *J Racial and Ethnic Health Disparities* June 2014 DOI: 10.1007/s40615-014-0026-2

http://www.fernjwebb.participoll.com/

Measures of Association

Review of Epidemiologic Measures

Used to describe how information (usually measured by variables) are associated or related to each other (variables)

Associations

Association:
The extent to which things occur together (non-directional)
OR
Statistical dependence between two variables:
(e.g., correlation between stages of change and weight)

Independent, x (risk factor, exposure, treatment [clinical trials])
Dependent, y (Outcome, event)

$$P(y) = x$$

In research, we would like to establish causal associations:
(uni-directional)

Measures of Association

Relative Risk
Risk Ratio
Odds Ratio

Disease/Outcome

Exposed	YES	NO	
YES	a	b	a + b
NO	c	d	c + d
	a + c	b + d	a + b + c + d

Analysis

Choose the Appropriate Statistic to Measure the Association based on:

- √ Type and number of independent variables:
  - Nominal, ordinal, interval, continuous/ratio
  - One variable or multiple

- √ Type and number of dependent variables:
  - Nominal, ordinal, interval, continuous/ratio
  - One variable or multiple

- √ Same for **<u>any type of design or study</u>**

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What is the primary independent variable?

A. Stages of Change

B. Weight Loss

C. Age, BMI, education, family health history, location, marital status, personal health

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What type of data is the **primary independent variable**?

A. Nominal

B. Ordinal

C. Interval

D. Ratio

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What is the dependent variable?

A. Stages of Change

B. Age, BMI, education, family health history, location, marital status, personal health

C. Weight Loss

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What type of data is the **dependent variable**?

A. Nominal

B. Ordinal

C. Interval

D. Ratio

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What type of data are the **other variables**?

A. Nominal

B. Ordinal

C. Ratio

D. Nominal, Ordinal & Ratio

http://www.fernjwebb.participoll.com/

Let's Practice!*

Examining the association between **stages of change** and **weight loss** (y/n). We will include age, body mass index [BMI], education, family health history (sum), healthy diet index (sum), location, marital status, personal health and life satisfaction in the model given their importance.

What statistic should we use to **measure this association**?

A. Analysis of Variance (ANOVA)

B. Multiple Regression

C. Logistic Regression

D. Chi-Square Test of Independence

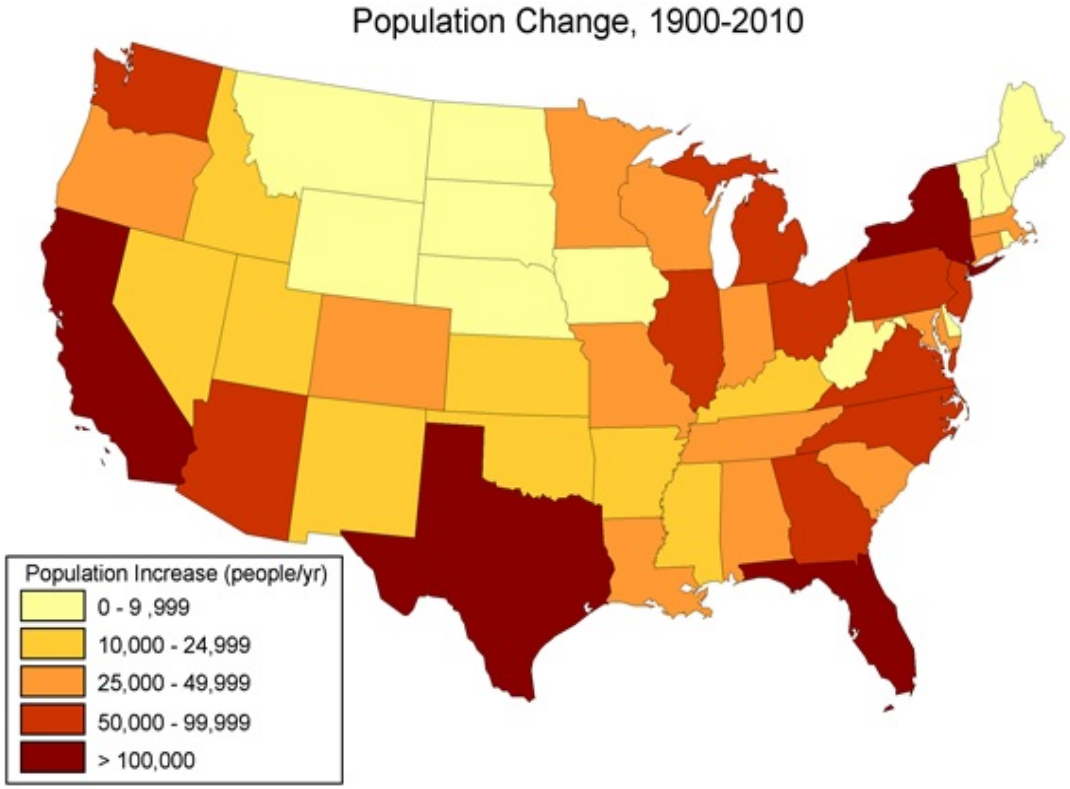
* *Hatcher L, Stepanski E. A step-by-step approach to using the SAS system for univariate and multivariate statistics. ISBN: 1-55544-634-5*

http://www.fernjwebb.participoll.com/

Analysis

Inferential Analysis

- How do data from our sample reflect truth in the population?



http://www.fernjwebb.participoll.com/

Inferential Analysis: Chance

Statistical methods that evaluate the role of chance is the same for **ANY/EVERY** study

- Hypothesis testing
- Estimation/Confidence Intervals

http://ocw.tufts.edu/Content/1/lecturenotes/194194

Hypothesis Testing

H_o = the null hypothesis. There is no association between stages of change and weight loss. Always start with the "null" ~ no difference!

H_A = the alternative hypothesis. There is an association between stages of change and weight loss.

There are four possible outcomes:

Null Hypothesis is:	"REALITY" True	"REALITY" False
Reject H_o	Type I error (P=α)	Correct 1- β
Fail to Reject H_o	Correct 1- α	Type II error (P=β)

Usually, α = .05
β = .20 or .10
Power = 1- β = .80 or .90

http://www.fernjwebb.participoll.com/

Estimation: Confidence Intervals

- **1.0 = no association**

If *p*- value > 0.05 (if set at 95%) or confidence interval (CI) includes 1
Fail to Reject H_o.

- **> 1.0 = Those with exposure have dependent variable/outcome more than those without exposure**

- **< 1.0 = Those with exposure have dependent variable/outcome less than those without exposure**

Relative Risk
Risk Ratio
Odds Ratio

If *p*- value < 0.05 (if set at 95%) or CI does not include 1
Statistically significant - Reject H_o.

***p*-value and CI should ALWAYS* give consistent findings!!**
*[* if based on same statistic]*

Let's Practice!*

Weight loss modeled as the dependent variable

**** These are fictitious data***

Which associations are **statistically significant**?

A. Stages of change, family health, personal health

B. Age, BMI, education, location

C. Healthy diet index, marital status, life satisfaction

D. A and C above

	Odds Ratio	Confidence Interval	*P*-value
Stages of Change	1.50	1.10-2.03	0.02
Age	1.00	0.97-1.04	0.08
BMI	1.01	0.96-1.06	0.07
Education	1.34	0.65-2.74	0.11
Family Health	2.10	1.87-3.39	0.03
Healthy Diet Index	3.10	2.81-4.27	0.001
Life Satisfaction	2.21	1.86-4.86	0.02
Location	1.01	0.78-1.32	0.21
Marital Status	0.76	0.36-0.92	0.01
Personal Health	0.73	0.65-0.96	0.01

http://www.fernjwebb.participoll.com/

Let's Practice!*

Weight loss modeled as the dependent variable

**** These are fictitious data***

Which associations are **not statistically significant**?

A. Stages of change, family health, personal health

B. Age, BMI, education, location

C. Healthy diet index, marital status, life satisfaction

D. A and C above

	Odds Ratio	Confidence Interval	*P*-value
Stages of Change	1.50	1.10-2.03	0.02
Age	1.00	0.97-1.04	0.08
BMI	1.01	0.96-1.06	0.07
Education	1.34	0.65-2.74	0.11
Family Health	2.10	1.87-3.39	0.03
Healthy Diet Index	3.10	2.81-4.27	0.001
Life Satisfaction	2.21	1.86-4.86	0.02
Location	1.01	0.78-1.32	0.21
Marital Status	0.76	0.36-0.92	0.01
Personal Health	0.73	0.65-0.96	0.01

http://www.fernjwebb.participoll.com/

Let's Practice!*

If you saw this table,

Which association might you question?

A. Stages of change

B. Education

C. Life Satisfaction

D. Personal Health

	Odds Ratio	Confidence Interval	*P*-value
Stages of Change	1.50	1.10-2.03	0.02
Age	1.00	0.97-1.04	0.08
BMI	1.01	0.96-1.06	0.07
Education	1.34	0.65-2.74	0.11
Family Health	2.10	1.87-3.39	0.03
Healthy Diet Index	3.10	2.81-4.27	0.001
Life Satisfaction	2.21	1.86-4.86	0.09
Location	1.01	0.78-1.32	0.21
Marital Status	0.76	0.36-0.92	0.01
Personal Health	0.73	0.65-0.96	0.01

http://www.fernjwebb.participoll.com/

Take Home Message

- Choose a measure of association based on data/variable type for independent and dependent variables!
- Use your “cheat sheet” – no need to guess or memorize!
- Consult with a biostatistician/statistical expert during the planning phase of your study ***before*** you finalize design and begin conducting your study!!