CPH Exam Review Webinar

Evidence-based Approaches to Public Health

Biostatistics

CPH Certified in Public Health

by National Board of Public Health Examiners

ASSOCIATION OF ASPPH SCHOOLS & PROGRAMS OF PUBLIC HEALTH

NATIONAL BOARD OF PUBLIC HEALTH EXAMINERS

Credentialing Public Health Leaders

CPH Study Resources

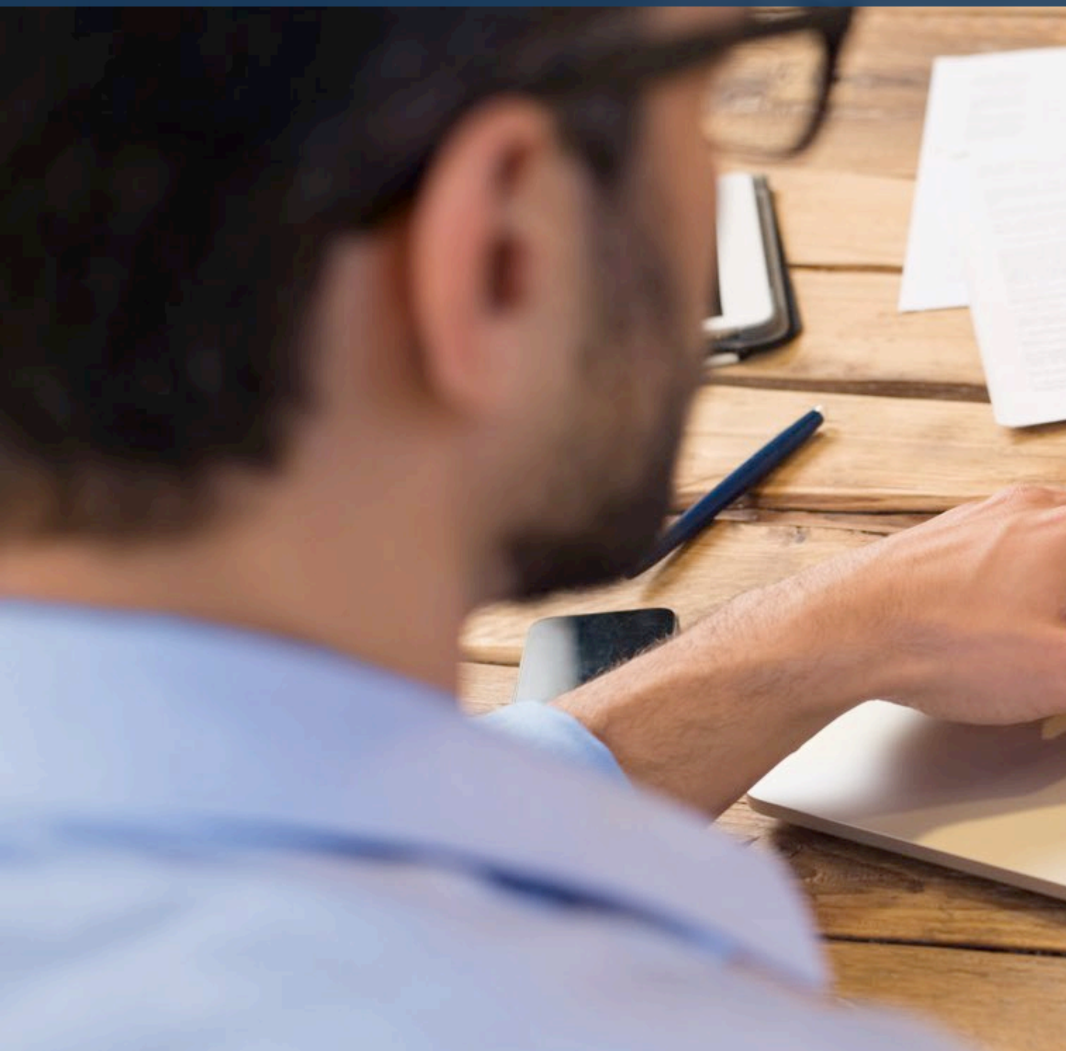
1. Content Outline
2. Sample Exam Questions
3. Practice Exams
4. Webinars
5. ASPPH Study Guide
6. APHA Study Guide

www.nbphe.org/cph-study-resources/

Content Outline

- Evidence-based Approaches to Public Health (10%)
- Communication (10%)
- Leadership (10%)
- Law and Ethics (10%)
- Public Health Biology and Human Disease Risk (10%)
- Collaboration and Partnership (10%)
- Program Planning and Evaluation (10%)
- Program Management (10%)
- Policy in Public Health (10%)
- Health Equity and Social Justice (10%)

Sample Exam Questions



Sample questions in
the format of
the CPH exam

Practice Exams

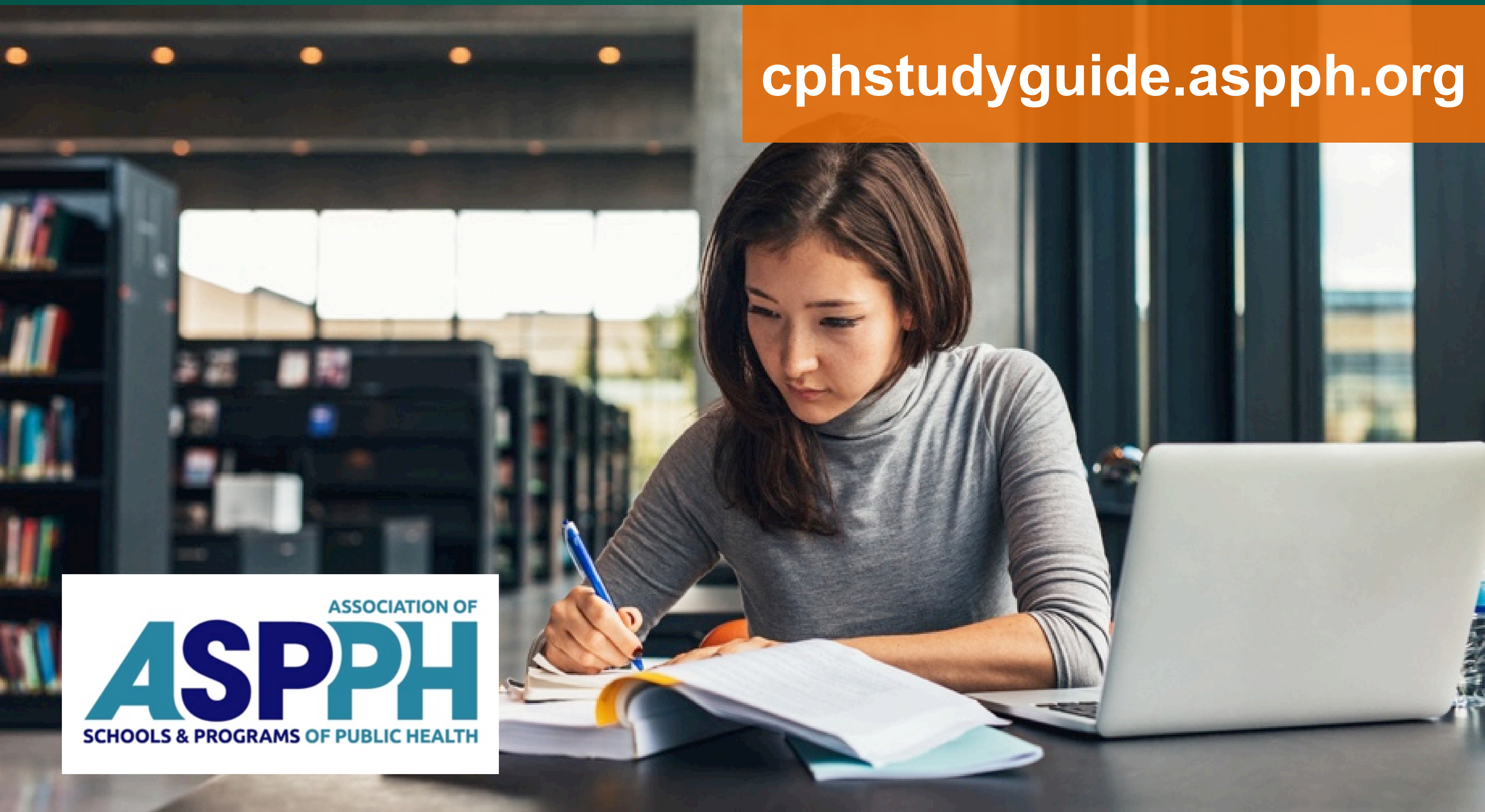
Online mini-exam of 50 questions from the CPH item-bank

Study Webinars

Additional Webinars
Lecture and Q&A

Today's webinar and all past webinars and presentations are posted on https://www.nbphe.org/cph-study-resources/

ASPPH CPH Study Guide



APHA Press Study Guide

APHA
AMERICAN PUBLIC HEALTH ASSOCIATION
For science. For action. For health.

Editors: Karen Liller, Jaime Corvin and Hari Venkatachalam
University of South Florida College of Public Health
Certified in Public Health Exam Review Guide
$41.95 APHA member /$51.95 non-member
eBook and print available via the APHA Bookstore at https://www.apha.org/publications-and-periodicals


Let’s Get Started!

Evidence-based Approaches to Public Health

Biostatistics

Danielle Fastring, PhD, MS, MPH
Assistant Teaching Professor
University of Southern Mississippi
College of Nursing and Health Professions
Department of Public Health

ASPPH CPH Exam Review Webinar Series

October 22, 2019

CPH Certified in Public Health
by National Board of Public Health Examiners

Association of ASPPH Schools & Programs of Public Health

National Board of Public Health Examiners
Credentialing Public Health Leaders

Learning Objectives

- Learning Objectives:
- Analyze quantitative and qualitative data
- Interpret results of data analysis for public health research, policy or practice

Abstract:

- This section will continue the discussion of evidence-based approaches to public health.
- Specifically, this section will provide a more detailed discussion of biostatistical concepts that will be critical to success as a public health professional.

Basic Concepts

- What is statistics?
  - Field of study concerned with
    - 1) the collection, organization, summarization, and analysis of data (descriptive statistics)
    - 2) the drawing of inferences about a body of data when only a part of the data is observed (inferential statistics)
- What is biostatistics?
  - A branch of statistics where the data analyzed are derived from biological sciences and medicine

Basic Concepts

- Biostatistics
  - Population:
    - the largest collection of entities for which we have an interest at a particular time
    - Populations are defined by our sphere of interest
    - Parameter: Characteristic or measure obtained from a population.
  - Sample
    - A subgroup or subset of the population
    - Statistic: Characteristic or measure obtained from a sample.
  - We compute statistics and use them to estimate parameters.

Basic Concepts

- Variable: Characteristic or attribute that can assume different values (it varies depending the person, place, or thing being characterized).
  - Ex. Height, weight, blood pressure, heart rate, level of satisfaction with wait time at an ER
  - Ex. Height is 64 inches (data = 64, variable = height)

Basic Concepts

- Variables can be quantitative or qualitative
  - Quantitative: one that can be measured; conveys information regarding amount (how many)
  - Qualitative: Variables that assume non-numerical values; conveys information about an attribute (race, marital status, ethnicity)

Basic Concepts

- Statistical Inference:
  - The procedure by which we reach a conclusion about a population on the basis of the information contained in a sample that has been drawn from that population
  - We generalize findings back to the population from which the sample was drawn
  - We want to learn more about a population and we need to draw a sample that is representative of that population. If we misrepresent the population in our sample, we introduce bias and can make the wrong inference

Measurement and Measurement Scales

- Measurement: The assignment of numbers to objects or events according to a set of rules
  - Nominal Scale: The lowest measurement scale. Used for naming or labeling. Although it can be done with numbers, the relationship between the numbers are not meaningful. (Categorical and Dichotomous Variables)
    - Ex. Marital Status, DL number, SS number

Measurement and Measurement Scales

- Ordinal Scale: Observations are ranked according to some criterion. We can order the measurements/categories but we don't know the distance between two ranks
  - Ex. Low, Med, High SES; unimproved, improved, much improved, Likert-type scale

Measurement and Measurement Scales

- Interval Scale: Level of measurement which classifies data that can be ranked and differences are meaningful. However, there is no meaningful zero, so ratios are meaningless. The scale is relative. (Continuous Variable)
  - ex. Temperature in º F or ºC (for these scales, 0 does not represent the absence of heat. The scale is relative. pH (a pH of 0 does not mean the absence of acidity) No true zero point for these variables.

Measurement and Measurement Scales

- Ratio Scale: Level of measurement which classifies data that can be ranked, differences are meaningful, and there is a true zero. True ratios exist between the different units of measure.
  - Ex. Height, length, Kelvin Temperature scale.
- Note: For a temperature scale to be a ratio scale, zero must NOT be arbitrary. If zero is defined as the temperature where molecular motion stops (absolute zero), then a ratio scale for temperature can be defined.
- The Kelvin temperature scale defines zero (0 K) this way. Absolute zero is -273.15 °C or -459.67 °F. Because the zero point for these two temperature scales is arbitrary, they are interval scale.

Ex. 1 Measurement Scale

- A researcher is designing a new questionnaire to examine patient stress levels on a scale of 0 – 5. What scale of measurement is being used for the outcome variable?

Review of Measures of Disease Frequency

- Count
- Ratio
- Proportion
- Rate

Measures of Disease Frequency

- Count: refers to the number of cases of a disease or other health phenomenon being studied.
  - Ex. College dorm students who had Hepatitis
- Proportion: A measure that states a count relative to the size of the group. The upper part (numerator) of the fraction is the piece, the lower part (denominator) is the whole.
  - Ex. College dorm students who had Hepatitis B/ all dorm students
- Ratio: Divide one number into another number, but the numerator does not have to be a subset of the denominator
  - Ex. College dorm students who had Hepatitis B / College dorm students who had Hepatitis A
- Rate: Similar to proportions and ratios, but includes a time component
  - % of dorm students who had Hepatitis B in 2018

Ex. 2 Measures of Disease Frequency

- Calculate the proportion of African-American male deaths among African-American and Caucasian boys aged 5 to 14 years.

A	B	Total (A + B)
Number of deaths among African-American boys	Number of deaths among Caucasian boys	Total
1,150	3,810	4,960
Proportion = A/(A + B) x 100 = (1,150/4,960) x100 = 23.2%		

Example 3: Measures of Disease Frequency

- Which of the following is expressed as a count?
- Which of the following is expressed as a proportion?
- Which of the following is expressed as a ratio?
- Which of the following is expressed as a rate?

A. Male Births / Female Births

B. Female Births

C. Female Births / Male + Female Births

D. Male Births + Female Births in 2019 / Population at midpoint of 2019

Review of Disease Distribution Terms

- Epidemic – Disease in excess of expected
  - Think : epidemic threshold
- Endemic – Habitual presence of disease within an area
  - Think: epidemic threshold
- Pandemic – Worldwide epidemic
  - Think: epidemic threshold + geography

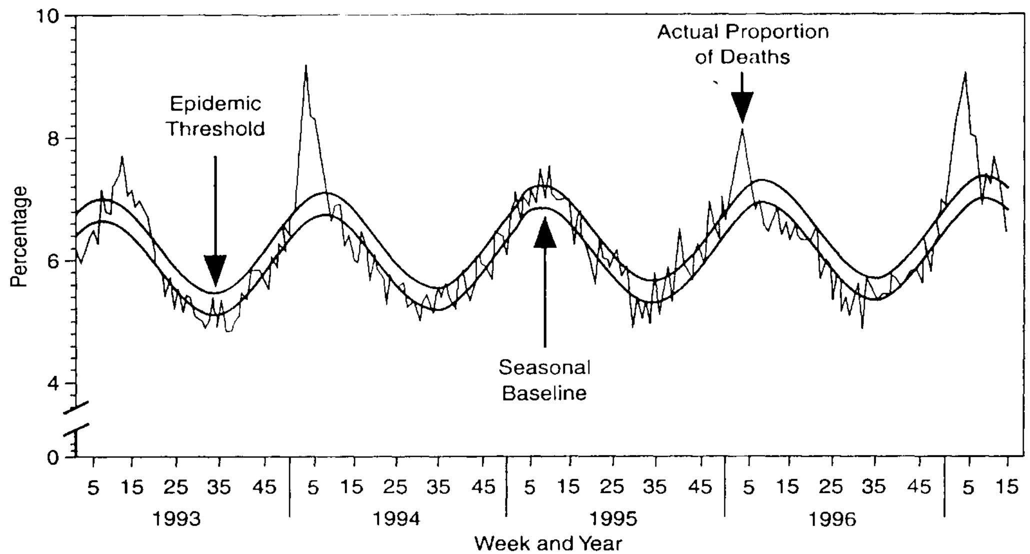


Figure 1–2 Weekly pneumonia and influenza (P&I) mortality as a percentage of all deaths in 122 cities—United States, January 3, 1993–April 5, 1997. *Source:* Reprinted from Influenza Activity, *MMWR*, Vol 46, p 327, 1997, Centers for Disease Control and Prevention.

Review of Study Design and Measures of Association

- Descriptive Studies:
  - Case studies / reports
  - Cross-Sectional Studies
  - Ecological Studies
- Analytical Studies:
  - Case-Control studies
  - Cohort Studies
  - Randomized Control Studies

Case Control vs Cohort

Cohort Study

- Begin with a group of people who are disease free at baseline
- Follow them over time and record exposures
- Classified on exposure
- Incident cases
- MOA: Relative Risk
- Rare exposures
- Prevalent diseases

Case Control

- Compare Disease + (Cases) to Disease – (Controls)
- Collect exposure data retrospectively
- Classified on disease status
- Prevalent cases
- MOA: Odds Ratio
- Rare Diseases

Interpreting RR and OR

- If RR or OR = 1: no association between exposure and outcome
- If OR > 1: exposure increases the risk of the outcome
- If OR < 1: exposure decreases the risk of the outcome
- OR and RR range from 0-1 (no negatives)

Interpreting RR and OR

- OR of 2.0: Think: Case Control Study..organized on disease status. Begin conclusion with outcome
  - The odds of having diabetes are 2 times higher among those who are sedentary compared to those who are not sedentary
  - Those who have diabetes are 2 times more likely to be sedentary than those without diabetes
- RR of 2.0: Think: Cohort Study…organized on exposure. Begin conclusion with exposure.
    - Those who are sedentary have 2 times the risk of diabetes as those who are not sedentary

Ex. 4 Positive Association

- A case-control study comparing ovarian cancer cases with community controls found an odds ratio of 2.0 in relation to exposure to radiation. Which is the correct interpretation of the measure of association?

A. Women exposed to radiation had 2.0 times the risk of ovarian cancer when compared to women not exposed to radiation

B. Women exposed to radiation had 2.0 times the risk of ovarian cancer when compared to women without ovarian cancer

C. Ovarian cancer cases had 2.0 times the odds of exposure to radiation when compared to controls

D. Ovarian cancer cases had 2.0 times the odds of exposure to radiation when compared to controls with other cancers

Ex. 5 Protective Association

- Interpret an OR of .75 if the exposure is eating kale and the outcome is obesity
  - Think…OR…Case Control..organized on disease status
  - Those who are obese are .75 times as likely to have eaten kale than those who are not obese
  - The odds of developing obesity are .75 times as high among those who ate kale compared to those who did not eat kale
- Interpret if it were a RR
  - Think…RR…Cohort study…organized on exposure
  - Those who ate kale are .75 times as likely to become obese than those who did not eat kale

Another Cohort Metric: Attributable Risk

- Whereas RR tells you the strength of an association, attributable risk tells you how much of the disease that occurs can be attributed to a certain exposure.
- We can calculate the attributable risk among exposed individuals or for an entire population.
- Background risk: the risk of non-exposed people is not zero.
  - Ex. We can attribute lung caner to smoking, but some people who don't smoke still get lung cancer. The risk in our unexposed population is background risk.

Attributable Risk Calculation

Bladder Cancer

Use of Artificial Sweetener

	Yes	No	Totals	Incidence per 1,000 per year
Yes	55	2445	2500	22
No	39	2961	3000	13

AR = (Incidence in exposed) – (Incidence in unexposed)

Incidence Rates: exposed → a/a+b = 55/2500 = 22 per 1,000

not exposed→ c/c+d = 39/3000 = 13 per 1000

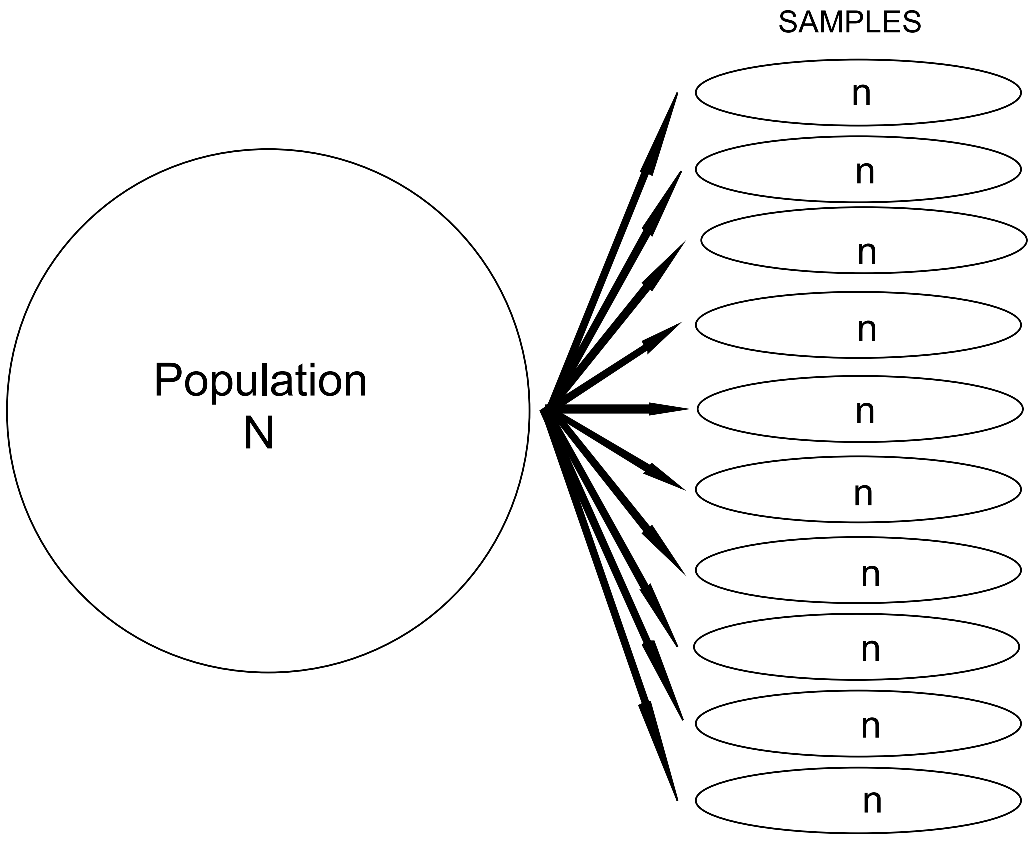
AR = (22-13)/ 1000 = 9 incident cases per 1,000 population of bladder cancer are attributed to use of artificial sweetener

If we had an effective educational campaign, we would hope to prevent 9 of the 22/1000 incident cases of bladder cancer that people who use AS would experience

CPH Certified in Public Health
by National Board of Public Health Examiners

Sampling and Probability Distributions

Sampling from a Population



Probability Sampling

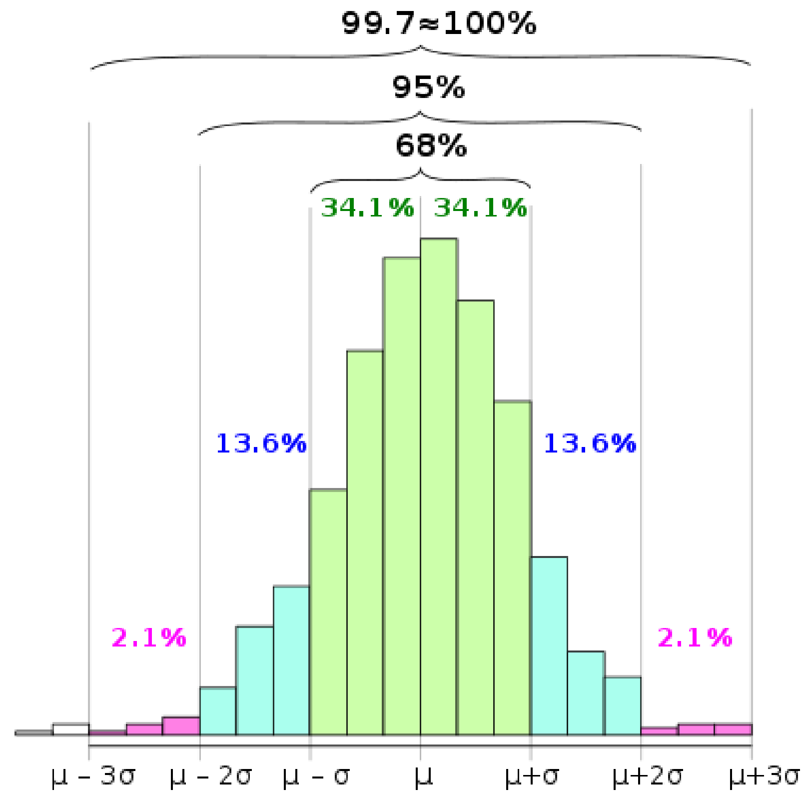
- Simple random sample
  - Enumerate all members of population N (sampling frame), select n individuals at random (each has the same probability of being selected)
- Systematic sample
  - Start with sampling frame; determine sampling interval (N/n); select first person at random from first (N/n) and every (N/n) thereafter
- Stratified sample
  - Organize population into mutually exclusive strata; select individuals at random within each stratum

Probability Distribution: Binomial Distribution

- Binomial distribution: model for discrete outcome
  - Process or experiment has 2 possible outcomes: success and failure
  - Replications of process are independent
  - P(success) is constant for each replication

Probability Distribution: Normal Distribution

- Notation: m=mean s=st. dev.
- Mean, Median, Mode are the same and located at the center of the distribution (not skewed)



Ex. 6 Normal Distribution

- Scores of a certain exam for a group of students were normally distributed with a mean of 100 and standard deviation 20. Based on the information provided what is the percentage of students who scored between 80 and 120?
  A. 50%
  B. 68%
  C. 75%
  D. 95%

Inferential Statistics: Confidence Interval

Statistical Inference

- Two statistical inference methods:
- Estimation: when the population parameter (e.g, population mean) is unknown, corresponding sample statistics (e.g., sample mean) are used to generate estimates.
- Hypothesis testing, an explicit statement or hypothesis is generated about the population parameter. Sample statistics are analyzed and determined to either support or reject the hypothesis about the parameter.
  - It is assumed that the sample drawn from the population is a random sample.

Estimation: from sample statistics to population parameters

- A sample mean can be used to estimate the population mean.
- If a second sample from the same population is obtained, the sample mean of the second sample is likely to be different from the first sample mean.
- How should we account for this variation?

Sampling Distribution

- Imagine that we repeat the random sampling process many times with the same sample size, we are going to have many different sample means.
- These many sample means constitute their own distribution: the sampling distribution.

Types of Estimation

- A point estimate for a population parameter is the "best" single number estimate of that parameter.
- A confidence interval estimate is a range of values for the population parameter with a level of confidence attached.

Confidence Level and Standard Error

- Confidence Level: reflects the likelihood that the confidence interval contains the true, unknown parameter (90%, 95% and 99%)

- Standard error reflects the variability of the sampling distribution of the sample statistic, for example, sample mean:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

  - Population standard deviation can be replaced by sample standard deviation.
  - Sample size has an impact on standard error.

Example: confidence intervals for population mean (m) based on large sample size

- For n > 30 (Find Z in Normal Distribution table)

$$\overline{X} \pm Z * \frac{s}{\sqrt{n}}$$

- The midpoint of the CI is $\overline{X}$

- $Z * \frac{s}{\sqrt{n}}$ is the margin of error
  - Z reflects the confidence level
  - $\frac{s}{\sqrt{n}}$ is the estimated standard error, reflects the variability of the sampling distribution.

	$\alpha = .05$	$\alpha = .01$
Two Tailed	1.960	2.576
One Tailed	1.645	2.326

Example: Systolic Blood Pressure

- In the Framingham Offspring Study (n=3534), the mean systolic blood pressure (SBP) was 127.3 with a standard deviation of 19.0. Generate a 95% confidence interval for the true mean SBP.

	$\alpha = .05$	$\alpha = .01$
Two Tailed	1.960	2.576
One Tailed	1.645	2.326

$$\overline{X} \pm Z \frac{s}{\sqrt{n}}$$

$$127.3 \pm 1.96 \frac{19.0}{\sqrt{3534}}$$

$$127.3 \pm 0.63$$

(126.7, 127.9)

Interpretation of CI in General

- Suppose that we would like to estimate a population mean using a 95% confidence level.
- Since the population mean is a fixed value, the confidence interval generated based on a specific sample may or may not contain the population mean.
- If we repeatedly compute a 95% CI with different random samples of the same size, 95% of the CIs will contain the true population mean.

Ex. 7 CI

- The point estimate and margin of error of the mean age of a group in a study is 35.5 ± .3. What would be the center value of the 95% Confidence Interval?
- A. 35.2
- B. 35.8
- C. 35.5
- Cannot be determined by the information given

Ex. 8 CI

- If the 95% confidence interval is found to be (100 – 300), which would be potential values for a 99% Confidence Interval?
  A. 50 – 350
  B. 100 – 300
  C. 150 – 250
  D. Cannot be determined from the given information

- What about a 90% CI?

Inferential statistics: Introduction to Hypothesis Testing

From Confidence Interval to Hypothesis Testing

- A confidence interval estimate is a range of values for the population parameter with a level of confidence attached.
- With hypothesis testing, an explicit statement or hypothesis is generated about the population parameter. Sample statistics are analyzed and used to determine whether we should reject the hypothesis about the parameter.

Hypothesis Testing

- Research hypothesis is generated about unknown population parameter
- Sample data are analyzed and determined to support or refute the research hypothesis

Null and Alternative Hypotheses

- Two rival hypotheses:
  – Null hypothesis: assumes that nothing is going on, usually carries equality
  – Alternative hypothesis: the “research hypothesis”. It is the one that reflects the researcher’s belief.
  – We write both in terms of the population (think: the population from which the sample was drawn)
- Two possible “conclusions”
  – Reject the null hypothesis
  – Fail to reject the null hypothesis

Hypothesis Testing Procedures

- Set up null and research hypotheses, select a
- Select test statistic
- Compute test statistic
- Calculated test statistic vs critical value
- Draw conclusion & summarize significance

Hypothesis Test (Z-test)

- Suppose that you know that the mean age of a certain population is 30 with a standard deviation of 4.47. We draw a sample of n=10 that has a sample mean($\bar{X}$) = 27. Can we conclude that the mean of this population is **different from** 30 years? Alpha (α) = .05

Developing a Hypothesis Statement

- We have 3 choices for a hypothesis statement
- 1) Non-Directional (key word = "difference")

  H_0: μ=30

  H_a: μ≠30

****Use key words to determine which set of hyp's you need**
****Write your H_a in the direction of key words then write your H_o to reflect everything else**

- 2) Directional: 2 choices
  - A) key word=more, greater, positive direction

    H_0: μ≤30

    H_a: μ>30

  - B) key word=less, smaller, negative direction

    H_0: μ≥30

    H_a: μ<30

Calculate the Test Statistic

- In our case, it is the Z statistic. Each situation has a different formula.

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad = \frac{27 - 30}{4.47 / \sqrt{10}}$$

- Z= -2.12
- This is your test statistic (Zcalc)

Determine the Critical Region

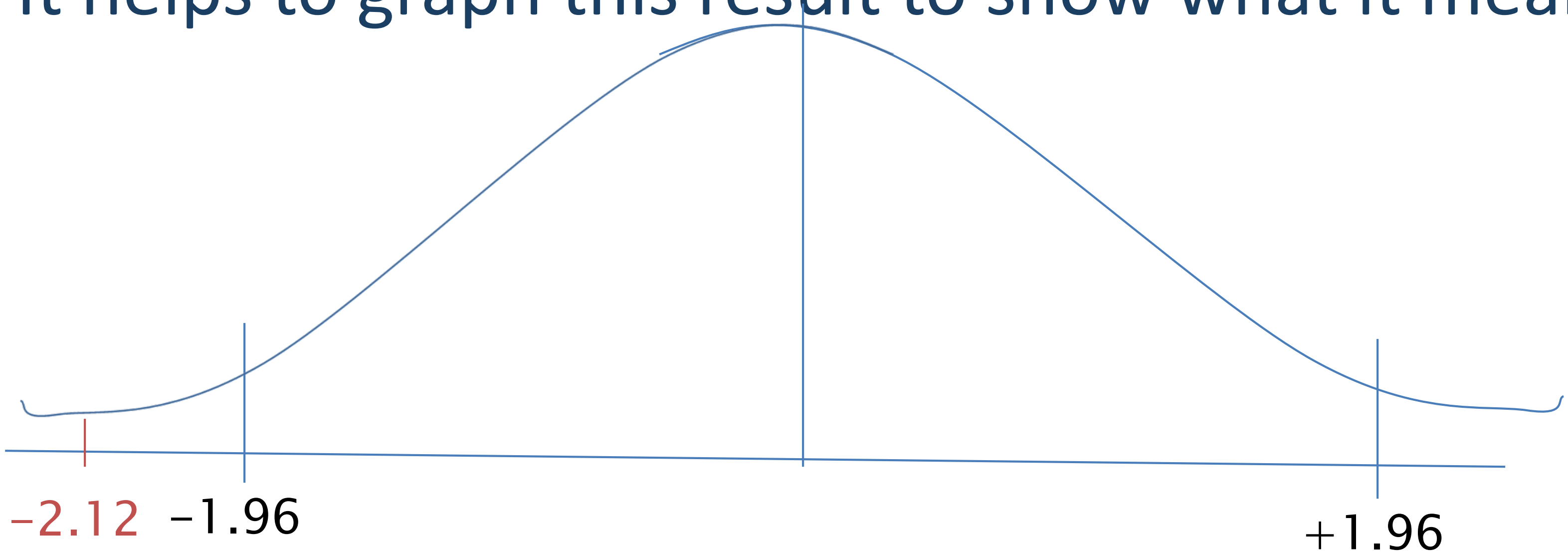
- Now we need to determine if our test statistic that we calculated is in the rejection region. We need to look up our critical value. Here are the critical values for Z:

	$\alpha = .05$	$\alpha = .01$
Two Tailed	1.960	2.576
One Tailed	1.645	2.326

- Our problem was two tailed, $\alpha = .05$

So, Zcrit = 1.960

- It helps to graph this result to show what it means



This is a graph of our critical region. Notice it goes in both directions (b/c our hyp was non directional) If our Zcalc falls in the tail of our rejection region, then we reject the null.

Where does our Zcalc fall?

Make a decision

- If |Z calc| > |Z crit| then we reject the null

|-2.12| > | ±1.96| so we must reject the null

Step 6: State your conclusion

- The population from which our sample was drawn has a mean age that is statistically significantly different than our reference population.
- The population from which or sample was drawn has a mean age that is statistically significantly different than 30.

Confidence Interval Z-test

- Use the formula for CI that corresponds to your test statistic.
- We need a CI for the Z test

$$CI = \bar{x} \pm Z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}$$

All of the info is given in the original problem

$\bar{X}$ = 27

σ = 4.47

n = 10

If you want 95% CI, the Z_{crit} you use corresponds to α=.05 (1.96)

If you want 99% CI the Z_{crit} you use corresponds to α=.01 (2.58)

Step 7: Cont.

$$CI = \bar{x} \pm Z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}$$

$$= 27 \pm (1.96)(4.47/\sqrt{10})$$

$=27 \pm (1.96)\ (1.414)$

$=27 \pm (2.77)$

$(27 + 2.77)\ ,\ (27 - 2.77)$

$(29.77,\ 24.23)$

We are 95% confident that the population from which the sample was drawn has a mean age that lies b/t 24.23 and 29.77

CPH Certified in Public Health
by National Board of Public Health Examiners

Directional hypothesis

- What if we had the same problem, but the question asked was: Can you conclude that the sample mean was statistically significantly less than 30. What changes?
- **State your hypothesis:** $H_0: \mu \geq 30 \quad H_A: \mu < 30$
- Calculate the test statistic: We would still get -2.12
- **Z_{crit}= -1.645** (we want less than, so neg)
- We still reject the null
- The population from which the sample was drawn, has a mean age that is statistically significantly less than the reference population

P-value

- The probability of observing the obtained data (or more extreme values) given the null hypothesis was true
- Calculate p-values based on the sampling distribution of the test statistic
  - Approximate with statistical tables
  - Get exact value with statistical computing package, e.g., SPSS
- Reject the null hypothesis if the p-value is lower than the alpha level (α).
  - Equivalently, reject the null if the value of the test statistic exceeds the critical value that corresponds to the alpha level.

Two Types of Error

- Since we only have information from the sample, we are subject to make mistakes. There are two types of errors.
- Type-I error: reject the null hypothesis when it is true.
  – Probability of making a Type-I error = α
- Type-II error: fail to reject the null hypothesis when it is false.
  – Probability of making a Type-II error = β

	Conclusion in Test of Hypothesis	
	Do Not Reject H_0	**Reject H_0**
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

- Power: the probability that we will correctly reject a false null. Power = 1-β
- This is largely dependent on effect size. If we have a large difference b/t two groups, then we can detect that difference more readily (all things being equal n, alpha, Beta) than if we have a smaller effect size.

Ex. 9 Power

- A research study with a statistical power of 60% and alpha is set at 0.05 reports finding no statistically significant difference (p=0.11). To interpret the findings of this study, if a true difference really exists then there is:
  A. An 89% probability that it would be detected.
  B. A 60% probability that it would be detected.
  C. A 40% probability that it would be detected.
  D. An 11% probability that it would be detected.

Inferential Statistics: The Most Commonly Used Tests

Chi-Square Test of Independence

- This test determines whether two categorical variables are independent
- Test statistic is formed by summarizing difference between observed values and expected values.

Example: Living Arrangement and Exercise Status

Is there a relationship between students' living arrangement and exercise status?

	Exercise Status			
	None	Sporadic	Regular	Total
Dormitory	32	30	28	90
On-campus Apt	74	64	42	180
Off-campus Apt	110	25	15	150
At Home	39	6	5	50
Total	255	125	90	470

Example: Living Arrangement and Exercise Status (cont'd)

Test statistic: $\chi^2 = \sum \frac{(O - E)^2}{E}$

Table entries are Observed (Expected) frequencies

	Exercise Status			
	None	Sporadic	Regular	Total
Dormitory	32 (90*255/470=48.8)	30 (23.9)	28 (17.2)	90
On-campus Apt	74 (97.7)	64 (47.9)	42 (34.5)	180
Off-campus Apt	110 (81.4)	25 (39.9)	15 (28.7)	150
At Home	39 (27.1)	6 (13.3)	5 (9.6)	50
Total	255	125	90	470

Example: Living Arrangement and Exercise Status (cont'd)

Calculate your test statistic:

$$\chi^2 = \frac{(32-48.8)^2}{48.8} + \frac{(30-23.9)^2}{23.9} + \frac{(28-17.2)^2}{17.2} + \ldots + \frac{(5-9.6)^2}{9.6}$$

$$\chi^2 = 60.5$$

Look up the critical value in a X^2 table : Df= (R-1)*(C-1) = (3)*(2)=6
Alpha= .05, Critical value = 12.59

Conclusion. Reject H_0 because 60.5 > 12.59. We have statistically significant evidence at α=0.05 to show that living arrangement and exercise status are not independent (they are associated).

Hypothesis Testing for Comparing Two Population Means

- The two independent t-test concerns the difference of two population means of continuous outcomes. Examples are:
  - Compare gender difference
  - Evaluate treatment effects

Example: Clinical Trial on Lower Cholesterol

- A clinical trial aims to assess the effectiveness of a new drug in lowering cholesterol. Patients are randomized to receive the new drug or placebo and total cholesterol is measured after 6 weeks on the assigned treatment.

	Sample Size	Mean	Std Dev
New Drug	15	195.9	28.7
Placebo	15	227.4	30.3

- Is there evidence of a statistically significant reduction in cholesterol for patients on the new drug?

Example: Clinical Trial on Lower Cholesterol (cont'd)

H0: m1≥m2

Ha: m1<m2 a=0.05

Test statistic $t = \dfrac{\overline{X}_1 - \overline{X}_2}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Decision rule, df=n1+n2-2 = 28

Reject H0 if t < -1.701

Example: Clinical Trial on Lower Cholesterol (cont'd)

Calculate your test statistic:

- Test statistic:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{195.9 - 227.4}{29.5\sqrt{\frac{1}{15} + \frac{1}{15}}} = -2.92$$

- Conclusion. Reject H0 because the absolute value of the test statistic |-2.92| > the absolute value of the critical value |-1.701|. We have statistically significant evidence at a=0.05 to show that the mean cholesterol level is lower in patients on treatment as compared to placebo.

Ex. 10 Hypothesis Testing

- The goal of an ANOVA statistical analysis is to determine whether or not:
  – The means of two samples are different
  – The means of more than two samples are different
  – The means of two populations are different
  – The means of more than two populations are different

Ex. 11 Hypothesis Testing

- National Physical Activity Guidelines for Americans suggest adults need to participate in at least 150 minutes per week in moderate intensity physical activity for substantial health benefits. A study was designed to test whether there is a difference in mean time (in minutes) spent in moderate intensity physical activity in adults with no, mild, moderate and severe depression. Time spent in moderate intensity physical activity is a continuous measure which can be assessed using an accelerometer. Which among the following is the most appropriate statistical technique to test the difference in time spent in moderate intensity physical activity between adults with no, mild, moderate and severe depression?

  A. Chi-Square test

  B. t-test

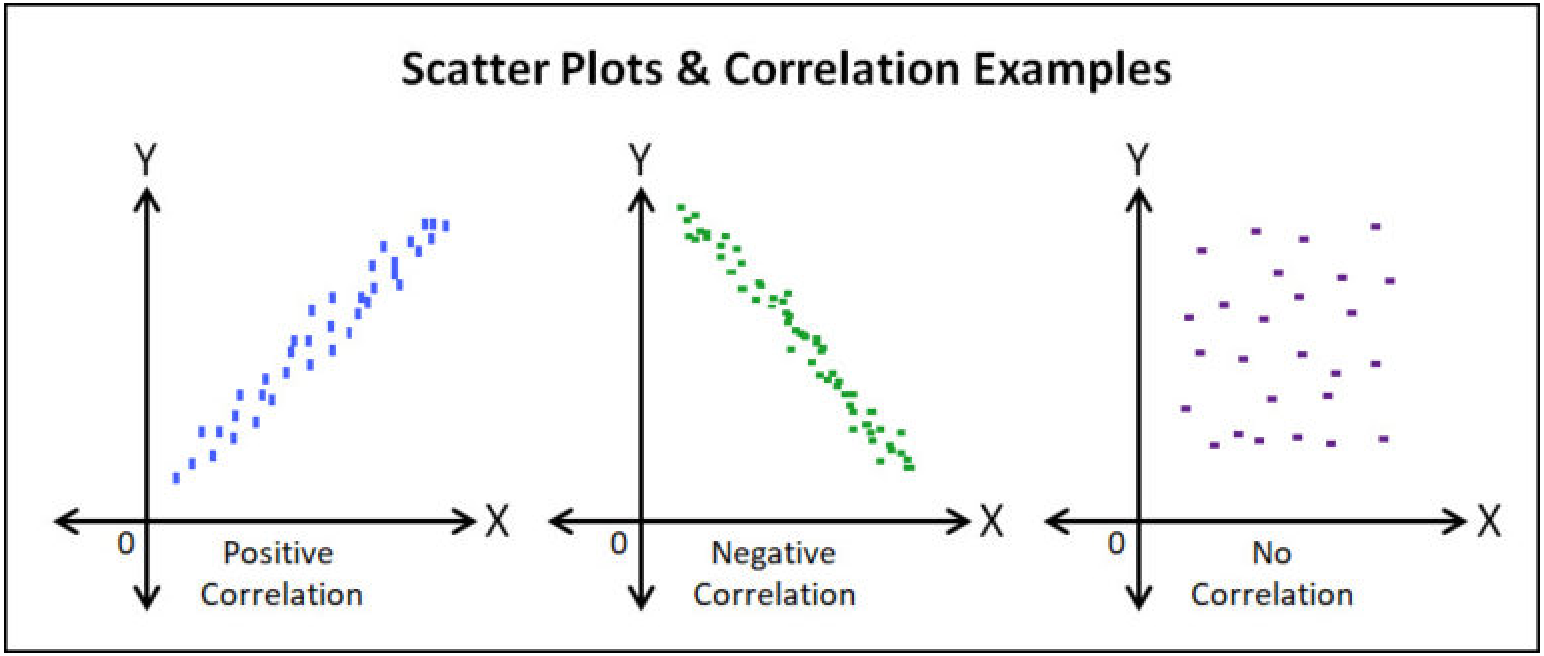
  C. ANOVA

  D. z-test

Correlation and Linear Regression Analysis

- Correlation
  - Concerned with measuring the strength of the relationship between variables
- Regression
  - Used to predict or estimate the value of one variable corresponding to a given value of another variable

Scatter Diagram



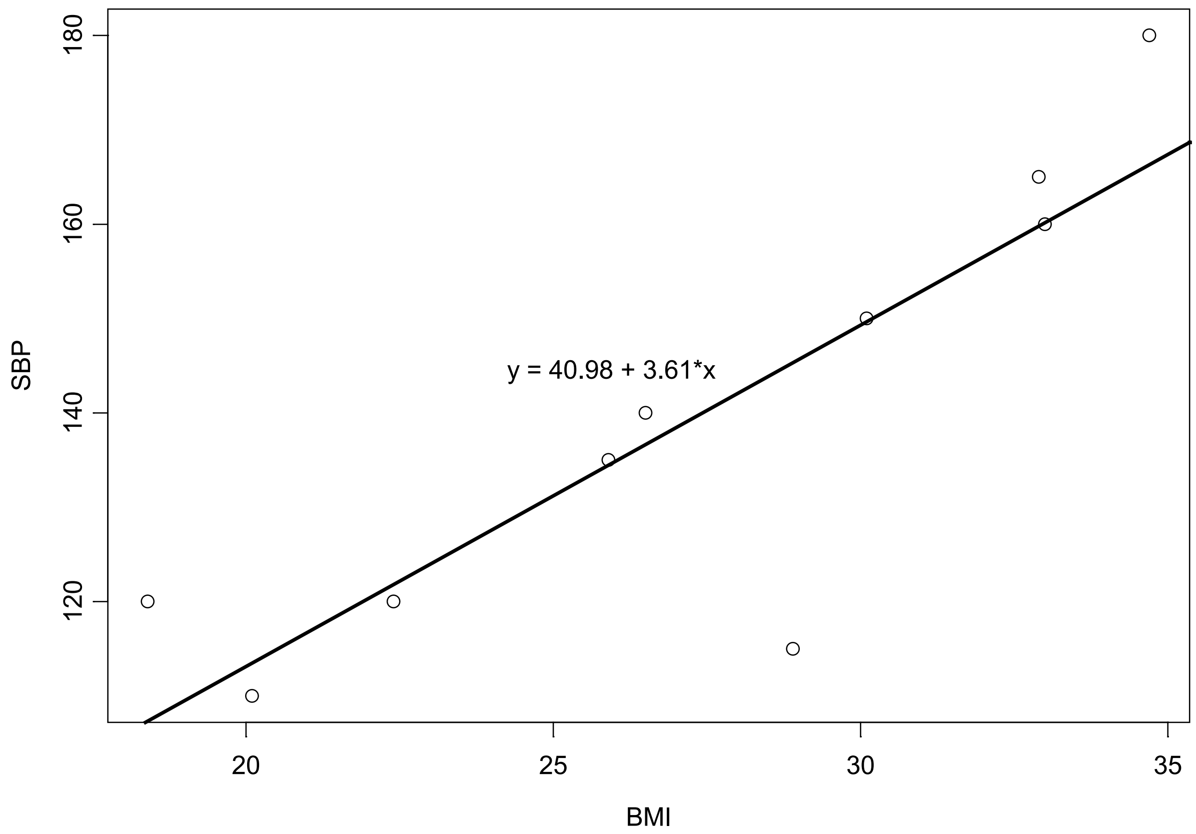
Correlation Coefficient

- Population correlation coefficient (r) is a measure of the strength of the linear relationship between x and y.
- $-1 < r < +1$
- Sign indicates nature of relationship (positive or direct, negative or inverse)
- Magnitude indicates strength
- Percent variation attributed to predictor variables is r^2

Simple Linear Regression

- Y = Dependent, Outcome variable (e.g., SBP)
- X = Independent, Predictor variable (e.g., BMI)
- Population regression equation:
- $Y = \beta_0 + \beta_1 X + \text{error}$
- β_0 is the intercept and β_1 is the slope

Simple Linear Regression Line Between BMI and Systolic Blood Pressure



Regression Equation and Its Applications

- Regression equation: $\hat{y} = 40.98 + 3.61\,x$
- How to use the equation?
- What is expected SBP for a male with BMI=20?

$$\hat{y} = 40.98 + 3.61\,(20) = 113.81$$

- Compare two males whose BMIs differ by one unit – how do SBPs compare?
  - Person with higher BMI will have SBP that is 3.61 units higher on average

Ex. 12 Regression

- The most useful visual tool for exploring relationships in bivariate data (paired measurements of two quantitative variables) is the:
  A. Histogram
  B. Choropleth Map
  C. Box and Whisker Plot
  D. Scatterplot

Ex. 13 Regression

- Assume that a linear regression analysis is performed with predictor variables age and socio-economic status, and the correlation coefficient r is 0.4. What percent of the variation in the outcome is explained by age and socio-economic status?
  A. 25%
  B. 40%
  C. 16%
  D. 8%

What questions do you have?

Reminder

CPH Study Resources

1. Content Outline
2. Sample Exam Questions
3. Practice Exams
4. Webinars
5. ASPPH Study Guide
6. APHA Study Guide

www.nbphe.org/cph-study-resources/