

STAT 113: Statistics and Society
Ellen Gundlach, Purdue University

(Chapters refer to Moore and Notz, Statistics: Concepts and Controversies, 8e)

Learning Objectives for Exam 1:

Unit 1, Part 1: Population vs. sample

- Explain and give examples of individuals and variables. (Ch 1)
- Calculate a proportion. (Ch 3)
- Distinguish between a population and a sample. (Ch 1)
- Distinguish between a parameter and a statistic. Identify parameters and statistics in stories. (Ch 1)
- Explain the advantages and disadvantages of taking a census. (Ch 1)
- Describe the problems that can occur with sample statistics and explain how to prevent them. (Ch 3)
 - Bias
 - Variability
- Identify the method of sampling used when given a story. Explain the advantages/disadvantages of each. (Ch 2, 4)
 - Convenience Sampling
 - Voluntary Response sampling
 - Simple random sampling
 - Stratified random sampling

Unit 1, Part 2: Sampling problems and surveys

- Define undercoverage. Explain the problems it causes and how to prevent it. Recognize undercoverage in a story and suggest groups that might be undercovered when taking a sample. (Ch 4)
- Describe random sampling error and how it differs from other types of error. (Ch 4)
- Define response error/bias and nonresponse. Recognize them in a story and suggest ways to fix them. (Ch 4)
- Identify important features of a survey presented in a story, including potential problems, and explain how to improve the survey. (Ch 4)

Unit 2, Part 1: Data collection, features of experiments

- Identify the method of data collection used when given a story. Explain how these methods are different. (Ch 1)
 - Anecdotal evidence
 - Observational study (including surveys)
 - Experiment
- Explain why experiments are better than observational studies. (Ch 5)
- Identify important features of an experiment from a story. (Ch 5)
 - Response variable
 - Explanatory variable
 - Treatment
- Explain in terms of a story the problems with conducting the simplest possible experiment. (Ch 5)
 - Lurking variables

- Placebo effect
- Bias by researchers
- Explain why randomized, comparative experiments are good. (Ch 6)
- Define whether the 3 principles of good experimental design are being used in a story. (Ch 6)
 - Control or comparison group
 - Randomization to treatment groups
 - Large sample size
- Define a double-blind experiment in terms of a story, and explain the benefits. (Ch 6)

Unit 2, Part 2: Types of experiments, clinical trials

- Outline an experiment in a diagram, identifying all the important features. (Ch 5)
- Identify the type of experimental design used in a story. (Ch 6)
 - Completely randomized design (CRD)
 - Randomized block design
 - Matched pairs design
- Explain what is required for a clinical trial with humans and what can go wrong. (Ch 7)
- Explain how new drugs get approved by the FDA.

Unit 3: Causation

- Differentiate between causation, common response, and confounding. Identify which of these terms is most appropriate for a particular story. (Ch 15)
- Diagram causation, common response, and confounding situations. (Ch 15)
- Suggest appropriate lurking variables which may be involved in common response or confounding for a story. (Ch 15)
- Explain what is necessary in order to show causation. (Ch 15)
- Explain whether there is a link between autism and vaccines and the history of this issue.

Unit 4: Ethics of experiments with humans and animals

- Define and identify in a story the 3 principles of ethical experiments with humans. (Ch 7)
 - Review board approval
 - Informed consent
 - Confidentiality
- Explain the difference between confidentiality and anonymity.
- Define and identify in a story the 3 Rs of ethical experiments with animals. (Not in the book.)
 - Replacement
 - Reduction
 - Refinement

Unit 5, Part 1: Measurements and numbers

- Determine whether a measurement is valid and/or reliable from a story. (Ch 8)
- Identify any sources of bias involved in a measurement for a story. (Ch 8)
- Explain how calculating an average can improve reliability over individual measurements. (Ch 8)
- Identify potential problems and unrealistic numbers presented in a research study. (Ch 9)
- Understand that peer review is important in evaluating the worth of experimental results.
- Explain how to estimate crowd size.
- Calculate a percent change. Know the limits of percent decreases. (Ch 9)

Unit 5, Part 2: Do we buy it?

- Evaluate the claims made by advertisers.
- Ask questions about how the data was collected.

Unit 6: Government statistics

- Explain what citizens need from their government statistics. (Ch 16)
- Explain why the US Census and the American Community Survey are important. (Ch 1 and 16)

Unit 7: Big data

- Define big data.
- Explain positive uses of big data.
- Explain potential ethical problems of big data.
- Know how your own data is being collected and used.

Learning Objectives for Exam 2:

Unit 8, Part 1: Graphs for categorical data

- Identify whether a variable is categorical or quantitative, and explain the difference in these types. (Ch 1)
- Create (by hand and with Excel) and interpret graphs appropriate for categorical variables. (Ch 10)
 - Pie charts
 - Bar graphs
- Use the guidelines to create good graphs and to identify problems with graphs created by others. (Ch 10)

Unit 8, Part 2: Graphs for a single quantitative variable

- Create (by hand) and interpret graphs appropriate for quantitative variables. (Ch 11)
 - Histograms (also by Excel)
 - Stemplots
 - Boxplots (more on this in Ch 12)
- Explain the differences and similarities between types of graphs and when they are appropriate. (Ch 10 and 11)
 - Bar graphs vs. histograms
 - Histograms vs. stemplots
- Describe the shape of the distribution of a quantitative variable as (Ch 11):
 - unimodal, bimodal, multimodal
 - symmetric, skewed right, skewed left
- Compare the quantitative variable graphs for several groups. (Ch 11 and 12)
- Create (by hand and with Excel) and interpret line graphs/time plots. (Ch 10)

Unit 9: Summary statistics for a single quantitative variable

- Calculate (by hand) and explain the differences between measures of center: (Ch 12)
 - Mean (also with Excel)
 - Median
 - Mode
- Calculate and interpret measures of spread: (Ch 12)
 - Standard deviation and variance (only with Excel, never by hand)

- Percentiles, quartiles, the 5-number summary, and the interquartile range (IQR) (all by hand)
- Use the 1.5 IQR rule to check for any outliers. (Not in the book.)
- Use the 5-number summary to draw a modified boxplot by hand, including showing any outliers.
- Explain when it is more appropriate to use the mean and standard deviation or the median and IQR to describe the center and spread, based on the shape of the distribution.

Unit 10: Normal distribution basics

- Identify data that is Normally distributed by looking at a histogram or stemplot. (Ch 13)
- Discuss how the shape/position of the Normal curve changes when the standard deviation increases/decreases or when the mean increases/decreases. (Ch 13)
- Define the standardized value or Z-score. (Ch 13)
- Calculate the Z-score to do comparisons between things measured on different scales. (Ch 13)
- Calculate probabilities and cut-off values using the 68%-95%-99% Empirical Rule.

Unit 11, Part 1: Scatterplots and correlation

- Create (by hand and with Excel) properly labeled scatterplots. (Ch 14)
- Interpret the scatterplot in terms of the story. (Ch 14)
- Describe the important features of a scatterplot. (Ch 14)
 - Form: linear or nonlinear
 - Direction: positive or negative (or no association)
 - Strength: strong, moderate, weak
 - Suspected outliers
- Differentiate between association and causation. (Ch 14 and 15)
- Define the correlation, explain the features of correlation, and estimate the correlation based on a scatterplot. (Ch 14)
- Calculate the correlation with Excel. (Ch 14)

Unit 11, Part 2: Regression

- Explain when least-squares regression is appropriate and why it should be used. (Ch 15)
- Explain what the slope, intercept, x , and y mean in terms of the story. (Ch 15)
- Calculate (using Excel) the least-squares regression line and R^2 from raw data. (Output will be provided on exam.) (Ch 15)
- Explain what R^2 means in terms of the story and how R^2 is related to the correlation. (Ch 15)
- Calculate predicted response variables using the regression line. (Ch 15)
- Differentiate between prediction and extrapolation. (Ch 15)
- Explain how outliers can affect the results of a regression line and correlation and how to check whether an outlier is influential. (Ch 15)

Unit 12, Part 1: Two-way tables: joint, marginal, and conditional distributions

- Explain when two-way tables are appropriate. (Ch 24)
- Calculate joint, marginal, and conditional distributions and percentages. (Ch 24)
- Identify from the wording of a question whether the joint, marginal, or conditional percentage would be most appropriate. (Ch 24)
- Draw bar graphs to show the distributions. (Ch 24)

Unit 12, Part 2: Interpreting medical test results

- Explain and calculate rates of true positives and true negatives.
 - Explain and calculate rates of false positives and false negatives.
 - Explain and calculate sensitivity and specificity.
 - Know when you should be worried about your test results.
-

Learning objectives for the Final Exam:

Unit 13, Part 1: Definitions of random variables and probability

- Explain how random events behave in the short run and in the long run. Explain how random and haphazard are not the same thing. (Ch 17)
- Perform basic probability calculations using die rolls and coin tosses. (Ch 17)
- Define probability, and apply the rules for probability. (Ch 17)
- Explain the law of averages and whether it is true. (Ch 17)
- Explain how personal probability differs from a scientific or experimental probability. (Ch 17)

Unit 13, Part 2: Probability models

- Define a probability model. Create a probability model for a particular story's events. (Ch 18)
- Explain whether a probability model is valid. (Ch 18)
- Apply the basic rules of probability to a story problem. (Ch 18)
- Calculate probabilities using a probability model. (Ch 18)

Unit 13, Part 3: Expected values

- Define expected value. (Ch 20)
- Calculate the expected value when given a probability model or a story. (Ch 20)
- Define the law of large numbers, and explain whether it is real. (Ch 20)
- Compare the law of large numbers to the law of averages. (Ch 17 and 20)
- Explain how casinos and insurance companies stay in business and make money. (Ch 20)

Unit 14, Part 1: Confidence intervals: the basics

- Define statistical inference and explain when it is used. (Ch 21)
- Explain what a confidence interval tells us and whether the results refer to the population or the sample. (Ch 21)
- Explain what type of error is covered in the margin of error. (Ch 21)
- Describe how increasing/decreasing the sample size or confidence level changes the margin of error. (Ch 21)
- Apply cautions for using confidence intervals: (Ch 21)
 - Need a simple random sample.
 - Data must be collected correctly.
 - For quantitative data, no outliers.

Unit 14, Part 2: Confidence intervals for a proportion

- Determine whether a story is better described with a proportion or a mean. (Ch 21)
- Use appropriate notation for proportions. (Ch 21)
- Calculate a confidence interval for a proportion. (Ch 21)

Unit 14, Part 3: Confidence intervals for a mean

- Determine whether a story is better described with a proportion or a mean. (Ch 21)

- Use appropriate notation for means. (Ch 21)
- Calculation a confidence interval for a mean. (Ch 21)

Unit 15, Part 1: Basics of hypothesis testing

- Understand what statistical significance means. (Ch 5, Ch 22, 23)
- Understand that the point of inference is to estimate what is true for the population using our sample results and variability.
- Understand that lack of significance does not mean that there is no effect, only that we do not have good evidence for an effect. Small samples often miss important effects that are really present in the population. (Ch 23)

Unit 15, Part 2: Cancer clusters

- Understand what statistical significance means. (Ch 5, Ch 22, 23)
- Explain what a cancer cluster is and how it is determined to exist in a region.