



# West Carteret High School



## Notes to Accompany AP Statistics

Spring 2017



# Table of Contents

## Preliminaries

|                             |   |
|-----------------------------|---|
| I. Course Introduction      | 3 |
| II. Course Content Overview | 7 |

## Part I – Analyzing Data

|                                          |    |
|------------------------------------------|----|
| 1. Exploring Data                        | 11 |
| 2. Describing Location in a Distribution | 19 |
| 3. Examining Relationships               | 27 |

## Part II – Producing Data

|                      |    |
|----------------------|----|
| 4. Designing Studies | 35 |
|----------------------|----|

## Part III – Probability and Random Variables

|                                       |    |
|---------------------------------------|----|
| 5. Probability: What Are the Chances? | 43 |
| 6. Random Variables                   | 47 |
| 7. Sampling Distributions             | 55 |

## Part IV - Inference

|                                                     |    |
|-----------------------------------------------------|----|
| 8. Estimating with Confidence                       | 59 |
| 9. Testing a Claim                                  | 65 |
| 10. Comparing Two Populations or Groups             | 71 |
| 11. Inference for Distributions of Categorical Data | 77 |
| 12. More about Regression                           | 83 |

## Appendices

|                                    |    |
|------------------------------------|----|
| Appendix A – Portfolio Information | 89 |
| Appendix B – Formulas and Tables   | 90 |
| Appendix C – Inference Summary     | 96 |

# I. Course Introduction

---

Background. An understanding of statistics is necessary for anyone to be a savvy consumer, an informed citizen of a democracy, and a successful student of many other disciplines. The AP Statistics course gives students an opportunity to build a deep conceptual understanding of graphical analysis and statistical inference and the roles they play in decision making.

Advanced Placement Course. This is a college-level class that will move at a fast pace. Students who take this course are assumed to be: **(1) capable of working at the college level**, and **(2) willing to work at the college level**. If you do not feel you meet these two requisites, I highly encourage you to change your schedule. There is a reason that six GPA quality points are awarded for Advanced Placement classes. That having been said, if your reason for taking this course is to enhance your GPA, you have come to the wrong place with the wrong instructor. In order to receive AP credit for this course, students must take the Advanced Placement Examination which will be administered on **Thursday, May 11 at 12 noon**.

How to Succeed in Mr. Mayo's AP Statistics Class. Years ago as an AP Calculus teacher, I put together the following advice to my students concerning how to approach the course. (If you have recently taken my Honors Precalculus course, it will sound familiar.) It has proven to be a successful model to approaching an AP mathematics course.

## **How to Succeed in Mr. Mayo's AP Statistics Class (With thanks to Dr. Dan Kennedy)**

My main goal in teaching Statistics is helping you be successful. I want each and every one of you to succeed and my assumption is that you would not be in this class if you did not want to succeed. The main question will be how *committed* to success will *you* be. Not only should students enrolled in this course be capable of operating at the college level but they must also be willing to operate at the college level.

One measure of success, unfortunately, will be your grade. As Paul Dressel stated in 1957, "... a grade (is) ... an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite amount of material." That having been said, I still need to prepare you for further studies in mathematics. Because of this fact, and other necessities such as you needing a GPA, etc., I will be grading your performance this year. As Dr. Dan Kennedy says, "A grade is an inadequate way to measure achievement, but a fairly realistic way to measure performance." In everything we do in life our performance is assessed. Teachers just take it a step further and assign a grade to performance.

Your grade in this class will be generally based upon quizzes, tests, a portfolio and a final examination. However, many other things will be necessary for you to be successful. These include your class work, homework, and study.

### **Class work**

You must keep a **notebook** which must be entirely devoted to AP Statistics. I suggest you get yourself a brand new thick 3-ring binder and fill it with a lot of notebook paper. Your notes will

be your notes just like they will be in your college courses. I expect you to bring your notebook to class each day along with your text, calculator and something to write with. Nothing brings tears to my eyes quicker than hearing a student asking another to borrow a pencil seconds before the bell rings. Quite simply, if you do not have these things with you when class starts, I reserve the right to ask you to leave due to your lack of preparedness.

I will not grade your notebook but it will become a key ingredient in your success. It should be organized by chapter and you should keep all completed quizzes, tests and midterm exams in it. Students should not merely take notes for the sake of taking notes. These notes should be what you use to refer to throughout the course. Organize them accordingly.

### **Homework**

There will be a homework assignment almost every night. Just like learning to hit a baseball, you must practice math in order to be able to understand it and use it. Doing your homework will be essential. I cannot overemphasize this. In my heart of hearts, I believe each one of you is intellectually capable of learning Statistics. I sometimes fear that not all of you possess the work ethic, however. Many of you are so intellectually capable that in past math courses you were successful without doing homework. I believe that most of you will find Statistics to be different. In my experience the students who did not do well were victims of not doing the work outside of class.

Additionally, most of the homework will not be a repeat of what we do in class. It will be designed to stretch what we cover in class and because of this I need you to complete it in order to be ready for the next day's lesson. Although I believe that homework is one of the most important ingredients in your success, I will only occasionally collect it and grade it. We will spend a great deal of time in class each day going over it. When we start class, I will randomly pick students to come to the board and brief the homework problems to the class. When you are chosen to do this, **I will not accept responses such as, "I could not do that one" or "I had to work (had a game) last night."**

I expect you to do everything in your power to complete the homework assignment. Before attempting the homework problems, you should reread the section of the text related to the homework as well as your notes. Then complete the problems. Dr. Kennedy describes the homework process as having two parts: (1) finding out how to do the problem; (2) doing it. He further says there should be no restrictions on part (1). If you know how to do it, do it. If you do not seek help wherever you can find it: your study buddies, the text book, your notes, my website, other websites, your mom, dad, sister or brother, or ME. If all else fails, call me at **622-8949** or email me at [colin.mayo@carteretk12.org](mailto:colin.mayo@carteretk12.org) and I will assist you.

Collaboration is another very important ingredient in your success. Once again, if you are not prepared to discuss your homework, I reserve the right to ask you to leave class because you are unprepared. I must caution you about part (2). Even though you may have collaborated, when push comes to shove, you should understand and be able to do the problems by yourself. Finally, if I feel homework is not being done to my satisfaction, I will implement sanctions that will assist the class in getting it done.

### **Reading the Text**

I have carefully chosen the text we are using. It is designed to be read. It is not just there to provide homework problems. Each night I expect you to read the next section to be covered in class. You may not understand everything that is in the section and that is OK. When you read a math text you should have pencil and paper to work the problems along with the text. Make note of topics and concepts you do not understand and be sure to get these taken care of in class. I have placed reading guides for each chapter on my website in School Loop for your use.

### **Class Participation**

To be successful you must participate in class. Whether we are working on the board, I am presenting a lesson, we are discussing a topic or you are working with your team I expect everyone to be a player. Ask questions of me and your fellow students. Provide input. Take an active part. The best way to learn math is to talk about it. Math is a team sport in my class and we will have no one sitting the bench.

### **Blitz**

Unless otherwise stated, Blitz attendance is **mandatory**. This will allow us to recoup time prior to the AP Exam.

### **Portfolio**

You will be assigned portfolio items for each chapter we study. They are based upon the skills you will need to be successful at the next level of mathematics. Once completed, your portfolio will be a set of notes in your own words that you will have to refer to once you take your next math course. Many students dubiously wait until the last minute before it is due to complete it. It is much better to complete the items as they are covered in class. This way you are not freaking out the night before it is due and, more importantly, by completing it along the way, you will better understand the material. This is not an optional project. If you do not submit it, you will not receive credit for the course. *Late or incomplete submissions will at most be awarded 70% of the grade.* More details for portfolio completion are provided in the portfolio description.

### **Seeking Extra Help**

It is better to seek help earlier rather than later. I am typically at school by 7:15 each morning and on most afternoons I stay until at least 4:30. I will assist you with whatever your questions are. Sometimes just a few minutes of one-on-one will quickly clear up challenges. If you need help with assignments, I am willing to assist by phone or email. **Do not text me.** I will usually answer email questions as if they were text messages.

### **Studying for Quizzes, Tests, Exams**

Wow! Where to begin? I believe studying mathematics is an art form. I do not think I really learned how to successfully study mathematics until I was in graduate school. It is not hard really, perhaps just different. It is my personal goal to help each of you to develop your own effective style of studying mathematics. If we can do this you will be highly prepared for your college studies. Again, I believe we learn mathematics by practicing. Work problems! Before doing so, however, you must know the context from which they come. You should know how the course material is sequenced and how the topics you are studying fit in? How do definitions and theorems relate to what you are studying? Use your portfolio to help you. That is why we are constructing it. Put yourself in my shoes. What did I stress? What do you think I believe is

important? Find your weaknesses and attack them. Study with someone or in a group. This saved me in many of my courses in graduate school. We all have our strengths and weaknesses but by working together we can do great things.

Finally, I want to reiterate that I really do want you to succeed. It is going to take a great deal of commitment on our parts to make this happen. I hope to foster this commitment in all that I do and will assist you in any way I can.

Materials. All students will need the following:

- Three-ring loose leaf binder for notes and study guides
- Loose leaf paper
- Pencils
- A composition book for your portfolio
- A composition book for your homework
- Graphing calculator (TI-83 or TI-84)

Grading. As was alluded to in the “How to Succeed ...” your grade will be computed on a cumulative-point basis. Graded items will include quizzes, tests, projects, and your portfolio. As was previously stated, the portfolio is not an optional assignment.

***\*\*\*Students who fail to submit a portfolio or assignments, will receive a course grade of Incomplete which will be removed when the assignments are submitted. Late assignments will receive at most a grade of 70%.\*\*\****

## II. Course Content Overview

---

The topics for AP Statistics are divided into four major themes:

- **Exploratory analysis** (20-30% of exam): Exploratory analysis of data makes use of graphical and numerical techniques to study patterns and departures from patterns.
- **Planning and conducting a study** (10-15% of exam): Data must be collected according to a well-developed plan if valid information is to be obtained.
- **Probability** (20-30% of exam): Probability is a tool used for anticipating what the distribution of data should look like under a given model.
- **Statistical Inference** (30-40% of exam): Statistical inference can be thought of as the process of selecting a reasonable model, including a statement in probability language, of how confident one can be about the selection.

The following topic outline is provided by The College Board to describe the major topics covered on the AP Statistics Exam.

### I. Exploring Data: Describing patterns and departures from patterns.

*Exploratory analysis of data makes use of graphical and numerical techniques to study patterns and departures from patterns. Emphasis should be placed on interpreting information from graphical and numerical displays and summaries.*

A. Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot)

1. Center and spread
2. Clusters and gaps
3. Outliers and other unusual features
4. Shape

B. Summarizing distributions of univariate data

1. Measuring center: median, mean
2. Measuring spread: range, interquartile range, standard deviation
3. Measuring position: quartiles, percentiles, standardized scores (z-scores)
4. Using boxplots
5. The effect of changing units on summary measures

C. Comparing distributions of univariate (dotplots, back-to-back stemplots, parallel boxplots)

1. Comparing center and spread: within groups, between group variation
2. Comparing clusters and gaps
3. Comparing outliers and other unusual features
4. Comparing shapes

#### D. Exploring bivariate data

1. Analyzing patterns in scatterplots
2. Correlation and linearity
3. Least-squares regression line
4. Residual plots, outliers and influential points
5. Transformations to achieve linearity: logarithmic and power transformations

#### E. Exploring categorical data

1. Frequency tables and bar charts
2. Marginal and joint frequencies for two-way tables
3. Conditional relative frequencies and associations
4. Comparing distributions using bar charts.

## **II. Sampling and Experimentation: Planning and conducting a study**

*Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained. This plan includes clarifying the question and deciding upon a method of data collection and analysis.*

#### A. Overview of methods of data collection

1. Census
2. Sample survey
3. Experiment
4. Observation study.

#### B. Planning and conducting surveys

1. Characteristics of a well-designed and well-conducted survey
2. Populations, samples and random selection
3. Sources of bias in sampling and surveys
4. Sampling methods, including simple random sampling, stratified random sampling and cluster sampling

#### C. Planning and conducting experiments

1. Characteristics of a well-designed and well-conducted experiment
2. Treatments, control groups, experimental units, random assignments and replication
3. Sources of bias and confounding, including placebo effect and blinding
4. Completely randomized designs
5. Randomized block design, including matched pairs design

#### D. Generalizability of results and types of conclusions that can be drawn from observational studies, experiments and surveys

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

*Probability is the tool used for anticipating what the distribution of data should look like under a given model.*

#### A. Probability

1. Interpreting probability, including long-run relative frequency interpretation
2. "Law of Large Numbers" concept
3. Addition rule, multiplication rule, conditional probability and independence
4. Discrete random variables and their probability distributions, including binomial and geometric
5. Simulation of random behavior and probability distributions
6. Mean (expected value) and standard deviation of a random variable, and linear transformations of a random variable.

#### B. Combining independent random variables

1. Notion of independence versus dependence
2. Mean and standard deviation for sums and differences of independent random variables

#### C. The normal distribution

1. Properties of the normal distribution
2. Using tables of the normal distribution
3. The normal distribution as a model for measurements

#### D. Sampling Distributions

1. Sampling distribution of a sample proportion
2. Sampling distribution of a sample mean
3. Central Limit Theorem
4. Sampling distribution of a difference between two independent sample proportions
5. Sampling distribution of a difference between two independent sample means
6. Simulation of sampling distributions
7. t-Distribution
8. Chi-square distribution

#### **IV. Statistical Inference: Estimating population parameters and testing hypotheses**

*Statistical inference guides the selection of appropriate models.*

##### **A. Estimation (point estimators and confidence intervals)**

1. Estimating population parameters and margins of error
2. Properties of point estimators, including unbiasedness and variability
3. Logic of confidence intervals, meaning of confidence intervals, and properties of confidence intervals
4. Large sample confidence interval for proportion
5. Large sample confidence interval for a difference between two proportions
6. Confidence interval for a mean
7. Confidence interval for a difference between two means (unpaired and paired)
8. Confidence interval for the slope of a least-squares regression line

##### **B. Tests of significance**

1. Logic of significance testing, null and alternative hypotheses; p-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power
2. Large sample test for proportion
3. Large sample test for a difference between two proportions
4. Test for a mean
5. Test for a difference between two means (unpaired and paired)
6. Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables)
7. Test for the slope of a least-squares regression line

# Chapter 1 – Exploring Data

---

**AP Standards.** The following AP Standards are covered in Chapter 1:

**I. Exploring Data: Describing patterns and departures from patterns.**

A. Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot) (1.2)

1. Center and spread
2. Clusters and gaps
3. Outliers and other unusual features
4. Shape

B. Summarizing distributions of univariate data (1.3)

1. Measuring center: median, mean
2. Measuring spread: range, interquartile range, standard deviation
3. Measuring position: quartiles, percentiles, standardized scores (z-scores)
4. Using boxplots

C. Comparing distributions of univariate (dotplots, back-to-back stemplots, parallel boxplots) (1.2, 1.3)

1. Comparing center and spread: within groups, between group variation
2. Comparing clusters and gaps
3. Comparing outliers and other unusual features
4. Comparing shapes

E. Exploring categorical data (1.1)

1. Frequency tables and bar charts
2. Marginal and joint frequencies in two-way tables
4. Comparing distributions using bar charts

---

## Key Vocabulary

- |                             |               |                       |
|-----------------------------|---------------|-----------------------|
| ▪ individual                | ▪ association | ▪ median              |
| ▪ variable                  | ▪ dotplot     | ▪ quartiles           |
| ▪ frequency table           | ▪ stemplot    | ▪ $Q_1$ , $Q_3$       |
| ▪ relative frequency table  | ▪ histogram   | ▪ IQR                 |
| ▪ distribution              | ▪ SOCS        | ▪ five-number summary |
| ▪ pie chart                 | ▪ outlier     | ▪ minimum             |
| ▪ bar graph                 | ▪ symmetric   | ▪ maximum             |
| ▪ two-way table             | ▪ $\Sigma$    | ▪ boxplot             |
| ▪ marginal distributions    | ▪ $\bar{x}$   | ▪ resistant           |
| ▪ conditional distributions | ▪ spread      | ▪ standard deviation  |
| ▪ side-by-side bar graph    | ▪ variability | ▪ variance            |

## Summary

### Section 1.1

- Data sets contain information on a number of *individuals*.
- A *variable* describes some characteristic of an individual.
- *Categorical* variables describe some characteristic of an individual and place individuals into categories while *quantitative* variables have numerical values that measure some characteristic of each individual.
- *W<sup>5</sup>HW* - When you meet a new set of data, ask yourself the following key questions:
  - *Who* are the individuals described in the data? How many individuals are there?
  - *What* are the variables? In what units are they recorded?
  - *Why* were the data gathered? Do we hope to answer some specific question?
  - *When, where, how, by who* were the data produced?
- The distribution of a categorical variable lists the categories and gives the count (freq table) or percent (relative freq table) of individuals that fall into each category.
- *Pie charts* and *bar graphs* display the distribution of a categorical variable. Bar charts can also be used to compare any set of quantities measured in the same variable.
- When examining any graph, **ask yourself, “What do I see?”**
- A *two-way table* of counts organizes data about two categorical variables. They are often used for large amounts of data by grouping outcomes into categories.
- The row totals and column totals of a two-way table give the *marginal distributions* of the two individual variables.
- There are two sets of *conditional distributions* for a two-way table: the distributions of the row variable for each value of the column variable, and the distributions of the column variable for each value of the row variable. *Side-by-side bar graphs* can be used to display conditional distributions.
- **4-Step Process** – Statistical problems should be organized using four steps: **(1) State, (2) Plan, (3) Do, and (4) Conclude.**
- To describe the *association* between the row and column variables, compare an appropriate set of conditional distributions. Remember that even a strong association between two categorical variables can be influenced by other variables lurking in the background.

### Section 1.2

- *Dotplots, stemplots, or histograms* can be used to show the distributions of *quantitative variables*.
- When examining any graph, look for an **overall pattern** and for notable **departures** from that pattern. **Shape, center, and spread** describe the overall pattern of the distribution of a quantitative variable. **Outliers** are observations that lie outside the overall pattern of a distribution. Always look for outliers and try to explain them. **Don’t forget your SOCS! Include context!**
- Some distributions have simple shapes such as symmetric or skewed. The number of modes is another aspect of shape. Not all distributions have a simple shape.

- Remember, histograms are for quantitative data; bar graphs are for categorical data. Also, be sure to use relative frequency histograms when comparing data sets of different sizes.

### Section 1.3

- A numerical summary of a distribution should report its **center** and **spread**, or variability.
- The **mean** and **median** describe the center of a distribution in different ways. The mean is the average of the observations and the median is the midpoint of the values.
- When you use the median to indicate the center of a distribution, describe its spread with **quartiles**. The first quartile has about one-fourth of the observations below it, and the third quartile has about three-fourths of the observations below it. An extreme observation is an outlier if it is smaller than  $Q_1 - (1.5 \times \text{IQR})$  or larger than  $Q_3 + (1.5 \times \text{IQR})$ .
- The **five-number summary** consists of the median, the quartiles, and the high and low extremes and provides a quick overall description of the distribution. The median describes the center and the quartiles and extremes describe the spread.
- **Boxplots** are based upon the five-number summary and are useful for comparing two or more distributions.
- The **variance**,  $s^2$  and especially its square root, the **standard deviation**  $s$ , are common measures of spread about the mean as center. The standard deviation is zero when there is no spread and gets larger as the spread increases.
- The median is a **resistant** measure of center because it is relatively unaffected by extreme observations. The mean is **nonresistant**. The quartiles are resistant but the standard deviation is not.
- The mean and standard deviation are strongly influenced by outliers or skewness. They are good descriptions for symmetric distributions and are most useful for the Normal distribution.
- The median and quartiles are not affected by outliers. The five-number summary is the preferred numerical summary for skewed distributions.
- Numerical summaries do not fully describe the shape of a distribution. *Always plot your data.*

## AP Exam Tips

- If you learn to distinguish categorical from quantitative variables now, it will pay big rewards later. The type of data determines what kinds of graphs and which numerical summaries are appropriate. You will be expected to analyze categorical and quantitative data effectively on the AP Exam.
  - When comparing distributions of quantitative data, it is not enough to list values for the center and spread of each distribution. You have to explicitly *compare* these values, using words like “greater than,” “less than” or “about the same as.”
  - If you are asked to make a graph on a free response question, be sure to label and scale your axes. Unless your calculator shows labels and scaling, do not just transfer a calculator screen shot to your paper.
  - You may be asked to determine whether a quantitative data set has outliers. Be prepared to state and use the rule for identifying outliers.
  - Use statistical terms carefully and correctly on the AP Exam. Do not say “mean” if you mean “median.” For that matter, do not use the word “average.” Range is a single number; so are  $Q_1$ ,  $Q_3$ , and IQR. Avoid colloquial use of language like “the outlier *skews* the mean.” Skewed is a shape. If you misuse a term, expect to lose some credit. Remember, as CJ Kolson used to say, “Professionals speak in professional terms.”
- 

## Chapter 1 Portfolio Items

- 1.1 **W<sup>5</sup>HW** – Explain the steps involved in W<sup>5</sup>HW.
- 1.2 **Four-Step Process** – Explain the 4-step process.
- 1.3 **SOCS** – Explain the meaning of SOCS.
- 1.4 **Measures of Center and Location** – List and discuss the various measures of center and location.
- 1.5 **Measures of Spread** – List and discuss the various measures of spread. Include a discussion of how to identify outliers.
- 1.6 **Categorical Data** – Using the *categorical* data set of your choice construct a bar graph to describe the distribution categorical data. Include a list of your data. Use SOCS to describe the data set from your graph.
- 1.7 **Quantitative Data**. Using the *quantitative* data set of your choice, create a graphical display of the data and use the display and SOCS to interpret it in terms of the shape, center, and spread of the distribution, as well as gaps and outliers. Include a list of your data. Use a variety of numerical measures to describe the distribution. These should include mean, median, quartiles, five-number summary, standard deviation, range and variance.

## Can you?

1. Identify the individuals and variables in a set of data?
2. Classify variables as categorical or quantitative? Identify units of measurement for a quantitative variable?
3. Make a bar graph of the distribution of a categorical variable to compare related quantities?
4. Recognize when a pie chart can and cannot be used?
5. Identify what makes some graphs deceptive?
6. From a two-way table of counts, answer questions involving marginal and conditional distributions?
7. Describe the relationship between two categorical variables by computing appropriate conditional distributions?
8. Construct bar graphs to display the relationship between two categorical variables?
9. Make a dotplot or stemplot to display small sets of data?
10. Describe the overall pattern (shape, center, spread) of a distribution and identify major departures from the pattern (like outliers)?
11. Identify the shape of a distribution from a dotplot, stemplot, or histogram as roughly symmetric or skewed? Identify the number of modes?
12. Interpret histograms?
13. Calculate and interpret measures of center (mean, median)?
14. Calculate and interpret measures of spread (IQR and standard deviation)?
15. Identify outliers using the  $1.5 \times \text{IQR}$  rule?
16. Compute a five-number summary and make a boxplot?
17. Select appropriate measures of center and spread?
18. Use appropriate graphs and numerical summaries to compare distributions of quantitative variables?



### 3. Making Histograms

- Enter data into L<sub>1</sub>
- Check [y=] to ensure no functions are housed there
- Press [2<sup>ND</sup>] [y=]

```
STAT PLOT
1:Plot1...Off
  [L1] [L2]
2:Plot2...Off
  [L1] [L2]
3:Plot3...Off
  [L1] [L2]
4↓PlotsOff
```

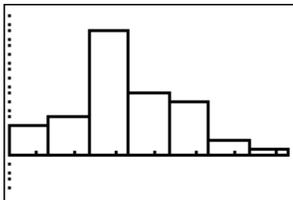
- Select 1:
- Turn Plot1 **ON**
- Choose the Histogram plot

```
Plot1 Plot2 Plot3
On Off
Type: [L1] [L2] [L3] [L4]
      [L5] [L6] [L7] [L8]
Xlist:L1
Freq:1
```

- Set the window to match the class intervals chosen by pressing [WINDOW] and entering information

```
WINDOW
Xmin=35
Xmax=75
Xscl=5
Ymin=-3
Ymax=15
Yscl=1
Xres=1
```

- Press [ZOOM] and select 9:ZoomStat





## Chapter 2 – Describing Location in a Distribution

---

**AP Standards.** The following AP Standards are covered in Chapter 2:

**I. Exploring Data: Describing patterns and departures from patterns.**

- B. Summarizing distributions of univariate data (2.1)
- 3. Measuring position: quartiles, percentiles, standardized scores (z-scores)
- 5. The effect of changing units on summary measures

**III. Anticipating Patterns**

- C. The normal distribution (2.2)
    - 1. Properties of the normal distribution
    - 2. Using tables of the normal distribution
    - 3. The normal distribution as a model for measurements
- 

**Key Vocabulary**

- percentiles
- cumulative relative frequency graphs
- z-scores
- transforming data
- density curves
- median of density curve
- transform data
- mean of density curve
- standard deviation of density curve
- Normal curves
- Normal distributions
- 68-95-99.7 rule
- $N(\mu, \sigma)$
- standard Normal distribution
- standard Normal table
- Normal probability plot
- $\mu$  mu
- $\sigma$  sigma

## Summary

### Section 2.1

- This chapter focuses on describing an individual value's location within a distribution of data and modeling distributions with density curves.
- Z-scores and percentiles provide easily calculated measures of relative standing for individuals.
- To **standardize** any observation  $x$ , subtract the mean of the distribution and then divide by the standard deviation:

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

The resulting z-score tells how many standard deviations  $x$  lies from the distribution mean.

- An **observation percentile** is the percent of the distribution that is at or below the value of the observation.
- A **cumulative relative frequency graph** allows us to examine location within a distribution. They begin by grouping the observations into equal-width classes. The completed graph shows the accumulating percent of observations as you move through the classes in increasing order.
- It is common to **transform data**, especially when changing units of measurement. When you add a constant  $a$  to all the values in a data set, measures of center (mean, median) and location (quartiles, percentiles) increase by  $a$ . Measures of spread do not change. When you multiply all the values in a data set by a constant  $b$ , measures of center and location are multiplied by  $b$ , while measures of spread are multiplied by  $|b|$ . Neither of these transformations changed the shape of the distribution.
- We can sometimes describe the overall pattern of a distribution by a **density curve**. Density curves come in assorted shapes. A density curve always remains on or above the horizontal axis and has total area 1 underneath it. An area of a density curve gives the proportion of observations that fall in a range of values.
- A density curve is an idealized description of the overall pattern of a distribution that smooths out the irregularities in the actual data. We write the **mean of a density curve** as  $\mu$  and the **standard deviation of a density curve** as  $\sigma$  to distinguish them from the mean  $\bar{x}$  and the standard deviation  $s_x$  of the actual data.
- **Chebyshev's inequality** gives us a useful rule of thumb for the percent of observations in any distribution that are within a number of standard deviations from the mean. In any distribution, the percent of observations falling within  $k$  standard deviations from the mean is at least

$$(100) \left(1 - \frac{1}{k}\right)$$

- The mean, median and quartiles of a density curve can be located by eye. The mean is the balance point of the curve. The median divides the area under the curve in half. The quartiles, with the median, divide the area under the curve into quarters. The standard deviation cannot be located by eye on most density curves.
- The mean and median are equal for symmetric density curves. The mean of a skewed curve is located farther toward the long tail than the median is.

## Section 2.2

- The **Normal distributions** are described by a special family of bell-shaped, symmetric density curves, called Normal curves. The mean and standard deviation completely specify a Normal distribution. The mean is the center of the curve and the standard deviation is the distance from the mean to the inflection points on either side.
- Normal distributions satisfy the **68-95-99.7 rule** which describes what percent of the observations lie within one, two, and three standard deviations from the mean.
- All Normal distributions are the same when measurements are standardized. If  $x$  is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ , then  $z = (x - \mu)/\sigma$  has the standard normal distribution with mean 0 and standard deviation 1.
- **Table A** gives the proportions of standard Normal observations that are less than  $z$  for many values of  $z$ . By standardizing, we can use Table A for any Normal distribution.
- To perform certain inference procedures in later chapters, we will need to know that the data come from populations that are approximately Normally distributed. To **assess Normality**, one can observe the shape of histograms, stemplots, and boxplots to see how well the data fit the 68-95-99.7 rule for Normal distributions. Another good method is to construct a **Normal probability plot**.

## AP Exam Tips

- Do not use “calculator speak” when showing work on free response questions. Writing **normalcdf(305, 325, 304, 8)** will *not* earn you credit for a Normal calculation. At the very least, you must indicate what each of those calculator inputs represents. For example, “I used normalcdf with lower bound 305, upper bound 325, mean 304, and standard deviation 8.” Better yet, sketch and label a Normal curve to show what you are finding.
  - Normal probability plots are not included on the AP Statistics course outline. However, these graphs are very useful tools for assessing Normality. You may use them on the AP Exam if you wish – just be sure that you know what you are looking for (linear pattern).
- 

## Chapter 2 Portfolio Items

2.1 Using an example, *explain* how to compute measures of relative standing for individual values in a distribution. This should include *standardized values (z-scores)* and *percentile ranks*.

2.2 Explain the properties of the *Normal distribution* and the *68-95-99.7 Rule*. Include examples.

2.3 Explain how to use tables to find (a) the proportion of values on an interval of the Normal distribution and (b) a value with a given proportion of observations above or below it. Include examples.

2.4 Explain how to use technology to find (a) the proportion of values on an interval of the Normal distribution and (b) a value with a given proportion of observations above or below it. Include examples.

2.5 Explain how to assess Normality. Include an example.

### Can you?

1. Use percentiles to locate individual values within distributions of data?
2. Interpret a cumulative relative frequency graph?
3. Find the standardized value (z-score) of an observation? Interpret z-scores in context?
4. Describe the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data?
5. Approximately locate the median (equal-area point) and the mean (balance point) on a density curve?
6. Use the 68-95-99.7 rule and symmetry to state what percent of the observations from a Normal distribution fall between two points when the points lie at the mean or one, two, or three standard deviations on either side of the mean?
7. Use the standard Normal distribution to calculate the proportion of values in a specified interval?
8. Use the standard Normal distribution to determine a z-score from a percentile?
9. Use Table A to find the percentile of a value from any Normal distribution and the value that corresponds to a given percentile?
10. Make an appropriate graph to determine if a distribution is bell-shaped?
11. Use the 68-95-99.7 rule to assess Normality of a data set?

1. Finding areas with normalcdf

- Press [2<sup>ND</sup>] [VARS] (DIST) and choose 2:normalcdf(

```

0:QUIT DRAW
1:normalPdf(
2:normalcdf(
3:invNorm(
4:invT(
5:tPdf(
6:tcdf(
7:χ²Pdf(
    
```

- Complete the command **normalcdf(** with the **lower bound**, **upper bound**, **mean**, **standard deviation** and press [ENTER]

Note: if there is no lower bound, use –E99; if there is no upper bound, use E9. The E key is [2<sup>nd</sup>] [, ] (EE)

```

normalcdf(125, E9
9, 100, 15
.0477903304
    
```

Note: if no mean and standard deviation are specified, the calculator defaults to a mean of 0 and standard deviation of 1 (standard Normal).

2. Finding values with invNorm

- Press [2<sup>ND</sup>] [VARS] (DIST) and choose 3:invNorm(

```

0:QUIT DRAW
1:normalPdf(
2:normalcdf(
3:invNorm(
4:invT(
5:tPdf(
6:tcdf(
7:χ²Pdf(
    
```

- Complete the command **invNorm(** with the **desired area**, **mean**, **standard deviation** and press [ENTER]

```

invNorm(.9, 100, 1
5
119.2232735
    
```

Note: if no mean and standard deviation are specified, the calculator defaults to a mean of 0 and standard deviation of 1 (standard Normal).

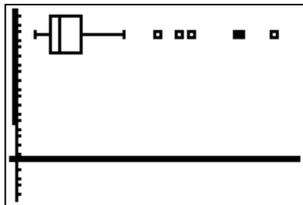
### 3. Normal probability plots

- Enter data in desired List
- Calculate **1-Var Stats**

```
1-Var Stats
x̄=141.8472222
Σx=10213
Σx²=2295469
Sx=109.2086335
σx=108.4475884
↓n=72
```

```
1-Var Stats
↑n=72
minX=43
Q1=82.5
Med=102.5
Q3=151.5
maxX=598
```

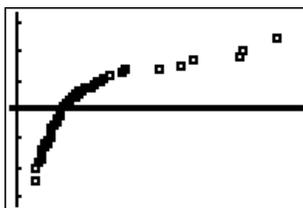
- Compare the mean and median. If they are close it suggests that distribution is fairly symmetric.
- Create a **boxplot**. Examine the plot for symmetry.



- To construct a Normal probability plot, define Plot1 like this:

```
Plot1 Plot2 Plot3
Off Off
Type: L1 L2 L3
Data List: L1
Data Axis: Y
Mark: ■ + .
```

- Press **[ZOOM]** and select **9:ZoomStat**



If the Normal probability plot is linear, it is reasonable that the data follow a Normal distribution.



## Chapter 3 – Examining Relationships

---

**AP Standards.** The following AP Standards are covered in Chapter 3:

### I. Exploring Data: Describing patterns and departures from patterns.

D. Exploring bivariate data (3.1, 3.2)

1. Analyzing patterns in scatterplots
  2. Correlation and linearity
  3. Least-squares regression line
  4. Residual plots, outliers and influential points
- 

## Summary

### Section 3.1

- Exploring bivariate data involves examining data to find relations between two quantitative variables.
- A scatterplot displays the individual relationship between two quantitative variables measured on the same individuals. Mark values of one variable on the x-axis and the values of the other variable on the y-axis. Plot each individual's data as a point on the graph.
- If we think that a variable  $x$  may help explain, predict, or even cause changes in another variable  $y$ , we call  $x$  an **explanatory variable** and  $y$  a **response variable**. Always plot the explanatory variable, if there is one, on the x-axis and the response variable on the y-axis.
- In examining a scatterplot, look for an overall pattern showing the **direction**, **form**, and **strength** of the relationship and look for **outliers** or other departures from the pattern.
- **Direction:** If the relationship has a clear direction, we speak of either a **positive association** (high values of two variables tend to occur together) or **negative association** (high values of one variable tend to occur with low values of the other variable).
- **Form:** Linear relationships, where the points show a straight-line pattern, are an important form of relationship. Curved relationships and clusters are other forms to watch for.
- **Strength:** The strength of a relationship is determined by how close the points in the scatterplot lie to a simple form such as a line.
- The **correlation**  $r$  measures the strength and direction of the linear relationship between two quantitative variables  $x$  and  $y$ . Although you can calculate a correlation for any scatterplot,  $r$  measures only straight-line relationships.
- Correlation indicates the direction of a linear relationship by its sign:  $r > 0$  for a positive association and  $r < 0$  for a negative association. Correlation always satisfies  $-1 < r < 1$  and indicates the strength of the relationship by how close it is to  $-1$  or  $1$ . Perfect correlation,  $r = \pm 1$ , occurs only when the points on a scatterplot lie exactly on a straight line.

- Remember these important facts about  $r$ : Correlation ignores the distinction between explanatory and response variables. The value  $r$  is not affected by changes in the unit of measurement of either variable. Correlation is not resistant, so outliers can greatly change the value of  $r$ .

### Section 3.2

- A **regression line** is a straight line that describes how a response variable  $y$  changes as an explanatory variable  $x$  changes. You can use a regression line to **predict** the value of  $y$  for any value of  $x$  by substituting this  $x$  into the equation of the line.
- The **slope  $b$**  of the regression line  $\hat{y} = a + bx$  is the rate at which the predicted response  $\hat{y}$  changes along the line as the explanatory variable  $x$  changes. Specifically,  $b$  is the *predicted* change in  $y$  when  $x$  increases by 1 unit.
- The  **$y$  intercept  $a$**  of a regression line  $\hat{y} = a + bx$  is the predicted response  $\hat{y}$  when the explanatory variable  $x = 0$ . This prediction has no statistical use unless  $x$  can actually take on values near 0.
- Avoid **extrapolation**, the use of a regression line for prediction using values of the explanatory variable outside the range of the data from which the line was calculated.
- The most common method of fitting a line to a scatterplot is least squares. The **least-squares regression line** is the straight line  $\hat{y} = a + bx$  that minimizes the sum of the squares of the vertical distances of the observed points from the line.
- The least-squares regression line of  $y$  on  $x$  is the line with slope  $b = r(s_y/s_x)$  and intercept  $a = \bar{y} - b\bar{x}$ . This line always passes through the point  $(\bar{x}, \bar{y})$ .
- You can examine the fit of a regression line by studying the **residuals**, which are the differences between the observed and predicted values of  $y$ . Be on the lookout for points with unusually large residuals and also for nonlinear patterns and uneven variation about the residual = 0 line in the **residual plot**.
- The **standard deviation of the residuals** measures the average size of the prediction errors (residuals) when using the regression line.
- The **coefficient of determination  $r^2$**  is the fraction of the variation in one variable that is accounted for by the least-squares regression on the other variable.
- Correlation and regression must be interpreted with caution. Plot the data to be sure the relationship is roughly linear and to detect outliers. Also look for **influential observations**, individuals that substantially change the correlation or the regression line. Outliers in  $x$  are often influential for the regression line.
- Most of all, be careful not to conclude there is a **cause-and-effect** relationship between two variables just because they are strongly associated. **ASSOCIATION DOES NOT IMPLY CAUSATION!**

## AP Exam Tips

- If you are asked to make a scatterplot on a free-response question, be sure to label and scale both axes. Do not copy unlabeled calculator graph directly onto your paper.
  - If you are asked to interpret a correlation, start by looking at a scatterplot of the data. Then be sure to address direction, form, strength, and outliers and put your answer in context.
  - There is no firm rule for how many decimal places to show for answers on the AP Exam. Give your answer correct to two or three nonzero decimal places. *Exception:* If you are using one of the tables, give the value shown in the table.
  - The formula sheet for the AP Exam uses different notation for the equations of the slope and y-intercept of the least squares regression line:  $b_1 = r \frac{s_y}{s_x}$  and  $b_0 = \bar{y} - b_1 \bar{x}$ . That is because the least squares regression line is written as  $\hat{y} = b_0 + b_1 x$ . We prefer our simpler version without the subscripts:  $\hat{y} = a + bx$ .
  - Students often have a hard time interpreting the value of  $r^2$  on AP Exam questions. They frequently leave out key words in the definition. Treat this as a fill in exercise. Write “\_\_\_\_% of the variation in (response variable name) is accounted for by the regression line.”
  - Do not forget to put a “hat” on the response variable when you write a regression equation. Calculator and computer output for regression usually do not do this. For example, you should write  $\widehat{Fat\ Gain} = 3.505 - 0.00344(NEA)$ .
- 

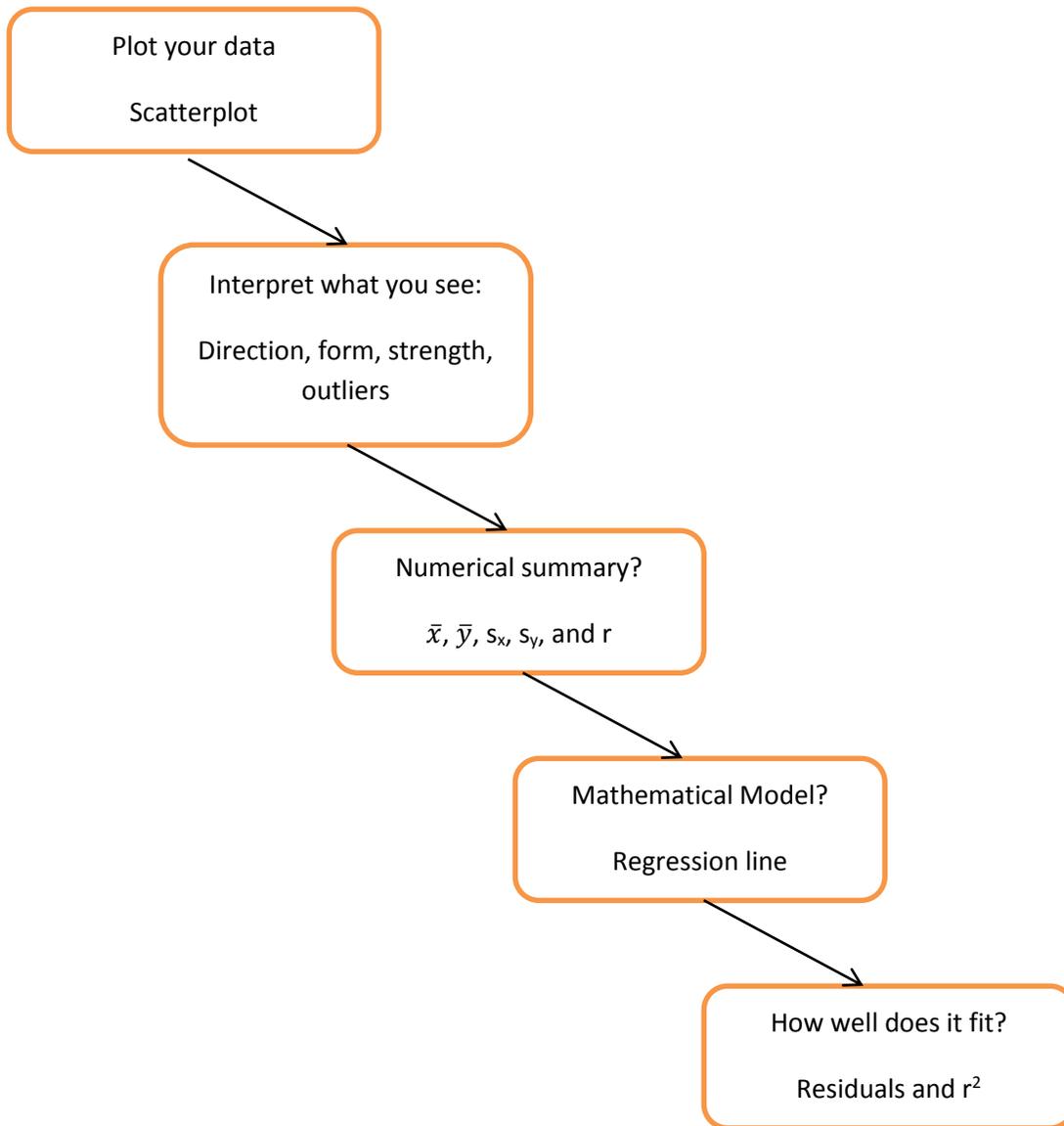
## Chapter 3 Portfolio Items

- 3.1 Explain how to construct and interpret a scatterplot for a set of bivariate data. Include an example.
- 3.2 Explain how to compute and interpret the correlation  $r$  between two variables. Include an example.
- 3.3 Explain how to construct a *least-squares regression line* and how to use it as a predictor model. Include an example.
- 3.4 Explain how to measure the quality of a regression line as a model for bivariate data. Include an example.
- 3.5 Explain how to find the slope and intercept of the least-squares regression line from the means and standard deviations of  $x$  and  $y$  and their correlation. Include an example. Explain the meaning of the slope and y-intercept in context of the problem.

(As a suggestion, you may want to use the same data set/example for all of the items.)

## Can you?

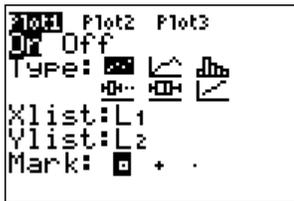
1. Describe why it is important to investigate relationships between variables?
2. Identify explanatory and response variables in situations where one variable helps to explain or influences the other?
3. Make a scatterplot to display the relationship between two quantitative variables?
4. Describe the direction, form, and strength of the overall pattern of a scatterplot?
5. Recognize outliers in scatterplots?
6. Know the basic properties of correlation?
7. Calculate and interpret correlation?
8. Explain how the correlation  $r$  is influenced by extreme observations?
9. Interpret the slope and y-intercept of a least-squares regression line?
10. Use the least-squares regression line to predict  $y$  for a given  $x$ ?
11. Explain the dangers of extrapolation?
12. Calculate and interpret residuals?
13. Explain the concept of least squares?
14. Use technology to find a least-squares regression line?
15. Find the slope and intercept of the least-squares regression line from the means and standard deviations of  $x$  and  $y$  and their correlation?
16. Construct and interpret residual plots to assess if a linear model is appropriate?
17. Use the standard deviation of the residuals to assess how well the line fits the data?
18. Use  $r^2$  to assess how well the line fits the data?
19. Identify the equation of a least-squares regression line from a computer output?
20. Explain why association does not imply causation?
21. Recognize how the slope, y-intercept, standard deviation of the residuals and  $r^2$  are influenced by extreme observations?



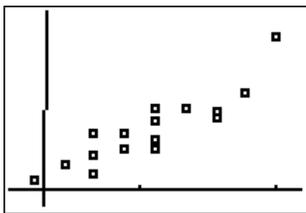
## Technology

### 1. Constructing a scatterplot

- Enter the explanatory variable in  $L_1$  and the response variable in  $L_2$ .
- Define a scatterplot in the statistics plot menu



- Press **[ZOOM]** and select **9:ZoomStat**



- **[TRACE]** can be used to determine data point values

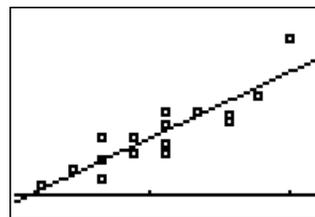
### 2. Least-squares regression

- Enter the data and construct a scatterplot.
- To determine the least-squares regression line: press **[STAT]**, choose **8:LinReg (a+bx)**.
- Complete the command to read **LinReg (a+bx) L1, L2, Y1**. ( $Y_1$  is found under **[VARS] <Y VARS> 1:Function**) and press **[ENTER]**.

```
LinReg(ax+b) L1,  
L2,Y1
```

```
LinReg  
y=ax+b  
a=.0179637619  
b=-.012700604  
r2=.7998407228  
r=.8943381479
```

- If  $r^2$  and  $r$  do not appear, press **[2<sup>nd</sup>] [0]** (CATALOG), scroll down to **DiagnosticOn** and press **[ENTER]**.
- Deselect all other equations in the **Y=** screen and press **[GRAPH]** to overlay the least-squares regression line on the scatterplot



### 3. Constructing residual plots

- Repeat all the steps in **2. Least-squares regression** above.
- Define  $L_3$  as **RESID**. This can be done by going to **[STAT] EDIT** and setting  $L_3 =$   $L_1$ RESID which is found by typing **[2<sup>nd</sup>] [STAT] (LIST)**, choosing **NAMES** and **RESID**.

```

2ND] CALC TESTS
1:Edit...
2:SortA(
3:SortD(
4:ClnList
5:SetUpEditor
    
```

| L1 | L2   | RES   | 3 |
|----|------|-------|---|
| 5  | .1   | ----- |   |
| 2  | .03  |       |   |
| 9  | .19  |       |   |
| 8  | .12  |       |   |
| 3  | .04  |       |   |
| 7  | .095 |       |   |
| 3  | .07  |       |   |

L3 = LRESID

```

NAMES OPS MATH
1:L1
2:L2
3:L3
4:L4
5:L5
6:L6
RESID
    
```

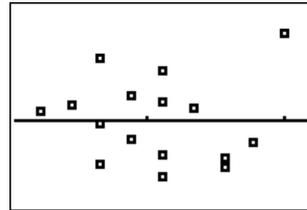
- This will populate  $L_3$  with the residuals that were computed when you performed the least-squares regression.

| L1 | L2   | L3     | 3 |
|----|------|--------|---|
| 5  | .1   | .00677 |   |
| 2  | .03  | .04103 |   |
| 9  | .19  | -.011  |   |
| 8  | .12  | -.0012 |   |
| 3  | .04  | -.018  |   |
| 7  | .095 | .02881 |   |
| 3  | .07  |        |   |

L3(1) = .0228817946...

- Turn off Plot1 and deselect the regression equation. Specify Plot2 with  $L_1$  as the XList and  $L_3$  as the YList.

- Press **[ZOOM]** and select **9:ZoomStat**



- The x-axis serves as a reference line, with points above the line corresponding to positive residuals and points below the line corresponding to negative residuals. **[TRACE]** can be used to find individual values.
- 1-Variable Stats** can be calculated on  $L_3$  to determine that the sum of the residuals is 0.

```

1-Var Stats
x̄=2.8125E-14
Σx=4.5E-13
Σx²=.005849655
Sx=.0197478351
σx=.0191207591
↓n=16
    
```

Note: in this case, the calculator gives the sum of the residuals as  $4.5 \times 10^{-13}$  which is its way of telling us it is 0.



## Chapter 4 – Designing Studies

---

**AP Standards.** The following AP Standards are covered in Chapter 4:

### II. Sampling and Experimentation: Planning and conducting a study

A. Overview of methods of data collection (4.1, 4.2)

1. Census
2. Sample survey
3. Experiment
4. Observation study.

B. Planning and conducting surveys (4.1)

1. Characteristics of a well-designed and well-conducted survey
2. Populations, samples and random selection
3. Sources of bias in sampling and surveys
4. Sampling methods, including simple random sampling, stratified random sampling and cluster sampling

C. Planning and conducting experiments (4.2)

1. Characteristics of a well-designed and well-conducted experiment
2. Treatments, control groups, experimental units, random assignments and replication
3. Sources of bias and confounding, including placebo effect and blinding
4. Completely randomized designs
5. Randomized block design, including matched pairs design

D. Generalizability of results and types of conclusions that can be drawn from observational studies, experiments and surveys. (4.3)

---

### Key Vocabulary

- |                             |                            |                                    |
|-----------------------------|----------------------------|------------------------------------|
| ▪ sample                    | ▪ strata                   | ▪ placebo effect                   |
| ▪ population                | ▪ undercoverage            | ▪ single blind experiment          |
| ▪ sample survey             | ▪ nonresponse              |                                    |
| ▪ voluntary response sample | ▪ response bias            | ▪ control group                    |
| ▪ confounded                | ▪ sampling frame           | ▪ completely randomized experiment |
| ▪ design                    | ▪ systematic random sample | ▪ randomized block design          |
| ▪ convenience sampling      | ▪ observational study      | ▪ matched pair design              |
| ▪ biased                    | ▪ experimental             | ▪ statistically significant        |
| ▪ simple random sample      | ▪ confounding              | ▪ replication                      |
| ▪ table of random digits    | ▪ lurking variable         | ▪ hidden bias                      |
| ▪ probability sample        | ▪ experimental units       | ▪ double-blind experiment          |
| ▪ stratified random sample  | ▪ subjects                 | ▪ block design                     |
| ▪ cluster sampling          | ▪ random assignment        | ▪ data ethics                      |
| ▪ inference                 | ▪ treatment                |                                    |
| ▪ margin of error           | ▪ factor                   |                                    |
|                             | ▪ level                    |                                    |

## Summary

### Section 4.1

- Designs for producing data are essential parts of statistics in practice.
- Random sampling and randomized comparative experiments are very important.
- **Sampling** selects a part of a population of interest to represent the whole.
- A **sample survey** selects a sample from the population of all individuals. We base conclusions about the population based upon data about the sample.
- **Random sampling** uses chance to select a sample.
- The **sampling method** refers to the method used to select the sample from the population. **Probability sampling methods** use impersonal chance to select a sample.
- **Multistage samples** select successively smaller groups within the population in stages, resulting in a sample consisting of clusters of individuals. Each stage may employ an SRS, a stratified sample, or another type of sample.
- **Larger samples** give more accurate results than smaller samples but might have greater opportunity costs associated with them.
- The basic probability sample is a **Simple Random Sample (SRS)**. An SRS gives every possible sample of a given size the same chance of being chosen.
- Choose an SRS by labeling the members of the population and using a **table of random digits** to select the sample. Technology can automate this process.
- To choose a **stratified random sample**, divide the population into **strata**, groups of similar individuals. Then choose a separate SRS from each stratum and combine them to form the full sample.
- To choose a **cluster sample**, divide the population into groups, or clusters. Randomly select some of these clusters. All of the individuals in the chosen cluster are then selected to be in the sample.
- Failure to use probability sampling often results in **bias**, or systematic errors in the way the sample represents the population. **Voluntary response** samples, in which the respondents choose themselves, and **convenience samples**, in which individuals easiest to reach are chosen, are particularly prone to large bias.
- **Sampling errors** come from the act of choosing a sample. Random sampling error and **undercoverage** are common types of sampling error. Undercoverage occurs when some members of the population are left out of the **sampling frame**, the list from which the sample is actually chosen.
- The most serious errors in most careful surveys, however, are **nonsampling errors**. These have nothing to do with choosing the sample – they are present even in a census. The single biggest problem for sample surveys is **nonresponse**: people cannot be contacted or refuse to answer. Incorrect answers by respondents can lead to **response bias**. Finally, the exact **wording of questions** has a big influence on answers.

## Section 4.2

- We can produce data intended to answer specific questions by **observational** studies or experiments. Experiments are distinguished from **observational studies** such as sample surveys by the active application of some treatment on the subjects of the experiment.
- Statistical studies often try to show that changing one variable (the explanatory variable) causes changes in another variable (the response variable). Variables are **confounded** when their effects on a response cannot be distinguished from each other. Observational studies and uncontrolled experiments often fail to show that changes in an explanatory variable actually cause changes in a response variable because the explanatory variable is confounded by other variables.
- In an experiment, one or more **treatments** are imposed on the **experimental units** or **subjects**. Each treatment is a combination of **levels** of the explanatory variables, which are called **factors**.
- The **design** of an experiment refers to the choice of treatments and the manner in which the experimental units or subjects are assigned to the treatments.
- The basic principles of statistical design of experiments are **control**, **replication**, and **randomization**.
- The simplest form of control is **comparison**. Experiments should compare two or more treatments in order to prevent **confounding** the effect of a treatment with other influences, such as lurking variables.
- **Replication** of treatments on many units reduces the role of chance variation and makes the experiment more sensitive to differences among the treatments.
- **Randomization** uses chance to assign subjects to treatments. Randomization creates treatment groups that are similar (except for chance variation) before the treatments are applied. Randomization and comparison together prevent bias, or systematic favoritism, in experiments.
- Randomization can be performed by giving numerical labels to the experimental units and using a **table of random digits** to choose treatment groups.
- Another form of control is to restrict randomization by forming **blocks** of experimental units. The units in each block are similar in some way that is important to the response. Randomization is then carried out separately within each block.
- **Matched pairs** are a common form of blocking for comparing just two treatments. In some matched pairs designs, each subject receives both treatments in a random order. In others, the subjects are matched in pairs as closely as possible and one subject receives one treatment and the other receives the other treatment.
- Good experiments require attention to detail as well as good statistical design. Many behavioral and medical experiments are **double-blind**. That is, neither the subjects nor those interacting know who is receiving which treatment. If one party knows and the other does not, then the experiment is **single-blind**.
- The **placebo effect** is a term that doctors use to describe the phenomenon where patients get better because they expect the treatment to work even though they have been administered a placebo or fake drug.

### Section 4.3

- Most statistical studies aim to make inferences that go beyond the data actually produced. **Inference about the population** requires that the individuals taking part in a study be randomly selected from the larger population. A well-designed experiment that randomly assigns treatments to experimental units allows for **inference about cause and effect**.
- **Lack of realism** in an experiment can prevent us from generalizing its results.
- In the absence of an experiment, good evidence of causation requires a strong association that appears consistently in many studies, a clear explanation for the alleged causal link, and careful examination of possible lurking variables.
- Studies involving humans must be screened in advance by an **institutional review board**. All participants must give their **informed consent** before taking part. Any information about the individuals in the study must be kept confidential.
- Remember that randomized comparative experiments can answer questions that cannot be answered without them. Also remember that “interests of the subject must always prevail over the interests of science and society.”

## AP Exam Tips

- If you are asked to describe how the design of a study leads to bias, you are expected to identify the *direction* of the bias. Suppose you are asked, “Explain how using a convenience sample of students in your statistics class to estimate the proportion of all high school students who own a graphing calculator could result in bias.” You might respond, “This sample would probably include a much higher proportion of students with graphing calculators than in the population at large because a graphing calculator is required for the statistics class. That is, this method would probably lead to an overestimate of the actual population proportion.”
  - If you are asked to identify a possible confounding variable in a given setting, you are expected to explain how the variable you choose (1) is associated with the explanatory variable and (2) affects the response variable.
  - If you are asked to describe the design of an experiment on the AP Exam, you will not get full credit for a diagram like Figure 4.5 (p. 246). You are expected to describe how the treatments are assigned to the experimental units and to clearly state what will be measured or compared. Some students prefer to start with a diagram and then add a few sentences. Others choose to skip the diagram and put their entire response in narrative form.
  - Do not mix the language of experiments and the language of sample surveys or other observational studies. You will lose credit for saying things like “use a randomized block design to select the sample for this survey” or “this experiment suffers from nonresponse error since some of the subjects dropped out during the study.”
- 

## Chapter 4 Portfolio Items

4.1 Explain the differences between *observational studies* and *experiments*; discuss the advantages of each; give examples.

4.2 Identify and give examples of different types of *sampling methods* including a clear definition of a *simple random sample (SRS)*.

4.3 Identify and give examples of *sources of bias* in sample surveys.

4.4 Identify and explain *the three basic principles of experimental design*.

4.5 Explain what is meant by a *completely randomized design*; include an example.

4.6 Explain the difference between the purposes of *randomization* and *blocking* in an experimental design; include examples.

4.7 Explain how to use random numbers from a table or technology to select a random sample.

## Can you?

1. Identify the population and sample in a sample survey?
2. Identify voluntary response samples and convenience samples? Explain how these bad sampling methods can lead to bias?
3. Describe how to use Table D to select a simple random sample (SRS)?
4. Distinguish a simple random sample from a stratified random sample or cluster sample? Give advantages and disadvantages of each sampling method?
5. Explain how undercoverage, nonresponse, and question wording can lead to bias in a sample survey?
6. Distinguish between an observational study and an experiment?
7. Explain how a lurking variable in an observational study can lead to confounding?
8. Identify the experimental units or subjects, explanatory variables (factors), treatments, and response variables in an experiment?
9. Describe a completely randomized design for an experiment?
10. Explain why random assignment is an important experimental design principle?
11. Describe how to avoid the placebo effect in an experiment?
12. Explain the meaning and the purpose of blinding in an experiment?
13. Explain in context what “statistically significant” means?
14. Distinguish between a completely randomized design and a randomized block design?
15. Know when a matched pairs experimental design is appropriate and how to implement such a design?
16. Determine the scope of inference for a statistical study?
17. Evaluate whether a statistical study has been carried out in an ethical manner?

## Technology

### Random number generation

- Press **[MATH]** <PRB> 5:randInt(

```
MATH NUM CPX PRB
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(
```

- Complete the entry on **5:randint(** with the lowest integer value, highest integer value, and number of digits you want.

```
randInt(1,10,5
```

- Press **[ENTER]** and the numbers will be produced.

```
randInt(1,10,5
(10 10 2 6 5)
```

- Numbers can also be produced and stored in a list by using the **[STO>]** key

```
randInt(1,10,5)→
L1
```



## Chapter 5 – Probability: What Are the Chances?

---

**AP Standards.** The following AP Standards are covered in Chapter 6:

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

A. Probability (5.1, 5.2, 5.3)

1. Interpreting probability, including long-run relative frequency interpretation
2. “Law of large numbers” concept
3. Addition rule, multiplication rule, conditional probability and independence
5. Simulation of random behavior and probability distributions

---

### Key Vocabulary

- |                        |                            |                               |
|------------------------|----------------------------|-------------------------------|
| ▪ law of large numbers | ▪ replacement              | ▪ intersection (and)          |
| ▪ probability          | ▪ event                    | ▪ conditional probability     |
| ▪ simulation           | ▪ $P(A)$                   | ▪ independent events          |
| ▪ two-way table        | ▪ complement $A^c$         | ▪ general multiplication rule |
| ▪ sample space         | ▪ disjoint                 | ▪ general addition rule       |
| ▪ $S = \{H, T\}$       | ▪ mutually exclusive event | ▪ multiplication rule         |
| ▪ tree diagram         | ▪ Venn diagram             |                               |
| ▪ probability model    | ▪ union (or)               |                               |

### Summary

- A chance process has outcomes that we cannot predict but nonetheless have a regular distribution in very many repetitions. The **law of large numbers** says that the proportion of times that a particular outcome occurs in many repetitions will approach a single number. The long-run relative frequency of a chance outcome is its **probability**. A probability is a number between 0 (never occurs) and 1 (always occurs).
- Probabilities describe only what happens in the long run. Short runs of random phenomena often do not look random because they do not show the regularity that in fact emerges in very many repetitions.
- There are times when actually carrying out an experiment, sample survey or operational study is too costly, too slow, or impractical. A carefully designed **simulation** can provide approximate answers to questions. A **probability model** can be used to calculate a theoretical probability.
- A simulation is used to imitate chance behavior and is most often performed with random numbers representing **independent trials**.
- **Steps of a simulation:** Follow the **4-Step Process**:
  - **State:** Formulate a question of interest about some chance process.
  - **Plan:** Describe how to use a chance device to imitate one repetition of the process. Explain clearly how to identify the outcomes of the chance process and what variable to measure.
  - **Do:** Perform many repetitions of the simulation.
  - **Conclude:** Use the results of your simulation to answer the question of interest.

- A **probability model** for a random phenomenon consists of a **sample space**  $S$  and assignment of probabilities.
- The **sample space**  $S$  is the set of all possible outcomes of the random phenomenon. Sets of outcomes are called **events**. A number  $P(A)$  is assigned to an event  $A$  as its probability.
- The **multiplication principle** says if you can do one task in  $n$  ways and another task in  $m$  ways, then both tasks can be done in  $n \times m$  ways.
- **Sampling with replacement** requires that objects selected from distinct choices be replaced before the next selection. Probabilities are the same for each draw. In **sampling without replacement**, probabilities change for each new selection.
- The **complement**  $A^c$  or  $A'$  of an event  $A$  consists of exactly all the outcomes that are not in  $A$ . Events  $A$  and  $B$  are **disjoint** (mutually exclusive) if they have no outcomes in common. Events  $A$  and  $B$  are **independent** if knowing that one event occurs does not change the probability we should assign to the other event.
- Any assignment of probability must obey the rules that state the basic properties of probability:
  - $0 \leq P(A) \leq 1$  for any event  $A$ ;
  - $P(S) = 1$  for the sample space  $S$ ;
  - **Addition rule:** If events  $A$  and  $B$  are **disjoint**, then  $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$ ;
  - **Complement rule:** For any event  $A$ ,  $P(A^c) = 1 - P(A)$ ;
  - **Multiplication rule:** If events  $A$  and  $B$  are **independent**, then  $P(A \text{ and } B) = P(A \cap B) = P(A)P(B)$ ;
  - **General addition rule:**  $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$ ;
  - **General multiplication rule:**  $P(A \text{ and } B) = P(A \cap B) = P(A)P(B|A)$
  - **Conditional probability:**  $P(B|A)$  of an event  $B$  given an event  $A$  is defined by
 
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
- A **Venn diagram**, together with the general addition rule, can be helpful for finding the probabilities of the union of two events or the joint probability.
- Constructing a table is a good approach to determining a conditional probability.
- In problems with several stages, draw a **tree diagram** or organize use of the multiplication and addition rules.

### AP Exam Tips

- On the AP Exam, you may be asked to describe how you will perform a simulation using rows of random digits. If so, provide a clear enough description of your simulation process for the reader to get the same results you did from *only* your written explanation.
- Many probability problems involve simple computations that you can do on your calculator. It may be tempting to just write down your final answer without showing any supporting work. Do not do it! A “naked answer,” even if it is correct, will usually earn no credit on a free-response question.
- On probability questions, you may usually choose whether to use words or symbols when showing your work. You can write statements like  $P(A|B)$  if events  $A$  and  $B$  are defined clearly, or you can use verbal equivalent such as  $P(\text{reads the New York Times} | \text{reads USA Today})$ . Use the approach that makes the most sense to you.

## Portfolio Items for Chapter 5

5.1 Explain how to perform a simulation using a table of random numbers; include an example and use the 4-Step Process.

5.2 Create a list of the basic probability rules.

5.3 Demonstrate how to use the basic probability rules to solve probability problems. Include examples.

5.4 Choosing the example of your choice write out the sample space and use it to answer probability questions.

5.5 Explain the concept of disjoint (mutually exclusive) events; include examples. Include a discussion of how probability rules might change for disjoint events.

5.6 Explain the concept of independent events; include examples. Include a discussion of how probability rules might change for independent events.

5.7 Explain how to determine if two events are independent; include examples.

5.8 Explain how to solve problems involving conditional probability; include examples.

### Can you?

1. Interpret probability as a long-run relative frequency?

2. Use simulation to model chance behavior?

3. Describe a probability model for a chance process?

4. Use basic probability rules, including the complement rule and the addition rule for mutually exclusive events?

5. Use a Venn diagram to model a chance process involving two events?

6. Use the general addition rule?

7. When appropriate, use a tree diagram to describe chance behavior?

8. Use the general multiplication rule to solve probability questions?

9. Determine whether two events are independent?

10. Find the probability that an event occurs using a two-way table?

11. When appropriate, use the multiplication rule for two independent events to compute probabilities?

12. Compute conditional probabilities?



## Chapter 6 – Random Variables

---

**AP Standards.** The following AP Standards are covered in Chapter 7:

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

A. Probability (6.1, 6.2, 6.3)

4. Discrete random variables and their probability distributions, including binomial and geometric

6. Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable

B. Combining independent random variables (6.2)

1. Notion of independence versus dependence

2. Mean and standard deviation for sums and differences of independent random variables

---

### Key Vocabulary

- random variable
- discrete random variable
- probability distribution
- mean of a random variable
- variance of a random variable
- probability density curve
- continuous random variable
- standard deviation
- binomial setting
- binomial random variable
- binomial distribution
- binomial coefficient
- binomial probability
- linear transformation
- normal approximation
- geometric setting
- geometric distribution
- geometric random variable
- Normal approximation
- geometric probability
- factorial
- expected value
- standard deviation
- $\mu_X$
- $\mu_Y$
- uniform distribution

## Summary

- A **random variable** is a variable taking numerical values determined by the outcome of a random phenomenon.
- The **probability distribution** of a random variable tells us what the possible values of the random variable are and what probabilities are assigned to those values.
- A **discrete random variable** has a countable number of values. They are usually the result of **counting**. The distribution assigns each value a probability between 0 and 1 such that the sum of all of the probabilities is 1. The probability of any event is the sum of the probabilities of all the values that make up the event.
- A **continuous random variable** takes on all values in some interval of numbers. They are usually the result of **measuring**. A density curve describes the probability distribution of a continuous random variable. The probability of any event is the area under the curve above the values that make up the event.
- **Normal distributions** are one type of continuous probability distribution.
- A probability distribution can be portrayed by drawing a histogram in the discrete case or by graphing the density curve in the continuous case.
- The probability distribution of a random variable  $X$  has a mean  $\mu_x$  and a standard deviation  $\sigma_x$ .
- The mean is the balance point of the probability histogram or density curve.
- If  $X$  is discrete with possible values  $x_i$  having probabilities  $p_i$ , the mean is the average of the values of  $X$ , each weighted by its probability:
  - $\mu_x = x_1p_1 + x_2p_2 + \dots + x_kp_k$
- The **variance**  $\sigma_x^2$  is the average squared deviation of the values of the variable from their mean. For a discrete random variable,
  - $\sigma_x^2 = (x_1 - \mu)^2p_1 + (x_2 - \mu)^2p_2 + \dots + (x_k - \mu)^2p_k$
- The **standard deviation**  $\sigma_x$  is the square root of the variance. The standard deviation measures the variability of the distribution about the mean.
- The mean and standard deviation of a continuous random variable can be computed from the density curve but to do so requires more advanced mathematics.
- The **law of large numbers** says that the average of the values of the  $X$  observed in many trials must approach  $\mu$ .
- Adding a constant  $a$  (which could be negative) to a random variable increases (or decreases) the mean of the random variable by  $a$  but does not affect its standard deviation or the shape of the probability distribution.
- A **linear transformation** of a random variable involves adding a constant  $a$ , multiplying by a constant  $b$ , or both. We can write a linear transformation of the random variable  $X$  in the form  $Y = a + bX$ . The shape, center, and spread of the distribution of  $Y$  are as follows:
  - **Shape:** Same as probability distribution of  $X$
  - **Center:**  $\mu_y = a + b\mu_x$
  - **Spread:**  $\sigma_y = |b|\sigma_x$
- If  $X$  and  $Y$  are any two random variables, then:  $\mu_{x+y} = \mu_x + \mu_y$  and  $\mu_{x-y} = \mu_x - \mu_y$
- If  $X$  and  $Y$  are *independent*, then:  $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$  and  $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2$

- Any linear combination of independent Normal random variables is also Normally distributed.
- A count of  $X$  successes has a binomial distribution in the **binomial setting**: there are  $n$  observations; the observations are independent of each other; each observation results in a success or a failure; and each observation has the same probability of success  $p$ .
- The **binomial probability** of observing  $k$  successes in  $n$  trials is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- The **binomial coefficient**  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  counts the number of ways  $k$  successes can be arranged among  $n$  observations.
- Given a random variable  $X$ , the **probability distribution function** (pdf) assigns a probability to each value of  $X$ . For each value of  $X$ , the **cumulative distribution function** (cdf) assigns the sum of the probabilities less than or equal to  $X$ .
- The **mean** (expected value) and **standard deviation** of the binomial count  $X$  are

$$\mu = np \quad \sigma = \sqrt{np(1-p)}$$

- The **Normal approximation** to the binomial distribution says that if  $X$  is a count having binomial distribution with parameters  $n$  and  $p$ , then when  $n$  is large,  $X$  is approximately with mean  $\mu = np$  and standard deviation  $\sigma = \sqrt{np(1-p)}$ . The approximation is used when  $np \geq 10$  and  $n(1-p) \geq 10$ .
- A count  $X$  of successes has a **geometric distribution** in the geometric setting if the following conditions are satisfied: each observation results in a success or a failure; observations are independent; each observation has the same probability  $p$  of success; and  $X$  counts the number of trials required to obtain the first success. The geometric random variable differs from the binomial because in the geometric setting the number of trials varies and the desired number of defined successes (1) is fixed in advance.
- If  $X$  has the geometric distribution with probability of success  $p$ , the **geometric probability** that  $X$  takes on any value is  $P(X = n) = (1-p)^{n-1}p$
- The **mean** (expected value) and **standard deviation** of a geometric count  $X$  are

$$\mu = 1/p \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

- The probability that it takes more than  $n$  trials to see the first success is  $P(X > n) = (1-p)^n$

## AP Exam Tips

- If the mean of a random variable has a non-integer value, but you report it as an integer, your answer will be marked incorrect.
  - When you solve problems involving random variables, start by defining the random variable of interest. For example, let  $X$  = the Apgar score of a randomly selected baby or let  $Y$  = the height of a randomly selected young woman. Then state the probability you are trying to find in terms of the random variable:  $P(X > 7)$ .
  - If you have trouble solving problems involving sums and differences of Normal random variables with algebraic methods, use the simulation strategy on page 376 to earn some (and possibly full) credit.
  - Do not rely on “calculator speak” when showing your work on free-response questions. Writing **binompdf(5,0.25,3)=0.08789** will *not* earn full credit. Show the binomial probability formula with the numbers entered into it. To specify the Binomial distribution, you must state  $p$  and  $n$ .
- 

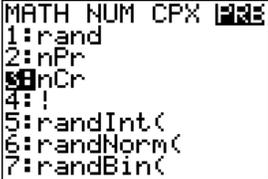
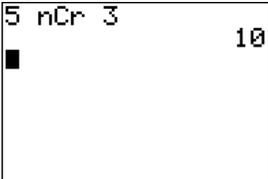
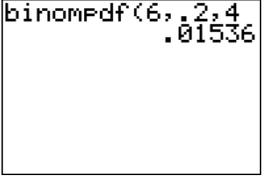
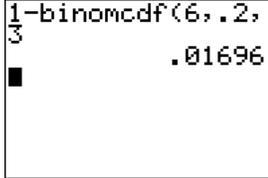
## Chapter 6 Portfolio Items

- 6.1 Explain what is meant by a *random variable*; include examples.
- 6.2 Explain what a *discrete random variable* is; include examples.
- 6.3 Explain what a *continuous random variable* is; include examples.
- 6.4 Explain what is meant by the *law of large numbers*.
- 6.5 Explain how to calculate the *mean and variance of a discrete random variable*; include an example.
- 6.6 Explain how to calculate the *mean and variance of distributions formed by combining two random variables*; include an example.
- 6.7 Explain what is meant by a *binomial setting* and *binomial distribution*; include discussion of necessary conditions and examples.
- 6.8 Explain how to use technology to solve probability questions in a binomial setting; include examples.
- 6.9 Explain how to compute the *mean and variance* of a binomial random variable; include examples.
- 6.10 Explain what is meant by a *geometric setting*; include discussion of necessary conditions and examples.
- 6.11 Explain how to solve probability questions in a geometric setting; include examples.
- 6.12 Explain how to calculate the *mean* of a geometric random variable; include examples.

## Can You?

1. Use a probability distribution to answer questions about possible values of a random variable?
2. Calculate the mean of a discrete random variable?
3. Interpret the mean of a random variable?
4. Calculate the standard deviation of a discrete random variable?
5. Interpret the standard deviation of a random variable?
6. Describe the effects of transforming a random variable by adding or subtracting a constant and multiplying or dividing by a constant?
7. Find the mean and standard deviation of the sum or difference of independent random variables?
8. Determine whether two random variables are independent?
9. Find probabilities involving the sum or difference of independent Normal random variables?
10. Determine whether the conditions for a binomial random variable are met?
11. Compute and interpret probabilities?
12. Calculate the mean and standard deviation of a binomial random variable and Interpret these values in context?
13. Find probabilities involving geometric random variables?

## Technology

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>1. Combinations</b></p> <ul style="list-style-type: none"> <li>[MATH] &lt;PRB&gt; 3:nCr</li> </ul>  <ul style="list-style-type: none"> <li>Example <math>\binom{5}{3}</math>, type 5 [MATH] &lt;PRB&gt; 3:nCr 3 [ENTER]</li> </ul>                                                     | <p><b>3. Binomial: Computing <math>P(X \leq \#)</math></b></p> <ul style="list-style-type: none"> <li>[2<sup>nd</sup>] [VARS] (DISTR) B:binomcdf(</li> </ul>  <p>Complete <b>binomcdf(</b> with number of trials (n), probability of success (p), number of successes (r) and [ENTER].</p>  <p>Note: this give the probability of <i>r</i> or fewer successes in n trials.</p> |
| <p><b>2. Binomial: Computing <math>P(X = \#)</math></b></p> <ul style="list-style-type: none"> <li>[2<sup>nd</sup>] [VARS] (DISTR) A:binompdf(</li> </ul>  <p>Complete <b>binompdf(</b> with number of trials (n), probability of success (p), number of successes (r) and [ENTER].</p>  | <p><b>4. Binomial: Computing <math>P(X &gt; \#)</math></b></p> <ul style="list-style-type: none"> <li>Type 1 - [2<sup>nd</sup>] [VARS] (DISTR) B:binomcdf(</li> </ul>  <p>Complete <b>binomcdf(</b> with number of trials (n), probability of success (p), number of successes (r) and [ENTER].</p> <p>Note: this gives the probability of the complement of <i>r</i> or fewer successes in n trials.</p>                                                      |

### 5. Geometric: Computing $P(X = \#)$

- [2<sup>nd</sup>] [VARS] (DISTR) E:geompdf(

```
DISTR DRAW
0:pdf(
A:binompdf(
B:binomcdf(
C:Poissonpdf(
D:Poissoncdf(
E:geometpdf(
F:geometcdf(
```

Complete **geompdf**( with probability of success (p) and number (n) and press [ENTER].

```
geometpdf(.2,3
.128
```

### 7. Geometric: Computing $P(X > \#)$

- Type 1 - [2<sup>nd</sup>] [VARS] (DISTR) F:geomcdf(

```
1-geometcdf(.2,3
.512
```

Complete **geomcdf**( with probability of success (p) and number (n) and press [ENTER].

Note: this gives the *complement* of the probability that the first success will occur *on or before the #th trial*.

### 6. Geometric: Computing $P(X \leq \#)$

- [2<sup>nd</sup>] [VARS] (DISTR) F:geomcdf(

```
DISTR DRAW
0:pdf(
A:binompdf(
B:binomcdf(
C:Poissonpdf(
D:Poissoncdf(
E:geometpdf(
F:geometcdf(
```

Complete **geomcdf**( with probability of success (p) and number (n) and press [ENTER].

```
geometcdf(.2,3
.488
```

Note: this gives the probability that the first success will occur *on or before the #th trial*.



## Chapter 7 – Sampling Distributions

---

**AP Standards.** The following AP Standards are covered in Chapter 7:

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

D. Sampling Distributions (7.1, 7.2, 7.3)

1. Sampling distribution of a sample proportion
2. Sampling distribution of a sample mean
3. Central Limit Theorem
6. Simulation of sampling distributions

---

### Key Vocabulary

- |                           |                              |                                                         |
|---------------------------|------------------------------|---------------------------------------------------------|
| ▪ parameter               | ▪ biased estimator           | ▪ sample proportion                                     |
| ▪ statistic               | ▪ unbiased estimator         | ▪ mean and standard deviation of sampling distributions |
| ▪ sampling variability    | ▪ bias                       | ▪ central limit theorem                                 |
| ▪ sampling distribution   | ▪ variability                |                                                         |
| ▪ population distribution | ▪ variability of a statistic |                                                         |

### Summary

#### Section 7.1

- A number that describes a population is called a **parameter**. To estimate an unknown parameter, use a **statistic** calculated from a sample. (p with p; s with s)
- A **population** distribution of a variable describes the value of the variable for all individuals in a population. A statistic produced from a probability sample or randomized experiment has a **sampling distribution** that describes how the statistic varies in repeated data production. The sampling distribution answers the question, “What would happen if we repeated the sample or experiment many times?” Formal statistical inference is based upon the sampling distributions of statistics.
- A statistic can be an **unbiased** estimator or a **biased** estimator. A statistic as an estimator of a parameter may suffer from **bias** or from high **variability**. Bias means the center (mean) of the sampling distribution is not equal to the true value of the parameter. The variability of the statistic is described by the spread of its sampling distribution.
- The **variability** of a statistic is described by the spread of its sampling distribution. Larger samples give smaller spread.
- Properly chosen statistics from randomized data production designs have no bias resulting from the way the sample is selected or the way the experimental units are assigned to treatments. The variability of the statistic is determined by the size of the sample or by the size of the experimental groups. Statistics from larger samples have less variability.
- When trying to estimate a parameter, choose a statistic with low or no bias and minimum variability. Do not forget to consider the shape of the sampling distribution before doing inference.

## Section 7.2

- When we want information about the **population proportion**  $p$  of individuals with some special characteristic, we often take an SRS and use the **sample proportion**  $\hat{p}$  to estimate the unknown parameter  $p$ .
- The sampling distribution of  $\hat{p}$  describes how the statistic varies in all possible samples from the population.
- The **mean** of the sampling distribution is equal to the population proportion  $p$ . That means that  $\hat{p}$  is an **unbiased estimator** of  $p$ .
- The **standard deviation** of the sampling distribution is  $\sqrt{p(1-p)/n}$  for an SRS of size  $n$ . This formula can be used if the population is at least 10 times as large as the sample (the *10% condition*).
- The standard deviation of  $\hat{p}$  gets smaller as the sample size  $n$  gets larger. Because of the square root, a sample four times as large is needed to cut the standard deviation in half.
- When the sample size  $n$  is large, the sampling distribution of  $\hat{p}$  is close to a Normal distribution with mean  $p$  and standard deviation  $\sqrt{p(1-p)/n}$ . In practice, use this Normal approximation when both  $np \geq 10$  and  $n(1-p) \geq 10$ .

## Section 7.3

- When we want information about the **population mean**  $\mu$  for some variable, we often take an SRS and use the **sample mean**  $\bar{x}$  to estimate the unknown parameter  $\mu$ .
- The **sampling distribution of  $\bar{x}$**  describes how the statistic  $\bar{x}$  varies in *all* possible samples from the population.
- The **mean** of the sampling distribution is  $\mu$ , so that  $\bar{x}$  is an unbiased estimator of  $\mu$ .
- The **standard deviation** of the sampling distribution of  $\bar{x}$  is  $\sigma/\sqrt{n}$  for an SRS of size  $n$  if the population has standard deviation  $\sigma$ . This formula can be used if the population is at least 10 times as large as the sample (the *10% condition*).
- Choose an SRS of size  $n$  from a population  $\mu$  and standard deviation  $\sigma$ . If the population distribution is Normal, then so is the sampling distribution of the sample mean  $\bar{x}$ .
- The **Central Limit Theorem** states that for large  $n$  the sampling distribution of  $\bar{x}$  is approximately Normal for any population with finite standard deviation  $\sigma$ . The mean and standard deviation of the Normal distribution are the mean  $\mu$  and the standard deviation  $\sigma/\sqrt{n}$  of  $\bar{x}$  itself.
- We can use the Normal distribution to calculate approximate probabilities for events involving  $\bar{x}$  whenever the *Normal condition* is met:
  - If the population distribution is Normal, so is the sampling distribution of  $\bar{x}$ .
  - If  $n \geq 30$ , the CLT tells us that the sampling distribution of  $\bar{x}$  will be approximately Normal in most cases.

## AP Exam Tips

- Terminology matters. Do not say “sample distribution” when you mean sampling distribution. You will lose credit on free-response questions for misusing statistical terms. Again, “Professionals speak in professional terms.”
  - Notation matters. The symbols  $\hat{p}$ ,  $\bar{x}$ ,  $p$ ,  $\mu$ ,  $\sigma$ ,  $\mu_{\hat{p}}$ ,  $\sigma_{\hat{p}}$ ,  $\mu_{\bar{x}}$ , and  $\sigma_{\bar{x}}$  all have specific and different meanings. Either use notation correctly – or do not use it at all. You can expect to lose credit if you use incorrect notation.
  - The Central Limit Theorem only applies to the sampling distribution of the sample means.
- 

## Chapter 7 Portfolio Items

7.1 Explain the difference between a parameter and a statistic.

7.2 Explain the meaning of sampling variability.

7.3 Explain what a sampling distribution of a statistic is.

7.4 Explain the meaning of the bias of a statistic.

7.5 Explain the meaning of the variability of a statistic. Include a discussion of the effects of sample size.

7.6 Explain what the sampling distribution of a sample proportion is and how to compute the mean and standard deviation for the sampling distribution of  $\hat{p}$ . Include a discussion of the rule of thumb that must be used to justify computation of the standard deviation in this manner.

7.7 Explain how to use a Normal approximation to the sampling distribution of  $\hat{p}$  to solve problems involving  $\hat{p}$ . Include a discussion of the rule of thumb that must be used to justify use of the Normal approximation.

7.8 Given the mean and standard deviation of a population, explain how to calculate the mean and standard deviation for the sampling distribution of a sample mean.

7.9 State the Central Limit Theorem and explain how to use it to solve probability problems for the sampling distribution of a sample mean. Include examples.

## Can you?

1. Distinguish between a statistic and a parameter?
2. Recognize the fact of sampling variability: a statistic will take different values when you repeat a sample or experiment?
3. Interpret a sampling distribution as describing the values taken by a statistic in all possible repetitions of a sample or experiment under the same conditions?
4. Distinguish between population distribution, sampling distribution, and the distribution of sample data?
5. Describe the bias and variability of a statistic in terms of the mean and spread of its sampling distribution?
6. Determine whether a statistic is an unbiased estimator of a population parameter?
7. Understand that the variability of a statistic is controlled by the size of the sample and that statistics from larger samples are less variable?
8. Recognize when a problem involves a sample proportion  $\hat{p}$  ?
9. Find the mean and standard deviation of the sampling distribution of a sample proportion  $\hat{p}$  for an SRS of size  $n$  from a population having population proportion  $p$ ?
10. Check whether the 10% and Normal conditions are met in a given setting?
11. Know the standard deviation (spread) of the sampling distribution of  $\hat{p}$  gets smaller at the rate  $\sqrt{n}$  as the sample size  $n$  gets larger?
12. Recognize when you can use the Normal approximation to the sampling distribution of  $\hat{p}$  ?
13. Use the Normal approximation to calculate probabilities that concern  $\hat{p}$  ?
14. Use the sampling distribution of  $\hat{p}$  to evaluate a claim about a population proportion.
15. Recognize when a problem involves the mean  $\bar{x}$  of a sample?
16. Find the mean and standard deviation of the sampling distribution of a sample mean  $\bar{x}$  from an SRS of size  $n$  when the mean  $\mu$  and standard deviation  $\sigma$  of the population are known?
13. Know that the standard deviation (spread) of the sampling distribution of the sample mean  $\bar{x}$  gets smaller at the rate rate  $\sqrt{n}$  as the sample size  $n$  gets larger?
14. Understand that  $\bar{x}$  has approximately a Normal distribution when the sample is large (Central Limit Theorem)?
15. Use the Central Limit Theorem and the Normal approximation to calculate probabilities involving  $\bar{x}$  ?

## Chapter 8 – Estimating with Confidence

---

**AP Standards.** The following AP Standards are covered in Chapter 10:

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

D. Sampling Distributions (8.3)

7. t-distribution

### IV. Statistical Inference: Estimating population parameters and testing hypotheses

A. Estimation (point estimators and confidence intervals) (8.1, 8.2, 8.3)

1. Estimating population parameters and margins of error
2. Properties of point estimators, including unbiasedness and variability
3. Logic of confidence intervals, meaning of confidence intervals, and properties of confidence intervals
4. Large sample confidence interval for proportion
6. Confidence interval for a mean
7. Confidence interval for a difference between two means (unpaired and paired)

### Key Vocabulary

- |                       |                               |                         |
|-----------------------|-------------------------------|-------------------------|
| ▪ point estimator     | ▪ independent                 | ▪ one-sample t interval |
| ▪ point estimate      | ▪ four step process           | ▪ robust                |
| ▪ confidence interval | ▪ level C confidence interval |                         |
| ▪ margin of error     | ▪ degrees of freedom          |                         |
| ▪ interval            | ▪ standard error              |                         |
| ▪ confidence level    | ▪ one -sample z interval      |                         |
| ▪ random              | ▪ t distribution              |                         |
| ▪ normal              | ▪ t-procedures                |                         |

### Summary

#### Section 8.1

- To estimate an unknown population parameter, start with a statistic that provides a reasonable guess. The chosen statistic is a **point estimator** for the parameter. The specific value of the point estimator that we use gives us a **point estimate** for the parameter.
- A **confidence interval** uses sample data to estimate an unknown population parameter with an indication of how accurate the estimate is and how confident we are that the result is correct.
- Any confidence interval has two parts: an interval computed from the data and a confidence level. The interval often has the form: **estimate  $\pm$  margin of error** or as stated on the AP Statistics formula sheet:

**Confidence interval: statistic  $\pm$  (critical value) • (standard deviation of statistic)**

- The **confidence level** states the probability that the method will give a correct answer. For a 95% confidence interval, in the long run 95% of your intervals will contain the true parameter value.
- Other things being equal, the **margin of error** of a confidence interval gets smaller as

- The confidence level decreases
- The sample size increases

## Section 8.2 Confidence Intervals for $p$

- Confidence intervals for a population proportion  $p$  when the data are an SRS of size  $n$  are based on the sample proportion  $\hat{p}$ . When  $n$  is large,  $\hat{p}$  has an approximately Normal distribution with mean  $p$  and standard deviation  $\sqrt{p(1-p)/n}$ .

- A **level C confidence interval** for  $p$  is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- The **critical value  $z^*$**  is chosen so that the area under the standard Normal distribution between  $-z^*$  and  $z^*$  is  $C$ . Because of the Central Limit Theorem, this interval is approximately correct for large samples where the population is not Normal.
- The inference procedure is approximately correct when these conditions are met:
  - **Random:** data were produced by random sampling or random assignment
  - **Normal:** the sample is large enough to satisfy  $n\hat{p} \geq 10$  and  $n(1-\hat{p}) \geq 10$
  - **Independent:** Individual observations are independent. When sampling without replacement, we check the *10% condition*: the population is at least 10 times as large as the sample.

*To construct a confidence interval:*

*Step 1: State*— What *parameter* do you want to estimate, and at what confidence level?

*Step 2: Plan* — Identify the appropriate inference *method*. Check *conditions*.

*Step 3: Do* — If conditions are met, carry out the inference procedure.

*Confidence interval = estimate  $\pm$  margin of error*

*Confidence interval = statistic  $\pm$ (critical value)(standard deviation of statistic)*

*Step 4: Conclude* — Interpret your results in the context of the problem.

***Three C's : Conclusion, Connection, and Context.***

To find the sample size needed to obtain a confidence interval with approximate margin of error  $m$  for proportion is found by setting

$$z^* \sqrt{\frac{p^*(1-p^*)}{n}} \leq m$$

and solving for  $n$ , where  $p^*$  is a guessed value for the sample proportion  $\hat{p}$ , and  $z^*$  is the critical value for the level of confidence you want. If you use  $p^* = 0.5$  in this formula, the margin of error of the interval will be less than or equal to  $m$  no matter what the value of  $\hat{p}$  is.

## Section 8.3 Estimating a Population Mean

- **Confidence intervals for the mean  $\mu$  of a Normal population** are based on the sample mean  $\bar{x}$  of an SRS. Because of the Central Limit Theorem, the resulting procedures are approximately correct for other population distributions when the sample is large.
- If we somehow know  $\sigma$ , we use the  $z$  critical value and the standard Normal distribution to help calculate confidence intervals. To find the sample size required to obtain a confidence interval with specified margin of error  $m$  for the mean  $\bar{x}$  is found by setting

$$z^* \frac{\sigma}{\sqrt{n}} \leq m$$

and solving for  $n$ , where  $z^*$  is the critical value for the desired level of confidence. Always round  $n$  up when you use this formula.

- In reality, we do not know the population standard deviation  $\sigma$ . Replace the standard deviation of  $\bar{x}$  with the **standard error**  $s/\sqrt{n}$  and use the **t-distribution** with  $n-1$  **degrees of freedom**.
- A **level C confidence interval (t interval)** for the mean  $\mu$  of a Normal population with unknown standard deviation  $\sigma$ , based on an SRS of size  $n$ , is given by

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Where  $t^*$  is the critical value so that the  $t$  curve with  $n-1$  degrees of freedom has area  $C$  between  $-t$  and  $t$ .

- The inference procedure is approximately correct when these conditions are met:
  - **Random:** data were produced by random sampling or random assignment
  - **Normal:** the sample is large enough to satisfy  $n\hat{p} \geq 10$  and  $n(1 - \hat{p}) \geq 10$
  - **Independent:** Individual observations are independent. When sampling without replacement, we check the *10% condition*: the population is at least 10 times as large as the sample.
- Follow the **4-Step Process – State, Plan, Do, Conclude** – whenever you are asked to construct and interpret a confidence interval for a population mean.
- Remember: **inference for proportions uses  $z$ ; inference for means uses  $t$** .
- The  $t$  procedures are relatively **robust** when the population is non-Normal, especially for large sample sizes. This means the probability calculations remain fully accurate when a condition for the use of the procedure is violated.

## AP Exam Tips

- On a given problem, you may be asked to interpret the confidence interval, the confidence level or both. Be sure you understand the difference: the confidence level describes the long-run capture rate of the method and the confidence interval gives a set of plausible values for the parameter.
  - If a free-response question asks you to construct and interpret a confidence interval, you are expected to do the entire four-step process. That includes clearly defining the parameter and checking conditions.
  - You may use your calculator to compute confidence intervals on the AP Exam. But there's a risk involved. If you just give the calculator answer and no work, you will get either full credit for the "Do" step (if the interval is correct) or no credit (if it is wrong). To be safe, you should show the calculation with the appropriate formula and then checking with your calculator. If you opt for the calculator-only method, be sure to name the procedure (e.g., one- proportion z-interval) and to give the interval (e.g., 0.514 to 0.606).
- 

## Chapter 8 Portfolio Items

8.1 Explain the meaning of *statistical inference*.

8.2 Give a formula for the basic form of all *confidence intervals*.

8.3 List and explain the four steps for constructing a confidence interval.

8.4 Using the example of your choice, explain how to construct a confidence interval for the mean  $\mu$  of a Normal population with known population standard deviation  $\sigma$ .

8.5 Explain how to find the sample size required to obtain a confidence interval for  $\mu$  of specified margin of error; include an example.

8.6 Using the example of your choice, explain how to construct a confidence interval for the mean  $\mu$  of a population when the population standard deviation is unknown.

8.7 Using the example of your choice, explain how to construct a confidence interval for a population proportion  $p$ .

8.8 Explain how to find the sample size required to obtain a confidence interval for  $p$  of specified margin of error; include an example.

Note: for items 8.4-8.8, be sure to use the 4-step process and write out formulas as appropriate even if you use a calculator to check answers.

## Can you?

1. Interpret a confidence level?
2. Interpret a confidence interval?
3. Understand that a confidence interval gives a range of plausible values for the parameter?
4. Understand why each of the three inference conditions—Random, Normal, and Independent—is Important?
5. Explain how practical issues like nonresponse, undercoverage, and response bias can affect the interpretation of a confidence interval?
6. Construct and interpret a confidence interval for a population proportion?
7. Determine critical values for calculating a confidence interval using a table or your calculator?
8. Carry out the steps in constructing a confidence interval for a population proportion: define the parameter; check conditions; perform calculations; interpret results in context?
9. Determine the sample size required to obtain a level  $C$  confidence interval for a population proportion with a specified margin of error?
10. Understand how the margin of error of a confidence interval changes with the sample size and the level of confidence  $C$ ?
11. Construct and interpret a confidence interval for a population mean?
12. Determine the sample size required to obtain a level  $C$  confidence interval for a population mean with a specified margin of error?
13. Carry out the steps in constructing a confidence interval for a population mean: define the parameter; check conditions; perform calculations; interpret results in context?
14. Determine sample statistics from a confidence interval?

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Estimating population mean <math>\sigma</math> known (data)</b></p> <ul style="list-style-type: none"> <li>Enter data in list</li> <li>Press <b>[STAT]</b> and choose <b>&lt;TESTS&gt;</b><br/><b>7:ZInterval</b></li> </ul> <pre>EDIT CALC TESTS 3↑2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval... 8:TInterval... 9↑2-SampZInt...</pre> <pre>ZInterval Inet: <b>DATA</b> Stats <math>\sigma</math>: 2.1 List: L1 Freq: <math>\square</math> C-Level: .95 Calculate</pre> <ul style="list-style-type: none"> <li>Choose <b>DATA</b>, enter population standard deviation and C-Level. Choose <b>Calculate</b></li> </ul> <pre>ZInterval (14.214, 16.696) <math>\bar{x}</math>=15.45454545 Sx=10.89370127 n=11</pre> | <p><b>Estimating population mean <math>\sigma</math> unknown</b></p> <ul style="list-style-type: none"> <li>When the population standard deviation is unknown, the t-procedure is used. The steps are the same as Z-Intervals except initially press <b>[STAT]</b> and choose <b>&lt;TESTS&gt;</b><br/><b>8:TInterval</b></li> <li>Make other entries as appropriate.</li> </ul>                                                                                                                                                                                                                                                             |
| <p><b>Estimating population mean <math>\sigma</math> known (statistics)</b></p> <ul style="list-style-type: none"> <li>Press <b>[STAT]</b> and choose <b>&lt;TESTS&gt;</b><br/><b>7:ZInterval</b></li> </ul> <pre>EDIT CALC TESTS 3↑2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval... 8:TInterval... 9↑2-SampZInt...</pre> <pre>ZInterval Inet: Data <b>STATS</b> <math>\sigma</math>: 2.1 <math>\bar{x}</math>: 15.45454545... n: 11 C-Level: .95 Calculate</pre> <ul style="list-style-type: none"> <li>Choose <b>Stats</b>, enter population standard deviation, sample mean, sample size and C-Level. Choose <b>Calculate</b></li> </ul> <pre>ZInterval (14.214, 16.696) <math>\bar{x}</math>=15.45454545 n=11</pre>    | <p><b>Estimating a population proportion p</b></p> <ul style="list-style-type: none"> <li>Press <b>[STAT]</b> and choose <b>&lt;TESTS&gt;</b> <b>A:1-PropZInt</b></li> </ul> <pre>EDIT CALC TESTS 8↑TInterval... 9:2-SampZInt... 0:2-SampTInt... 1↑1-PropZInt... B:2-PropZInt... C: <math>\chi^2</math>-Test... D↓<math>\chi^2</math>GOF-Test...</pre> <p>Enter number of successes <b>x</b>, sample size <b>n</b>, and desired confidence level. Highlight <b>Calculate</b> and <b>[ENTER]</b></p> <pre>1-PropZInt x: 246 n: 439 C-Level: .95 Calculate</pre> <pre>1-PropZInt (.51393, .60679) <math>\hat{p}</math>=.5603644647 n=439</pre> |

## Chapter 9 – Testing a Claim

---

**AP Standards.** The following AP Standards are covered in Chapter 9:

### IV. Statistical Inference: Estimating population parameters and testing hypotheses

B. Tests of significance (9.1, 9.2, 9.3)

1. Logic of significance testing, null and alternative hypotheses; p-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power
2. Large sample test for a proportion
4. Test for a mean

---

### Key Vocabulary

- |                          |                             |                       |
|--------------------------|-----------------------------|-----------------------|
| ▪ Significance test      | ▪ significance level        | ▪ Type I Error        |
| ▪ Null Hypothesis        | ▪ one-sample z test         | ▪ Type II Error       |
| ▪ Alternative Hypotheses | ▪ test statistic            | ▪ Power               |
| ▪ One sided alternative  | ▪ one sample t test         | ▪ Degrees of freedom  |
| ▪ Two sided alternative  | ▪ paired data               | ▪ t-distribution      |
| ▪ p-value                | ▪ four-step process         | ▪ paired t procedures |
| ▪ $\alpha$ level         | ▪ statistically significant |                       |

### Summary

#### Section 9.1

- A **significance test** assesses the evidence provided by data against a **null hypothesis  $H_0$**  in favor of an **alternative hypothesis  $H_a$** .
- The hypotheses are stated in terms of population parameters. Often,  $H_0$  is a statement of no change or no difference.  $H_a$  says that a parameter differs from its null hypothesis value in a specified direction (**one-side alternative**) or in either direction (**two-sided alternative**).
- The reasoning of a significance test is as follows. Suppose that the null hypothesis is true. If we repeated our data production many times, would we often get data inconsistent with  $H_0$  as the data we actually have? If the data is unlikely when  $H_0$  is true, they provide evidence against  $H_0$ .
- The **P-value** of a test is the probability, computed supposing  $H_0$  is true, that the statistic will be a value least extreme as that actually observed in the direction of the specified  $H_a$ .
- Small P-values indicate strong evidence against  $H_0$ . To calculate a P-value, we must know the sampling distribution of the test statistic when  $H_0$  is true. There is no universal rule for how small a P-value in a significance test provides convincing evidence against a null hypothesis.
- If the P-value is smaller than a specified value  $\alpha$  (called the **significance level**), the data are **statistically significant** at a level  $\alpha$ . In that case, we can **reject  $H_0$** . (**P-value low –  $H_0$  must go**). If the P-value is greater than or equal to  $\alpha$ , we **fail to reject  $H_0$** .
- A **Type I error** occurs if we reject  $H_0$  when it is in fact true. A **Type II error** occurs if we fail to reject  $H_0$  when it is actually false.
- In a fixed level  $\alpha$  significance test, the probability of a Type I error is the significance level  $\alpha$ .

- The **power** of a significance test against a specific alternative is the probability that the test will reject  $H_0$  when the alternative is true. Power measures the ability of the test to detect an alternative value of the parameter. For a specific alternative,  $P(\text{Type II Error}) = 1 - \text{power}$ .
- When you plan a statistical study, plan the inference as well. In particular, ask what sample size you need for successful inference. Increasing the size of the sample increases the power (reduces the probability of Type II error) when the significance level remains fixed. We can also increase the power of a test by using a higher significance level (say,  $\alpha = 0.10$  instead of  $\alpha = 0.05$ ), which means increasing the risk of a Type I error.

## Section 9.2

- As with confidence intervals, you should verify that the three conditions – Random, Normal, and Independent – are met before you can carry out a significance test.
- Significance tests for  $H_0: p = p_0$  are based on the **test statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

With P-values calculated from the standard Normal distribution.

- The **one-sample z test for proportion** is approximately correct when (1) the data were produced by random sampling, (2) the population is at least 10 times as large as the sample, and (3) the sample is large enough to satisfy  $np_0 \geq 10$  and  $n(1 - p_0) \geq 10$  (that is, the expected numbers of successes and failures are both at least 10).
- Follow the **4-step process** when you are asked to carry out a significance test:
  - **State:** What *hypotheses* do you want to test, and at what significance level? Define any parameters you use.
  - **Plan:** Choose the appropriate inference *method*. Check *conditions*.
  - **Do:** If the conditions are met, perform *calculations*.
    - Compute the **test statistic**
    - Find the **P-value**
- Confidence intervals provide additional information that significance tests do not – namely, a range of plausible values for the true population parameter  $p$ . A two-sided test of  $H_0: p = p_0$  at significance level  $\alpha$  gives the same conclusion as a  $100(1 - \alpha)\%$  confidence interval.

## Section 9.3

- Significance test for the mean  $\mu$  of a Normal distribution are based on the sampling distribution of the sample mean  $\bar{x}$ . Due to the Central Limit Theorem, the resulting procedures are approximately correct for other population distributions when the sample is large.
- If we somehow know  $\sigma$ , we can use a z test statistic and the standard Normal distribution to perform calculations. In practice, we typically do not know  $\sigma$ . Then we use the **one-sample t statistic**

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

with P-values calculated from the t distribution with  $n-1$  degrees of freedom.

- The **one-sample t test** is approximately correct when
  - **Random:** The data were produced by random sampling or a randomized experiment.
  - **Normal:** The population is Normal **or** the sample size is large ( $n \geq 30$ ).
  - **Independent:** Individual observations are independent. When sampling without replacement, check that the population is at least 10 times as large as the sample.
- Confidence intervals provide additional information that significance tests do not – namely, a range of plausible values for the parameter  $\mu$ . A two-sided test of  $H_0: \mu = \mu_0$  at a significance level  $\alpha$  gives the same conclusion as a  $100(1 - \alpha)\%$  confidence interval for  $\mu$ .
- Analyze **paired data** by first taking the difference within each pair to produce a single sample. Then use one-sample t procedures.
- Very small differences can be highly significant (small P-value) when a test is based on a large sample. A statistically significant difference need not be practically important. Plot the data to display the difference you are seeking, and use confidence intervals to estimate the actual values of the parameters.

---

### AP Exam Tips

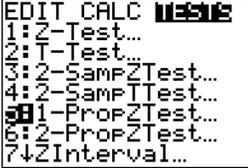
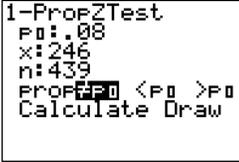
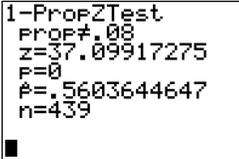
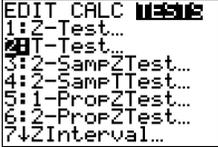
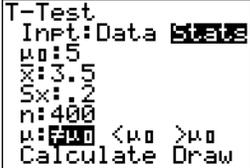
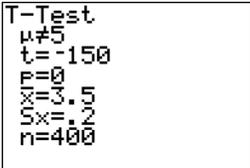
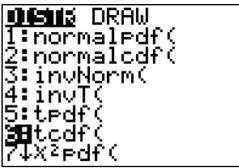
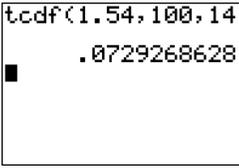
- The conclusion to a significance test should always include three components: (1) an explicit comparison of the P-value to a stated significance level **or** an interpretation of the P-value as a conditional probability, (2) a decision about the null hypothesis: reject or fail to reject  $H_0$ , and (3) an explanation of what the decision means in context.
- When a significance test leads to a fail to reject  $H_0$  decision, as in this example, be sure to interpret the results as “we do not have enough evidence to conclude  $H_a$ .” Saying anything that sounds like you believe  $H_0$  is (or might be) true will lead to a loss of credit. Do not write text-message-type responses like “FTR the  $H_0$ .”
- You can use your calculator to carry out the mechanics of a significance test on the AP Exam. But there is a risk involved. If you just give the calculator answer and no work, and one or more of your values is incorrect, you will probably get no credit for the “Do” step. We recommend doing the calculation with the appropriate formula and then checking it with your calculator. If you opt for the calculator-only method, be sure to name the procedure (one proportion z test) and report the test statistic ( $z = 1.15$ ) and P-value (0.1243).
- Remember: if you just give a calculator result and no work and one or more of your values are wrong, you probably will not get any credit for the “Do” step. Do the calculation with the appropriate formula and then check with your calculator. If you opt for the calculator-only method, name the procedure (t test) and report the test statistic ( $t = -0.94$ ), degrees of freedom ( $df = 14$ ) and P-value (0.1809).

## Chapter 9 Portfolio Items

- 9.1 Explain how to state correct hypotheses for a significance test about a population mean or proportion. Include an example.
  - 9.2 Explain what a P-value represents. Use a real-life example.
  - 9.3 Explain the meaning and relationship between the significance level of a test,  $P(\text{Type II error})$  and power.
  - 9.4 List the conditions necessary to carry out a test about a population proportion.
  - 9.5 Explain how to conduct a significance test about a population proportion. Include an example and be sure to check necessary conditions.
  - 9.6 List the conditions necessary to carry out a test about a population mean?
  - 9.7 Explain how to conduct a significance test about a population mean. Include an example and be sure to check necessary conditions.
  - 9.8 Explain how to recognize paired data and explain how to use a one sample  $t$  procedure to perform a significance test. Include an example.
- 

### Can you?

1. State correct hypotheses for a significance test about a population proportion or mean?
2. Interpret  $P$ -values in context?
3. Interpret a Type I error and a Type II error in context, and give the consequences of each?
4. Understand the relationship between the significance level of a test,  $P(\text{Type II error})$ , and power?
5. Check conditions for carrying out a test about a population proportion?
6. If conditions are met, conduct a significance test about a population proportion?
7. Use a confidence interval to draw a conclusion for a two-sided test about a population proportion?
8. Check conditions for carrying out a test about a population mean?
9. If conditions are met, conduct a one-sample  $t$  test about a population mean  $\mu$ ?
10. Use a confidence interval to draw a conclusion for a two-sided test about a population mean?
11. Recognize paired data and use one sample  $t$  procedures to perform significance tests for such data?

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>One-proportion z test</b></p> <ul style="list-style-type: none"> <li>Press [STAT] then choose &lt;TESTS&gt;<br/><b>5:1-PropZTest</b></li> </ul>  <ul style="list-style-type: none"> <li>On the 1-PropZTest screen, enter the hypothesized proportion <math>p_0</math>, number of successes <math>X</math>, and random sample size <math>n</math>. Also specify alternative hypothesis. Select &lt;Calculate&gt; and [ENTER]</li> </ul>   <ul style="list-style-type: none"> <li>If you select the &lt;Draw&gt; option, you will see the area of rejection</li> </ul> | <p><b>One-sample t test</b></p> <ul style="list-style-type: none"> <li>If using data, enter sample data into <math>L_1</math></li> <li>Press [STAT] then choose &lt;TESTS&gt;<br/><b>2:T-Test</b></li> </ul>  <ul style="list-style-type: none"> <li>Enter hypothesized mean, sample mean, sample standard deviation and sample size. Choose alternative hypothesis. Select &lt;Calculate&gt; and [ENTER].</li> </ul>   <ul style="list-style-type: none"> <li>If you select the &lt;Draw&gt; option, you will see the area of rejection</li> </ul> |
| <p><b>Computing P-values from t distributions</b></p> <ul style="list-style-type: none"> <li>Press [2<sup>nd</sup>] [VAR] (DISTR) and choose <b>tcdf(</b></li> </ul>  <ul style="list-style-type: none"> <li>Enter (lower bound, upper bound, df)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |



## Chapter 10 – Comparing Two Populations or Groups

**AP Standards.** The following AP Standards are covered in Chapter 10:

### IV. Statistical Inference: Estimating population parameters and testing hypotheses

- A. Estimation (point estimators and confidence intervals) (10.1, 10.2)
5. Large sample confidence interval for a difference between two proportions
  7. Confidence interval for a difference between two means (unpaired and paired)
- B. Tests of significance (10.1, 10.2)
3. Large sample test for a difference between two proportions
  5. Test for a difference between two means (unpaired and paired)

### Key Vocabulary

- difference between two proportions
- two sample z interval for proportions
- two sample z test for difference between two proportions
- two sample z statistic
- two sample t statistic
- pooled combined sample proportion
- standard error
- randomization distribution
- paired t-test
- two sample t test for means
- two sample t interval for means
- difference between two means
- pooled two sample t statistic

### Inference about proportions

|                        |                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Estimate</b></p> | <p><b>Two-sample z interval for <math>p_1-p_2</math> (2-PropZInt)</b></p> <p>An approximate level C confidence interval for <math>p_1-p_2</math> is</p> $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$ <p>where <math>z^*</math> is the standard Normal critical value.</p> | <p><b>Random:</b> The data are produced by a random sample of size <math>n_1</math> from population 1 and a random sample of size <math>n_2</math> from population 2 or by two groups of size <math>n_1</math> and <math>n_2</math> in a randomized experiment.</p> <p><b>Normal:</b> The counts of “successes” and “failures” in each sample or group -- <math>n_1p_1</math>, <math>n_1(1-p_1)</math>, <math>n_2p_2</math>, <math>n_2(1-p_2)</math> – are at least 10.</p> <p><b>Independent:</b> Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples (the <i>10% condition</i>).</p> |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Test</b> | <p><b>Two-sample z test for <math>p_1-p_2</math> (2-PropZTest)</b></p> <p><b>Significance tests of <math>H_0: p_1 - p_2 = 0</math> use the <b>pooled (combined) sample proportion</b></b></p> $\hat{p}_c = \frac{\text{count of successes in both samples combined}}{\text{count of individuals in both samples combined}}$ $= \frac{X_1 + X_2}{n_1 + n_2}$ <p>The <b>two-sample z test for <math>p_1-p_2</math></b> uses the test statistic</p> $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$ <p>with P-values calculated from the standard Normal distribution.</p> | <p><b>Random:</b> The data are produced by a random sample of size <math>n_1</math> from population 1 and a random sample of size <math>n_2</math> from population 2 or by two groups of size <math>n_1</math> and <math>n_2</math> in a randomized experiment.</p> <p><b>Normal:</b> The counts of “successes” and “failures” in each sample or group -- <math>n_1p_1</math>, <math>n_1(1-p_1)</math>, <math>n_2p_2</math>, <math>n_2(1-p_2)</math> – are at least 10.</p> <p><b>Independent:</b> Observations and independent samples or groups; <i>10% condition</i> if sampling without replacement</p> |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### Inference about means

|                 |                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                     |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Estimate</b> | <p><b>Two –sample t interval for <math>\mu_1-\mu_2</math> (2-SampTInt)</b></p> $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ <p>df = min(<math>n_1 - 1</math>, <math>n_2 - 1</math>)</p> | <p><b>Random:</b> Data from random samples or randomized experiment</p> <p><b>Normal:</b> Population distributions Normal or large samples (<math>n_1 \geq 30</math>, <math>n_2 \geq 30</math>)</p> <p><b>Independent:</b> Observations and independent samples or groups; <i>10% condition</i> if sampling without replacement</p> |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|             |                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                     |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Test</b> | <p><b>Two-sample t test for <math>\mu_1-\mu_2</math> (2-SampTTest)</b></p> $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ <p>df = min(<math>n_1 - 1</math>, <math>n_2 - 1</math>)</p> | <p><b>Random:</b> Data from random samples or randomized experiment</p> <p><b>Normal:</b> Population distributions Normal or large samples (<math>n_1 \geq 30</math>, <math>n_2 \geq 30</math>)</p> <p><b>Independent:</b> Observations and independent samples or groups; <i>10% condition</i> if sampling without replacement</p> |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## AP Exam Tips

- You may use your calculator to compute a confidence interval on the AP Exam but there is risk involved. If you just give the calculator answer and no work, you will either get full credit for the “Do” step (if the interval is correct) or no credit (if it is wrong). If you opt for the calculator method, be sure to name the procedure (e.g., two proportion z-interval) and to give the interval (e.g., 0.223 to 0.297).
  - When checking the Normal condition on an AP Exam question involving inference about means, be sure to include a graph. Do not expect credit for describing a graph that you made on your calculator but did not put on your paper.
  - When a significance test leads to a fail to reject  $H_0$  decision, as in the previous example, be sure to interpret the results as “We do not have enough evidence to conclude  $H_a$ .” Saying anything that sounds like you believe  $H_0$  is (or might be) true will lead to a loss of credit.
- 

## Chapter 10 Portfolio Items

10.1 Explain how to determine whether a problem requires inference about comparing means or proportions. Include examples.

10.2 Explain how to recognize from the design of a study whether a one-sample t, paired t, or two-sample t procedure is needed. Include examples.

10.3 Explain how to calculate and interpret a confidence interval for the difference between two proportions using the two-sample z statistic. Include an example and consideration of necessary conditions. Verify your answer with technology.

10.4 Explain how to perform a two-proportion z test to test the null hypothesis that two population proportions in two distinct populations are equal. Include an example and consideration of necessary conditions. Verify your answer with technology.

10.5 Explain how to calculate and interpret a confidence interval for the difference between two means using the two-sample t statistic with conservative degrees of freedom. Include an example and consideration of necessary conditions. Verify your answer with technology.

10.6 Explain how to use the two-sample t test with conservative degrees of freedom to test the null hypothesis that two populations have equal means. Include an example and consideration of necessary conditions. Verify your answer with technology.

### Can you?

1. Describe the characteristics of the sampling distribution of  $\hat{p}_1 - \hat{p}_2$ ?
2. Calculate probabilities using the sampling distribution of  $\hat{p}_1 - \hat{p}_2$ ?
3. Determine whether the conditions for performing inference for proportions are met?
4. Construct and interpret a confidence interval to compare two proportions?
5. Perform a significance test to compare two proportions?
6. Interpret the results of inference procedures in a randomized experiment?
7. Describe the characteristics of the sampling distribution of  $\bar{x}_1 - \bar{x}_2$  ?
8. Calculate probabilities using the sampling distribution of  $\bar{x}_1 - \bar{x}_2$  ?
9. Determine whether the conditions for performing inference for means are met?
10. Use two-sample  $t$  procedures to compare two means based on summary statistics?
11. Use two-sample  $t$  procedures to compare two means from raw data?
12. Interpret standard computer output for two-sample  $t$  procedures?
13. Perform a significance test to compare two means?
14. Check conditions for using two-sample  $t$  procedures in a randomized experiment?
15. Interpret the results of inference procedures in a randomized experiment?
16. Determine the proper inference procedure to use in a given setting?

Technology

|                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Confidence interval for a difference in proportions</b></p> <p>[STAT] &lt;TESTS&gt; B:2-PropZInt</p> <pre> EDIT CALC <b>TESTS</b> 1:Z-Test... 2:T-Test... 3:2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval...                 </pre> <pre> 2-PropZTest x1:584 n1:800 x2:1096 n2:2253 P1:<del>0.72</del> &lt;P2 &gt;P2 Calculate Draw                 </pre> | <p><b>Confidence interval for a difference in means</b></p> <p>[STAT] &lt;TESTS&gt; 0:2-SampTInt</p> <pre> 2-PropZTest x1:584 n1:800 x2:1096 n2:2253 P1:<del>0.72</del> &lt;P2 &gt;P2 Calculate Draw                 </pre> <pre> 2-SampTInt ↑n1:30 x̄2:23.7 Sx2:17.5 n2:30 C-Level: .95 Pooled:<input checked="" type="checkbox"/> Yes Calculate                 </pre>               |
| <p><b>Significance test for a difference in proportions</b></p> <p>[STAT] &lt;TESTS&gt; 6:2-PropZTest</p> <pre> 2-PropZTest x1:584 n1:800 x2:1096 n2:2253 P1:<del>0.72</del> &lt;P2 &gt;P2 Calculate Draw                 </pre> <pre> 2-PropZTest x1:584 n1:800 x2:1096 n2:2253 P1:<del>0.72</del> &lt;P2 &gt;P2 Calculate Draw                 </pre>                                          | <p><b>Significance test for a difference in means</b></p> <p>[STAT] &lt;TESTS&gt; 4:2-SampTTest</p> <pre> EDIT CALC <b>TESTS</b> 1:Z-Test... 2:T-Test... 3:2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval...                 </pre> <pre> 2-SampTTest Inpt:Data <b>██████</b> x̄1:34.53 Sx1:14.26 n1:30 x̄2:23.7 Sx2:17.5 ↓n2:30                 </pre> |



## Chapter 11 – Inference for Distributions of Categorical Data

---

**AP Standards.** The following AP Standards are covered in Chapter 11:

### III. Anticipating Patterns: Exploring random phenomena using probability and simulation

D. Sampling Distributions (11.1)

8. Chi-square distribution

### IV. Statistical Inference: Estimating population parameters and testing hypotheses

B. Tests of significance (11.1)

6. Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables)

---

### Key Vocabulary

- |                                       |                            |                                                |
|---------------------------------------|----------------------------|------------------------------------------------|
| ▪ one way table                       | ▪ chi square distribution  | ▪ chi square test for homogeneity              |
| ▪ chi-square test for goodness of fit | ▪ degrees of freedom       | ▪ chi square test for association/independence |
| ▪ chi-square statistic                | ▪ chi-square distribution  |                                                |
| ▪ expected count                      | ▪ components of chi-square |                                                |
| ▪ observed count                      | ▪ cell counts              |                                                |
|                                       | ▪ r x c table              |                                                |

### Summary

#### The Chi-Square Test for Goodness of Fit

- A **one-way table** is often used to display the distribution of a categorical variable for a sample of individuals.
- The **chi-square goodness of fit test** tests the null hypothesis that a categorical variable has a specified distribution.
- The test compares the **observed count** in each category with the counts that would be expected if  $H_0$  were true. The **expected count** for any category is found by multiplying the specified proportion of the population distribution in that category by the sample size.
- The **chi-square statistic** is

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \sum \frac{(O - E)^2}{E}$$

where the sum is over all possible categories.

- The test compares the statistic  $\chi^2$  with critical values from the **chi-square distribution** with degrees of freedom  $df = \text{number of categories} - 1$ . Large values of  $\chi^2$  are evidence against  $H_0$ , so the P-value is the area under the chi-square density curve to the right of  $\chi^2$ .
- The chi-square distribution is an approximation to the sampling distribution of the statistic  $\chi^2$ . You can safely use this approximation when all expected cell counts are at least 5 (**Large Sample Size condition**).

- Be sure to check the Random, Large Sample Size, and Independent conditions are met before performing a chi-square goodness of fit test.
- If the test finds a statistically significant result, do a follow-up analysis that compares the observed and expected counts and that looks for the largest **components** of the chi-square statistic.

### Chi-square test for homogeneity

- **Finding expected counts:** The expected count in any cell of a two-way table when  $H_0$  is true is

$$\text{expected count} = \frac{(\text{row total})(\text{column total})}{\text{table total}}$$

- Suppose the Random, Large Sample Size, and Independent conditions are met. You can use the **chi-square test for homogeneity** to test

$H_0$ : There is no difference in the distribution of a categorical variable for several populations or treatments

$H_a$ : There is a difference in the distribution of a categorical variable for several populations or treatments

Start by finding the expected counts. Then calculate the chi-square statistic

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Where the sum is over all cells (not including totals) in the two-way table. If  $H_0$  is true, the chi-square statistic has approximately a chi-square distribution with  $df = (\text{number of rows} - 1)(\text{number of columns} - 1)$ . The P-value is the area to the right of the chi-square statistic under the corresponding chi-square density curve.

- If the test finds a statistically significant result, do a follow-up analysis that compares the observed and expected counts and that looks for the largest components of the chi-square statistic.

### Chi-square test for association/independence

- If data are produced using a single random sample from a population of interest, then each observation is classified according to two categorical variables. The **chi-square test of association/independence** tests the null hypothesis that there is no association between the two categorical variables in the population of interest. Another way to state the null hypothesis is  $H_0$ : The two categorical variables are independent in the population of interest.

## AP Exam Tips

- In the “Do” step, you are not required to show every term in the chi-square statistic. Writing the first few terms of the sum and last term, separated by ellipses, is considered as “showing work.” We suggest you do this and then use your calculator.
  - If you have trouble distinguishing between the two types of chi-square tests for two-way tables, you are better off just saying, “chi-square test” than choosing the wrong type. Better yet, learn how to tell the difference.
- 

## Chapter 11 Portfolio Items

11.1 Describe the situation where the *chi-square test for goodness of fit* is appropriate.

11.2 Explain how to conduct a *chi-square test for goodness of fit*. Include an example and considerations for necessary conditions. Verify your answer with technology.

11.3 Explain how to determine which observations contribute most to the total value if a chi-square statistic turns out to be significant. Include an example.

11.4 Using the words *populations* and *categorical variables*, describe the major difference between *homogeneity of populations* and *independence*.

11.5 Explain how to use a chi-square test for homogeneity to determine whether the distribution of a categorical variable differs for several populations or treatments. Include an example and consideration of necessary conditions.

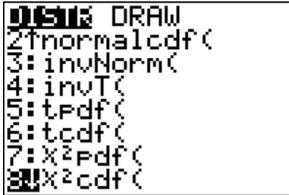
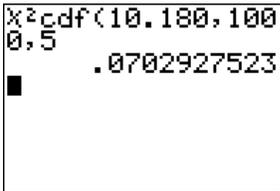
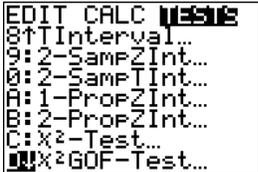
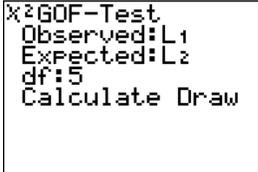
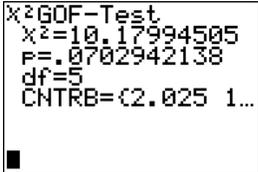
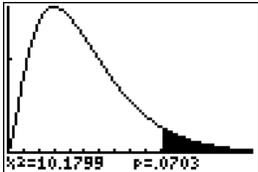
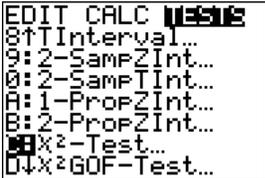
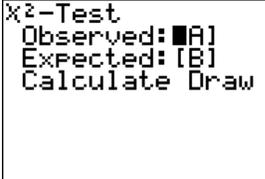
11.6 Explain how to use a chi-square test of association/independence to determine whether there is convincing evidence of an association between two categorical variables. Include an example and consideration of necessary conditions.

11.7 Explain how to examine individual components of the chi-square statistic as part of a follow-up analysis. Include an example.

## Can you?

1. Compute expected counts, conditional distributions, and contributions to the chi-square statistic?
2. Check the Random, Large sample size, and Independent conditions before performing a chi-square test for goodness-of-fit?
3. Use a chi-square goodness-of-fit test to determine whether sample data are consistent with a specified distribution of a categorical variable?
4. Examine individual components of the chi-square statistic as part of a follow-up analysis?
5. Check the Random, Large sample size, and Independent conditions before performing a chi-square test for homogeneity?
6. Use a chi-square test for homogeneity to determine whether the distribution of a categorical variable differs for several populations or treatments?
7. Interpret computer output for a chi-square test based on a two-way table?
8. Examine individual components of the chi-square statistic as part of a follow-up analysis?
9. Show that the two-sample z test for comparing two proportions and the chi-square test for a 2-by-2 two-way table give equivalent results?
10. Check the Random, Large sample size, and Independent conditions before performing a chi-square test for association/independence?
11. Use a chi-square test of association/independence to determine whether there is convincing evidence of an association between two categorical variables?
12. Interpret computer output for a chi-square test based on a two-way table?
13. Examine individual components of the chi-square statistic as part of a follow-up analysis?
14. Distinguish between the three types of chi-square tests?

Technology

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Finding P-values for chi-square tests</b></p> <p>[2<sup>nd</sup>] [VARS] (DISTR) 8: <math>\chi^2</math>cdf(</p>  <p>Complete <b>8: <math>\chi^2</math>cdf(</b> with the chi-square value, a very large number (1000) and the df. [ENTER]</p>                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <p><b>Chi-square goodness of fit test</b></p> <p>Enter observed counts and expected counts into two separate lists, L1 and L2.</p> <p>[STAT] &lt;TESTS&gt; D: <math>\chi^2</math>GOF-Test</p>     | <p><b>Chi-square test for two-way tables</b></p> <p>Enter the observed counts in matrix [A]<br/>         [2<sup>nd</sup>][X<sup>-1</sup>](MATRIX) &lt;EDIT&gt; 1:A<br/>         Enter the dimensions and the counts</p> <p>[STAT] &lt;TESTS&gt; C: <math>\chi^2</math>-Test</p>   <p>To see the expected counts, go to the home screen and ask for a display of matrix [B].</p> |



## Chapter 12 – More about Regression

---

**AP Standards.** The following AP Standards are covered in Chapter 12:

**I. Exploring Data: Describing patterns and departures from patterns.**

D. Exploring bivariate data (12.2)

5. Transformations to achieve linearity: logarithmic and power transformations

**IV. Statistical Inference: Estimating population parameters and testing hypotheses**

A. Estimation (point estimators and confidence intervals) (12.1)

8. Confidence interval for the slope of a least-squares regression line

B. Tests of significance (12.1)

7. Test for the slope of a least-squares regression line

---

### Key Vocabulary

- sample regression line
- true regression line
- t interval for slope
- standard error of slope
- t test for slope
- standardized test statistic
- standard error
- exponential model
- power model
- logarithmic model

## Summary

### Least-squares regression

- Least-squares regression fits a straight line of the form  $\hat{y} = a + bx$  to data to predict a response variable  $y$  from an explanatory variable  $x$ . Inference in this setting uses the **sample regression line** to estimate or test a claim about the **population (true) regression line**.
- The conditions for regression inference are
  - **Linear:** The actual relationship between  $x$  and  $y$  is linear. For any fixed value of  $x$ , the mean response  $\mu_y$  falls on the population (true) regression line  $\mu_y = \alpha + \beta x$ .
  - **Independent:** Individual observations are independent.
  - **Normal:** For any fixed value of  $x$ , the response  $y$  varies according to the Normal distribution.
  - **Equal variance:** The standard deviation of  $y$  (call it  $\sigma$ ) is the same for all values of  $x$ .
  - **Random:** The data are produced from a well-designed random sample or randomized experiment.
- The slope  $b$  and intercept  $a$  of the least-squares line estimate the slope  $\beta$  and the intercept  $\alpha$  of the population (true) line. To estimate  $\sigma$ , use the standard deviation  $s$  of the residuals.
- Confidence intervals and significance tests for the slope  $\beta$  of the population regression line are based on a  $t$  distribution with  $n - 2$  degrees of freedom.
- The **t interval for the slope**  $\beta$  has the form  $b \pm t^* SE_b$ , where the **standard error of the slope** is  $SE_b = \frac{s}{s_x \sqrt{n-1}}$ .
- To test the null hypothesis  $H_0: \beta = \text{hypothesized value}$ , carry out a **t test for slope**. This test uses the statistic  $t = \frac{b - \beta_0}{SE_b}$ . The most common null hypothesis is  $H_0: \beta = 0$ , which says that there is no linear relationship between  $x$  and  $y$  in the population.

### Transformations

- Nonlinear relationships between two quantitative variables can sometimes be changed into linear relationships by **transforming** one or both of the variables. Transformation is particularly effective when there is reason to think that the data are governed by some nonlinear mathematical model.
- When theory or experience suggests that the relationship between two variables follows a **power model** of the form  $y = ax^p$ , there are two transformations involving powers and roots that can linearize a curved pattern in a scatterplot. (1) Raise the values of the explanatory variable  $x$  to the power  $p$ , then look at the graph of  $(x^p, y)$ . (2) Take the  $p$ th root of the values of the response variable  $y$ , then look at a graph of  $(x, \sqrt[p]{y})$ .
- In a linear model of the form  $y = a + bx$ , the values of the response variable are predicted to increase by a constant amount  $b$  for each increase of 1 unit in the explanatory variable. For an **exponential model** of the form  $y = ab^x$ , the predicted values of the response variable are multiplied by an additional factor of  $b$  for each increase of one unit in the explanatory variable.
- A useful strategy for straightening a curved pattern in a scatterplot is to take the **logarithm** of one or both variables. To achieve linearity when the relationship between two variables follows an exponential model, plot the logarithm (base 10 or base  $e$ ) of  $y$  against  $x$ . When a power

model describes the relationship between two variables, a plot of  $\log y$  or  $\ln y$  versus  $\log x$  or  $\ln x$  should be linear.

- Once we transform the data to achieve linearity, we can fit a least-squares regression line to the transformed data and use this linear model to make predictions.

---

### AP Exam Tips

- The AP Exam formula sheet gives  $\hat{y} = b_0 + b_1x$  for the equation of the sample (estimated) regression line. We will stick with our simpler notation,  $\hat{y} = a + bx$ , which is also used by TI calculators. Just remember: the coefficient of  $x$  is always the slope, no matter what symbol is used.
- When you see a list of data values on an exam question, do not just start typing the data into your calculator. Read the question first. Often, additional information is provided that makes it unnecessary for you to enter the data at all. This can save you valuable time on the AP Exam.

---

### Chapter 12 Portfolio Items

12.1 Identify the conditions necessary to perform *inference for regression*. Using a set of data, explain how to check the conditions for performing inference for regression are present.

12.2 Explain what is meant by the *standard error about the least-squares line*.

12.3 Explain how to compute a *confidence interval for the slope of the regression line*. Include an example.

12.4 Explain how to conduct a *test of the hypothesis that the slope of the regression line is 0* (or that the correlation is 0) in the population. Include an example.

12.5 Explain how to use transformations involving powers and roots to achieve linearity for a relationship between two variables. Create a least-squares regression line from the transformed data and use it to make a prediction. Include an example.

12.6. Explain how to use transformations involving logarithms to achieve linearity for a relationship between two variables. Include an example.

12.7 Explain how to determine which of several transformations does a better job of producing a linear relationship. Include examples.

### Can you/Do you?

1. Make a scatterplot to show the relationship between an explanatory and a response variable?
2. Use a calculator or software to find the correlation and the least-squares regression line?
3. Recognize the regression setting: a straight-line relationship between an explanatory and response variable?
4. Recognize which type of inference you need in a particular regression setting?
5. Inspect the data to recognize situations in which inference is not safe: a nonlinear relationship, influential observations, strongly skewed residuals in a small sample, or nonconstant variation of the data points about the regression line?
6. Explain in any specific regression setting the meaning of the slope  $\beta$  of the true regression line?
7. Understand computer output for regression and from the output, find the slope and intercept of the least-squares line, their standard errors, and the standard error about the line?
8. Use computer output to carry out tests and calculate confidence intervals for  $\beta$ ?
  1. Check conditions for performing inference about the slope  $\beta$  of the population regression line?
  2. Interpret computer output from a least-squares regression analysis?
  3. Construct and interpret a confidence interval for the slope  $\beta$  of the population regression line?
  4. Perform a significance test about the slope of a population regression line?
5. Use transformations involving powers and roots to achieve linearity for a relationship between two variables?
  6. Make predictions from a least-squares regression line involving transformed data?
  7. Use transformations involving logarithms to achieve linearity for a relationship between two variables?
  8. Make predictions from a least-squares regression line involving transformed data?
  9. Determine which of several transformations does a better job of producing a linear relationship?

**1. Checking conditions for linear regression**

- After performing linear regression, plot residuals as detailed in **Chapter 3**.
- Check to confirm sum of residuals is 0. Another way to do this is to use the **SUM(** command and the variable **RESID**. Type **[2<sup>nd</sup>] [STAT] (LIST) <MATH> 5:SUM(**

```
NAMES OPS MATH
1:min(
2:max(
3:mean(
4:median(
5:sum(
6:Prod(
7↓stdDev(
```

**[ENTER]** and then type **[2<sup>nd</sup>] [STAT] (LIST) <NAMES>** and choose **RESID** and **[ENTER]**

```
sum( LRESID
-1.04E-11
█
```

**2. Linear regression t-test**

- Enter data into **L<sub>1</sub>** and **L<sub>2</sub>**.
- Press **[STAT] <TESTS> F:LinRegTTest**

```
EDIT CALC TESTS
B†2-PropZInt...
C:X²-Test...
D:X²GOF-Test...
E:2-SampFTest...
F:LinRegTTest...
G:LinRegTInt...
H:ANOVA(
```

Press **[ENTER]**

```
LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
β & ρ: EQ <0 >0
RegEQ: █
Calculate
```

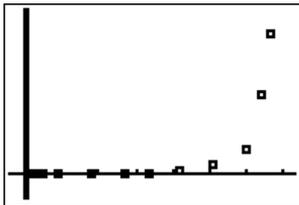
- In the LinRegTTest screen, specify **L<sub>1</sub>** for Xlist and **L<sub>2</sub>** for Ylist and **≠0** for the hypothesized slope. Highlight the command **Calculate** and press **[ENTER]**.
- The linear regression t test results take two screens to present

```
LinRegTTest
y=a+bx
β≠0 and ρ≠0
t=3.006991527
P=.0147877882
df=9
↓a=-6554334.304
█
```

```
LinRegTTest
y=a+bx
β≠0 and ρ≠0
↑b=884184.8667
s=10129460.95
r²=.5011638967
r=.7079293021
```

## Modeling exponential growth

- Enter the exponential data into  $L_1$  and  $L_2$ . Plot shows exponential growth



- Define  $L_3$  as the natural logarithm of  $L_2$ .

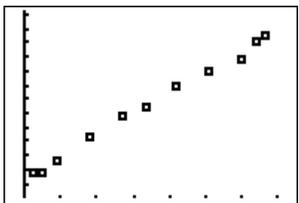
| L1 | L2     | L3 | 3 |
|----|--------|----|---|
| 1  | 2250   |    |   |
| 2  | 2500   |    |   |
| 4  | 5000   |    |   |
| 8  | 29000  |    |   |
| 12 | 120000 |    |   |
| 15 | 275000 |    |   |
| 19 | 1.18E6 |    |   |

$L_3 = \ln(L_2)$

| L1 | L2     | L3          | 3 |
|----|--------|-------------|---|
| 1  | 2250   | 7.718685495 |   |
| 2  | 2500   | 7.824       |   |
| 4  | 5000   | 8.5172      |   |
| 8  | 29000  | 10.275      |   |
| 12 | 120000 | 11.695      |   |
| 15 | 275000 | 12.525      |   |
| 19 | 1.18E6 | 13.981      |   |

$L_3(1) = 7.718685495...$

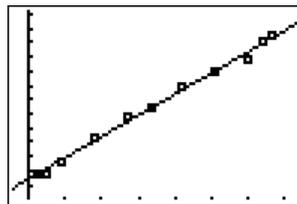
- Make a scatterplot of  $L_1$  and  $L_3$ .



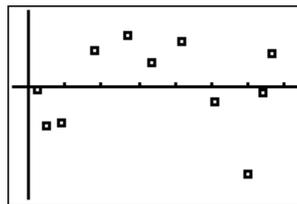
- Perform least-squares regression on the transformed data and graph the plot again with the regression line.

```
LinReg(ax+b) L1,
L3, Y1
```

```
LinReg
y=ax+b
a=.331612825
b=7.407787705
r^2=.9948661832
r=.9974297886
```



- Construct and inspect the residual plot.



- To use the model to predict, define  $Y_2 = e^{(Y_1)}$  then enter x-values in  $Y_2( )$  and [ENTER]. For example, to predict a y-value for an x of 33, enter  $Y_2(33)$ .

## Appendix A - Portfolio Information

---

**Objective:** AP Statistics students will construct a portfolio that will assist in their preparation for the AP Statistics Exam.

**Format:** Your portfolio should be housed in a **composition book**. There must be a **tab** for each chapter. Your name should be clearly written on the outside of the front cover.

**Portfolio Items:** A list of portfolio items is included in each chapter description in “West Carteret High School Notes to Accompany AP Statistics.”

**Procedures:** For each topic the following procedures will be followed.

1. Write the portfolio item and number from the topic list at the top of the page. (The item can be abbreviated but must be understandable.) For example,

|                                                                               |
|-------------------------------------------------------------------------------|
| 1.1 <b>W<sup>5</sup>HW</b> – Explain the steps involved in W <sup>5</sup> HW. |
|-------------------------------------------------------------------------------|

2. State any definitions that relate to the topic, if applicable.
3. In the case of a problem type item, explain in words the procedure for completing the problem related to the topic. Number the steps used. **Include examples in your entry.** The examples should be worked out completely and should include graphs and diagrams as necessary.
4. Start each portfolio topic on a new page. Use the backs of journal pages or you may run out of space.
5. Your entries *must be hand-written*. Do not cut and paste information from handouts into your portfolio.
6. **Each student is responsible for his own work.** It should go without saying that you are not to copy or paraphrase from someone else’s portfolio. This is considered *cheating* and will be dealt with as such.
7. Your portfolio will be due at the end of each chapter; however, the best way to attack it is to complete the topics as we cover them. Waiting until the last minute will prove to be a disaster.

**Missing, Late, Incomplete Portfolios:** The portfolio is a requirement of this course. If it is not submitted, a grade of INCOMPLETE will be awarded for the course and will be removed when the portfolio is submitted. Portfolios submitted incomplete will be returned and will receive a maximum of 70% once resubmitted complete. Late portfolios will receive a maximum of 70%.

# Appendix B - Formulas and Tables

The following list of formulas and tables will be furnished to students taking the AP Statistics Exam.

## Formulas for AP Statistics Exam

### I. Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

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

$$\hat{y} = b_0 + b_1x$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1} \sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

$$b_1 = r \frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

### II. Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$E(X) = \mu_x = \sum x_i p_i$$

$$\text{Var}(X) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$$

If  $X$  has a binomial distribution with parameters  $n$  and  $p$ , then:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu_x = np$$

$$\sigma_x = \sqrt{np(1-p)}$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If  $\bar{x}$  is the mean of a random sample of size  $n$  from an infinite population with mean  $\mu$  and standard deviation  $\sigma$ , then:

$$\mu_{\bar{x}} = \mu$$

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

### III. Inferential Statistics

$$\text{Standardized test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

$$\text{Confidence interval: statistic} \pm (\text{critical value}) \cdot (\text{std. deviation of statistic})$$

#### Single-Sample

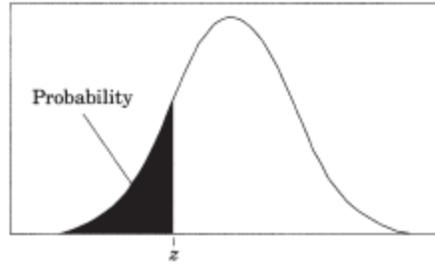
| Statistic         | Standard Deviation of Statistic |
|-------------------|---------------------------------|
| Sample Mean       | $\frac{\sigma}{\sqrt{n}}$       |
| Sample Proportion | $\sqrt{\frac{p(1-p)}{n}}$       |

#### Two-Sample

| Statistic                        | Standard Deviation of Statistic                          |
|----------------------------------|----------------------------------------------------------|
| Difference of sample means       | $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
|                                  | Special case when $\sigma_1 = \sigma_2$                  |
|                                  | $\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$            |
| Difference of sample proportions | $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ |
|                                  | Special case when $p_1 = p_2$                            |
|                                  | $\sqrt{p(1-p)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$     |

$$\text{Chi-square test statistic} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

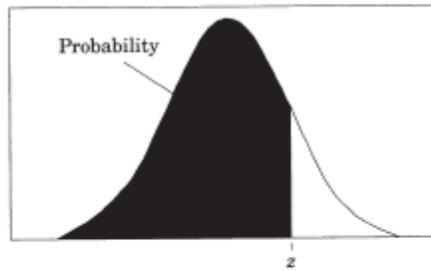
Table entry  
for  $z$  is the  
probability  
lying below  $z$ .



**Table A** Standard normal probabilities

| $z$  | .00   | .01   | .02   | .03   | .04   | .05   | .06   | .07   | .08   | .09   |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| -3.4 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0002 |
| -3.3 | .0005 | .0005 | .0005 | .0004 | .0004 | .0004 | .0004 | .0004 | .0004 | .0003 |
| -3.2 | .0007 | .0007 | .0006 | .0006 | .0006 | .0006 | .0006 | .0005 | .0005 | .0005 |
| -3.1 | .0010 | .0009 | .0009 | .0009 | .0008 | .0008 | .0008 | .0008 | .0007 | .0007 |
| -3.0 | .0013 | .0013 | .0013 | .0012 | .0012 | .0011 | .0011 | .0011 | .0010 | .0010 |
| -2.9 | .0019 | .0018 | .0018 | .0017 | .0016 | .0016 | .0015 | .0015 | .0014 | .0014 |
| -2.8 | .0026 | .0025 | .0024 | .0023 | .0023 | .0022 | .0021 | .0021 | .0020 | .0019 |
| -2.7 | .0035 | .0034 | .0033 | .0032 | .0031 | .0030 | .0029 | .0028 | .0027 | .0026 |
| -2.6 | .0047 | .0045 | .0044 | .0043 | .0041 | .0040 | .0039 | .0038 | .0037 | .0036 |
| -2.5 | .0062 | .0060 | .0059 | .0057 | .0055 | .0054 | .0052 | .0051 | .0049 | .0048 |
| -2.4 | .0082 | .0080 | .0078 | .0075 | .0073 | .0071 | .0069 | .0068 | .0066 | .0064 |
| -2.3 | .0107 | .0104 | .0102 | .0099 | .0096 | .0094 | .0091 | .0089 | .0087 | .0084 |
| -2.2 | .0139 | .0136 | .0132 | .0129 | .0125 | .0122 | .0119 | .0116 | .0113 | .0110 |
| -2.1 | .0179 | .0174 | .0170 | .0166 | .0162 | .0158 | .0154 | .0150 | .0146 | .0143 |
| -2.0 | .0228 | .0222 | .0217 | .0212 | .0207 | .0202 | .0197 | .0192 | .0188 | .0183 |
| -1.9 | .0287 | .0281 | .0274 | .0268 | .0262 | .0256 | .0250 | .0244 | .0239 | .0233 |
| -1.8 | .0359 | .0351 | .0344 | .0336 | .0329 | .0322 | .0314 | .0307 | .0301 | .0294 |
| -1.7 | .0446 | .0436 | .0427 | .0418 | .0409 | .0401 | .0392 | .0384 | .0375 | .0367 |
| -1.6 | .0548 | .0537 | .0526 | .0516 | .0505 | .0495 | .0485 | .0475 | .0465 | .0455 |
| -1.5 | .0668 | .0655 | .0643 | .0630 | .0618 | .0606 | .0594 | .0582 | .0571 | .0559 |
| -1.4 | .0808 | .0793 | .0778 | .0764 | .0749 | .0735 | .0721 | .0708 | .0694 | .0681 |
| -1.3 | .0968 | .0951 | .0934 | .0918 | .0901 | .0885 | .0869 | .0853 | .0838 | .0823 |
| -1.2 | .1151 | .1131 | .1112 | .1093 | .1075 | .1056 | .1038 | .1020 | .1003 | .0985 |
| -1.1 | .1357 | .1335 | .1314 | .1292 | .1271 | .1251 | .1230 | .1210 | .1190 | .1170 |
| -1.0 | .1587 | .1562 | .1539 | .1515 | .1492 | .1469 | .1446 | .1423 | .1401 | .1379 |
| -0.9 | .1841 | .1814 | .1788 | .1762 | .1736 | .1711 | .1685 | .1660 | .1635 | .1611 |
| -0.8 | .2119 | .2090 | .2061 | .2033 | .2005 | .1977 | .1949 | .1922 | .1894 | .1867 |
| -0.7 | .2420 | .2389 | .2358 | .2327 | .2296 | .2266 | .2236 | .2206 | .2177 | .2148 |
| -0.6 | .2743 | .2709 | .2676 | .2643 | .2611 | .2578 | .2546 | .2514 | .2483 | .2451 |
| -0.5 | .3085 | .3050 | .3015 | .2981 | .2946 | .2912 | .2877 | .2843 | .2810 | .2776 |
| -0.4 | .3446 | .3409 | .3372 | .3336 | .3300 | .3264 | .3228 | .3192 | .3156 | .3121 |
| -0.3 | .3821 | .3783 | .3745 | .3707 | .3669 | .3632 | .3594 | .3557 | .3520 | .3483 |
| -0.2 | .4207 | .4168 | .4129 | .4090 | .4052 | .4013 | .3974 | .3936 | .3897 | .3859 |
| -0.1 | .4602 | .4562 | .4522 | .4483 | .4443 | .4404 | .4364 | .4325 | .4286 | .4247 |
| -0.0 | .5000 | .4960 | .4920 | .4880 | .4840 | .4801 | .4761 | .4721 | .4681 | .4641 |

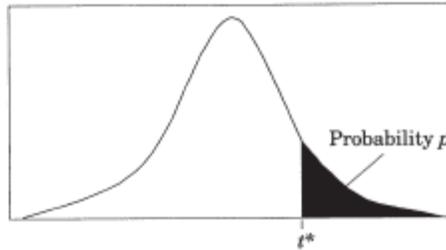
Table entry  
for  $z$  is the  
probability  
lying below  $z$ .



**Table A** (Continued) Standard normal probabilities

| $z$ | .00   | .01   | .02   | .03   | .04   | .05   | .06   | .07   | .08   | .09   |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

Table entry for  $p$  and  $C$  is the point  $t^*$  with probability  $p$  lying above it and probability  $C$  lying between  $-t^*$  and  $t^*$ .

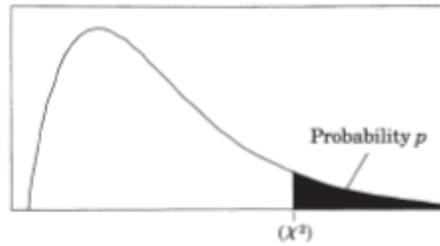


**Table B**  $t$  distribution critical values

| df       | Tail probability $p$ |       |       |       |       |       |       |       |       |       |       |       |
|----------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|          | .25                  | .20   | .15   | .10   | .05   | .025  | .02   | .01   | .005  | .0025 | .001  | .0005 |
| 1        | 1.000                | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2        | .816                 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3        | .765                 | .978  | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4        | .741                 | .941  | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5        | .727                 | .920  | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6        | .718                 | .906  | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.969 |
| 7        | .711                 | .896  | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8        | .706                 | .889  | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9        | .703                 | .883  | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10       | .700                 | .879  | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11       | .697                 | .876  | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12       | .695                 | .873  | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13       | .694                 | .870  | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14       | .692                 | .868  | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15       | .691                 | .866  | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16       | .690                 | .865  | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17       | .689                 | .863  | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18       | .688                 | .862  | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.611 | 3.922 |
| 19       | .688                 | .861  | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20       | .687                 | .860  | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21       | .686                 | .859  | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22       | .686                 | .858  | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23       | .685                 | .858  | 1.060 | 1.319 | 1.714 | 2.069 | 2.177 | 2.500 | 2.807 | 3.104 | 3.485 | 3.768 |
| 24       | .685                 | .857  | 1.059 | 1.318 | 1.711 | 2.064 | 2.172 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25       | .684                 | .856  | 1.058 | 1.316 | 1.708 | 2.060 | 2.167 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26       | .684                 | .856  | 1.058 | 1.315 | 1.706 | 2.056 | 2.162 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27       | .684                 | .855  | 1.057 | 1.314 | 1.703 | 2.052 | 2.158 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28       | .683                 | .855  | 1.056 | 1.313 | 1.701 | 2.048 | 2.154 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29       | .683                 | .854  | 1.055 | 1.311 | 1.699 | 2.045 | 2.150 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30       | .683                 | .854  | 1.055 | 1.310 | 1.697 | 2.042 | 2.147 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |
| 40       | .681                 | .851  | 1.050 | 1.303 | 1.684 | 2.021 | 2.123 | 2.423 | 2.704 | 2.971 | 3.307 | 3.551 |
| 50       | .679                 | .849  | 1.047 | 1.299 | 1.676 | 2.009 | 2.109 | 2.403 | 2.678 | 2.937 | 3.261 | 3.496 |
| 60       | .679                 | .848  | 1.045 | 1.296 | 1.671 | 2.000 | 2.099 | 2.390 | 2.660 | 2.915 | 3.232 | 3.460 |
| 80       | .678                 | .846  | 1.043 | 1.292 | 1.664 | 1.990 | 2.088 | 2.374 | 2.639 | 2.887 | 3.195 | 3.416 |
| 100      | .677                 | .845  | 1.042 | 1.290 | 1.660 | 1.984 | 2.081 | 2.364 | 2.626 | 2.871 | 3.174 | 3.390 |
| 1000     | .675                 | .842  | 1.037 | 1.282 | 1.646 | 1.962 | 2.056 | 2.330 | 2.581 | 2.813 | 3.098 | 3.300 |
| $\infty$ | .674                 | .841  | 1.036 | 1.282 | 1.645 | 1.960 | 2.054 | 2.326 | 2.576 | 2.807 | 3.091 | 3.291 |

Confidence level  $C$

Table entry for  $p$  is the point ( $\chi^2$ ) with probability  $p$  lying above it.



**Table C**  $\chi^2$  critical values

| df  | Tail probability $p$ |       |       |       |       |       |       |       |       |       |       |
|-----|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | .25                  | .20   | .15   | .10   | .05   | .025  | .02   | .01   | .005  | .0025 | .001  |
| 1   | 1.32                 | 1.64  | 2.07  | 2.71  | 3.84  | 5.02  | 5.41  | 6.63  | 7.88  | 9.14  | 10.83 |
| 2   | 2.77                 | 3.22  | 3.79  | 4.61  | 5.99  | 7.38  | 7.82  | 9.21  | 10.60 | 11.98 | 13.82 |
| 3   | 4.11                 | 4.64  | 5.32  | 6.25  | 7.81  | 9.35  | 9.84  | 11.34 | 12.84 | 14.32 | 16.27 |
| 4   | 5.39                 | 5.99  | 6.74  | 7.78  | 9.49  | 11.14 | 11.67 | 13.28 | 14.86 | 16.42 | 18.47 |
| 5   | 6.63                 | 7.29  | 8.12  | 9.24  | 11.07 | 12.83 | 13.39 | 15.09 | 16.75 | 18.39 | 20.51 |
| 6   | 7.84                 | 8.56  | 9.45  | 10.64 | 12.59 | 14.45 | 15.03 | 16.81 | 18.55 | 20.25 | 22.46 |
| 7   | 9.04                 | 9.80  | 10.75 | 12.02 | 14.07 | 16.01 | 16.62 | 18.48 | 20.28 | 22.04 | 24.32 |
| 8   | 10.22                | 11.03 | 12.03 | 13.36 | 15.51 | 17.53 | 18.17 | 20.09 | 21.95 | 23.77 | 26.12 |
| 9   | 11.39                | 12.24 | 13.29 | 14.68 | 16.92 | 19.02 | 19.68 | 21.67 | 23.59 | 25.46 | 27.88 |
| 10  | 12.55                | 13.44 | 14.53 | 15.99 | 18.31 | 20.48 | 21.16 | 23.21 | 25.19 | 27.11 | 29.59 |
| 11  | 13.70                | 14.63 | 15.77 | 17.28 | 19.68 | 21.92 | 22.62 | 24.72 | 26.76 | 28.73 | 31.26 |
| 12  | 14.85                | 15.81 | 16.99 | 18.55 | 21.03 | 23.34 | 24.05 | 26.22 | 28.30 | 30.32 | 32.91 |
| 13  | 15.98                | 16.98 | 18.20 | 19.81 | 22.36 | 24.74 | 25.47 | 27.69 | 29.82 | 31.88 | 34.53 |
| 14  | 17.12                | 18.15 | 19.41 | 21.06 | 23.68 | 26.12 | 26.87 | 29.14 | 31.32 | 33.43 | 36.12 |
| 15  | 18.25                | 19.31 | 20.60 | 22.31 | 25.00 | 27.49 | 28.26 | 30.58 | 32.80 | 34.96 | 37.70 |
| 16  | 19.37                | 20.47 | 21.79 | 23.54 | 26.30 | 28.85 | 29.63 | 32.00 | 34.27 | 36.46 | 39.25 |
| 17  | 20.49                | 21.61 | 22.98 | 24.77 | 27.59 | 30.19 | 31.00 | 33.41 | 35.72 | 37.95 | 40.79 |
| 18  | 21.60                | 22.76 | 24.16 | 25.99 | 28.87 | 31.53 | 32.35 | 34.81 | 37.16 | 39.42 | 42.31 |
| 19  | 22.72                | 23.90 | 25.33 | 27.20 | 30.14 | 32.85 | 33.69 | 36.19 | 38.58 | 40.88 | 43.82 |
| 20  | 23.83                | 25.04 | 26.50 | 28.41 | 31.41 | 34.17 | 35.02 | 37.57 | 40.00 | 42.34 | 45.31 |
| 21  | 24.93                | 26.17 | 27.66 | 29.62 | 32.67 | 35.48 | 36.34 | 38.93 | 41.40 | 43.78 | 46.80 |
| 22  | 26.04                | 27.30 | 28.82 | 30.81 | 33.92 | 36.78 | 37.66 | 40.29 | 42.80 | 45.20 | 48.27 |
| 23  | 27.14                | 28.43 | 29.98 | 32.01 | 35.17 | 38.08 | 38.97 | 41.64 | 44.18 | 46.62 | 49.73 |
| 24  | 28.24                | 29.55 | 31.13 | 33.20 | 36.42 | 39.36 | 40.27 | 42.98 | 45.56 | 48.03 | 51.18 |
| 25  | 29.34                | 30.68 | 32.28 | 34.38 | 37.65 | 40.65 | 41.57 | 44.31 | 46.93 | 49.44 | 52.62 |
| 26  | 30.43                | 31.79 | 33.43 | 35.56 | 38.89 | 41.92 | 42.86 | 45.64 | 48.29 | 50.83 | 54.05 |
| 27  | 31.53                | 32.91 | 34.57 | 36.74 | 40.11 | 43.19 | 44.14 | 46.96 | 49.64 | 52.22 | 55.48 |
| 28  | 32.62                | 34.03 | 35.71 | 37.92 | 41.34 | 44.46 | 45.42 | 48.28 | 50.99 | 53.59 | 56.89 |
| 29  | 33.71                | 35.14 | 36.85 | 39.09 | 42.56 | 45.72 | 46.69 | 49.59 | 52.34 | 54.97 | 58.30 |
| 30  | 34.80                | 36.25 | 37.99 | 40.26 | 43.77 | 46.98 | 47.96 | 50.89 | 53.67 | 56.33 | 59.70 |
| 40  | 45.62                | 47.27 | 49.24 | 51.81 | 55.76 | 59.34 | 60.44 | 63.69 | 66.77 | 69.70 | 73.40 |
| 50  | 56.33                | 58.16 | 60.35 | 63.17 | 67.50 | 71.42 | 72.61 | 76.15 | 79.49 | 82.66 | 86.66 |
| 60  | 66.98                | 68.97 | 71.34 | 74.40 | 79.08 | 83.30 | 84.58 | 88.38 | 91.95 | 95.34 | 99.61 |
| 80  | 88.13                | 90.41 | 93.11 | 96.58 | 101.9 | 106.6 | 108.1 | 112.3 | 116.3 | 120.1 | 124.8 |
| 100 | 109.1                | 111.7 | 114.7 | 118.5 | 124.3 | 129.6 | 131.1 | 135.8 | 140.2 | 144.3 | 149.4 |

Table D Random digits

| Line |       |       |       |       |       |       |       |       |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 101  | 19223 | 95034 | 05756 | 28713 | 96409 | 12531 | 42544 | 82853 |
| 102  | 73676 | 47150 | 99400 | 01927 | 27754 | 42648 | 82425 | 36290 |
| 103  | 45467 | 71709 | 77558 | 00095 | 32863 | 29485 | 82226 | 90056 |
| 104  | 52711 | 38889 | 93074 | 60227 | 40011 | 85848 | 48767 | 52573 |
| 105  | 95592 | 94007 | 69971 | 91481 | 60779 | 53791 | 17297 | 59335 |
| 106  | 68417 | 35013 | 15529 | 72765 | 85089 | 57067 | 50211 | 47487 |
| 107  | 82739 | 57890 | 20807 | 47511 | 81676 | 55300 | 94383 | 14893 |
| 108  | 60940 | 72024 | 17868 | 24943 | 61790 | 90656 | 87964 | 18883 |
| 109  | 36009 | 19365 | 15412 | 39638 | 85453 | 46816 | 83485 | 41979 |
| 110  | 38448 | 48789 | 18338 | 24697 | 39364 | 42006 | 76688 | 08708 |
| 111  | 81486 | 69487 | 60513 | 09297 | 00412 | 71238 | 27649 | 39950 |
| 112  | 59636 | 88804 | 04634 | 71197 | 19352 | 73089 | 84898 | 45785 |
| 113  | 62568 | 70206 | 40325 | 03699 | 71080 | 22553 | 11486 | 11776 |
| 114  | 45149 | 32992 | 75730 | 66280 | 03819 | 56202 | 02938 | 70915 |
| 115  | 61041 | 77684 | 94322 | 24709 | 73698 | 14526 | 31893 | 32592 |
| 116  | 14459 | 26056 | 31424 | 80371 | 65103 | 62253 | 50490 | 61181 |
| 117  | 38167 | 98532 | 62183 | 70632 | 23417 | 26185 | 41448 | 75532 |
| 118  | 73190 | 32533 | 04470 | 29669 | 84407 | 90785 | 65956 | 86382 |
| 119  | 95857 | 07118 | 87664 | 92099 | 58806 | 66979 | 98624 | 84826 |
| 120  | 35476 | 55972 | 39421 | 65850 | 04266 | 35435 | 43742 | 11937 |
| 121  | 71487 | 09984 | 29077 | 14863 | 61683 | 47052 | 62224 | 51025 |
| 122  | 13873 | 81598 | 95052 | 90908 | 73592 | 75186 | 87136 | 95761 |
| 123  | 54580 | 81507 | 27102 | 56027 | 55892 | 33063 | 41842 | 81868 |
| 124  | 71035 | 09001 | 43367 | 49497 | 72719 | 96758 | 27611 | 91596 |
| 125  | 96746 | 12149 | 37823 | 71868 | 18442 | 35119 | 62103 | 39244 |
| 126  | 96927 | 19931 | 36809 | 74192 | 77567 | 88741 | 48409 | 41903 |
| 127  | 43909 | 99477 | 25330 | 64359 | 40085 | 16925 | 85117 | 36071 |
| 128  | 15689 | 14227 | 06565 | 14374 | 13352 | 49367 | 81982 | 87209 |
| 129  | 36759 | 58984 | 68288 | 22913 | 18638 | 54303 | 00795 | 08727 |
| 130  | 69051 | 64817 | 87174 | 09517 | 84534 | 06489 | 87201 | 97245 |
| 131  | 05007 | 16632 | 81194 | 14873 | 04197 | 85576 | 45195 | 96565 |
| 132  | 68732 | 55259 | 84292 | 08796 | 43165 | 93739 | 31685 | 97150 |
| 133  | 45740 | 41807 | 65561 | 33302 | 07051 | 93623 | 18132 | 09547 |
| 134  | 27816 | 78416 | 18329 | 21337 | 35213 | 37741 | 04312 | 68508 |
| 135  | 66925 | 55658 | 39100 | 78458 | 11206 | 19876 | 87151 | 31260 |
| 136  | 08421 | 44753 | 77377 | 28744 | 75592 | 08563 | 79140 | 92454 |
| 137  | 53645 | 66812 | 61421 | 47836 | 12609 | 15373 | 98481 | 14592 |
| 138  | 66831 | 68908 | 40772 | 21558 | 47781 | 33586 | 79177 | 06928 |
| 139  | 55588 | 99404 | 70708 | 41098 | 43563 | 56934 | 48394 | 51719 |
| 140  | 12975 | 13258 | 13048 | 45144 | 72321 | 81940 | 00360 | 02428 |
| 141  | 96767 | 35964 | 23822 | 96012 | 94591 | 65194 | 50842 | 53372 |
| 142  | 72829 | 50232 | 97892 | 63408 | 77919 | 44575 | 24870 | 04178 |
| 143  | 88565 | 42628 | 17797 | 49376 | 61762 | 16953 | 88604 | 12724 |
| 144  | 62964 | 88145 | 83083 | 69453 | 46109 | 59505 | 69680 | 00900 |
| 145  | 19687 | 12633 | 57857 | 95806 | 09931 | 02150 | 43163 | 58636 |
| 146  | 37609 | 59057 | 66967 | 83401 | 60705 | 02384 | 90597 | 93600 |
| 147  | 54973 | 86278 | 88737 | 74351 | 47500 | 84552 | 19909 | 67181 |
| 148  | 00694 | 05977 | 19664 | 65441 | 20903 | 62371 | 22725 | 53340 |
| 149  | 71546 | 05233 | 53946 | 68743 | 72460 | 27601 | 45403 | 88692 |
| 150  | 07511 | 88915 | 41267 | 16853 | 84569 | 79367 | 32337 | 03316 |

**Appendix C - Inference Summary**

