

Final Study Guide
Math 160 Section 3

1 Exploring Data

1.1 Chapter 1 - Picturing Distributions With Graphs

1. In the context of statistics, what is data? What are individuals? What are variables?
2. What are categorical variables? What are quantitative variables? Give examples of each.
3. What is the distribution of a variable? Name graphs that can be used to display the distributions of a) categorical and b) quantitative variables.
4. Make a stemplot (split stems) or histogram of the following data, and describe its distribution (give number of peaks, skewness).

24, 34, 27, 28, 12, 1, 13, 24, 16, 20, 31, 31, 23, 24, 30, 36, 9, 32, 20

1.2 Chapter 2 - Describing Data With Numbers

1. What is a measure of center? What is a measure of spread? What does resistance mean?
2. Indicate with a vertical line the locations of the median, the first quartile, and the second quartile in the following data sets:

2 3 3 4 7 8 8 11
20 33 35 41 55 57 60 73 74

3. What is the five-number summary?
4. What are the formulas defining the mean and the standard deviation?
5. Why do we divide by $n - 1$ rather than n in the formula for the standard deviation?
6. How can we find all these statistics using a calculator? Calculate them for the data sets appearing in this and the previous section.
7. When is it preferable to use the five-number summary rather than the mean and standard deviation? Which (five-number summary or mean and standard deviation) should we use for each of the data sets already given?

1.3 Chapter 4 - Scatterplots And Correlation

1. What are explanatory and response variables?
2. Here are data for the fuel consumption of a British Ford Escort at different speeds. Speed is measured in kilometers per hour, and fuel consumption is measured in liters of gasoline used per hundred kilometers of travel.

Speed	10	20	30	40	50	60	70	80	90	100
Fuel	21.00	13.00	10.00	8.00	7.00	5.90	6.30	6.95	7.57	8.27
Speed	110	120	130	140	150					
Fuel	9.03	9.87	10.79	11.77	12.83					

Make a scatterplot of this data. Describe the form, direction, and strength of the relationship between speed and fuel consumption.

3. What does the correlation coefficient r measure? What are its possible values, and what do they indicate?

- Coffee is exported from many third-world countries. When coffee prices are high, farmers clear forest to plant more coffee trees. Here are five years' data on the prices paid to coffee growers in Indonesia and the percent of forest area lost in a national park lying in a coffee-producing region:

Price (cents/pound)	29	40	54	55	72
Forest lost (percent)	0.49	1.59	1.69	1.82	3.10

Make a scatterplot of these data. Calculate the correlation by hand. Then, calculate the correlation using your calculator.

1.4 Chapter 5 - Regression

- What is a regression line?
- What is special about the least-squares regression line?
- What are the formulas used to calculate the least-squares regression line?
- For the coffee data above, which variable is explanatory? Calculate the least-squares regression line by hand and using your calculator.
- Does the least-squares regression line change if we change which variable is explanatory?
- What is one point that the least-squares regression line always passes through?
- What is one way to tell whether an explanatory variable is a good predictor of a response variable?
- What is a residual? Calculate the residuals for the coffee data above, by hand and using your calculator.
- How do outliers affect a least-squares regression line?
- Explain what a lurking variable is, and give an example of a situation where one appears.

1.5 Chapter 6 - Two-Way Tables

- What are marginal and conditional distributions?

The next three questions refer to the following two-way table, which gives data reflecting peoples' attitudes about the relative quality of recycled products. Buyers of recycled products, and people who do not buy recycled products, were asked to compare the quality of recycled products to non-recycled products.

	Higher Quality	Same Quality	Lower Quality
Buyers	20	7	9
Nonbuyers	29	25	43

- What are the marginal distributions of perceived product quality, and of buyer preference?
- What is the conditional distribution of perceived product quality among buyers of recycled products?
- Is there a difference between buyers and nonbuyers in the perceived quality of recycled products? Use a graph to demonstrate your conclusions.
- State Simpson's Paradox.

2 Producing Data

2.1 Chapter 8 - Sampling

- What is the difference between an experiment and an observational study? Which must we perform if we want to show that one phenomenon causes another?

2. What does it mean to confound two variables? Give an example of a situation where one variable is confounded for another.
3. In the following situations, describe the population and the sample.
 - (a) A study wants to examine the income of individuals with a college degree. Using census data, 1000 college graduates are chosen at random and their income is recorded.
 - (b) You wish to study the reliability of your scale. You weigh yourself ten times.
 - (c) How long are ferrets? You go to your local pet store and measure the lengths of all the ferrets there - 14 ferrets total.
 - (d) You are out at a bar with some friends. Frustrated that you can't get a date, you ask them whether you're really so unattractive.
4. What is bias in an observational study? What is bias in an experiment?
5. Which of the following studies are biased? Why?
 - (a) At the beach, you hand out a survey about religious beliefs.
 - (b) At your church, you hand out a survey about religious beliefs.
 - (c) You choose a random sample of 150 telephone numbers from the phone book. You call each number and ask the residents if they support the police using deadly force.
6. What are the steps to choosing an SRS?
7. Why is taking an SRS so important in statistics?
8. Besides the SRS, name two other sampling designs. How do these sampling designs reduce systematic variation?

2.2 Chapter 9 - Experiments

1. What is the difference between a factor and a treatment?
2. In an experiment to test the effects of caffeine on information retention, a statistics class is randomly divided into six groups. Each group drinks one of three sizes of cups of coffee - grande, venti, and small or whatever - and memorizes a list of 5 or 10 words. What are the factors in this experiment, and how many treatments are there?
3. A sociology professor is interested in whether having multiple cats makes you crazier than having one cat. He recruits fifty cat owners, and divides them into two groups - those with one cat, and those with more cats. Then, he gives all the participants the Minnesota Multiphasic Personality Inventory, and compares the results. Is this an experiment or an observational study? Why?
4. What features of a randomized comparative experiment make it more able to discern the effects of a given treatment than other types of experiments?
5. What is statistical significance?
6. How do double-blind, block, and matched-pairs designs eliminate systematic variation in experimental conditions?
7. A pharmaceutical company develops a product called "liquid earplugs". These are eardrops which decrease the volume of noises around you. You want to compare the noise-damping effects of these eardrops to traditional earplugs. How do you make the conditions of the experiment similar and realistic for both treatment and control groups?

3 Probability

3.1 Chapter 10 - Introducing Probability

1. What is meant by the statement, "the probability of the valence electron of hydrogen having 'up' spin is one-half"?
2. What is meant by the statement, "the probability of dying is one"?
3. When do we say a phenomenon is random?
4. What are the components of a probability model?
5. Write down a probability model to describe the game of Russian Roulette. (What is the sample space, what are the events, and what are their probabilities?)
6. Name two things wrong with the following statement. "Dude, I wouldn't play with those guys if I were you. The probability of you getting shut down is like 100, the probability of you making a basket is like .9, and the probability of you making a fool of yourself is like half."
7. In a probability model, the probability of the entire sample space is what? The probability of each event is what kind of number?
8. If two events are disjoint, what is the probability that either of them occurs?
9. In Dungeons and Dragons, one sometimes must roll a three-sided die. If the probability of rolling a three is .4, and the probabilities of the other two sides are equal, what is the probability model describing this die?
10. Choose a real number between 7 and 19, with the uniform distribution. What is the probability that the number is less than 10? What is the probability that the number is exactly 10?

3.2 Chapter 12 - General Rules of Probability

1. What is meant by independence? Which of the following pairs of events is independent?
 - (a) Choosing one red card from a deck of cards, and choosing a second red card.
 - (b) Rolling one 3 on a six-sided die, and then rolling another 3 on the same die.
 - (c) Passing a man on the street, and then passing another man (instead of a woman).
2. What is the multiplication rule for independent events?
3. Draw a Venn diagram illustrating the general addition rule.
4. Only 40% of people love people. Of the people who love people, 70% are extremely lucky. Of people who don't love people, only 37% are extremely lucky.
 - (a) Draw a tree diagram to illustrate this situation.
 - (b) What percentage of people who love people are also lucky?
 - (c) What percentage of people (all people) are unlucky?
 - (d) State the general multiplication rule. How did you use the general multiplication rule in the above calculations?