

Probability & Statistics 40: Midterm Exam Review

Chapter 1:

1. In one of the first attempts to discover the speed of light, Simon Newcomb in 1882 made 66 measurements of the time light takes to travel between the Washington Monument and his laboratory on the Potomac River. Why did Newcomb repeat his measurement 66 times and then take the average of the 66 as his final result?

- (a) Averaging several measurements reduces any bias that is present in his instruments.
- (b) The average of several measurements is more reliable (less variable) than a single measurement.
- (c) Even if a measuring process is not valid, averaging several measurements made by this process will be valid.
- (d) Both (a) and (b) but not (c).

2. Professor Iconu has developed a new college entrance test. Any such test must have several versions because some people take the test more than once. Unfortunately, it turns out that the same person often gets very different scores depending on which version of the test is offered. The test suffers from

- (a) large bias.
- (b) confounding.
- (c) low accuracy.
- (d) low reliability

3. During a visit to the doctor, you are weighed on a very accurate scale. You are weighed five times and the five readings are essentially the same. When being weighed, you are wearing all of your clothes and a very heavy pair of hiking boots. As a measure of your weight without clothes, the reading on the scale is

- (a) unbiased and reliable
- (b) unbiased and unreliable
- (c) biased and reliable
- (d) biased and unreliable

4. Which of the following statements do you think could possibly be true?

- (a) The number of students enrolled at Ohio State University is about 2 million.
- (b) A basketball team made 110% of its free throws in a game last week.
- (c) The temperature will be 195 degrees (Fahrenheit) tomorrow in Chicago.
- (d) More than 30 million people live in California.

5. A company used to give IQ tests to all job applicants. This is now illegal because IQ is not related to the performance of workers in all the company's jobs. That is, IQ as a measure of future performance on the job is

- (a) biased.
- (b) invalid.
- (c) inaccurate.
- (d) unreliable.

6. You measure the age (years), weight (pounds), and marital status (single, married, divorced, or widowed) of 1400 women. How many variables did you measure?

- (a) 1400
- (b) one
- (c) two
- (d) three

7. "In American History, 20 students failed. Only 11 students failed World History. American History must be a more difficult course than World History." This statement is misleading because the measurement "number of students who fail" used as a surrogate for "difficulty of course" is

- (a) inaccurate.
- (b) unreliable.
- (c) invalid.
- (d) confounded.

8. Following are data on the populations and numbers of death row prisoners for several states.

State	Population (thousands)	Death Row Prisoners
California	28,168	247
Florida	12,377	204
Illinois	11,544	120
Nevada	1,060	45

Which state has the highest number of death row prisoners *relative to the size of its population*?

- (a) California
- (b) Florida
- (c) Illinois
- (d) Nevada

9. It is hard to measure "intelligence." Let's do it the easy way: measure height in inches, and call the result "intelligence." Not only is this method easy, it gives the same number every time we repeat the measurement on the same person. Measuring intelligence this way is

- (a) not reliable and not valid.
- (b) highly reliable but not valid.
- (c) valid, but not reliable.
- (d) both valid and highly reliable.

10. The Dow Jones Industrial Average (DJIA) is the most common measure of stock market prices. Suppose that the DJIA starts at 9000 points and drops 300 points. This is a decrease of

- (a) 0.033% (b) 3.3% (c) 33.3% (d) 333%

11. A 300 point drop in the DJIA was a big drop when the DJIA was at 2000 and a much smaller drop when the DJIA reached 9000. The percent by which stock prices fall is a more meaningful measure. The lesson here is that

- (a) rates are often more meaningful than counts
(b) it is easy to make a mistake calculating a percent
(c) you have to beware of roundoff error.
(d) there are lies, damned lies, in statistics.

12. The net asset value of a mutual fund has increased from \$27 on December 31 to \$33 now. The percent increase in value is about

- (a) 22%. (b) 18%. (c) 1.2%. (d) 122%.

13. Professor Ziegenfuss of the Geology Department has ordered a new instrument which is supposed to measure the iron content of iron ore. After the instrument arrives, he uses it to measure the iron content of five test samples of ore, all of which are known to be exactly 16% iron. The numbers given by the machine on these five test samples are 5%, 3%, 28%, 16%, and 25%. Based on these measurements, one can conclude that the new instrument

- (a) is biased.
(b) is not reliable.
(c) is broken.
(d) is confounded.

14. A police department gives job applicants a test of their knowledge of modern popular music. Experience shows that those who score well on this test tend to become lousy police officers. As a measure of future job performance, the music test scores

- (a) are not reliable.
(b) are biased.
(c) have predictive validity.
(d) are invalid.

Chapter 2:

1. Scotland is considering independence from England. An opinion poll showed that 51% of Scots favor "independence." Another poll taken at the same time showed that only 34% favored being "separate" from England. The reason these results differ by so much is that

- (a) samples will usually differ just by chance due to random sampling.
(b) the wording of questions has a big effect on poll results.
(c) more follow-up efforts reduced the nonresponse rate of the second poll.
(d) the sample sizes are different, so the margins of error are different.

2. The telephone company says that 62% of all residential phone numbers in Los Angeles are unlisted. A telephone survey contacts a random sample of 1000 Los Angeles telephone numbers, of which 58% are unlisted. In this setting,

- (a) 62% is a parameter and 58% is a statistic
(b) 58% is a parameter and 62% is a statistic
(c) 62% and 58% are both parameters
(d) 58% and 62% are both statistics

3. The student newspaper runs a weekly question that readers can answer online or by campus mail. One question was "Do you think the college is doing enough to provide student parking?" Of the 136 people who responded, 79% said "No." The number 79% is a

- (a) margin of error (b) parameter (c) reliability (d) statistic

4. If we applied the quick method to the poll in the previous question, we would obtain this 95% confidence interval:

- (a) $79\% \pm 11.7\%$ (b) $79\% \pm 7.3\%$ (c) 136 ± 79 (d) $79\% \pm 8.6\%$

5. The newspaper poll in the previous problem does not give a trustworthy estimate of student opinion because of

- (a) bias due to nonresponse
(b) bias due to undercoverage
(c) bias due to the suggestive wording of the question
(d) bias due to relying on voluntary response

6. On January 6, just after the National Basketball Association labor dispute was settled, the Gallup Poll asked a random sample of 671 adults "How much have you missed watching NBA basketball since the dispute started?" 60% answered "Not at all." Gallup says that the margin of error for this result is plus or minus 4 percentage points. This means that
- (a) we can be 95% confident that between 56% and 64% of all adults did not miss watching NBA games
 - (b) we can be certain that between 56% and 64% of all adults did not miss watching NBA games
 - (c) in many samples, all the results would fall between 56% and 64%
 - (d) we are 95% confident that if we take one more sample the result will fall between 56% and 64%

A recent Gallup poll asked "Do you consider pro wrestling to be a sport, or not?" Of the people asked, 81% said "No." Here is what Gallup says about the accuracy of this poll:

The results below are based on telephone interviews with a randomly selected national sample of 1,028 adults, 18 years and older, conducted August 16-18, 1999. For results based on this sample, one can say with 95 percent confidence that the maximum error attributable to sampling and other random effects is plus or minus 3 percentage points. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.

The next seven questions concern this situation.

7. The population for this poll appears to be

- (a) all adults, 18 years and older.
- (b) 95% of adults, 18 years and older.
- (c) the 1028 adults who were interviewed.
- (d) 95% of the 1028 adults who were interviewed.

8. The sample for this poll is

- (a) all adults, 18 years and older.
- (b) 95% of adults, 18 years and older.
- (c) the 1028 adults who were interviewed.
- (d) 95% of the 1028 adults who were interviewed.

9. Which of these sources of possible errors in the poll result are covered by the margin of error of plus or minus three points?

- (a) The poll left out people without telephones.
- (b) Some people chosen for the sample refused to answer.
- (c) Some people did not tell the truth because they were embarrassed to admit they like pro wrestling.
- (d) None of these.

10. Which of these sources of possible errors in the poll result are examples of nonsampling errors?

- (a) The poll left out people without telephones.
- (b) Some people chosen for the sample refused to answer.
- (c) Some people did not tell the truth because they were embarrassed to admit they like pro wrestling.
- (d) (b) and (c) but not (a).

11. Which of these is a correct confidence statement based on this Gallup poll?

- (a) We are 95% confident that between 78% and 84% of all adults think that pro wrestling is not a sport.
- (b) We are 95% confident that between 78% and 84% of the 1028 people interviewed think that pro wrestling is not a sport.
- (c) There is a 95% chance that the opinions of the 1028 people interviewed fairly represent the opinions of all adults.
- (d) In many samples, 95% will find that 81% of the people interviewed think pro wrestling is not a sport.

12. In all, 151 people in the sample of 1028 adults said they were fans of pro wrestling. Gallup asked these 151 people, "Who is your favorite pro wrestler?" Twenty-four percent said Steve Austin was their favorite. Gallup gave a margin of error for this result. This margin of error is

- (a) plus or minus three percentage points, just as for the overall poll.
- (b) less than plus or minus three points because the sample for this question is smaller.
- (c) greater than plus or minus three points because the sample for this question is smaller.
- (d) less than plus or minus three points because the population of wrestling fans is smaller than the population of all adults.

13. Applying the quick method, we find that the margin of error for 95% confidence changes as follows when the sample size drops from 1028 to 151:

- (a) 8.1% to 3.1%.
- (b) 3.1% to 8.1%.
- (c) 3% to 2%.
- (d) 3% to 5%.

14. When you drop your pencil point blindly into the middle of a table of random digits, what is the chance that the three digits to the right of where you land will be 999?
- (a) 1 in 100, because every three-digit group has the same chance to come up.
 - (b) 1 in 1000, because every three-digit group has the same chance to come up.
 - (c) no chance, because 999 is not a random group of digits.
 - (d) can't say -- it is completely random.
15. Your statistics class has 30 students. You want to call an SRS of 5 students from your class to ask where they use a computer for the online exercises. You label the students 01, 02, . . . , 30. You enter the table of random digits at this line:
- 14459 26056 31424 80371 65103 62253 22490 61181
- Your SRS contains the students labeled
- (a) 14, 45, 92, 60, 56
 - (b) 14, 31, 03, 10, 22
 - (c) 14, 03, 10, 22, 22
 - (d) 14, 03, 10, 22, 06
16. You take an SRS of size 500 from the 37,000 students at Purdue University. You then take an SRS of size 500 from the 4,400,000 adults in the state of Indiana. The margin of error in a 95% confidence statement for the Indiana sample is
- (a) the same as for the Purdue sample because both are samples of size 500.
 - (b) smaller than for the Purdue sample because the population is much larger.
 - (c) larger than for the Purdue sample because the population is much larger.
 - (d) either larger or smaller than for the Purdue sample because it changes at random when we take a sample.
17. When we take a census, we attempt to collect data from
- (a) a stratified random sample
 - (b) every individual selected in a simple random sample
 - (c) every individual in the population
 - (d) a voluntary response sample
18. To reduce the variability of estimates from a simple random sample, you should
- (a) use a smaller sample.
 - (b) increase the bias.
 - (c) use a count, not a percent.
 - (d) use a larger sample.
19. Which of the following sources of error is included in the margin of error
- (a) chance variation in choosing a random sample.
 - (b) errors in entering the data into the computer.
 - (c) some of the subjects did not understand the questions.
 - (d) all of the above.
20. We divide the class into two groups: first year students and others. We then take random samples from each group. This is an example of
- (a) simple random sampling
 - (b) clustered sampling
 - (c) multistage sampling
 - (d) stratified random sampling
21. Some common sources of nonsampling error in samples of human populations are
- (a) using voluntary response samples; some subjects lie.
 - (b) some subjects lie; some subjects can't be contacted.
 - (c) some subjects can't be contacted; drawing a sample from names in a telephone directory.
 - (d) Both (b) and (c).

Chapter 3:

The next three questions concern this situation: Does using a cell phone while driving make an accident more likely? Researchers compared telephone company and police records to find 699 people who had cell phones and were also involved in an auto accident. Using phone billing records, they compared cell phone use in the period of the accident with cell phone use the same period on a previous day. Result: the risk of an accident was 4 times higher when using a cell phone.

1. This study is

- (a) a randomized comparative experiment.
- (b) an experiment, but without randomization.
- (c) a simple random sample.
- (d) an observational study, but not a simple random sample.

2. The explanatory variable in this study is

- (a) whether or not the subject had an auto accident.
- (b) whether or not the subject was using a cell phone.
- (c) the risk of an accident.
- (d) whether or not the subject owned a cell phone.

3. An example of a lurking variable that might affect the results of this study is:

- (a) whether or not the subject had an auto accident.
- (b) whether or not the subject was using a cell phone.
- (c) whether or not the subject was talking to a passenger in the car.
- (d) whether or not the subject owned a cell phone.

4. A study of a drug to prevent hair loss showed that 86% of the men who took it maintained or increased the amount of hair on their heads. But so did 42% of the men in the same study who took a placebo instead of the drug. This is an example of

- (a) a sampling error: the study should not have included men whose hair grew without the drug
- (b) the placebo effect: a treatment often works if you believe that it will work
- (c) an error in calculating percentages
- (d) failure to use the double-blind idea

5. Confounding often defeats attempts to show that one variable causes changes in another variable. Confounding means that

- (a) this was an observational study, so cause and effect conclusions are not possible
- (b) the effects of several variables are mixed up, so we cannot say which is causing the response
- (c) we don't know which is the response variable and which is the explanatory variable
- (d) we would get widely varied results if we repeated the study many times

The next six questions concern this situation: Want to stop smoking? Nicotine patches may help, and so may taking a drug that fights depression. A report in a recent issue of the *New England Journal of Medicine* describes a study of what works best. Here is part of the summary:

Use of nicotine replacement therapies and the antidepressant bupropion helps people stop smoking. We conducted a double-blind, placebo-controlled comparison of sustained-release bupropion (244 subjects), a nicotine patch (244 subjects), bupropion and a nicotine patch (245 subjects), and placebo (160 subjects) for smoking cessation.

Results. The abstinence rates at 12 months were 15.6 percent in the placebo group, as compared with 16.4 percent in the nicotine patch group, 30.3 percent in the bupropion group, and 35.5 percent in the group given bupropion and the nicotine patch.

6. How many treatments did this experiment compare?

- (a) two.
- (b) three.
- (c) four.
- (d) can't tell from the information given.

7. The response variable in this experiment is

- (a) the combination of drug (bupropion or placebo) and nicotine patch.
- (b) 893 people who want to quit smoking.
- (c) bupropion.
- (d) whether or not a subject was able to abstain from smoking for a year.

8. One group received a placebo. Why not just give this group no treatment at all?

- (a) It is not ethical to give no treatment at all in this setting.
- (b) Just thinking you are getting a treatment may have an effect, and we want to see if the real treatments do better than this.
- (c) A placebo is the same thing as no treatment at all.
- (d) Subjects would be disappointed if not given a pill.

9. The experiment was "double-blind." This means that
- (a) neither the subjects nor the people who worked with them knew whether they were taking bupropion or placebo.
 - (b) the subjects did not know that the treatments were intended to reduce their smoking.
 - (c) the subjects did not know whether they were taking bupropion or placebo.
 - (d) subjects were not allowed to see cigarette ads.
10. The subjects of the study included both men and women. All of the subjects were randomly assigned among all the treatments with the use of a table of random digits. This design is called
- (a) a simple random sample
 - (b) a completely randomized design.
 - (c) a matched pairs design.
 - (d) a block design.
11. The subjects of the study included both men and women. If the men and women were separately assigned to treatments, using the table of random digits twice, the design would be
- (a) a simple random sample
 - (b) a completely randomized design.
 - (c) a matched pairs design.
 - (d) a block design.
12. The most important advantage of experiments over observational studies is
- (a) a well designed experiment can give good evidence that the treatments actually cause the response.
 - (b) an experiment can compare two or more groups.
 - (c) we can use randomization to avoid bias in designing an experiment.
 - (d) we can study the relationship between two or more explanatory variables.
13. Ethical standards for randomized, controlled clinical trials include
- (a) not asking subjects to agree to participate without first informing them of the nature of the study and possible risks and benefits.
 - (b) insuring that each subject knows which treatment he or she received.
 - (c) allowing subjects to decide whether or not to be in the control group
 - (d) never testing drugs which have not been proven to be completely safe.
14. The reason that block designs are sometimes used in experimentation is to
- (a) prevent the placebo effect.
 - (b) allow double blinding.
 - (c) eliminate confounding with another factor.
 - (d) eliminate sampling variability.

The next three questions refer to a hypothetical experiment whose purpose is to determine whether the nutritional benefits of Little Chocolate Doughnuts can have a beneficial effect on the exam performance of Stat 001 students.

15. Suppose all students who attend Stat 001 lectures during Fall Term 1997 are given packages of Little Chocolate Donuts at the beginning of each lecture, which they are encouraged to eat during the first five minutes of class. Students who attend Stat 001 lectures during Spring Term 1998 will not be given doughnuts. Unfortunately, any systematic difference between Fall Term students and Spring Term students on the exams might be due to the fact that the spring term teacher is more skillful than the fall term teacher, rather than the doughnuts. This is an example of
- (a) confounding.
 - (b) the placebo effect.
 - (c) stratification.
 - (d) response error.
16. In the Spring Term, Stat 001 lectures will be given at 10:30 A.M. rather than at 8:30 A.M. The time of lecture (8:30 or 10:30) is
- (a) the response variable.
 - (b) an explanatory variable.
 - (c) a lurking variable.
 - (d) a stratum.
17. The response variable in this study is
- (a) the doughnut variable.
 - (b) time of lecture (8:30 or 10:30).
 - (c) the lecturer.
 - (d) exam performance.
18. The basic ethical requirements for any study of human subjects are
- (a) comparison, randomization, and replication.
 - (b) approval by a review board, informed consent, confidentiality of data.
 - (c) subjects are anonymous, subjects are randomly chosen, subjects cannot be harmed.
 - (d) data production, data analysis, inference.

19. Studies with human subjects must be approved in advance by an Institutional Review Board. The Board's main purpose is to
- be sure that the study is scientifically interesting.
 - be sure that the study uses good statistical techniques.
 - be sure that the study will have some benefit to society.
 - be sure that the subjects of the study are safe.

CHAPTER 7

1. In government data, a household consists of all occupants of a dwelling unit. Choose an American household at random and count the number of people it contains. Here is the assignment of probabilities for your outcome:

Number of persons	1	2	3	4	5	6	7
Probability	0.25	0.32	???	???	0.07	0.03	0.01

The probability of finding 3 people in a household is the same as the probability of finding 4 people. These probabilities are marked ??? in the table of the distribution. The probability that a household contains 3 people must be

- 0.68
- 0.32
- 0.16
- 0.08
- between 0 and 1, and we can say no more.

2. Which of the following statements about a table of random digits is true?

- If each line contains 40 digits, there will be exactly 4 zeros in every line.
- The probability that there are exactly 4 zeros in a line of 40 digits is exactly 0.5.
- The number of zeros in a line of 40 digits will vary, but on the average there will be 4 zeros per line.
- There can never be 4 zeros in a row because that pattern isn't random.
- Both (c) and (d) are true.

3. A friend rolls cheap dice many times. He reports that the probabilities of the possible outcomes are about as follows:

Outcome	1	2	3	4	5	6
Probability	0.2	0.2	0.2	0.1	0.1	0.2

Is this a legitimate probability model?

- Yes.
- No -- the faces must all have the same probability.
- No -- the 3 and 4 faces are opposite each other, so they must have the same probability.
- No -- the total probability for all faces is wrong.
- No -- not all the values given are possible values for a probability.

Choose an American household at random and ask how many cars and trucks that household owns. Here are the probabilities as of 1997:

Number of vehicles	0	1	2	3	4	5
Probability	0.04	0.25	0.45	0.18	0.06	0.02

4. This is a legitimate assignment of probabilities because it satisfies these rules:

- all the probabilities are between 0 and 1.
- all the probabilities are between -1 and 1.
- the sum of all the probabilities is exactly 1.
- Both (a) and (c).
- Both (b) and (c).

5. What is the probability that a randomly chosen household owns more than one motor vehicle?

- 0.96
- 0.71
- 0.26
- 0.25

6. Dice have six faces, showing 1 to 6 pips (spots). If a die is balanced, all six faces are equally likely. What must be the probability of each face?

- 1/10, or 0.10.
- 1/6, or 0.167.
- 2/10, or 0.20.
- could be any number between 0 and 1.

The casino game craps is based on rolling two dice. Here is the assignment of probabilities to the sum of the numbers on the up faces when two dice are rolled:

Outcome	2	3	4	5	6	7	8	9	10	11	12
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

7. The most common bet in craps is the "pass line." A pass line bettor wins immediately if either a 7 or an 11 comes up on the first roll. This is called a "natural." What is the probability of a natural?

- 2/36
- 6/36
- 8/36
- 12/36
- 20/36

8. Gigi has rolled a natural on four straight tosses of the dice. This excites the gamblers standing around the table. They should know that
- (a) Gigi has a hot hand, so she is more likely to roll another natural.
 - (b) The law of averages says that Gigi is now less likely to roll another natural.
 - (c) Rolls are independent, so the chance of rolling another natural has not changed.
 - (d) Four straight naturals are almost impossible, so the dice are probably loaded.

9. If I toss a fair coin five times and the outcomes are TTTTT, then the probability that tails appears on the next toss is
- (a) 0.5
 - (b) less than 0.5
 - (c) greater than 0.5
 - (d) 0
 - (e) 1

10. If a coin has 0.6 probability coming up tails, the probability that it comes up heads is
- (a) 0.5
 - (b) -0.2
 - (c) 0.4
 - (d) 0.6
 - (e) 1.0

CHAPTER 8

1. An exam has 40 multiple-choice questions, each with 5 choices. Only 1 of the 5 choices for each question is correct. If you used a table of random digits to randomly choose your answer on all questions, about how many answers would you expect to get correct?
- (a) 40
 - (b) 0
 - (c) 20
 - (d) 8
 - (e) 50

2. A basketball player makes 47% of her shots from the field during the season. To simulate whether a shot hits or misses you would assign random digits as follows:

- (a) One digit simulates one shot; 4 and 7 are a hit, other digits are a miss.
- (b) One digit simulates one shot; odd digits are a hit and even digits are a miss.
- (c) Two digits simulate one shot; 00 to 47 are a hit and 48 to 99 are a miss.
- (d) Two digits simulate one shot; 00 to 46 are a hit and 47 to 99 are a miss.
- (e) Two digits simulate one shot; 00 to 45 are a hit and 46 to 99 are a miss.

3. Use the correct choice from the previous question and these random digits to simulate 10 shots:
82734 71490 20467 47511 81676 55300 94383 14893

How many of these 10 shots are hits?

- (a) 2
- (b) 3
- (c) 4
- (d) 5
- (e) 6

4. You want to estimate the probability that the player makes 5 or more of 10 shots. You simulate 10 shots 25 times and get the following numbers of hits:

5 7 5 4 1 5 3 4 3 4 5 3 4 4 6 3 4 1 7 4 5 5 6 5 7

What is your estimate of the probability?

- (a) 5/25, or 0.20
- (b) 11/25, or 0.44
- (c) 12/25, or 0.48
- (d) 16/25, or 0.64
- (e) 19/25, or 0.76

5. Use the same simulation (25 trials with the results given in the previous exercise) to estimate the expected number of hits in 10 shots. Your estimate is:

- (a) 4 out of 10 shots
- (b) 4.4 out of 10 shots
- (c) 4.6 out of 10 shots
- (d) 5 out of 10 shots

6. In government data, a family consists of two or more persons who live together and are related by blood or marriage. Choose an American family at random and count the number of people it contains. Here is the assignment of probabilities for your outcome:

Number of persons	2	3	4	5	6	7
Probability	0.42	0.23	0.21	0.09	0.03	0.02

What is the probability that the family you choose has more than 2 people?

- (a) 0.35
- (b) 0.42
- (c) 0.58
- (d) 1.00
- (e) Between 0 and 1, and we can say no more.

7. Using the probabilities in the previous question, what is the expected size of the family you draw?

- (a) 2 people
- (b) 3 people
- (c) 3.14 people
- (d) 3.5 people
- (e) 4.5 people

8. Computer voice recognition software is getting better. Some companies claim that their software correctly recognizes 98% of all words spoken by a trained user. To simulate recognizing a single word when the probability of being correct is 0.98, you would use random digits as follows:

- (a) two digits simulate one word; 00 to 97 mean \correct."
- (b) two digits simulate one word; 00 to 98 mean \correct."
- (c) one digit simulates one word; 0 to 9 mean \correct."
- (d) three digits simulate one word; 001 to 098 mean \correct."

9. The program of the previous exercise recognizes words (or not) independently. To simulate the program's performance on 10 words, use your method from the previous problem and these random digits:

60970 70024 17868 29843 61790 90656 87964 18883

The number of words correct out of the 10 is

- (a) 10 (b) 9 (c) 8 (d) 2

10. A gambler who keeps placing \$1 bets on roulette will, after a very large number of bets, find that his average winnings per bet are close to \$0.947. (The house keeps the other \$0.053 per bet.) The statistical term for the number \$0.947 is

- (a) the probability of winning a bet. (b) the bias of a bet. (c) a random number. (d) the expected value of a bet.

Chapter 1:

1. (b)
2. (d)
3. (c)
4. (d)
5. (b)
6. (d)
7. (c)
8. (d)
9. (b)
10. (b)
11. (a)
12. (a)
13. (b)
14. (c)

Chapter 2:

1. (b)
2. (a)
3. (d)
4. (d)
5. (d)
6. (a)
7. (a)
8. (c)
9. (d)
10. (d)
11. (a)

12. (c)
13. (b)
14. (b)
15. (d)
16. (a)
17. (c)
18. (d)
19. (a)
20. (d)
21. (b)

Chapter 3:

1. (d)
2. (b)
3. (c)
4. (b)
5. (b)
6. (c)
7. (d)
8. (b)
9. (a)
10. (b)
11. (d)
12. (a)
13. (a)
14. (c)
15. (a)
16. (c)
17. (d)

18. (b)
19. (d)

Chapter 7:

1. (c)
2. (c)
3. (a)
4. (d)
5. (b)
6. (b)
7. (c)
8. (c)
9. (a)
10. (c)

Chapter 8:

1. (d)
2. (d)
3. (c)
4. (c)
5. (b)
6. (c)
7. (c)
8. (a)
9. (b)
10. (d)

Prob & Stats - Midterm Review Part 2

1) In a randomized comparative experiment, there will be 20 subjects split into 4 even groups.

a) Describe how you would use a random digit table to assign the subjects to each group.

b) Starting at line 120 of a random digit table, select the subjects for the first group.

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

2) A carnival booth game tries to attract customers by saying that the percentage of winners out of the last 10 customers was 63%. Is this possible? Explain either way.

3) The Gallop poll interviewed 1600 random adults. Of the sample, 288 of them said that they jog regularly.

a) What is the population of interest?

b) What is the parameter of interest?

c) What is the sample?

d) What is the sample statistic?

e) Using the quick method for 95% confidence, what is the margin of error?

4) Define and understand the following vocabulary words as they relate to measurements (CH 1.2):

a) Reliable

b) Validity

c) Bias

5) A Pharmaceutical company wants to test the effectiveness of 2 different drugs for pain relief after surgery. The hospital participating in the study allows each patient to choose which drug they will take after their surgery.

a) Explain why confounding makes the results of this study useless.

b) Given 50 patients who are willing to try either one of the 2 drugs proposed by the pharmaceutical company, outline an experimental design to compare the effectiveness of the two different pain relievers.

6) A report in the April 26, 2001 New England Journal of Medicine studied a new treatment for children with a severe anxiety disorder. The study was a randomized double-blind comparative experiment. Data from the study showed that 76% of the children treated with the new drug had a reduced anxiety level. Of the children who were given a placebo, 29% had a reduced anxiety level. Almost none of the patients in the study exhibited an increase in anxiety levels.

a) Explain the meaning of the word "placebo" in the above description. Then discuss why it was important to administer a placebo as part of the design of the experiment.

b) Explain what is meant by "double-blind" in the above description. Then discuss why it was important to make this experiment double-blind.

6) Do the people want a tax cut? Before the 2000 presidential election, the candidates debated what to do with the large government surplus. The Pew Center asked two questions of random samples of adults. Both said that Social Security would be "fixed". Here are the uses suggested for the remaining surplus.

Should the money be used for a tax cut, or should it be used to fund new government programs?

Should the money be used for a tax cut, or should it be spent on programs for education, the environment, health care, crime-fighting and military defense?

a) One of the questions drew 60% favoring a tax cut; the other, only 22%. Which wording pulls respondents toward a tax cut? Why?

b) Is the "wording problem" described above a source of sampling error or nonsampling error? Would this error be included in a margin of error for either poll. Explain.

7) A bag contains 3 different colors of marbles in it. It has 4 red marbles, 2 yellow marbles and 5 purple marbles.

a) What is the probability that a person picks a red marble?

b) What is the probability that a person picks a red marble *or* a yellow marble?

c) What if a person chooses two marbles. What is the probability that he/she picks a red marble *and* a purple marble?

8) A household is a group of people living together at the same address. The following probability model can be used to predict the number of people in a household chosen at random.

# of People	1	2	3	4	5	6	7
Probability	.25	.32	.17	.16	.07	.02	.01

- a) What is the probability that a household selected will have less than 4 people?
- b) What is the probability that a household selected will have either 3 or 4 people?
- c) What is the probability that a household selected will not have 2 people?
- d) What is the probability that a household selected will have more than 5 people?

9) A grocery chain runs a prize game by giving each customer a ticket that may win a prize when a box is scratched. Printed on the ticket are the following probabilities for a customer who shops once a week"

Amount Won	\$500	\$50	\$10	\$0
Probability	0.01	0.05	0.20	0.74

- a) What is the expected value of a customer's winnings in this game?
- b) What is the probability that a person who wins will win more than the expected value?
- c) Explain how digits could be assigned to simulate the amount won.
- d) What is the probability that a household selected will have more than 5 people?

Prob & Stats - Midterm Review Part 2

1) In a randomized comparative experiment, there will be 20 subjects split into 4 even groups.

a) Describe how you would use a random digit table to assign the subjects to each group.

5 people per group

Assign digits 00-19 to represent the 20 subjects

b) Starting at line 120 of a random digit table, select the subjects for the first group.

16 04 19 07 10

120	35476	55972	39421	65850	02266	35435	43742	1037
121	71487	09984	29077	14863	61683	47052	62224	61025
122	13873	81598	95052	90908	73592	75186	87136	95761

2) A carnival booth game tries to attract customers by saying that the percentage of winners out of the last 10 customers was 63%. Is this possible? Explain either way.

$$\frac{15}{100} = \frac{\%}{100}$$

$$\frac{6.3}{10} \neq \frac{63}{100}$$

No because you can't have 6.3 people win

3) The Gallop poll interviewed 1600 random adults. Of the sample, 288 of them said that they jog regularly.

a) What is the population of interest? all adults

b) What is the parameter of interest? do they jog

c) What is the sample? 1600

d) What is the sample statistic? $\frac{288}{1600} = 18\%$

e) Using the quick method for 95% confidence, what is the margin of error?

$$\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{1600}} = .025 = 2.5\% \quad 15.5\% \text{ to } 20.5\%$$

4) Define and understand the following vocabulary words as they relate to measurements (CH 1.2):

a) Reliable

b) Validity

c) Bias

5) A Pharmaceutical company wants to test the effectiveness of 2 different drugs for pain relief after surgery. The hospital participating in the study allows each patient to choose which drug they will take after their surgery.

a) Explain why confounding makes the results of this study useless.

b) Given 50 patients who are willing to try either one of the 2 drugs proposed by the pharmaceutical company, outline an experimental design to compare the effectiveness of the two different pain relievers.

Double-blind experiment

50 $\left\{ \begin{array}{l} 25 \text{ Drug A} \\ \text{randomly} \\ 25 \text{ placebo} \end{array} \right.$ *randomly choose the 2 groups*

6) A report in the April 26, 2001 New England Journal of Medicine studied a new treatment for children with a severe anxiety disorder. The study was a randomized double-blind comparative experiment. Data from the study showed that 76% of the children treated with the new drug had a reduced anxiety level. Of the children who were given a placebo, 29% had a reduced anxiety level. Almost none of the patients in the study exhibited an increase in anxiety levels.

a) Explain the meaning of the word "placebo" in the above description. Then discuss why it was important to administer a placebo as part of the design of the experiment.

b) Explain what is meant by "double-blind" in the above description. Then discuss why it was important to make this experiment double-blind.

6) Do the people want a tax cut? Before the 2000 presidential election, the candidates debated what to do with the large government surplus. The Pew Center asked two questions of random samples of adults. Both said that Social Security would be "fixed". Here are the uses suggested for the remaining surplus.

Should the money be used for a tax cut, or should it be used to fund new government programs?

Should the money be used for a tax cut, or should it be spent on programs for education, the environment, health care, crime-fighting and military defense?

a) One of the questions drew 60% favoring a tax cut; the other, only 22%. Which wording pulls respondents toward a tax cut? Why?

b) Is the "wording problem" described above a source of sampling error or nonsampling error? Would this error be included in a margin of error for either poll. Explain.

NON sampling errors are not accounted for in the margin of error

7) A bag contains 3 different colors of marbles in it. It has 4 red marbles, 2 yellow marbles and 5 purple marbles.

a) What is the probability that a person picks a red marble? $\frac{4}{11}$

b) What is the probability that a person picks a red marble or a yellow marble?
 $\frac{4}{11} + \frac{2}{11} = \frac{6}{11}$ add

c) What if a person chooses two marbles. What is the probability that he/she picks a red marble and a purple marble?

without replacement
mult $\frac{4}{11} \cdot \frac{5}{10} = \frac{20}{110} = \frac{4}{22} = .18$

8) A household is a group of people living together at the same address. The following probability model can be used to predict the number of people in a household chosen at random.

# of People	1	2	3	4	5	6	7
Probability	.25	.32	.17	.16	.07	.02	.01

a) What is the probability that a household selected will have less than 4 people?

$$.25 + .32 + .17 = .74$$

b) What is the probability that a household selected will have either 3 or 4 people?

$$.17 + .16 = .33$$

c) What is the probability that a household selected will not have 2 people?

$$1 - .32 = .68$$

d) What is the probability that a household selected will have more than 5 people?

$$.02 + .01 = .03$$

9) A grocery chain runs a prize game by giving each customer a ticket that may win a prize when a box is scratched. Printed on the ticket are the following probabilities for a customer who shops once a week"

Amount Won	\$500	\$50	\$10	\$0
Probability	0.01	0.05	0.20	0.74

$5 + 2.5 + 2 + 0 = \$9.50$

a) What is the expected value of a customer's winnings in this game?

$$\$9.50$$

b) What is the probability that a person who wins will win more than the expected value?

$$.26$$

c) Explain how digits could be assigned to simulate the amount won.

$$\begin{array}{l} \$500 / 00 \\ \$50 / 01-04 \\ \$10 / 05-19 \\ \$0 / 20-99 \end{array}$$

d) What is the probability that a household selected will have more than 5 people?