Probability & Statistics 40: Midterm Exam Review 2017-2018

**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
(a) 0.68 (b) 0.32 (c) 0.16 (d) 0.08 (e) between 0 and 1, and we can say no more.

2. Which of the following statements about a table of random digits is **true**?
(a) If each line contains 40 digits, there will be exactly 4 zeros in every line.
(b) The probability that there are exactly 4 zeros in a line of 40 digits is exactly 0.5.
(c) The number of zeros in a line of 40 digits will vary, but on the average there will be 4 zeros per line.
(d) There can never be 4 zeros in a row because that pattern isn't random.
(e) 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?
(a) Yes.
(b) No -- the faces must all have the same probability.
(c) No -- the 3 and 4 faces are opposite each other, so they must have the same probability.
(d) No -- the total probability for all faces is wrong.
(e) 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:
(a) all the probabilities are between 0 and 1. (b) all the probabilities are between -1 and 1.
(c) the sum of all the probabilities is exactly 1. (d) Both (a) and (c). (e) Both (b) and (c).

5. What is the probability that a randomly chosen household owns more than one motor vehicle?
(a) 0.96 (b) 0.71 (c) 0.26 (d) 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?
(a) 1/10, or 0.10. (b) 1/6, or 0.167. (c) 2/10, or 0.20. (d) 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?
(a) 2/36 (b) 6/36 (c) 8/36 (d) 12/36 (e) 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 4**

1. A company database contains the following information about each employee: age, date hired, sex (male or female), ethnic group (Asian, black, Hispanic, etc.), job category (clerical, management, technical, etc.), yearly salary. Which of the following lists of variables are *all* categorical?

(a) age, sex, ethnic group. (b) sex, ethnic group, job category. (c) ethnic group, job category, yearly salary.
(d) yearly salary, age. (e) age, date hired.

2. Were the extinctions that occurred in the last ice age more frequent among species of animals with large body sizes? A researcher gathers data on the average body mass (in kilograms) of all species known to have existed at that time. These measurements are values of

(a) a categorical variable. (b) a quantitative variable. (c) an invalid variable. (d) a margin of error.

3. In the situation of the previous question, what are the explanatory and response variables?

(a) There is no explanatory-response distinction in this situation.
(b) Explanatory: body mass of a species. Response: whether the species went extinct.
(c) Explanatory: the ice age. Response: whether a species went extinct.
(d) Explanatory: whether a species went extinct. Response: the body mass of the species.
(e) Explanatory: the ice age. Response: the body mass of a species.

4. An example of a categorical variable is

(a) the name of the college a student attends. (b) a student's weight in kilograms.
(c) a student's class rank, such as 25th out of 364. (d) a student's sex (male or female). (e) Both (a) and (d).

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:

| Year | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
|---|---|---|---|---|---|---|---|---|---|---|
| Return | 31.7 | −3.1 | 30.5 | 7.6 | 10.1 | 1.3 | 37.6 | 23.0 | 33.4 | 28.6 |

**The next five questions** are related to this situation.

5. The median return during this period is

(a) 5.5 (b) 20.07 (c) 23.0 (d) 25.8 (e) 28.6

6. The third quartile of these returns is

(a) 7.6 (b) 30.5 (c) 31.1 (d) 31.7 (e) 33.4

7. The mean return is

(a) 20.07 (b) 20.69 (c) 22.3 (d) 25.8 (e) 33.4

8. The standard deviation of the returns is

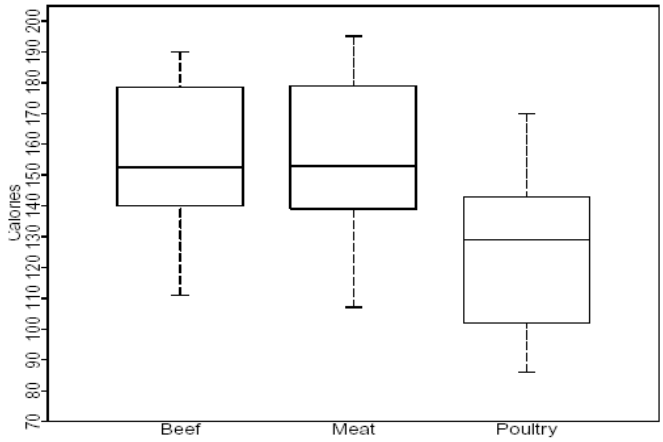
(a) 13.75 (b) 13.98 (c) 14.74 (d) 20.07 (e) 25.8

9. You have similar data on returns on common stocks for all years since 1945. To show clearly how returns have changed over time, your best choice of graph is

(a) a bar graph (b) a line graph (c) a pie chart (d) a histogram (e) a scatterplot

10. According to the student newspaper, the mean salary of male full professors in the School of Management is $117,302. The median of these salaries
(a) would be lower, because salary distributions are skewed to the left.
(b) would be lower, because salary distributions are skewed to the right.
(c) would be higher, because salary distributions are skewed to the left.
(d) would be higher, because salary distributions are skewed to the right.

Here are boxplots of the number of calories in 20 brands of beef hot dogs, 17 brands of meat hot dogs, and 17 brands of poultry hot dogs.




11. The main advantage of boxplots over stemplots and histograms is
(a) boxplots make it easy to compare several distributions, as in this example
(b) boxplots show more detail about the shape of the distribution
(c) boxplots use the five-number summary, whereas stemplots and histograms use the mean and standard deviation
(d) boxplots show skewed distributions, whereas stemplots and histograms show only symmetric distributions

12. This plot shows that
(a) all poultry hot dogs have fewer calories than the median for beef and meat hot dogs
(b) about half of poultry hot dog brands have fewer calories than the median for beef and meat hot dogs
(c) hot dog type is not helpful in predicting calories, because some hot dogs of each type are high and some of each type are low
(d) most poultry hot dog brands have fewer calories than most beef and meat hot dogs, but a few poultry hot dogs have more calories than the median beef and meat hot dog

13. We see from the plot that the median number of calories in a beef hot dog is about
(a) 190 (b) 179 (c) 153 (d) 139 (e) 129

14. The box in each boxplot marks
(a) the full range covered by the data
(b) the range covered by the middle half of the data
(c) the range covered by the middle three-quarters of the data
(d) the span one standard deviation on each side of the mean

15. The calorie counts for the 17 poultry brands are:
129 132 102 106 94 102 87 99 170 113 135 142 86 143 152 146 144
The median of these values is
(a) 129 (b) 132 (c) 130.5 (d) 121 (e) 170

**CHAPTER 5**

Suppose that the BAC of male students at a particular college who drink 5 beers varies from student to student according to a normal distribution with mean 0.08 and standard deviation 0.01. **The next three questions** use this information.

1. The middle 95% of students who drink 5 beers have BAC between
(a) 0.07 and 0.09 (b) 0.06 and 0.10 (c) 0.05 and 0.11 (d) 0.04 and 0.12

2. What percent of students who drink 5 beers have BAC above 0.08 (the legal limit for driving in most states)?
(a) 2.5% (b) 5% (c) 16% (d) 32% (e) 50%

3. What percent of students who drink 5 beers have BAC above 0.10 (the legal limit for driving other states)?
(a) 2.5% (b) 5% (c) 16% (d) 32% (e) 50%

4. SAT scores are normally distributed with mean 500 and standard deviation 100. Julie scores 650. Her standard score is
(a) 150 (b) 15 (c) 1.5 (d) 0.15

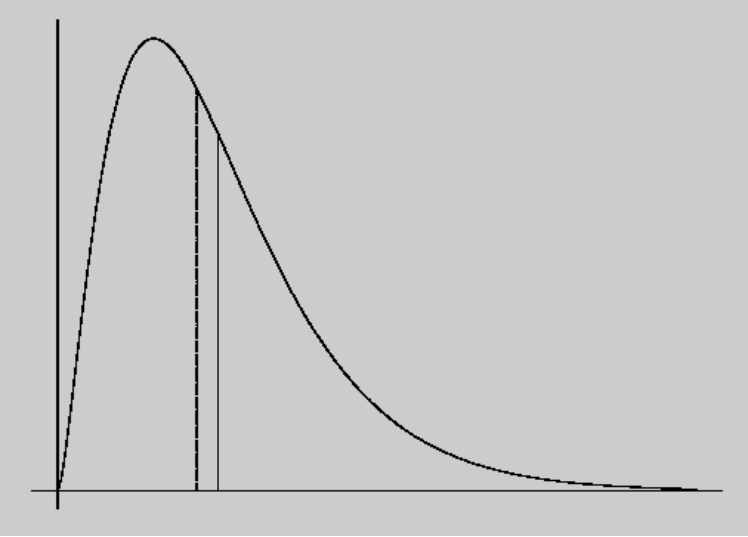
**The next four questions** use this information: The length of pregnancy isn't always the same. In pigs, the length of pregnancies varies according to a normal distribution with mean 114 days and standard deviation 5 days.

5. What range covers the middle 95% of pig pregnancies?
(a) 109 to 119 days (b) 104 to 124 days (c) 99 to 129 days (d) 94 to 134 days

6. What percent of pig pregnancies are longer than 114 days?
(a) 16% (b) 34% (c) 50% (d) 84%

7. What percent of pig pregnancies are longer than 109 days?
(a) 16% (b) 34% (c) 50% (d) 84%

8. The median length of a pig pregnancy is
(a) 119 days. (b) 114 days. (c) 109 days.
(d) between 109 and 119 days, but can't be more specific.
(e) greater than 114 days, but can't be more specific.

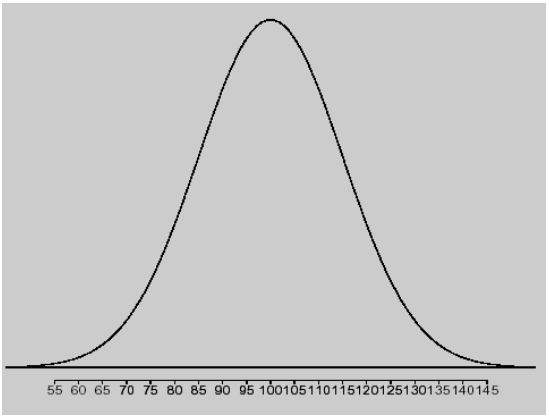


9. Two measures of center are marked on the density curve above.
(a) The median is at the solid line and the mean is at the dashed line.
(b) The median is at the dashed line and the mean is at the solid line.
(c) The mode is at the dashed line and the median is at the solid line.
(d) The mode is at the solid line and the median is at the dashed line.

10. Some people buy the stock of small companies. The Russell 2000 index, which tracks the price of such shares, was 648 on July 15, 1999. On October 15, the index was 415. What percent decrease is this?
(a) 156% (b) 64% (c) 56% (d) 36%

11. The mean of any density curve is
(a) the point where the curvature of the curve changes.
(b) the point at which the curve reaches its highest value.
(c) the point at which the curve would balance if made of solid material.
(d) the point with half the area under the curve to its left and to its right.

12. Fifty percent of the observations in any distribution will be between
(a) the quartiles (b) the mean plus or minus one standard deviation
(c) the mean plus or minus two standard deviations (d) the mean plus or minus three standard deviations
(e) the mean and the standard deviation




13. The mean of the normal curve to the right is
(a) 80 (b) 90 (c) 100 (d) 110 (e) 120

14. The standard deviation of the normal curve to the right is
(a) 5 (b) 10 (c) 15 (d) 20 (e) 25

15. If you know the mean and standard deviation of a distribution, do you know the complete shape of the distribution?
(a) Yes, always. (b) Yes if the distribution is normal, but not in general.
(c) Yes if the distribution is symmetric, but not in general. (d) No, never.

**CHAPTER 6**

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:

| Year | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
|---|---|---|---|---|---|---|---|---|---|---|
| Return | 31.7 | −3.1 | 30.5 | 7.6 | 10.1 | 1.3 | 37.6 | 23.0 | 33.4 | 28.6 |

**The next three questions** are related to this situation.

1. The correlation of U.S. stock returns with overseas stock returns during these years was $r = 0.44$. This tells you that
(a) when U.S. stocks rose, overseas stocks also tended to rise, but the connection was not very strong
(b) when U.S. stocks rose, overseas stocks rose by almost exactly the same amount
(c) when U.S. stocks rose, overseas stocks tended to fall, but the connection was not very strong
(d) there is almost no relationship between changes in U.S. stocks and changes in overseas stocks
(e) nothing, because this is not a possible value of $r$

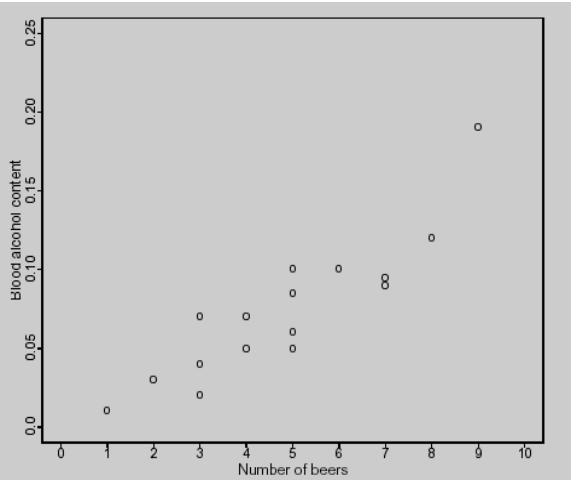
2. If $x$ is the return on U.S. stocks and $y$ is the return on overseas stocks in the same year, the least-squares regression line for predicting $y$ from $x$ is $y = -2.7 + 0.47x$. You think U.S. stocks will have a return of 10% in 1999. Using this regression line, you predict that the return on overseas stocks will be
(a) 7.4% (b) -2.23% (c) 2% (d) 3.17%

3. Stock returns are measured in percent. What are the units of the mean, the median, the quartiles, the standard deviation, and the correlation between U.S. and overseas returns?
(a) all are measured in percent.
(b) all are measured in percent except the standard deviation, which is measured in squared percent.
(c) all are measured in percent except the correlation, which is a number that has no units.
(d) all are measured in percent except the correlation, which is measured in squared percent.

4. Consider a large number of countries around the world. There is a positive correlation between the number of Nintendo games per person $x$ and the average life expectancy $y$. Does this mean that we could increase the life expectancy in Rwanda by shipping Nintendo games to that country?
(a) Yes: the correlation says that as Nintendos go up, so does life expectancy.
(b) No: if the correlation were negative we could accept that conclusion, but this correlation is positive.
(c) Yes: positive correlation means that if we increase $x$, then $y$ will also increase.
(d) No: the positive correlation just shows that richer countries have both more Nintendos and higher life expectancies.
(e) It makes no sense to calculate correlation between these variables.

5. Suppose that the correlation between the scores of students on Exam 1 and Exam 2 in a statistics class is $r = 0.7$. One way to interpret $r$ is to say what percent of the variation in Exam 2 scores can be explained by the straight line relationship between Exam 2 scores and Exam 1 scores. This percent is about
(a) 84% (b) 70% (c) 49% (d) 30%

6. What can we say about the relationship between a correlation $r$ and the slope $b$ of the least-squares line for the same set of data?
(a) $r$ is always larger than $b$ (b) $r$ and $b$ always have the same sign (+ or -)
(c) $b$ is always larger than $r$ (d) $b$ and $r$ are measured in the same units




7. One student drank 9 beers. You see from the scatterplot that his BAC was about
(a) 0.19 (b) 9 (c) 19 (d) 0.05

8. The scatterplot shows
(a) a weak negative relationship (b) a moderately strong negative relationship (c) almost no relationship
(d) a weak positive relationship (e) a moderately strong positive straight-line relationship between number of beers and BAC.

9. A plausible value of the correlation between number of beers and blood alcohol content, based on the scatterplot, is
(a) $r = -0.9$ (b) $r = -0.3$ (c) $r$ close to 0 (d) $r = 0.3$ (e) $r = 0.9$

**Ch 7**
1. c
2. c
3. a
4. d
5. b
6. b
7. c
8. c
9. a
10. c

**Ch 8**
1. d
2. d
3. c
4. c
5. b
6. c
7. c
8. a
9. b
10. d

**Ch 4**
1. b
2. b
3. b
4. e
5. d
6. d
7. a
8. c
9. b
10. d
11. a
12. d
13. c
14. b
15. a

**Ch 5**
1. b
2. e
3. a
4. c
5. b
6. c
7. d
8. b
9. b
10. d
11. c
12. a
13. c
14. c
15. b

**Ch 6**
1. a
2. c
3. c
4. d
5. c
6. b
7. a
8. e
9. e