STAT*2040 W14

Test 1 (White Version)

February 7 2014

1. The colour of the first page of this examination booklet (the cover sheet) is:

   (a) White **
   (b) Yellow

2. Suppose we draw a simple random sample of four observations, and find that the values are:

   $$4, 17, 3, 1$$

   What is the standard deviation of these values?

   $s = 7.274384$. *This is a very easy question. There are many examples of this type of question in the notes, suggested exercises, and Maple TA quizzes, so I'm not giving a full solution. (See, for example, questions 6, 7, 21, 22, 23 of the suggested exercises for descriptive statistics.) Using your calculator's standard deviation function is the best way to calculate this value.*

   (a) 5.3
   (b) 6.3
   (c) 7.3 **
   (d) 8.3
   (e) 9.3

3. Of the following options, which one best represents the proper interpretation of the standard deviation?

   As discussed in class, and implied by the formula, the standard deviation is the square root of the average squared distance from the mean.

   (a) It is the average distance from the mean.
   (b) It is the square of the average distance from the mean.
   (c) It is the square root of the average squared distance from the mean. **
   (d) It is the square root of the average distance from the mean.
   (e) It is the square root of the sum of the squared observations.

4. Suppose a researcher is interested in estimating the average weight of wild rabbits in a certain area. They set several cage traps in the area and capture 15 wild rabbits. The 15 rabbits are weighed and they are found to have a mean weight of 3.6 kg. Suppose that, in reality, the average weight of all wild rabbits in the area is 4.2 kg.

   This question is very similar to Question 12 of the Gathering Data exercises.

   Consider the following statements:

   I. 3.6 is the value of a statistic. *True, since it is a value based on sample data.*
   II. 4.2 is the value of a parameter. *True, since it is the true value for the entire population.*
   III. The 15 rabbits are a stratified random sample from the population of interest. *False. There is some randomness involved, but it most definitely does not satisfy the conditions of a stratified random sample.)*

Which of these statements are false?

(a) Just I.
(b) Just II.
(c) Just III.**
(d) I and II.
(e) All of them.

5. Of the following scenarios, which one would best be investigated by a randomized experiment rather than an observational study?

   This question is *very* similar to Question 27 of the suggested exercises for the Gathering Data chapter.

   There is only one of these scenarios in which an experiment would be plausible. (In any experiment, the researchers impose conditions on the experimental units.) In the other situations we would simply be observing pre-existing variables (they would be observational studies).

   (a) An investigation into whether men who have cheated on their spouse are more likely to be heroine users than men who have not cheated on their spouse.
   (b) An investigation into whether men are more likely than women to use guns in suicide attempts.
   (c) An investigation into whether a new drug is more effective than a standard drug in reducing blood pressure in men with high blood pressure. **
   (d) An investigation into differences in the the heights of male and female University of Guelph students.
   (e) An investigation that compares the heights of skyscrapers in different cities.

6. Suppose $P(A) = 0.45$, $P(B) = 0.35$, and $P(A \cap B) = 0.15$. What is $P(A^c|B)$?

   It's best to draw a Venn diagram for this one. $P(A^c|B) = \frac{P(A^c \cap B)}{P(B)}$. $P(B) = P(A^c \cap B) + P(A \cap B)$ *(draw the Venn diagram).* $0.35 = 0.15 + P(A^c \cap B) \implies P(A^c \cap B) = 0.20$. *So:* $P(A^c|B) = \frac{P(A^c \cap B)}{P(B)} = \frac{0.20}{0.35} = 0.571$.

   (a) 0.43
   (b) 0.45
   (c) 0.55
   (d) 0.57**
   (e) 0.59

7. Suppose a sample data set has a mean of 150, a standard deviation of 23, and the largest observation in this data set has a z-score of 2.6. If each observation in the data set is multiplied by 2 and then 5 is added, what will be the z-score of the largest observation in this transformed data set?

   z-scores won't change under a linear transformation. This was illustrated in Questions 2 and 3 of Lab 2 (and can be figured out if you know the effect of a linear transformation on the mean and standard deviation).

   (a) 1.3
   (b) 2.6**
   (c) 5.2

(d) 10.2
(e) It is impossible to tell without further information.

8. Which one of the following statements is true?

Question 36 from the descriptive statistics exercises.

(a) If a variable x is measured in seconds, then the units of the standard deviation are seconds^2. *False. The units of the standard deviation are the same as the units of the variable.*
(b) For any data set, the value of the 12.5$^{\text{th}}$ percentile is exactly half the value of Q_1. *False. A percentile is a value of a variable. There is no reason why the 12.5th percentile would need to be half the value of the 25th percentile. For example, the 12.5th percentile of the heights of Canadian men is roughly 167 cm, and the 25th percentile is about 171 cm.*
(c) The standard deviation can be negative. *False. Since it's the square root of the sum of squared terms, $s \geq 0$.*
(d) All of the values in the five-number summary are always positive. *False. The five-number summary is: Minimum, Q_1, Median, Q_3, Maximum. All of these quantities can be negative.*
(e) None of the above. **

9. Which one of the following statements is true?

(a) If $P(A) = P(B)$, then A and B are independent. *False.*
(b) If $P(A) > P(B)$, then $P(A|B) \geq P(B|A)$. **
True. $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$. *If* $P(A) > P(B)$, *then* $\dfrac{P(A \cap B)}{P(B)} > \dfrac{P(A \cap B)}{P(A)}$, *unless they both equal 0.*
(c) If A and B are independent, then $P(A \cup B) = P(A) + P(B)$. *False.*
(d) If $P(A|B) = 1$, then $P(B|A) = 1$. *False.*
(e) None of the above.

10. A pen of rabbits contains 8 American Blue rabbits and 14 British Giant rabbits. If 6 of these rabbits are randomly selected without replacement, what is the probability that no more than 1 American Blue rabbit is selected?

Basic hypergeometric question. If we let X represent the number of American Blue rabbits selected, we need to find $P(X \leq 1) = P(X = 0) + P(X = 1) = \dfrac{\binom{8}{0}\binom{14}{6}}{\binom{22}{6}} + \dfrac{\binom{8}{1}\binom{14}{5}}{\binom{22}{6}}$.

(a) 0.17
(b) 0.21
(c) 0.25**
(d) 0.29
(e) 0.33

11. Consider the following probability distribution for a random variable X.

What is $E(X^2)$?

We did a similar example in class, and there are similar questions in the suggested exercises.

$$E(X^2) = \sum x^2 p(x) = (-1)^2 \cdot 0.2 + 2^2 \cdot 0.2 + 10^2 \cdot 0.6 = 61.0.$$

x	-1	2	10
$p(x)$	0.2	0.2	0.6

(a) 3.7
(b) 6.2
(c) 13.4
(d) 38.4
(e) 61.0 **

12. Suppose X is a binomial random variable with parameters n and p. Which one of the following statements is true?

    (a) X can take on negative values. *False. X is a count of the number of successes. It must be at least 0.*
    (b) The expected value of X is its most likely value. *False. An expectation is a mean – it doesn't have to be a possible value of X.*
    (c) $P(X = 0)$ is always less than $P(X = 1)$. *False. Simple counterexample: $n = 1$, $p = 0.01$. $P(X = 0) = 0.99, P(X = 1) = 0.01$.*
    (d) X can take on one of $n - 1$ possible values. *False. A binomial random variable can take on one of $n + 1$ possible values: $0, 1, 2, \ldots, n$.*
    (e) None of the above. **

13. Suppose:

    - X is a random variable with $\mu_X = 3.2$ and $\sigma_X = 8.5$.
    - Y is a random variable with $\mu_Y = 4.7$ and $\sigma_Y = 2.2$.
    - X and Y are independent.

    What is the standard deviation of $X - Y$?

    We did a nearly identical example in class, and it's essentially the same as Question 6c of the discrete random variable suggested exercises.

    Since X and Y are independent, the variance of $X - Y$ is the sum of the variances: $Var(X - Y) = Var(X) + Var(Y) = 8.5^2 + 2.2^2 = 77.09$, and the standard deviation is the square root of the variance: $\sqrt{77.09} = 8.78$.

    (a) 6.3
    (b) 8.5
    (c) 8.8**
    (d) 10.7
    (e) 14.1

14. Suppose X is a binomial random variable with a very large n and a very small p. The exact values of n and p are unknown, but it is known that $E(X) = 5$. $P(X = 2)$ is closest to which one of the following?

For large n and small p, the binomial distribution can be well approximated by the Poisson distribution with $\lambda = np = E(X)$. $P(X = 2) \approx \frac{5^2 e^{-5}}{2!} = 0.084$.

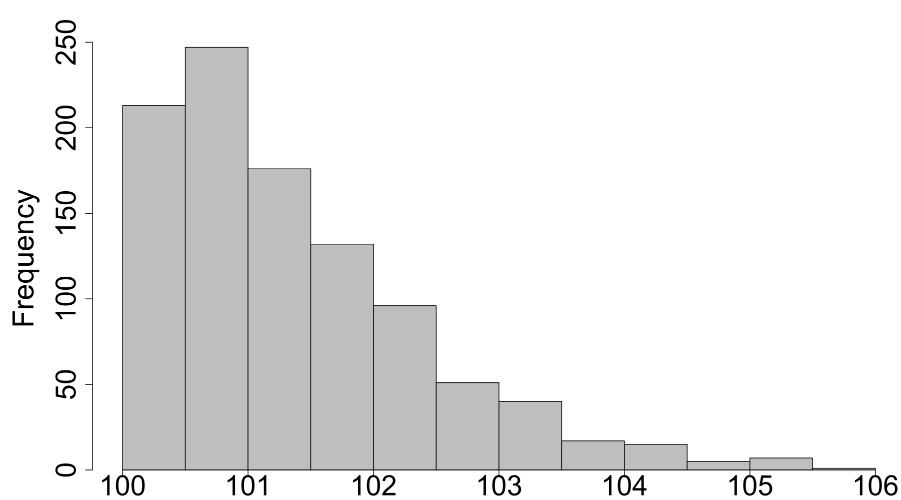
(a) 0.06
(b) 0.08**
(c) 0.10
(d) 0.12
(e) 0.14

15. Suppose we randomly select a Canadian adult. Let A be the event that the person is a heart surgeon. Let B be the event that the person has at least one university degree. Which one of the following statements is true?

Very similar to Question 24 in the suggested exercises for probability. Heart surgeons are much more likely to have a university degree than a randomly selected Canadian, so $P(B|A) > P(B)$. *You don't need to know the actual probabilities to answer this question, but in case you're interested:* $P(B|A) = 1$, $P(B) \approx 0.25$.

(a) $P(A|B) < P(A)$
(b) $P(B|A) > P(B)$ **
(c) A and B are independent.
(d) A and B are mutually exclusive.
(e) None of the above.

16. Consider the following frequency histogram.



Which one of the following statements is false?

(a) The standard deviation is less than 5. *True. The standard deviation will be a little greater than the average distance from the mean, which is much less than 5. The standard deviation would be *much* less than 5. (Although this distribution is skewed, the empirical rule can also give us a rough indication of the value of the standard deviation.)*

(b) The IQR is less than 5. *$IQR = Q_3 - Q_1$. Very roughly, this would be approximately $102 - 100.5$, or about 1.5.*
(c) The mean is less than 105. *True.*
(d) Q_1 is less than the standard deviation. *False. Q_1 is a percentile, which is a value of the variable. *All* percentiles for this data are greater than 100. Is the standard deviation less than 100? Yes, yes it is.*
(e) The mean is greater than the median. *True.*

17. Suppose a single six-sided die is rolled once. Let A be the event that an even number is rolled. Let B be the event that the number rolled is a 4 or a 5. Let C be the event that the number is a 1 or a 5.

Consider the following 3 statements:

I. A and B are independent events.
II. A and C are independent events.
III. B and C are independent events.

This one is very similar to the example I used when I first brought up the notion of independence in class. It's also very similar to question 6 of the suggested exercises for probability.

$P(A) = \frac{3}{6}, P(B) = \frac{2}{6}, P(C) = \frac{2}{6}$.

If B occurs (the number is a 4 or a 5), what is the probability the number is even? $\frac{1}{2}$. So, $P(A|B) = \frac{1}{2}$. Since $P(A|B) = P(A)$, A and B are independent.

If B occurs (the number is a 4 or a 5), what is the probability the number is a 1 or a 5? $\frac{1}{2}$. So, $P(C|B) = \frac{1}{2}$. Since $P(C|B) \neq P(C)$, B and C are not independent.

If C occurs (the number is a 1 or a 5), what is the probability the number is even? $\frac{0}{2} = 0$. So, $P(A|C) = 0$. Since $P(A|C) \neq P(A)$, A and C are not independent.

Which of these statements are true?

(a) Just I. **
(b) Just II.
(c) Just III.
(d) All of the statements are true.
(e) All of the statements are false.

18. An important electronic system relies on a very fragile component that has a 0.43 probability of failure. Since this failure probability is very high, designers put in a total of 4 of these components, connected in parallel. The system will work as long as at least one of the four components works.

What is the probability the electronic system works? Assume independence between components.

I did a nearly identical question in class, and question 43 in the probability exercises is nearly identical.

$$\begin{aligned}
P(\text{The system works}) &= P(\text{At least one component works}) \\
&= 1 - P(\text{All 4 components fail}) \\
&= 1 - 0.43^4 \\
&= 0.965812
\end{aligned}$$

(a) 0.956
(b) 0.966**
(c) 0.976
(d) 0.986
(e) 0.996

19. In a certain area, 8% of a large population of raccoons have rabies. If 10 raccoons from this area are randomly selected, what is the probability at least 9 of them **do not** have rabies?

There are 2 valid approaches here:

1. Let X be the number of raccoons that do not have rabies. Then X is a binomial random variable with $n = 10$ and $p = 0.92$, and we need to find:

$$\begin{aligned} P(X \geq 9) &= P(X = 9) + P(X = 10) \\ &= \binom{10}{9} 0.92^9 0.08^1 + \binom{10}{10} 0.92^{10} 0.08^0 \\ &= 0.3777291 + 0.4343885 \\ &= 0.8121 \end{aligned}$$

2. Alternatively, let X be the number of raccoons that have rabies. Then X is a binomial random variable with $n = 10$ and $p = 0.08$, and we need to find:

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{10}{0} 0.08^0 0.92^{10} + \binom{10}{1} 0.08^1 0.92^9 \\ &= 0.4343885 + 0.3777291 \\ &= 0.8121 \end{aligned}$$

(a) Less than 0.0001.
(b) 0.38
(c) 0.76
(d) 0.81**
(e) 0.92

20. In a certain area, 8% of a large population of racoons have rabies. Suppose we randomly sample raccoons from this area. What is the conditional probability that the first raccoon with rabies occurs on the 12th raccoon sampled, given that the first 10 raccoons sampled do not have rabies?

There are different approaches here. The easiest way:

If the first 10 raccoons do not have rabies, then all we need is for the 11th to not have rabies, and the 12th to have it. The probability of this is just $0.92 \times 0.08 = 0.0736$.

Harder way:

$$
\begin{aligned}
P(X = 12|X > 10) &= \frac{P(X = 12 \cap X > 10)}{P(X > 10)} \\
&= \frac{P(X = 12)}{P(X > 10)} \\
&= \frac{(1 - .08)^{11} 0.08}{(1 - 0.08)^{10}} \\
&= 0.92 \times 0.08 = 0.0736.
\end{aligned}
$$

(a) 0.03
(b) 0.04
(c) 0.06
(d) 0.07**
(e) 0.08

Sample Variance: $s^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1}$. Equivalent alternative formula: $s^2 = \dfrac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$

Sample z-score: $z = \dfrac{x - \bar{x}}{s}$

If we transform the data using the linear transformation $x^* = a + bx$:
$\bar{x}^* = a + b\bar{x}$, $s_{x^*} = |b|s_x$, $s^2_{x^*} = b^2 s^2_x$

Probability

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

$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

The conditional probability of A, given B has occurred is $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

Two events A and B are independent if and only if:

$P(A|B) = P(A), P(B|A) = P(B), P(A \cap B) = P(A)P(B)$.

Expected Value and Variance of a Discrete Random Variable

$E(X) = \mu = \sum xp(x)$, $\sigma^2 = E[(X-\mu)^2] = \sum (x-\mu)^2 p(x)$.
Alternative method: $E[(X-\mu)^2] = E(X^2) - [E(X)]^2$.

Properties of Expectation and Variance

$E(a + bX) = a + bE(X)$, $\sigma^2_{a+bX} = b^2\sigma^2_X$, $\sigma_{a+bX} = |b|\sigma_X$

If X and Y are two random variables then $E(X+Y) = E(X) + E(Y)$ and
$E(X-Y) = E(X) - E(Y)$.

If X and Y are independent: $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$ and $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$

Discrete Probablity Distributions

Binomial distribution: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. $\mu = np, \sigma^2 = np(1-p)$.

Hypergeometric distribution: $P(X = x) = \dfrac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$. $\mu = n\frac{a}{N}$.

Poisson distribution: $P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}, \lambda = \mu = \sigma^2$.

Geometric distribution: $P(X = x) = (1-p)^{x-1}p$. $\mu = \frac{1}{p}$, $\sigma^2 = \frac{1-p}{p^2}$.