STAT*2040 W15

Test 1 Solutions (White Version)

Examiner: Jeremy Balka

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

   (a) White *** (Not worth marks)
   (b) Yellow

2. Which one of the following statements is FALSE? (If A–D are all true, answer option E.)

   (a) A statistic is a numerical characteristic of a sample.
   (b) A parameter is a numerical characteristic of a population.
   (c) The value of a statistic can be negative.
   (d) The value of a statistic can be less than the value of the parameter it estimates.
   (e) None of the above. ***

3. Suppose that I wish to estimate the average weight of male students currently enrolled in STAT*2040. I ask the next 8 male students that come to see me during my office hours if they would step on a scale and have their weight measured. Six of these students consent to having their weight measured, and their average weight is 77.0 kg. Suppose that, unknown to me, the average weight of all male students currently enrolled in STAT*2040 is 74.2 kg.

   Which one of the following statements is FALSE? (If A–D are all true, answer option E.)

   (a) The population consists of all male students currently enrolled in STAT*2040.
   (b) 74.2 is the value of a parameter.
   (c) 77.0 is the value of a statistic.
   (d) The sample of 6 students is a simple random sample from the population of interest. ***
   (e) None of the above.

4. Suppose that a sample of 5 observations has a mean of 10 and a standard deviation of 20. If each of these 5 observations is multiplied by -4 and then 100 is added, what are the mean and standard deviation of the 5 transformed values?

   This is a basic linear transformation. If we let x represent the original observations and x^* represent the transformed values, then

   $$x^* = 100 - 4x$$

   $$\bar{x}^* = 100 - 4\bar{x} = 100 - 4 \cdot 10 = 60$$

   Recall that an additive constant does not affect any measure of variability. If we multiply the observations by a constant, the standard deviation gets multiplied by the absolute value of that constant:

   $$s_{x^*} = |-4| s_x = 4 \cdot 20 = 80$$

   (a) The transformed values have a mean of 60 and a standard deviation of 20.
   (b) The transformed values have a mean of 60 and a standard deviation of 80. ***

(c) The transformed values have a mean of 60 and a standard deviation of 180.
(d) The transformed values have a mean of 60 and a standard deviation of 320.
(e) The transformed values have a mean of 140 and a standard deviation of 20.

5. A Facebook friend of a relative of mine recently shared a link to web page entitled "Studies Prove Without Doubt That Unvaccinated Children Are Far Healthier Than Their Vaccinated Peers." As part of the evidence in support of this statement, the web page cites a 1992 study by the Immunization Awareness Society (IAS), in which a survey was sent out to their members. The survey was voluntary and involved questions about the health of children. A total of 245 surveys were returned (completed by members of the IAS and their friends and associates), involving 226 vaccinated children and 269 unvaccinated children. The survey results indicated that unvaccinated children had far fewer health problems.

   Regardless of one's beliefs about vaccination, it is not reasonable to read this study and draw the conclusion that unvaccinated children are far healthier. There are many reasons why it is not reasonable to reach this conclusion, including 3 of the following statements. Which one of the following is *not* a reason why this is not a reasonable conclusion?

   (a) Respondents with healthy unvaccinated children might be more likely to complete the survey.
   (b) Respondents with unvaccinated children might be more likely to exaggerate the health of these children.
   (c) The survey was completed only by members of the IAS and some of their friends and associates, which may bias the results.
   (d) The sample sizes are unequal, so no statistical inference procedures can be carried out. *** *We don't need equal sample sizes to carry out inference procedures.*

6. A sample of 4 mature green sea urchins had their weights measured (in grams). The results are illustrated in the following stemplot:

```
The decimal point is at the |

21 | 12
22 | 89
```

   (If you have done the required reading on stem plots, you should be able to determine all of the values from the plot. As a small hint, the smallest value is 21.1 grams.)

   What is the standard deviation of the weights of the 4 sea urchins? (Choose the closest value.)

   The 4 values are 21.1, 21.2, 22.8, and 22.9. These 4 values have a standard deviation of 0.98.

   (a) 0.82 grams.
   (b) 0.98 grams. ****
   (c) 1.12 grams.
   (d) 1.25 grams.
   (e) 1.38 grams.

7. Suppose we need to make up a sample data set of four numbers that lie between 100 and 200 (inclusive, and repeats are allowed). For example, we could pick 137, 137, 141, 200, or 100, 100, 100, 187.

   Which one of the following statements is TRUE?

   (a) The 4 values 100,150,150,200 would have the greatest possible standard deviation. *False. The greatest standard deviation is achieved when the values are 100,100,200,200.*
   (b) The 4 values 200,200,200,200 would have the greatest possible range. *False. These values have the smallest possible range.*
   (c) If the median of the sample of 4 values is 200, the mean must also equal 200. *False. Counterexample: 0,200,200,200 has a median of 200 and a mean of 150.*
   (d) The sample cannot possibly have both the greatest possible median and greatest possible standard deviation.**** *True. The greatest possible median can only be achieved when at least 3 of the values are 200. The greatest standard deviation is achieved when the values are 100,100,200,200. They cannot both occur.*
   (e) None of the above.

8. Samples of 40 mature male and 38 mature female green sea urchins had their heights measured. The results are illustrated in Figure 1.

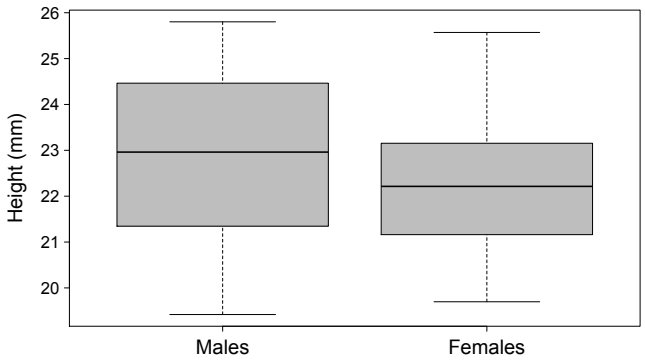


Figure 1: Heights (mm) of samples of mature green sea urchins.

   Which one of the following statements is TRUE?

   (a) The value of the interquartile range (IQR) for males is less than the minimum for females.**** *True. The IQR for males is approximately 3.2 mm, and the minimum for females is approximately 19.5 mm.*
   (b) The third quartile (Q_3) for males is less than the third quartile for females. *False.*
   (c) Each boxplot shows 2 outliers. *False. There are no outliers.*
   (d) The standard deviation of male height and the standard deviation of female height are both greater than 5 mm. *False. Males and females both have standard deviations that are less than 2 mm.*
   (e) Both distributions show strong left skewness. *False. The distributions are roughly symmetric.*

9. Suppose that green sea urchins are, to a reasonable approximation, randomly and independently distributed on a seabed at an average rate of 2.5 per 100 square metres. What is the probability that a randomly selected 50 square metre portion of the seabed has at least 2 green sea urchins? (Choose the closest value.)

   The distribution of sea urchins over a given area is going to have (approximately) a Poisson distribution. (Given that they are distributed randomly and independently.) The mean number per 100 square metres is 2.5, so the mean number per 50 square metres is 1.25. If we let X represent the number of green sea urchins in a randomly selected 50 square metre area, then X has a Poisson distribution with $\lambda = 1.25$.

   $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - [\frac{1.25^0 e^{-1.25}}{0!} + \frac{1.25^1 e^{-1.25}}{1!}] = 0.3554.$

   (a) 0.27
   (b) 0.36 ***
   (c) 0.41
   (d) 0.71
   (e) 0.87

10. In a certain area, the body weight of mature female green sea urchins has a mean of 21.7 grams and a standard deviation of 6.4 grams. If 10 mature female green sea urchins are randomly and independently selected from this area, what is the standard deviation of their total weight? (Choose the closest value.)

    The variance of the sum of independent random variables is the sum of their individual variances. $Var(X_1 + X_2 + \ldots + X_{10}) = Var(X_1) + Var(X_2) + \ldots Var(X_{10}) = 6.4^2 \cdot 10 = 409.6$. *The standard deviation is the square root of the variance:* $\sqrt{409.6} = 20.24$.

    (a) 12.2 grams
    (b) 20.2 grams.***
    (c) 32.0 grams.
    (d) 57.0 grams.
    (e) 64.0 grams.

11. A bucket contains 8 immature sea urchins and 24 mature sea urchins. If 6 sea urchins are randomly selected without replacement from this bucket, what is the probability that exactly 2 immature sea urchins are selected? (Choose the closest value.)

    Let X be the number of immature sea urchins selected. X has the hypergeometric distribution:

    $P(X = 2) = \frac{\binom{8}{2}\binom{24}{4}}{\binom{32}{6}} = 0.3283.$

    (a) 0.27
    (b) 0.30
    (c) 0.33***
    (d) 0.36
    (e) 0.39

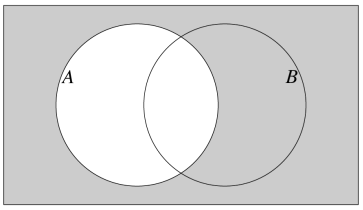
12. Two balanced six-sided dice are rolled. What is the probability that the two numbers that come up sum to 4?

    There are $6 \times 6 = 36$ *equally likely outcomes when 2 dice are rolled. 3 of these ((1,3), (3,1) and (2,2)) have a sum of 4.*

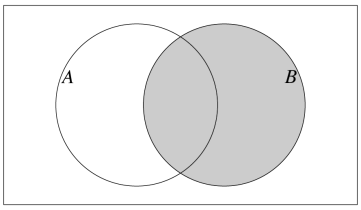
    (a) $\frac{2}{36}$
    (b) $\frac{3}{36}$ ****
    (c) $\frac{4}{36}$
    (d) $\frac{5}{36}$
    (e) $\frac{6}{36}$

13. Suppose $P(A) = 0.30$, $P(B) = 0.80$, and $P(A \cap B) = 0.11$. What is $P(A^c \cup B)$?

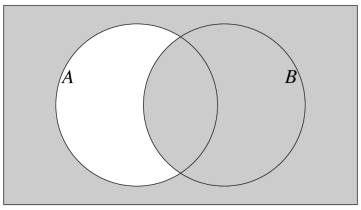
    It's best to draw Venn diagrams for this one.



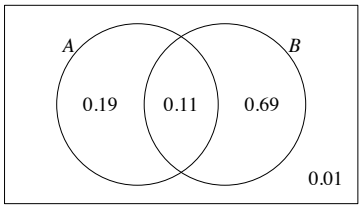
(a) A^c



(b) B



(c) $A^c \cup B$



(d) The probability of the shaded area is: $P(A^c \cup B) = 0.11 + 0.69 + 0.01 = 0.81$.

    (a) 0.79
    (b) 0.81 ***
    (c) 0.91
    (d) 0.99
    (e) 1.50

14. Which one of the following statements is FALSE? (If A–D are all true, choose option E.)

    (a) If $P(A) = 0.6$ and $P(B) = 0.7$, then $0.3 \leq P(A \cap B) \leq 0.6$. *True. The probability of the intersection cannot be greater than the lesser of* $P(A)$ *and* $P(B)$ *(0.6). And since* $P(A \cup B) \leq 1$*, the probability of the intersection cannot be less than 0.6 + 0.7 - 1 = 0.3.*
    (b) If $P(A|B) = 0$, then A and B are mutually exclusive. *True.* $P(A|B) = 0 \implies P(A \cap B) = 0 \implies$ *A and B are mutually exclusive.*
    (c) For any event A, A and its complement are mutually exclusive. *True. By definition,* A^c *is the set of all outcomes that are not in A.*

(d) If $P(A) = 0.4$, $P(B) = 0.4$, and $P(A \cup B) = 0.64$, then A and B are independent. *True. The given probabilities imply that $P(A \cap B) = 0.16$. Since $P(A \cap B) = P(A) \cdot P(B)$, A and B are independent.*

(e) None of the above. ****

15. Suppose we randomly select a Canadian province. Let A be the event that the province is Newfoundland. Let B be the event that the province is east of Ontario. (If you don't know where the province of Newfoundland is located in Canada, please ask a TA.) Which one of the following statements is TRUE?

Since Newfoundland lies to the east of Ontario, $P(B|A) = 1$.

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

16. Consider the following probability distribution of a random variable X. (One of the probabilities has been replaced by a question mark.)

x	100	200	300	400
$p(x)$	0.1	0.2	0.5	?

What is $P(X > 231.7|X < 378.4)$?

$P(X > 231.7|X < 378.4) = \frac{P(X>231.7 \cap X<378.4)}{P(X<378.4)} = \frac{P(X=300)}{P(X=100)+P(X=200+P(X=300)} = \frac{0.5}{0.1+0.2+0.5} = \frac{5}{8}$.

(a) 0.535
(b) 0.565
(c) 0.625 ***
(d) 0.80
(e) 0.875

17. Consider the following probability distribution of a random variable X.

x	20	40	80
$p(x)$	0.1	0.2	0.7

What is $E(\sqrt{X})$? (Choose the closest value.)

$E(\sqrt{X}) = \sum \sqrt{x}p(x) = \sqrt{20} \cdot 0.1 + \sqrt{40} \cdot 0.2 + \sqrt{80} \cdot 0.7 = 7.973$.

(a) 7.9
(b) 8.0 ***
(c) 8.1
(d) 8.2
(e) 8.3

18. Approximately 56% of Canadian males between the ages of 20 and 39 have a body mass index (BMI) that is less than 30.0 kg/m^2. If 12 Canadian males in this age group are randomly selected, what is the probability that no more than 10 have a BMI that is less than 30? (Choose the closest value.)

    Let X represent the number of sampled males that have a BMI less than 30. Then X has a binomial distribution with $n = 12$ and $p = 0.56$.

    $P(X \leq 10) = 1 - P(X \geq 11) = 1 - [P(X = 11) + P(X = 12)] = 1 - (0.008968 + 0.000951) = 0.990.$

    (a) 0.039
    (b) 0.961
    (c) 0.985
    (d) 0.990 ***
    (e) 0.999

19. Which one of the following statements is TRUE?

    (a) For any given n, the variance of a binomial random variable is greatest when $p = 1$. *False.* $\sigma^2 = np(1-p) = 0$ *if* $p = 1$.
    (b) The mean of a binomial random variable can be negative. *False. A binomial random variable is a count of the number of successes in n trials, and as such it cannot take on negative values and it cannot have a negative mean.*
    (c) If X is a binomial random variable, then $P(X = 1) > P(X = 0)$. *False.* $P(X = 1)$ *can be greater than, less than, or equal to* $P(X = 0)$.
    (d) The mean of a binomial random variable X can be less than the most likely value of X. *** *True. This will typically be the case when there is some left skewness. For example, if* $n = 10$ *and* 0.79*, then the most likely value of X is 8 and the mean is 7.9. Or consider a very simple example with* $n = 1$ *and* $p = 0.6$*. Here the most likely value of X is 1, and the mean is 0.6.*
    (e) None of the above.

20. Suppose researchers wish to investigate the effects of different levels of zinc and two types of diet on the growth of fathead minnow larvae. They design an experiment in which 200 fathead minnow larvae (of similar size) are randomly assigned to two groups (100 to each group). Group A receives a zinc level of 500 ppm in their water, is fed Diet 1, and is raised in Tank 1. Group B receives a zinc level of 1500 ppm in their water, is fed Diet 2, and is raised in Tank 2. After several weeks, the weight of the minnows is measured.

    Of the following options, which is the biggest problem with setting up the experiment in this way?

    (a) The levels of zinc, the diets, and the tanks are confounded. *** *We say that these 3 explanatory variables are confounded because it is impossible to separate their effects on the response variable. If there are observed differences between the groups, it will be impossible to say whether it was due to the zinc, the diets, the tanks, or some combination or interaction of these effects.*
    (b) There are many lurking variables that were not taken into account.
    (c) The minnow larvae were not all identical at the start of the experiment.
    (d) It would have been best to carry out this investigation with an observational study rather than an experiment.
    (e) The researchers should have used more than 2 diets.

21. Tantius et al. (2014) investigated tensile properties of human umbilical cords. In one part of the study, 23 human umbilical cords were stretched until they broke, and their elongation (% increase in length) at the breaking point was recorded. The elongation percentages are illustrated in Figure 2.

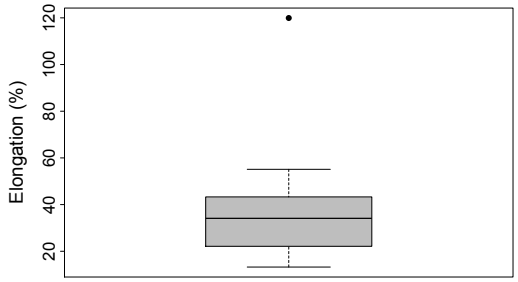


Figure 2: Boxplot of elongation percentage at breaking point.

Which one of the following statements is FALSE?

(a) The distribution is approximately symmetric, other than the one large outlier.
(b) If the outlier were removed from the calculations, the mean would decrease.
(c) If the outlier were removed from the calculations, the standard deviation would decrease.
(d) If the outlier's value was changed to 80 instead of 120, the variance would decrease.
(e) If the outlier's value was changed to 80 instead of 120, the median would decrease. *** *The outlier would still be the largest value. Since the median is the middle value in the data set (or the average of the two middle values), it would not change if the largest value was decreased a bit.*

22. The following table is based on a poll in the U.S. in which 1000 randomly selected adults were asked if they approved or disapproved of the way Barack Obama is handling his job as president.

	Republicans	Democrats	Independents
Approve	60	361	139
Disapprove	270	49	121

If a person is randomly selected from this group of 1000, what is the probability they say they approve, given they are a Democrat? (Choose the closest value.)

$P(\text{Approve}|\text{Democrat}) = \frac{P(\text{Approve} \cap \text{Democrat})}{P(\text{Democrat})} = \frac{361/1000}{410/1000} = \frac{361}{410}$.

(a) 0.41
(b) 0.56
(c) 0.76
(d) 0.88 ***
(e) 0.92