STAT*2040 W13

Test 1 (Solutions to the White Version)

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

   White

   (a) White
   (b) Yellow

2. A study surveyed 2,000 postmenopausal women. The women were asked several questions, including: "Do you practice hormone replacement therapy?" and "Do you have heart disease?" It was found that women who practiced hormone replacement therapy were found to have a lower risk of heart disease.

   Consider the following statements.

   I. This is an experiment, not an observational study. *False. It's a survey (a type of observational study).*
   II. The response variable is post-menopausal women. *False.*
   III. This type of study can yield very strong evidence of a causal link between heart disease and hormone replacement therapy in post-menopausal women. *False. Surveys do not yield very strong evidence of a causal link.*

   Which of these statements are true?

   (a) Just I.
   (b) I and III.
   (c) II and III.
   (d) All of them.
   (e) None of them.**

3. Which one of the following statements is true? (Assume for the purposes of this question that the events do not have probability exactly equal to 0 or 1.)

   (a) Mutually exclusive events can be independent. *False.*
   (b) The probability of the intersection of two events can be greater than the probability of their union. *False.*
   (c) If $P(A|B) = 1$ then $P(A \cap B) = 1$ *False.*
   (d) The sample mean is always greater than the sample variance. *False.*
   (e) Percentiles can be negative.** *True. The 75th percentile, for example, is the *value of the variable* that has 75% of the observations to the left. Since variables can possibly take on negative values, percentiles can be negative.*

4. Consider the following sample of 4 observations: $-2, 4, 6, 10$. What is the value of the sample variance? (Choose the closest value.)

   $s^2 = 25$

   (a) 3

(b) 5
(c) 25**
(d) 35
(e) 53

5. Which one of the following statements is true?

   (a) A parameter is a numerical characteristic of a sample. *False. A parameter is a numerical characteristic of a population.*
   (b) A statistic is a numerical characteristic of a population. *False. A statistic is a numerical characteristic of a sample.*
   (c) A statistic is usually exactly equal to the parameter is estimates. *False. Statistics have sampling distributions, vary from sample to sample, and are not usually exactly equal to the parameter they estimate.*
   (d) The value of a statistic stays constant from sample to sample. *False. Statistics vary from sample to sample.*
   (e) None of the above. **

6. Suppose a sample data set has a perfectly symmetric distribution with a mean of 0 and a standard deviation of 1. If each of the observations is multiplied by 18, then 1 is added, what are the values of the mean, median, and standard deviation of the new observations?

   Let x represent the original variable. Then multiplying by 18 and adding 1 is the linear transformation: $x^* = 1 + 18x$. The new mean will be $\bar{x}^* = 1 + 18 \times 0 = 1$, and the new standard deviation will be $18 \times 1 = 18$ (the additive constant does not affect measures of variability). Since the distribution is perfectly symmetric, the mean and median are equal.

   (a) All 3 quantities are 18.
   (b) All 3 quantities are 1.
   (c) The mean and median are both 1, and the standard deviation is 18. **
   (d) The mean and median are 1, and the standard deviation is 19.
   (e) The mean is 1, the standard deviation is 18, but it is impossible to determine the median.

7. Which one of the following statements is true?

   (a) The standard deviation can be negative. *False.*
   (b) The standard deviation can be greater than the variance.** *True. s will be greater than s^2 when $0 < s < 1$.*
   (c) The standard deviation is always greater than Q_1. *False. s is a measure of variability, whereas Q_1 is a measure of position. There is no relationship between them.*
   (d) The units of the standard deviation are the square of the units of the original variable. *False. The standard deviation has the same units as the original variable.*
   (e) The standard deviation cannot be less than 1. *False.*

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

   (a) In simple random sampling from a finite population, each member of the population has the same chance of being selected in the sample.

(b) In simple random sampling from a finite population, each possible sample of size n has the same chance of being selected.
(c) Simple random sampling is always done *with* replacement. ** *False. Simple random sampling is usually done without replacement.*
(d) Stratified random sampling is a type of sampling that has some advantages over simple random sampling in some situations.
(e) None of the above.

9. Suppose A and B are two events such that $P(A) = 0.4$, $P(B) = 0.3$ and $P(A \cup B) = 0.65$. What is $P(A \cap B^c)$? (Choose the closest value.)

   Drawing a Venn diagram may help to illustrate that $A \cap B^c$ is "just" A (A but not B).

   $P(A \cup B) = P(A) + P(B) - P(A \cap B) \implies 0.65 = 0.4 + 0.3 - P(A \cap B) \implies P(A \cap B) = 0.05.$

   $P(A \cap B^c) = 0.4 - .05 = 0.35.$

   (a) 0.28
   (b) 0.30
   (c) 0.35**
   (d) 0.40
   (e) 0.95

10. Let A be the event that a randomly selected North American is a Canadian citizen. Let B be the event that a randomly selected North American is currently in Newfoundland. Which one of the following statements is true?

    (a) $P(A|B) < P(A)$. *False. The information that the person is in Newfoundland makes it *much* more likely they are Canadian, so P(A|B) > P(A).*
    (b) $P(B|A) < 0.5$. ** *True. Far less than half of Canadian citizens are currently in Newfoundland.*
    (c) A and B are mutually exclusive. *False. It's possible to be both a Canadian Citizen and be in Newfoundland.*
    (d) A and B are independent. *False.*
    (e) $P(B|A) = P(A \cap B)$. *False. This would only be true if $P(A) = 1$.*

11. Consider the boxplots given in Figure 1. Which one of the following statements is true?

    (a) Sample A shows some signs of left skewness. *False. It has some right-skewness.*
    (b) Sample B is has a standard deviation that is greater than 3. *False. The range is a little less than 3, and the standard deviation would be much less.*
    (c) The IQR for Sample B is greater than 2. *False. The IQR would be a little less than 1.*
    (d) There are 4 outliers in the plot, plus 1 extreme outlier. *False. There is only 1 outlier.*
    (e) For Sample A, $Q_3 + 1.5 \times$ IQR is less than 5. ** *True. $Q_3 + 1.5 \times IQR$ will be less than the outlier.*

12. Is there a relationship between fatty fish consumption and the rate of prostate cancer? A study followed 6272 Swedish men for 30 years. They were categorized according to their fish consumption, and to whether they developed prostate cancer. The following table summarizes the results.

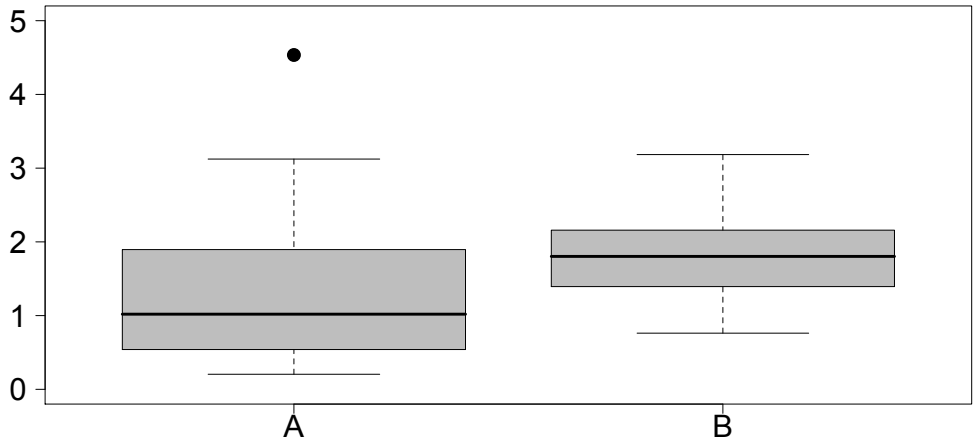

Figure 1: Two boxplots.

	Fish consumption			
	Never/seldom	Small	Moderate	Large
Prostate cancer	14	201	209	42
No prostate cancer	110	2420	2769	507

If one of these 6272 men is randomly selected, what is the probability their fish consumption was large or moderate, given they got prostate cancer?

$\frac{209+42}{14+201+209+42} \approx 0.5386266$

(a) 0.52
(b) 0.54**
(c) 0.56
(d) 0.58
(e) 0.60

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

x	12	14	16	18
$p(x)$	?	0.20	0.20	0.20

One of the probabilities in the table is missing. Find the missing probability, then calculate the mean of X. What is the mean of X?

Since the probabilities must sum to 1, the missing probability must be 0.40. $E(X) = 12 \cdot 0.40 + 14 \cdot 0.20 + 16 \cdot 0.20 + 18 \cdot 0.20 = 14.4$.

(a) 14.4**
(b) 14.6
(c) 14.8

(d) 15.0
(e) 15.2

14. The adult literacy rate in Gabon is approximately 88%. If 15 adults in Gabon are randomly selected, what is the probability exactly 13 are literate? (Choose the closest value.)

    This is a binomial problem with $n = 15$, $p = 0.88$. $P(X = 13) \approx 0.287$

    (a) 0.21
    (b) 0.23
    (c) 0.25
    (d) 0.27
    (e) 0.29**

15. The adult literacy rate in Gabon is approximately 88%. If 10 adults in Gabon are randomly selected, what is the probability at least 2 are not literate? (Choose the closest value.)

    There are a few ways of working through this one. One way is to let X be the number of people who are not literate. We need to find $P(X \geq 2)$, where X has a binomial distribution with $n = 10$, $p = 0.12$. $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

    (a) 0.12
    (b) 0.17
    (c) 0.34**
    (d) 0.42
    (e) 0.62

16. Which one of the following statements is true?

    (a) If $P(A|B) = 0$, A and B are independent. *False.*
    (b) If A and B are independent, $P(A \cup B) = P(A) + P(B)$. *False.* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    (c) If A and B are mutually exclusive, $P(A \cap B) = P(A) + P(B)$. *False. If A and B are mutually exclusive,* $P(A \cap B) = 0$.
    (d) If $P(A|B) = 1$, then $P(B|A) = 1$. *False. Counterexample: Roll a die. Let A be the even numbers, B be the number 2.*
    (e) If $P(A) = P(B)$, then $P(A|B) = P(B|A)$.** *True, since* $P(A|B) = \frac{P(A \cap B)}{P(B)}$ *and* $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

17. John has a crush on Stephanie. A friend of John's is having a party on Friday night. If Stephanie goes to the party, the probability that John will also go is 0.98. If Stephanie does not go to the party, the probability that John goes is 0.06. If the probability that Stephanie does not go to the party is 0.30, what is the probability that John goes? (Choose the closest value.)

    It may be best to draw a tree diagram for this type of problem. $P(\textit{John goes}) = 0.70 \times 0.98 + 0.30 \times 0.06 = 0.704$.

    (a) 0.06
    (b) 0.69
    (c) 0.70**
    (d) 0.76

(e) 0.82

18. In a group of 25 deer mice, 5 carry hantavirus. If 3 of these 25 mice are randomly selected without replacement, what is the probability exactly 1 carries hantavirus? (Choose the closest value.)

    Basic hypergeometric problem. We are picking 3 mice from 25. From the 5 with hantavirus we must pick exactly 1. From the 20 without hantavirus we must pick exactly 2. Hence the correct probability is $\frac{\binom{5}{1}\binom{20}{2}}{\binom{25}{3}} = 0.413$

    (a) 0.28
    (b) 0.34
    (c) 0.38
    (d) 0.41**
    (e) 0.44

19. A certain genetic defect affects 0.1% of the population. A test is available for this defect. If someone has the defect, the test will fail to detect it with probability 0.02. If somebody does not have the defect, the test will give a false positive with probability 0.05. Given a randomly selected person tests positive for this defect, what is the probability they have the defect? (Choose the closest value.)

    It's probability best to draw a tree diagram here. $P(\text{Defect}|\text{+ve test}) = \frac{P(\text{Defect}\cap\text{+ve test})}{P(\text{+ve test})} = \frac{0.001\times0.98}{0.001\times0.98+0.999\times0.05} = 0.0192$

    (a) 0.01
    (b) 0.02**
    (c) 0.03
    (d) 0.04
    (e) 0.98

20. Of the following 5 variables, which one would be best modelled by a Poisson distribution?

    (a) The amount of water in a randomly selected bottle of water. *False. The amount of water would have a continuous distribution. Most definitely not Poisson.*
    (b) The weight of a randomly selected University of Guelph Student. *False. Student weights would have a continuous distribution. Most definitely not Poisson.*
    (c) The number of red cards when 2 cards are randomly drawn with replacement from a standard deck. *False. It would have a binomial distribution.*
    (d) The number of moose-car collisions in Northern Ontario on a randomly selected day in June.** *Best answer. The Poisson distribution would likely provide a good approximation. At least to a good approximation, it is reasonable to think that these events are occurring randomly and independently of one another.*
    (e) The number of cars in a randomly selected 100 square metre portion of the University of Guelph campus on a Wednesday at 2:15 pm. *False. The probability of getting 0 cars would be very high (when we miss a parking lot), and the probability of getting a large number of cars would be high (when we hit a parking lot). It wouldn't have a Poisson distribution, for similar reasons to the "clumped" plant example given in class.*

21. Which one of the following statements is true?

(a) The mean of a binomial random variable cannot be less than 1. *False. The mean is np, and this can be less than 1.*

(b) The mean of a Poisson random variable cannot be less than 1. *False. The mean can be less than 1.*

(c) The variance of a binomial random variable can be less than its mean.** *True. The variance is $np(1-p)$ and the mean is np. Unless p is 0, the variance will be less than the mean.*

(d) A binomial random variable can take on negative values. *False. A binomial random variable is a count of the number of successes. The count cannot be negative.*

(e) The expected value of a binomial random variable must be equal to one of its possible values. *False. Counterexample: Toss a fair coin once and let X represent the number of heads. Here the possible values are 0 and 1, and the mean is 0.5.*