University of Guelph
Department of Mathematics and Statistics

STAT*2040
Statistics I

Test 1 (White version)

Answers and Solutions
February 5 2016

Examiner: Jeremy Balka

This exam is 70 minutes in duration

Name:

ID:

Signature:

Please read the instructions:

1. Fill out your name and ID number above.

2. When the examination starts, make sure your question paper is complete. You should have 19 multiple choice questions, along with a formula sheet. The first question is just a bookkeeping question, and does not count for marks, but please fill it in to ensure your exam is properly graded.

3. Do all rough work on this paper.

4. You are allowed to bring in a calculator, and pens and pencils.

5. There is only **one** correct answer for each question. Fill in only one bubble for each question.

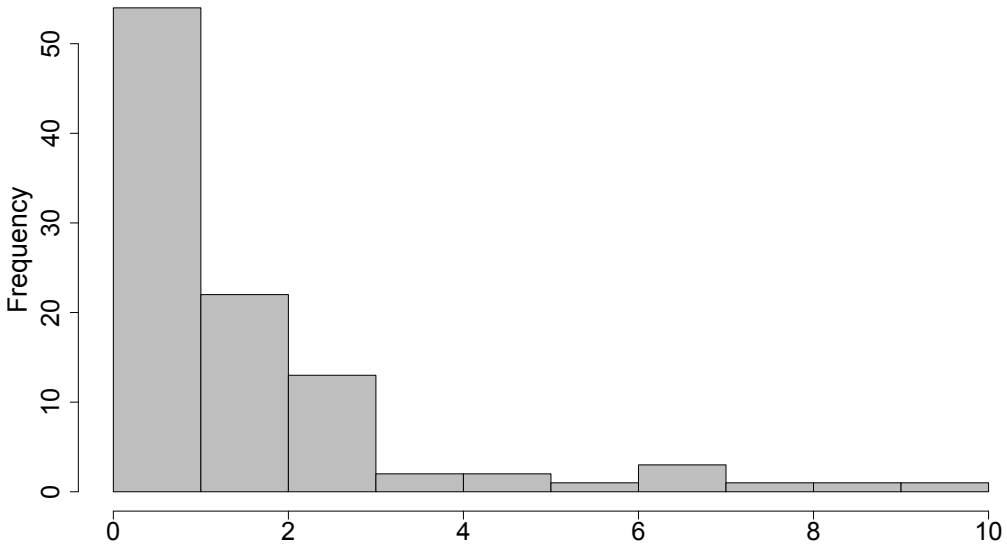
6. Fill out the computer answer sheet in pencil as you go. *There will be no extra time given at the end of the exam to fill in the sheet.*

7. The answers given in the exam are often rounded versions of the correct answer. Choose the closest value.

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

   (a) White **
   (b) Yellow

2. Consider the following frequency histogram, which illustrates a sample of $n = 100$ observations.



   Which one of the following statements is true?

   (a) The distribution is left skewed. *False. This distribution is right skewed.*
   (b) The median is greater than 1. *False. There are 100 observations, and more than 50 lie between 0 and 1. This implies that the median lies between 0 and 1.*
   (c) The mean would be less than the median. *False. For a right skewed distribution like this, the mean will be greater than the median.*
   (d) Q_3 is greater than 3. *False. More than 80 observations are less than 3, and so Q_3 (the value of the variable with 75% of the observations to the left) must be less than 3.*
   (e) None of the above. **

3. Which one of the following statements is TRUE?

   (a) The sample mean can never be negative.
   (b) The sample variance can never be negative. **
   (c) A statistic is a numerical characteristic of a population.
   (d) A parameter is a numerical characteristic of a sample.
   (e) None of the above.

4. Suppose that a professor wants to draw a sample from their class of 80 students. For each of the 80

students in the class, the professor flips a coin. If the coin lands on heads, the student is included in the sample. If the coin lands on tails, the student is left out of the sample.

Consider the following statements:

I. Each student in the class has the same chance of being included in the sample. *True.*

II. This sampling method is called stratified random sampling. *False.*

III. This sampling method will result in a simple random sample of size $n = 40$. *False. The sample size is not fixed at 40. This method would result in a sample size of anywhere between 0 and 40.*

Which of these statements are true?

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

5. Researchers investigated possible differences in the total cholesterol levels in the blood of male and female students at a large university. Total cholesterol (mg/dl) was measured on 26 male and 22 female student volunteers, with the following results.

	Males	Females
Sample mean	171.4	173.8
Sample standard deviation	32.9	34.1
Sample size	26	22

Which one of the following statements is true?

(a) This is an experiment, and not an observational study. *False. The groups (males, females) were pre-existing. The researchers took measurements, but did not impose the conditions.*
(b) Since the sample sizes are unequal, no conclusions can be drawn from this study. *False. There is no need for the sample sizes to be equal.*
(c) The total cholesterol level is a lurking variable. *False. The total cholesterol level is the response variable, not an unmeasured lurking variable.*
(d) The 48 students are a simple random sample from the students of the university. *False. They are student volunteers, and not a simple random sample from the University.*
(e) None of the above.**

6. Which one of the following statements is true?

(a) If $P(A) = 0.40$ and $P(B) = 0.40$, then $P(A \cap B) = 0.16$. *False. This is only the case if A and B are independent.*

(b) If $P(A) = 0.40$, $P(B) = 0.40$, and $P(A \cup B) = 0.80$, then A and B are independent. *False. Here A and B are mutually exclusive and not independent.*

(c) If $P(A) = 0.50$, and the probability that B occurs given A occurs is 0.50, then $P(A \cap B) = 0.50$. *False. Here $P(B|A = \frac{P(A \cap B)}{P(A)}) = 0.50$, which implies $P(A \cap B) = 0.25$.*

(d) If $P(A) = 0$ and $P(B) = 0.25$, then A and B are independent.** *True. If $P(A) = 0$ then $P(A \cap B) = 0$ and $P(A|B) = 0$.*

(e) None of the above.

7. A sample of 4 lizards had their tail lengths measured (in cm). The results are illustrated in the following stemplot:

```
The decimal point is at the |

7 | 45
8 | 28
```

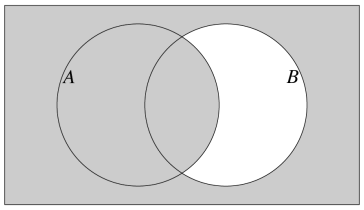
(If you have done the required reading on stem plots, you should be able to determine all of the values from the plot. As a small hint, the smallest value is 7.4 cm.)

The 4 values are 7.4, 7.5, 8.2, 8.8, which have a standard deviation of $s = 0.6551081$.

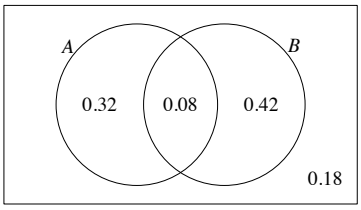
What is the standard deviation of the tail lengths of the 4 lizards? (Choose the closest value.)

(a) 0.55 cm
(b) 0.59 cm
(c) 0.66 cm.**
(d) 1.05 cm.
(e) 1.16 cm.

8. Suppose $P(A) = 0.40$, $P(B) = 0.50$, and $P(A \cap B) = 0.08$. What is $P(A \cup B^c)$?



(a) $A \cup B^c$



(b)

(a) 0.08
(b) 0.58**
(c) 0.72
(d) 0.82
(e) 0.88

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

x	10	40	180
$p(x)$	0.2	0.3	0.5

$$\begin{aligned} E(X) &= \sum xp(x) \\ &= 10 \cdot 0.2 + 40 \cdot 0.3 + 180 \cdot 0.5 \\ &= 104 \end{aligned}$$

$$\begin{aligned} E(X^2) &= \sum xp(x) \\ &= 10^2 \cdot 0.2 + 40^2 \cdot 0.3 + 180^2 \cdot 0.5 \\ &= 16700 \end{aligned}$$

$$\begin{aligned} \sigma^2 &= E(X^2) - [E(X)]^2 \\ &= 16700 - 104^2 \\ &= 5884 \end{aligned}$$

$$\sigma = \sqrt{5884} = 76.7$$

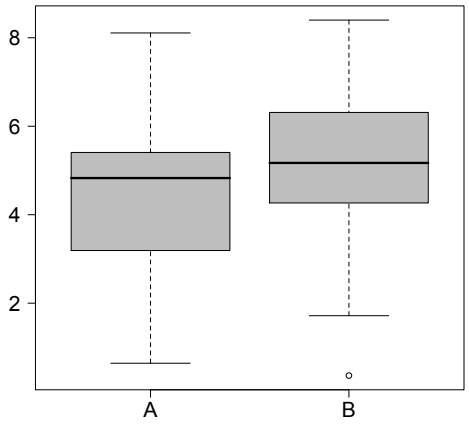
What is the value of the standard deviation of X? (Choose the closest value.)

(a) 21.7
(b) 31.6
(c) 58.2
(d) 76.7**
(e) 90.7

10. The following boxplots illustrate the distributions of samples A and B, where each sample has 40 observations.

Which one of the following statements is FALSE?

(a) The value of the interquartile range (IQR) for sample A is less than 4. *True*
(b) The first quartile (Q_1) for sample B is less than the median of sample A. *True*
(c) If the outlier from sample B were removed from the calculations, then the mean and standard deviation of sample B would increase.** *False. The mean would increase but the standard deviation would decrease.*



(d) The standard deviation of sample A is less than 10. *True.*

(e) Both distributions show strong right skewness.** *False.*

11. A certain type of surgery at a large hospital is always performed by one of two surgeons. Surgeon A is very careful and follows all protocols, while Surgeon B is lax and not very careful. Surgeon A performs 65% of this type of surgery, and there is a surgical site infection in 10% of their surgeries. Surgeon B performs 35% of this type of surgery, and there is a surgical site infection in 40% of their surgeries. Suppose we randomly select a surgery of this type at this hospital. Given the surgery results in a surgical site infection, what is the probability Surgeon B performed the surgery?

    This is a Bayes' Theorem problem. Most people find it easiest to solve these types of problems with a tree diagram (see Figure 1).

    Let B represent the event that Surgeon B performs the surgery, and I represent the event that the surgery results in a surgical site infection.

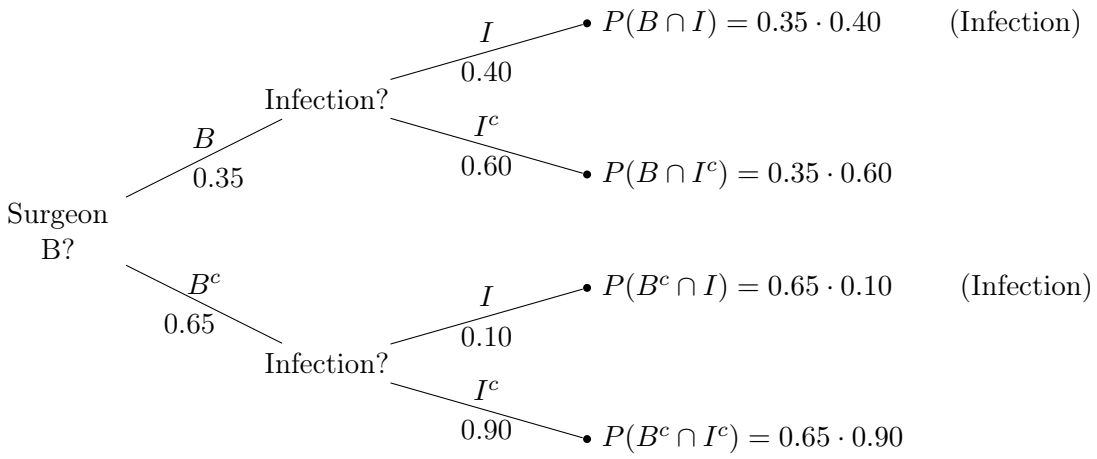


Figure 1: The tree diagram for Question 11.

We need to find $P(B|I) = \frac{P(B \cap I)}{P(I)} = \frac{0.35 \cdot 0.40}{0.35 \cdot 0.40 + 0.65 \cdot 0.10} = 0.6829$.

(a) 0.16
(b) 0.32
(c) 0.61
(d) 0.68**
(e) 0.72

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

x	2	4	6	8	10
$p(x)$	0.2	0.2	0.1	0.1	?

What is $P(X > 4.1 | X < 8.2)$? (Choose the closest value.)

$$\begin{aligned} P(X > 4.1 | X < 8.2) &= \frac{P(X > 4.1 \cap X < 8.2)}{P(X < 8.2)} \\ &= \frac{P(X = 6) + P(X = 8)}{P(X = 2) + P(X = 4) + P(X = 6) + P(X = 8)} \\ &= \frac{0.1 + 0.1}{0.2 + 0.2 + 0.1 + 0.1} \\ &= \frac{1}{3} \end{aligned}$$

(a) 0
(b) 0.16
(c) 0.20
(d) 0.33**
(e) 0.5

13. Suppose we have a sample data set of distance measurements, in metres. All of the observations are positive, and the observations are not all equal. Which one of the following statements is FALSE? (If A—D are all true, answer option E.)

(a) The standard deviation has units of metres.
(b) The mean has units of metres.
(c) The z-score of the smallest observation has units of metres.** *False. Z-scores are unitless.*
(d) The z-score of the smallest observation would be negative.
(e) The 10th percentile would be positive.

14. Many cities in the United States have buyback programs for handguns, in which the police department pays people to turn in guns. The guns are then destroyed. Is there a difference between the distribution of the size of guns turned in during buyback programs and the distribution of the size of guns used in homicides and suicides? A study investigated this question, using data from a gun buyback program and police records in Milwaukee. The results are illustrated in the following table.

Gun Calibre	Buybacks	Homicides	Suicides	Total
Small	719	75	40	834
Medium	182	202	72	456
Large	20	40	13	73
Other	20	52	0	72
Total	941	369	125	1435

Suppose one of these 1435 guns is randomly selected. Given the gun is of small or medium calibre, what is the probability that it was used in a homicide? (Choose the closest value.)

There are $834 + 456 = 1290$ small or medium calibre guns. Of these, $75 + 202 = 277$ were used in a homicide. So $P(\text{Homicide} \mid \text{Small or Medium}) = \frac{277}{1290} = 0.2147$.

(a) 0.21**
(b) 0.26
(c) 0.31
(d) 0.36
(e) 0.41

15. Suppose we are about to roll an ordinary six-sided die once. Let F be the event that we roll a 1 or a 2 ($F = \{1, 2\}$). F is independent of which one of the following events?

$P(F) = \frac{1}{3}$. We need to find an event such that $P(F|Event) = P(F) = \frac{1}{3}$.

(a) $A = \{5, 6\}$. *A and F share no sample points, so $P(F|A) = 0$ and A and F are not independent.*
(b) $B = \{2, 6\}$. *Of B's two sample points, the number 2 is shared with F, so $P(F|B) = \frac{1}{2} \neq P(F)$, and B and F are not independent.*
(c) $C = \{1, 2, 6\}$ *Of C's three sample points, the numbers 1 and 2 are shared with F, so $P(F|C) = \frac{2}{3} \neq P(F)$, and C and F are not independent.*
(d) $D = \{1, 3, 6\}$** *Of D's three sample points, the number 1 is shared with F, so $P(F|D) = \frac{1}{3} = P(F)$, and D and F are independent.*
(e) $E = \{1\}$ *E shares its only sample point (the number 1) with F, so $P(F|E) = 1 \neq P(F)$, and E and F are not independent.*

16. 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 129, 129, 132, 200, or 100, 100, 100, 143.

    Which one of the following statements is FALSE?

    (a) The 4 values 100,100,200,200 would have the greatest possible standard deviation. *True.*
    (b) The 4 values 200,200,200,200 would have the greatest possible mean. *True.*
    (c) If the median of the sample of 4 values is 200, the mean must also equal 200.** *False. The median will be 200 if 3 or more of the values are 200, and this does not necessarily lead to a mean of 200. For example, if the values are 100, 200, 200, 200 then the median is 200 but the mean is 175.*
    (d) The sample with the greatest possible mean has the smallest possible standard deviation. *True. The greatest possible mean is achieved if the values are 200,200,200,200, which leads to a standard deviation of 0.*
    (e) The smallest possible value of the median is greater than the smallest possible value of the variance. *True. The smallest possible median is 100 and the smallest possible variance is 0.*

17. Urn 1 contains 5 red balls and 3 black balls. Urn 2 contains 12 red balls and 6 black balls. Three balls are drawn without replacement from Urn 1, and three balls are drawn with replacement from Urn 2. What is the probability that none of the balls drawn are black? (Choose the closest value.)

    $P(\text{No black}) = P(\text{All red}) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{12}{18} \cdot \frac{12}{18} \cdot \frac{12}{18} = 0.0529.$

    (a) 0.032
    (b) 0.048
    (c) 0.053**
    (d) 0.058
    (e) 0.072

18. Suppose that a sample of 5 observations has a mean of 15 and a standard deviation of 8. If each of these 5 observations is multiplied by -2 and then 50 is added, what are the mean and standard deviation of the 5 transformed values?

    New mean $= 50 - 2 \times$ Old mean $= 50 - 2 \times 15 = 20$. The additive constant does not affect the variability:

    New standard deviation $= |-2| \times$ Old standard deviation $= 2 \times 8 = 16$.

    (a) The transformed values have a mean of 20 and a standard deviation of 16.**
    (b) The transformed values have a mean of 20 and a standard deviation of -16.
    (c) The transformed values have a mean of 20 and a standard deviation of 34.
    (d) The transformed values have a mean of -30 and a standard deviation of 34.
    (e) The transformed values have a mean of -30 and a standard deviation of 16.

19. Which one of the following statements is FALSE?

(a) A well-designed randomized experiment can give strong evidence of a causal relationship between the explanatory and response variables.
(b) Lurking variables are more of a concern in observational studies than in experiments.
(c) If two explanatory variables are said to be confounded, that means it is impossible to separate their effects on the response variable.
(d) Observational studies never include any sort of randomization. *False. Observational studies often involve drawing a random sample from a larger population.*
(e) None of the above.