STAT*2040 W15

Test 1 (White Version)

February 6 2015

This page is included to satisfy the U of G's privacy policy. Please fill out your first and last names on this page, but *not your ID number or any other information.*

Last Name:

First Name:

On the next page, please fill out your name and ID number.

University of Guelph
Department of Mathematics and Statistics

STAT*2040
Statistics I

Test 1 (White version)
February 6 2015

Examiner: Jeremy Balka

This exam is 80 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 22 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. 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?

   (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.

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.)

   (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.
   (b) The 4 values 200,200,200,200 would have the greatest possible range.
   (c) If the median of the sample of 4 values is 200, the mean must also equal 200.
   (d) The sample cannot possibly have both the greatest possible median and greatest possible standard deviation.
   (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.

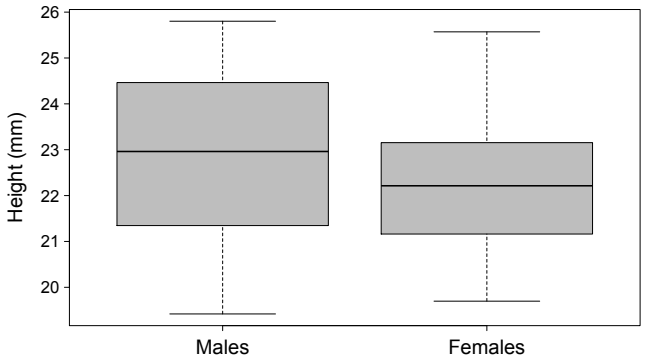

Figure 1: Heights (mm) of samples of mature green sea urchins.

   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.
   (b) The third quartile (Q_3) for males is less than the third quartile for females.
   (c) Each boxplot shows 2 outliers.
   (d) The standard deviation of male height and the standard deviation of female height are both greater than 5 mm.
   (e) Both distributions show strong left skewness.

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.)

   (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.)

    (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.)

    (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?

    (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 $P(A^c \cup B)$?

    (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 $0.3 \leq P(A \cap B) \leq 0.6$.
    (b) If $P(A|B) = 0$, then A and B are mutually exclusive.
    (c) For any event A, A and its complement are mutually exclusive.
    (d) If $P(A) = 0.4$, $P(B) = 0.4$, and $P(A \cup B) = 0.64$, then 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?

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

(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.)

(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.)

    (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 $p = 1$.
    (b) The mean of a binomial random variable can be negative.
    (c) If X is a binomial random variable, then $P(X = 1) > P(X = 0)$.
    (d) The mean of a binomial random variable X can be less than the most likely value of X.
    (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.
    (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.

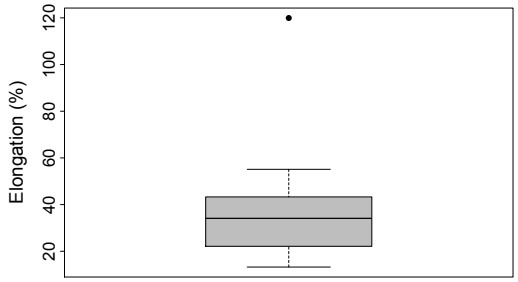

Figure 2: Boxplot of elongation percentage at breaking point.

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.

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.)

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

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 for the ith observation: $z_i = \dfrac{x_i - \bar{x}}{s}$

If we transform the data using the linear transformation $x^* = a + bx$, then:

$$\bar{x}^* = a + b\bar{x}, s_{x^*} = |b|s_x, s_{x^*}^2 = b^2 s_x^2$$

Probability

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

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

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

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

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

The Expected Value and Variance of Discrete Random Variables

$E(X) = \mu = \sum xp(x)$.

$\sigma^2 = E[(X-\mu)^2] = \sum (x-\mu)^2 p(x)$.

A handy relationship: $E[(X-\mu)^2] = E(X^2) - [E(X)]^2$.

Properties of Expectation and Variance

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

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

If X and Y are independent: $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ and $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$

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