University of Guelph
Department of Mathematics and Statistics

STAT*2040
Statistics I

Test 2 (White version)

# Solutions

March 17 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. According to the Canadian Health Measures Survey, the hand grip strength of Canadian females in the 20 to 39 age group is approximately normally distributed with a mean of 53.0 kg and a standard deviation of 12.0 kg. Suppose 20 females in this age group are randomly selected. What is the probability their mean grip strength exceeds 55.0 kg? (Choose the closest value.)

$$\begin{aligned} P(\bar{X} > 55.0) &= P(Z > \frac{55.0 - 53.0}{12.0/\sqrt{20}}) \\ &= P(Z > 0.745356) \\ &\approx 0.23 \end{aligned}$$

   (a) 0.17
   (b) 0.23 ***
   (c) 0.33
   (d) 0.43
   (e) 0.58

3. Which one of the following statements is true?

   (a) The sample standard deviation is always less than the sample mean.
   (b) The sample mean is always greater than the sample median.
   (c) If $A$ and $B$ are two events such that $P(A|B) > P(A)$, then $P(B|A) > P(B)$.*** *If* $P(A|B) > P(A)$ *then* $P(A \cap B) > P(A)P(B)$ *and* $P(B|A) > P(B)$.
   (d) If $A$ and $B$ are mutually exclusive events, then $A$ and $B$ are independent.
   (e) None of the above.

4. 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.)

   Copy and paste from Test 1. $P(\text{No black}) = P(\text{All red}) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{12}{18} \cdot \frac{12}{18} \cdot \frac{12}{18} = 0.0529$.

   Now that we know the binomial and hypergeometric distributions, we could have also used:

$$\frac{\binom{5}{3}\binom{3}{0}}{\binom{8}{3}} \times \binom{3}{3}(\frac{12}{18})^3(\frac{6}{18})^0 = 0.0529$$

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

5. A random sample of 50 Canadian males between the ages of 20 and 39 had several physical characteristics measured. One measurement was their hip circumference. The sample mean was

found to be 102.0 cm, with a resulting 95% confidence interval for $\mu$ of (98.7, 105.3). Of the following options, which one is the most appropriate interpretation of this interval?

(a) We can be 95% confident that the sample mean hip circumference of the 50 males in the study lies between 98.7 and 105.3 cm.
(b) We can be 95% confident that the true mean hip circumference of Canadian males in that age group lies between 98.7 and 105.3 cm. *****
(c) 95% of Canadian males in that age group have a hip circumference that lies between 98.7 and 105.3 cm.
(d) In repeated sampling, 95% of confidence intervals calculated in this manner will contain the sample mean hip circumference.
(e) All of the above.

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

(a) The sampling distribution of a statistic is the probability distribution of the statistic.
(b) The mean of the sampling distribution of a statistic is called the statistic's *standard error.* *** *False. The standard error of a statistic is the estimate of the standard deviation of the statistic's sampling distribution.*
(c) The variance of the sampling distribution of the sample mean decreases as the sample size increases.
(d) The mean of the sampling distribution of a statistic can be negative.
(e) None of the above.

7. Approximately 30% of births in Ontario are carried out by Caesarean section. In 10 randomly selected births in Ontario, what is the probability that no more than 2 were carried out by Caesarean section? (Choose the closest value.)

*We need to find* $P(X \leq 2)$ *where* $X$ *has a binomial distribution with* $n = 10$ *and* $p = 0.3$.

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.0282 + 0.1211 + 0.2335 = 0.3828$

(a) 0.13
(b) 0.23
(c) 0.35
(d) 0.38 ***
(e) 0.43

8. Which one of the following quantities can be negative?

(a) The mean of a binomial random variable.
(b) The variance of a hypergeometric random variable.
(c) The mean of a continuous uniform random variable.***
(d) $P(Z < a)$, where $Z$ is a standard normal random variable and $a$ is a constant.
(e) None of the above.

9. Four values are sampled independently from the standard normal distribution. What is the probability that the largest of the 4 values is less than 1.50?

*The largest value will be less than 1.50 if and only if all 4 values are less than 1.50. The probability that any one of the 4 values is less than 1.5 is* $P(Z < 1.5) = 0.9332$. *The probability that all 4 values are less than 1.50 is* $0.9332^4 = 0.758$.

(a) 0.07
(b) 0.76 ***
(c) 0.82
(d) 0.93
(e) 0.85

10. Suppose we are about to draw a sample from a normally distributed population where both the population mean and population standard deviation are unknown. We intend to find a confidence interval for the population mean. Consider the following statements:

I. All else being equal, an increase in the sample size would result in a decrease in the margin of error. *True*

II. All else being equal, an increase in the confidence level would result in a decrease in the margin of error. *False.*

III. All else being equal, an increase in the sample variance would result in a decrease in the margin of error. *False*

Which of these statements are true?

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

11. Which one of following statements is true?

(a) If we sample a large number of observations from a uniform distribution, a histogram of these observations will look approximately normal. *False. A histogram of a large number of observations sampled from a distribution will look like that distribution.*

(b) If we are sampling a large number of observations from the normal distribution, then the sample mean will have (approximately) a uniform distribution.

(c) If we are sampling a large number of observations from a uniform distribution, then the sampling distribution of the sample mean will be approximately normal. *** *True, due to the central limit theorem.*

(d) As the sample size increases, the sample mean tends to increase.

(e) None of the above.

12. Suppose we are about to sample from a normally distributed population, where the population mean is unknown but the population standard deviation is known. Suppose we intend on constructing a confidence interval for the population mean using the formula:

$$\bar{X} \pm 1.22\frac{\sigma}{\sqrt{n}}$$

The confidence level of the interval is closest to which one of the following?

*This probability is the area between* $-1.22$ *and* $1.22$ *under the standard normal curve:*

$$\begin{aligned} P(-1.22 < Z < 1.22) &= P(Z < 1.22) - P(Z < -1.22) \\ &= 0.8888 - 0.1112 \\ &= 0.7776 \end{aligned}$$

(a) 67%
(b) 78%***
(c) 89%
(d) 95%
(e) 99%

13. Which one the following statements is true?

(a) In practice, when carrying out inference procedures for $\mu$, $\sigma$ is usually known and thus we usually use $z$ procedures instead of $t$ procedures. *False. $\sigma$ is almost always unknown.*
(b) The variance of $\mu$ decreases as the sample size increases. *False. $\mu$ is a fixed value and has no variance.*
(c) If the assumptions of a procedure used to calculate a 90% confidence interval are violated, then the true confidence level of the interval will be greater than 90%. *False. It will differ from 90%, but will typically be lower.*
(d) A $t$ distribution with 5000 degrees of freedom has lower variance than the standard normal distribution. *False. The $t$ distribution has greater variance than the standard normal distribution (but at 5000 degrees of freedom the distributions will be very similar).*
(e) None of the above. ***

14. Suppose we sample 5 observations from a normally distributed population, and find that the sample mean and sample standard deviation are 24 and 16, respectively. Which one of the following is the appropriate margin of error of a 95% confidence interval for $\mu$?

*Since the standard deviation is based on sample data, we need to use the $t$ value in the interval calculation. The 95% margin of error is* $t_{0.025}\frac{s}{\sqrt{n}} = 2.776\frac{16}{\sqrt{5}} = 19.86$.

(a) 7.2
(b) 11.3
(c) 14.0
(d) 19.9 ****
(e) 21.6

15. Suppose we are sampling from a normally distributed population where the population mean is unknown, but the population standard deviation is known to be 10. We want to draw a sample and then test the null hypothesis $H_0$: $\mu = 0$ against a two-sided alternative. We draw a sample of 4 observations and find that the sample mean is 12.5. What is the $p$-value of the test?

*$\sigma$ is known here, so we will be using a $Z$ test.* $Z = \frac{12.5-0}{10/\sqrt{4}} = 2.5$. *The $p$-value is double the area in the tail, beyond 2.5.* $P(Z \geq 2.5) = 0.0062$ *and thus the $p$-value is 0.0124.*

(a) 0.006
(b) 0.012 ***
(c) 0.024
(d) 0.044
(e) 0.088

16. Which one of the following statements is true?

(a) If a null hypothesis is rejected at the 0.05 significance level, then it would also be rejected at the 0.10 significance level. ***
(b) If a null hypothesis is rejected at $\alpha = 0.05$, then it would also be rejected at $\alpha = 0.01$.
(c) If a null hypothesis is not rejected at the 0.05 significance level, then it will definitely not be rejected at the 0.10 significance level.
(d) The significance level of a test is $1 - p$-value.
(e) None of the above.

17. Estradiol is an estrogen that is biologically active in both males and females. Suppose it had long been conjectured that the true mean estradiol concentration in the blood of Caucasian Canadian males between the ages of 25 and 29 years was 50 pg/ml. A researcher wished to investigate this by drawing a sample and testing the appropriate null hypothesis against a two-sided alternative. 40 Caucasian Canadian males in this age group volunteered for a study and had their estradiol concentration measured. (Suppose it is reasonable to assume that these 40 males can be considered a random sample of Caucasian Canadian males in this age group.) The researcher used R to do the calculations of the hypothesis test, and found the following output.

```
One Sample t-test
data:  estradiol
t = -3.5814, df = 39, p-value = 0.0009354
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 42.64553 47.95447
sample estimates:
mean of x
     45.3
```

Of the following options, which one is the most appropriate conclusion to the hypothesis test?

*The p-value is small here (less than 0.001), so there is very strong evidence against the null hypothesis that $\mu = 50$. Note also that the sample mean is less than 50, and the entire 95% confidence interval for $\mu$ lies to the left of 50.*

(a) There is very strong evidence that the sample mean estradiol concentration in the 50 males in the study is less than 50 pg/ml.
(b) There is very strong evidence that the true mean estradiol concentration in the blood of Caucasian Canadian males in this age group is less than 50 pg/ml. ***
(c) There is very strong evidence that the true mean estradiol concentration in the blood of Caucasian Canadian males in this age group is greater than 45.3 pg/ml.
(d) There is not strong evidence that the true mean estradiol concentration in the blood of Caucasian Canadian males in this age group differs from 50 pg/ml.

(e) There is not strong evidence that the true mean estradiol concentration in the blood of Caucasian Canadian males in this age group differs from 45.3 pg/ml.

18. Suppose we wish to carry out a test of $H_0$: $\mu = 10$ against a two-sided alternative at a significance level of $\alpha = 0.05$. Suppose that, unknown to us, in reality $\mu = 10$. We draw a sample, find a $p$-value of 0.002, and make the appropriate conclusion to the hypothesis test. Which one of the following statements is true?

*Since the p-value is less than $\alpha$, we would reject the null hypothesis. But the null hypothesis is true here, so we rejected a true null hypothesis. This is called a Type I error.*

(a) We made a Type I error. ****
(b) We made a Type II error.
(c) We made both Type I and Type II errors.
(d) Whether or not we made a Type II error is unknown, but we know we did not make a Type I error.
(e) None of the above.

19. Suppose that the number of fatal crashes per year involving U.S. commercial airlines has a mean of approximately 1.1. To a reasonable approximation, fatal crashes can be thought of as occurring randomly and independently. What is the probability that there is at least 1 fatal crash involving U.S. commercial airlines in the next year? (Choose the closest value.)

*The number of airline crashes would follow (approximately) a Poisson distribution.*

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{1.1^0 e^{-1.1}}{0!} = 0.667.$$

(a) 0.25
(b) 0.33
(c) 0.55
(d) 0.58
(e) 0.67 ***