

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm – 5:00 pm
April 16, 2013

Page 1 of 11

Section Number: keys

Student Name: _____

Student ID: _____

Examination Instructions

1. This exam is composed of three parts. There are 10 True/False questions in Part I; 15 multiple choice questions in Part II. Part III has six open-ended questions.
2. You should be certain that you have 13 distinct pages, including this cover page. There are 31 problems.
3. The True/False type questions are worth one mark per question and each multiple choice question is worth two marks. The marks for the open ended questions are indicated in each question.
4. PLEASE make sure to read each question carefully!
5. The total marks for this exam is 100.
6. Use the scrap paper provided as needed.
7. Non-programmable calculators are permitted.
8. Two-sided letter size (8.5×11) handwritten formula sheet is allowed.
9. Please show all of your work. It is YOUR responsibility to convince us that you know what you are doing. Clarity, completeness, and organization count.
10. If you have any questions, feel free to ask.
11. Good luck!

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm – 5:00 pm
April 16, 2013

Page 2 of 11

Part I: True/False

1. Statistical inference is the process of making an estimate, prediction, or decision about a population based on sample data.

Answer: T

2. A distribution is skewed to the right if a greater proportion of the measurements lie to the left of the peak value.

Answer: F

3. Two events, A and B are mutually exclusive if $P(A \cup B) = P(A) + P(B)$.

Answer: T

4. The ratio of the variance to the mean of a Binomial random variable is less than that of a Poisson random variable.

Answer: T

5. If X denotes a continuous random variable, $P(X = c) = 0$ for every number c .

Answer: T

6. If a random sample of n measurements is selected from a population with mean μ , the expected value of the sample mean is always equal to μ regardless of the sample size n .

Answer: T

7. An estimator is a random variable calculated from a random sample that provides either a point estimate or an interval estimate for some population parameter.

Answer: T

8. The p-value of a statistical test is the smallest value of the significance level α for which the null hypothesis cannot be rejected.

Answer: F

9. In regression analysis, the dependent variable is a variable whose value is known and is being used to explain or predict the value of another variable.

Answer: F

10. Regardless of the degrees of freedom, every t distribution is symmetric around 0.

Answer: T

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm – 5:00 pm
April 16, 2013

Page 3 of 11

Part II: Multiple Choice

1. Which of the following refers to numbers that indicate the spread or scatter of observations in a data set? Answer: (c)
 - (a) measures of centre
 - (b) measures of location
 - (c) measures of variability
 - (d) measures of shape
2. Which of the following represents a disadvantage of using the sample range to measure dispersion? Answer: (b)
 - (a) It produces spreads that are too large.
 - (b) The largest or smallest observation (or both) may be an outlier.
 - (c) The sample range is not measured in the same units as the data.
 - (d) The sample range is measured in the same units as the data.
3. If two data sets have the same interquartile range, which of the following characteristics do both of those sets also have? Answer: (b)
 - (a) Both sets will have the same median.
 - (b) The distances from the first to third quartiles in both sets will be the same.
 - (c) Both sets will have the same range.
 - (d) The distances from the median to third quartile in both sets will be the same.
4. What is any subset of the sample space called? Answer: (a)
 - (a) an event
 - (b) an experiment
 - (c) a mutually exclusive event
 - (d) independent events
5. Which of the following statements is a property of the binomial distribution? Answer: (a)
 - (a) The binomial distribution tends to be more symmetric as the probability of success p approaches 0.5.
 - (b) As the number of trials increase, the expected value of the random variable decreases.
 - (c) As the number of trials increase for a given probability of success, the binomial distribution becomes more skewed.
 - (d) The variance of a binomial random variable decreases as the probability of success p approaches 0.5.
6. The number of traffic accidents per day on a certain section of highway is thought to be Poisson distributed with a mean to equal 2.19. Which of the following best approximates the probability of no accidents occurring on this section of highway during a one-day period? Answer: (d)
 - (a) 0.457
 - (b) 0.318
 - (c) 0.296
 - (d) 0.112

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)

2:00 pm – 5:00 pm
April 16, 2013

Final Exam

Page 4 of 11

7. If z is a standard normal random variable, how does the area between $z = 0.0$ and $z = 1.25$ compare to the area between $z = 1.25$ and $z = 2.5$?

Answer: (b)

- (a) The latter area will be larger than the former.
- (b) The latter area will be smaller than the former.
- (c) The two areas are the same.
- (d) The latter area is half of the former.

8. What is a numerical descriptive measure calculated from a sample called?

Answer: (b)

- (a) a parameter
- (b) a statistic
- (c) a population
- (d) a sampling distribution

9. Why do those who engage in estimation insist on random sampling, rather than convenience sampling or judgment sampling? Answer: (c)

- (a) because random sampling avoids the errors inherent in matched-pairs sampling
- (b) because random sampling avoids the errors inherent in work sampling
- (c) because random sampling eliminates the systematic error or bias that arises in nonrandom sampling
- (d) because random sampling is less expensive than convenience or judgment sampling

10. What is the type of sample statistic that is used to make inferences about a given type of population parameter?

Answer: (a)

- (a) the estimator of that parameter
- (b) the confidence level of that parameter
- (c) the confidence interval of that parameter
- (d) the point estimate of that parameter

11. Which of the following is a property of the sampling distribution of sample proportion \hat{p} ? Answer: (c)

- (a) An increase in the sample size n will result in an increase in the standard error of \hat{p} .
- (b) The mean of \hat{p} is different from the population proportion p .
- (c) The sampling distribution will be approximately normal provided that $np > 5$ and $nq > 5$.
- (d) For a fixed sample size n , the standard error of \hat{p} is minimal when $p = 0.5$.

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)

2:00 pm – 5:00 pm
April 16, 2013

Final Exam

Page 5 of 11

12. Which of the options below provides the best interpretation of a 95% confidence interval estimate of the population mean μ ? Answer: (C)
- (a) There is a 95% probability that the population mean μ will lie between the lower confidence limit (LCL) and the upper confidence limit (UCL).
 - (b) We are 95% confident that 5% the values of the sample means \bar{x} will result in a confidence interval that includes the population mean μ .
 - (c) If we repeatedly draw samples of the same size from the same population, 95% of the values of the sample means \bar{x} will result in a confidence interval that includes the population mean μ .
 - (d) We are 95% confident that we have selected a sample whose range of values does not contain the population mean μ .
13. Under which of the following circumstances is it IMPOSSIBLE to construct a confidence interval for the population mean? Answer: (C)
- (a) a non-normal population with a large sample size and an unknown population variance
 - (b) a normal population with a large sample size and a known population variance
 - (c) a non-normal population with a small sample size and an unknown population variance
 - (d) a normal population with a small sample size and an unknown population variance
14. Given the least-squares regression line $\hat{y} = 5 - 2x$, what may be said about the relationship between the two variables x, y ? Answer: (b)
- (a) The relationship between x and y is positive.
 - (b) The relationship between x and y is negative.
 - (c) As x increases, so does y .
 - (d) As x decreases, so does y .
15. Which of the following is NOT an assumption for the simple linear regression model? Answer: (A)
- (a) The distribution of the error terms will be skewed to left or right depending on the values of the dependent variable.
 - (b) The error terms have equal variances for all values of the independent variable.
 - (c) The error terms are independent of each other.
 - (d) The mean of the dependent variable for all levels of the independent variable can be connected by a straight line.

Part III: Show all of your work neatly in the blank space provided below each question.

1. A urn contains five balls, two red and three blue. Randomly select two balls sequentially without replacement, and define

A : second ball is red

B : first ball is blue

- (a) (4 marks) Find $P(A)$ and $P(B)$.

Sol)
$$P(A) = P(\text{Red on first and Red on Second}) + P(\text{Blue on first and Red on Second})$$
$$= \left(\frac{2}{5}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{2}{5}$$

$$P(B) = \frac{3}{5}.$$

- (b) (4 marks) Find $P(A \cap B)$ and $P(A|B)$.

Sol)
$$P(A|B) = \frac{2}{4} = \frac{1}{2}.$$

$$P(A \cap B) = P(A|B)P(B) = \left(\frac{1}{2}\right)\left(\frac{3}{5}\right) = \frac{3}{10}.$$

- (c) (2 marks) Are A and B independent? Explain.

Sol) No, A and B are dependent since

$$P(A) = \frac{2}{5} \neq \frac{1}{2} = P(A|B).$$

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm - 5:00 pm
April 16, 2013

Page 7 of 11

2. A recent study demonstrated that only 20% of the people in Vancouver feel that its mass transit system is adequate.

(a) (3 marks) If 20 people from Vancouver are selected at random, find the probability that more than 1 feel its mass transit system is adequate?

$$\begin{aligned}\text{sol)} \quad P(X > 1) &= 1 - P(X \leq 1) = 1 - P(X=0) - P(X=1) \\ &= 1 - (0.8)^{20} - 20 \cdot (0.2)(0.8)^{19} \\ &= 0.9308\end{aligned}$$

(b) (3 marks) Of these 20 people, how many should we expect feel that its mass transit system is adequate?

$$\text{sol)} \quad \mu = np = 20(0.2) = 4.$$

(c) (4 marks) If 250 people from Vancouver are selected at random, find the probability that 65 or less feel that its mass transit system is adequate by making use of the normal approximation.

$$\begin{aligned}\text{sol)} \quad P(X \leq 65) & \quad \mu = 250(0.2) = 50 \\ & \quad \sigma = \sqrt{250(0.2)(0.8)} = 6.325 \\ & \approx P\left(Z \leq \frac{65.5 - 50}{6.325}\right) \\ & = P(Z \leq 2.45) = 0.993.\end{aligned}$$

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm - 5:00 pm
April 16, 2013

Page 8 of 11

3. Suppose the duration of trouble-free operation of a new vacuum cleaner is normally distributed with a mean of 530 days and standard deviation of 100 days.
- (a) (3 marks) What is the probability that a randomly selected vacuum cleaner will work for at least 730 days without trouble?

$$\begin{aligned}\text{Sol)} \quad P(X \geq 730) &= P\left(Z \geq \frac{730 - 530}{100}\right) \\ &= P(Z \geq 2) = 0.0228.\end{aligned}$$

- (b) (2 marks) Describe the sampling distribution of the sample mean for a random sample of 25 vacuums.

$$\begin{aligned}\text{Sol)} \quad n=25, \quad \bar{x} &\text{ follows a normal distribution} \\ &\text{with mean } \mu=530 \text{ and the standard} \\ &\text{deviation } \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{25}} = 20.\end{aligned}$$

- (c) (5 marks) If 25 random vacuum cleaners are selected, within what limits would you expect the sample average to lie (symmetrically about μ) with probability 0.95?

$$\begin{aligned}\text{Sol)} \quad P(\mu - 1.96 \cdot SE < \bar{x} < \mu + 1.96 \cdot SE) &= 0.95 \\ (\Rightarrow) \quad 490.8 = 530 - (1.96)(20) < \bar{x} < 530 + (1.96)(20) &= 569.2 \\ &\text{with probability } 0.95.\end{aligned}$$

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm - 5:00 pm
April 16, 2013

Page 9 of 11

4. Independent random samples of $n_1 = 1265$ and $n_2 = 1688$ observations were selected from binomial populations 1 and 2, and $x_1 = 849$ and $x_2 = 910$ successes were observed.

- (a) (2 marks) What is the best point estimator for the difference $(p_1 - p_2)$ of two binomial proportions?

Sol) The best estimator is

$$\hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}, \begin{cases} x_1 \sim \text{Bin}(n_1, p_1) \\ x_2 \sim \text{Bin}(n_2, p_2) \end{cases}$$

- (b) (2 marks) Find a point estimate for $(p_1 - p_2)$ and calculate the 95% margin of error for this point estimate.

Sol) The point estimate is $\hat{p}_1 - \hat{p}_2 = \frac{849}{1265} - \frac{910}{1688} = 0.671 - 0.539 = 0.132$

$$\text{The 95\% ME} = 1.96 \sqrt{\frac{(0.671)(0.329)}{1265} + \frac{(0.539)(0.461)}{1688}} = 0.035$$

- (c) (4 marks) Construct a 98% confidence interval for $(p_1 - p_2)$.

$$0.132 - 2.33(0.0179) < p_1 - p_2 < 0.132 + 2.33(0.0179)$$

\Leftrightarrow

$$0.0903 < p_1 - p_2 < 0.1737$$

- (d) (2 marks) Based on the confidence interval in part (c), is there sufficient evidence to indicate a difference in the proportions? Explain.

Sol) Yes, since the interval is strictly positive.
It is unlikely that $p_1 - p_2 = 0$.

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm - 5:00 pm
April 16, 2013

Page 10 of 11

5. (10 marks) A sample of 40 speedometers of a particular brand is obtained, and each is calibrated to check for accuracy at 55 mph. The resulting sample average and sample standard deviation are 53.8 and 1.3, respectively. Does the sample information suggest that the true mean for this brand of speedometers is not accurate at 55 mph? Use a level .01 test.

Sol) $n = 40$

$$\bar{x} = 53.8$$

$$s = 1.3$$

$$H_0: \mu = 55$$

$$H_a: \mu \neq 55$$

$$z\text{-stat} = \frac{53.8 - 55}{(1.3/\sqrt{40})} = -5.838$$

since the test statistic is less than the critical value
 $-z_{0.005} = -2.58$, we reject the null hypothesis.

University of Regina
Mathematics and Statistics
STAT 160-001 (Bae) STAT 160-002 (Fallat)
Final Exam

2:00 pm - 5:00 pm
April 16, 2013

Page 11 of 11

6. The following $n = 10$ observations are a sample from a normal population:

5.1, 2.6, 3.4, 6.3, 2.8, 4.5, 5.9, 1.7, 3.7, 8.0

(a) (3 marks) Find the mean and standard deviation of this sample.

$$\text{sol)} \quad \bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 4.4$$

$$s = \sqrt{\left(\sum_{i=1}^{10} x_i^2 - \frac{1}{10} \left(\sum_{i=1}^{10} x_i \right)^2 \right) / (10-1)} = 1.9408.$$

(b) (4 marks) Find a 99% confidence interval for the population mean μ .

$$\text{sol)} \quad \bar{x} - t_{0.005} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.005} \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow 4.4 - (3.25) \frac{(1.9408)}{\sqrt{10}} < \mu < 4.4 + (3.25) \frac{(1.9408)}{\sqrt{10}}$$

$$\Leftrightarrow 2.406 < \mu < 6.3938$$

(c) (3 marks) Test $H_0: \mu = 4$ versus $H_a: \mu \neq 4$. Use $\alpha = 0.01$.

$$\text{sol)} \quad t\text{-statistic} = \frac{4.4 - 4.0}{(1.9408)/\sqrt{10}} = 0.6520.$$

Since the t -statistic is less than the critical value

$t_{0.005} = 3.25$, we do not reject the null hypothesis.

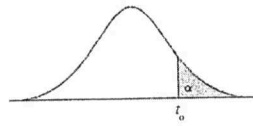


TABLE 4
Critical Values
of t
page 727

| df | $t_{0.100}$ | $t_{0.050}$ | $t_{0.025}$ | $t_{0.010}$ | $t_{0.005}$ | df |
|----------|-------------|-------------|-------------|-------------|-------------|----------|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 1 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 2 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 3 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 4 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 6 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 7 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 8 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 9 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 10 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 11 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 12 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 13 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 14 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 15 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 16 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 17 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 18 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 19 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 20 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 21 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 22 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 23 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 24 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 25 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 26 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 27 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 28 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 29 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | ∞ |

Source: From "Table of Percentage Points of the t -Distribution," *Biometrika* 32 (1941):300. Reproduced by permission of the *Biometrika* Trustees.

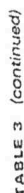


TABLE 3 (continued)

TABLE 3 (continued)