UNIVERSITY OF REGINA
Department of Mathematics and Statistics

Statistics 160
Zhao (001), Bae (002), Fallat (003)

Final Exam, Fall 2015 (December 18$^{\text{th}}$, 2015)

Time: 180 minutes **Full Name:**________________________________

Pages: 11 **Student Number:**____________________________

INSTRUCTIONS

1. Indicate your **section number** by circling one of the three above.
2. **Do not** open the exam until you are instructed to do so.
3. This exam has a total of 100 marks. The marks allocated for each question are found to the left of the question.
4. **All work and answers** are to be placed **on the right side pages of this exam** below the question. If you require more space for an answer, work on the left side (facing) page and **indicate** on the question there is work to be found there.
5. The left side pages are to be used as **scrap paper**. They are provided for **rough work and checking only** and will not be graded unless you expressly indicate there is work to be found there.
6. Only a non-programmable **calculator** and a **two-sided letter size** (8.5×11) **handwritten formula sheet** are permitted during the exam.
7. Please show all of your work. It is your responsibility to convince me that you know what you are doing. Clarity, completeness, and organization count.
8. One person at a time may be given permission for a brief washroom break.
9. **Good luck**!

For instructor use only:

Question:	1	2	3	4	5	6	7	Total
Marks:	60	6	6	8	7	8	5	100
Score:								

Marks

(60) 1. This part consists of 20 multiple choice questions. Each question is worth 3 marks. Select the correct answer by **circling** the appropriate choice for each question.

1) Suppose $P(A) = 0.80$, $P(B) = 0.70$ and $P(A \cup B) = 0.94$. Which one of the following statements correctly defines the relationship between events A and B?

   A. Events A and B are independent, but not mutually exclusive. ✓

   B. Events A and B are mutually exclusive, but not independent.

   C. Events A and B are neither mutually exclusive nor independent.

   D. Events A and B are both mutually exclusive and independent.

2) The average number of traffic accidents on a certain section of highway is two per week. The number of accidents is assumed to follow a Poisson distribution. Find the probability of no accident on this section of highway during a 1-week period.

   A. 0.0001

   B. 0.1353 ✓

   C. 0.2707

   D. 0.3679

3) Which of the following statements is a property of the binomial distribution?

   A. The variance of a binomial distribution is the same as the mean.

   B. As the number of trials increases for a given success probability, the binomial distribution becomes more skewed.

   C. For a given number of trials, the variance of a binomial distribution gets its maximum when the success probability equals 0.5. ✓

   D. The success probability of a binomial distribution is the same as the ratio of its variance and the mean.

4) Which of the following conditions is required for the normal approximation to the binomial probabilities to be adequate?

   A. The binomial distribution has to be symmetric or close to symmetric.

   B. As long as n is large we can use a normal distribution to approximate the binomial probabilities.

   C. As long as both $np > 5$ and $nq > 5$.✓

   D. The normal distribution needs to be the standard normal with $\mu = 0$ and $\sigma = 1$.

5) Consider a binomial random variable X with $n = 20$ and $p = 0.6$. Using the correction for continuity, approximate $P(X > 9)$.

   A. 0.8186

   B. 0.8729 ✓

   C. 0.9147

   D. 0.9452

6) Let Z be a standard normal random variable. Find a z_0 such that $P(-z_0 < Z < z_0) = 0.8262$.

A. 0.8262

B. 0.9131

C. 1.36 ✓

D. -1.36

7) Suppose you want to select a sample of size $n = 2$ from a population containing $N = 4$ objects, say the four objects are identified as x_1, x_2, x_3 and x_4, using simple random sampling. What is the probability that x_1 and x_4 are both selected in the sample?

A. 1/6 ✓

B. 1/4

C. 1/12

D. 1/16

8) According to the Central Limit Theorem if a random sample of size $n = 80$ is selected from a binomial distribution with population proportion $p = 0.25$, then the sampling distribution of the sample proportion $\hat{p}$ is approximately normally distributed. Find the standard deviation of $\hat{p}$.

A. 0.0023

B. 0.0484 ✓

C. 0.1875

D. 0.4330

9) From a sample of 200 items, 12 items are defective. In this case, what will be the point estimate for the true proportion of defective items?

A. 0.06 ✓

B. 0.12

C. 12

D. 16.67

10) Which of the options below provides the best interpretation of a 95% confidence interval estimate of the population mean μ?

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 $\overline{X}$ will result in a confidence interval that includes the population mean μ.

C. Under repeated sampling, we expect that 95% of the confidence intervals constructed to contain the true 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.

11) In a two-tailed test of the population mean, the null hypothesis will be rejected at an α level of significance if which of the following conditions hold:

A. $|Z_{obs}| \geq z_{\alpha}$

B. $|Z_{obs}| < -z_{\alpha/2}$

C. $-z_{\alpha} < Z_{obs} < z_{\alpha}$

D. $|Z_{obs}| \geq z_{\alpha/2}$ ✓

12) If a hypothesis is rejected at the 0.05 level of significance, then what can be deduced from this decision?

A. the hypothesis must be rejected at any level.

B. the hypothesis must be rejected at the 0.02 level.

C. the hypothesis must not be rejected at the 0.02 level.

D. the hypothesis may or may not be rejected at the 0.02 level. ✓

13) Which of the following is NOT a required assumption for constructing a confidence interval estimate of the difference between two population means when the sample sizes are small?

A. The populations are normally distributed.

B. The population variances are equal.

C. The sample sizes are equal. ✓

D. The samples are selected independently at random from the populations.

14) Which sampling distribution is used to make inferences about a single population variance?

A. a normal distribution

B. a t distribution

C. a χ^2 distribution ✓

D. an F distribution

15) Which of the following is an assumption behind the paired-difference t test for the difference between two population means?

A. The populations are normally distributed. ✓

B. The sample variances are equal.

C. The sample means are equal.

D. The two samples are independent.

16) What are the assumptions for analysis of variance test and estimation procedures?

A. The observations within each population do not need to be normally distributed.

B. The observations within each population may not have a common variance as long as they are normally distributed.

C. The observations within each population have a common variance.

D. The observations within each population are normally distributed with a common variance. ✓

17) Which of the following is NOT an assumption for the simple linear regression model?

A. the distribution of error terms will be skewed to the left or right depending on x. ✓

B. the error terms have equal variances for all values of x

C. the error terms are independent on one another

D. the error terms all have mean 0.

18) A regression analysis between sales (y in \$1,000) and advertising (x in \$100) resulted in the following line of best fit: $\hat{y} = 82 + 7x$. If advertising costs were \$900, what could we reasonably expect the amount of sales (in dollars) to be?

A. \$6,382

B. \$82,063

C. \$88,300

D. \$145,000 ✓

19) If each element in a population is classified into one and only one of k categories (with $k > 2$), which of the following best describes this kind of population?

A. it is a normal population

B. it is a multinomial population ✓

C. it is a chi-squared population

D. it is a binomial population

20) In a χ^2 goodness-of-fit test with 5 degrees of freedom and a significance level of 0.05. Which of the following computed values of the χ^2 test statistic will lead to a rejection of the null hypothesis?

A. 7.814

B. 8.952

C. 10.78

D. 17.61 ✓

Show all of your work neatly at the blank space below each question.

2. Two major league baseball teams, A and B, are scheduled to play a championship series: the winner is the first team to win two games in a total that cannot exceed three games. The event that A wins any one game is independent of the event that A wins any other, and probability that A wins any one game is equal to 0.6.

(2) (a) List the simple events in the experiment. (Hint: Event E_1 : AA represents an event that A wins two games in a row.)

(sol) AA, ABA, ABB, BAA, BAB, BBB

(2) (b) Let X equal the total number of games in the championship series; that is, $x = 2, 3$. Construct the probability distribution for X.

(sol) $p(2) = P(AA) + P(BB) = 0.6^2 + 0.4^2 = 0.52$,
$p(3) = P(ABA) + P(ABB) + P(BAA) + P(BAB) = 1 - p(2) = 0.48$.

(2) (c) Calculate the mean and variance of X.

(sol) $\mu = 2(0.52) + 3(0.48) = 2.48$, $\sigma^2 = 2^2(0.52) + 3^2(0.48) - (2.48)^2 = 0.2496$.

3. A restaurant in a certain resort polled 100 guests as to whether or not they arrived by car or by bus. The result was 70 by car and 30 by bus.

(3) (a) Construct a 90% confidence interval for the true proportion of all guests who arrive by bus.

(sol) $0.3 \pm (1.645)(0.0458)$

(3) (b) If the restaurant wanted to obtain a narrower estimate so that its error of estimate is within 0.05, with a 90% confidence, how many guests should be polled?

(sol) $n = (0.3)(0.7)(1.645/0.05)^2 \approx 228$.

4. Two independent random samples of sizes $n_1 = n_2 = 5$ are selected from each of two normal populations:

Sample from Population 1:	4.8	5.2	5.0	4.9	5.1
Sample from Population 2:	5.0	4.7	4.9	4.8	4.9

(3) (a) Find a 95% confidence interval for σ_1^2, the population variance of Population 1.
(sol) $df = n_1 - 1 = 4$. $\chi^2_{0.025} = 11.1433, \chi^2_{0.975} = 0.4844$. $s_1^2 = 0.025$.
$UCB = (n_1 - 1)s_1^2/\chi^2_{0.975} = 0.206441$. $LCB = (n_1 - 1)s_1^2/\chi^2_{0.025} = 0.008974$.

(2) (b) Test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_a : \sigma_1^2 > \sigma_2^2$. Use $\alpha = 0.05$.
(sol) $F = s_1^2/s_2^2 = 0.025/0.013 = 1.923 < 6.39 = F_{0.05}$. Thus we do not reject the null hypothesis.

(3) (c) Test $H_0 : (\mu_1 - \mu_2) = 0$ against $H_a : (\mu_1 - \mu_2) > 0$. Use $\alpha = 0.05$.
(sol) By the result of Part (b), we assume the variances are equal. The pooled estimate s^2 is $s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2) = 0.1315$. The test statistic is $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s\sqrt{1/n_1 + 1/n2}} = (5 - 4.86)/\sqrt{(0.1315)(1/5 + 1/5)} = 0.6104$. The critical value is $t_{0.05} = 1.860$. Since the observed test statistic is smaller than the critical value, we do not reject the null hypothesis.

5. A clinical psychologist wished to compare three methods for reducing hostility levels in university students using a certain psychological test (HLT). High scores on this test were taken to indicate great hostility. Eleven students who got high and nearly equal scores were used in the experiment. Five were selected at random from among the 11 problem cases and treated by method A, three were taken at random form the remaining 6 students and treated by method B, and the other 3 students were treated by method C. All treatments continued throughout a semester, when the HLT test was given again. The results are shown in the table.

Method	Scores on the HLT				
A	73	83	76	68	80
B	54	74	71		
C	79	95	87		

(4) (a) Perform an analysis of variance for this experiment by completing the following ANOVA table.

Source	*df*	SS	MS	*F*
Treatments	$k-1=$	SST =	MST =	$F=$
Error	$n-k=$	SSE =	MSE =	
Total	$n-1=$	Total SS =		

(sol)

Source	*df*	SS	MS	*F*
Treatments	$k-1=2$	SST = 641.88	MST = 320.94	$F=5.15$
Error	$n-k=8$	SSE = 498.67	MSE = 62.33	
Total	$n-1=10$	Total SS = 1140.55		

(3) (b) Do the data provide sufficient evidence to indicate a difference in mean student response to the three methods after treatment? Use $\alpha=0.05$.

(sol) $H_0:\mu_1=\mu_2=\mu_3$. $H_a:$ not H_0. The rejection region is $F>4.46=F_{0.05}(df_1=2, df_2=8)$. Since the observed test statistic $5.15>4.46$, we reject the null hypothesis.

6. Suppose the following six points are provided:

x	1	2	3	4	5	6
y	9.7	6.5	6.4	4.1	2.1	1.0

Assuming the following summaries: $\Sigma_i x_i = 21$, $\Sigma_i y_i = 29.8$, $\Sigma_i x_i^2 = 91$, $\Sigma_i y_i^2 = 199.52$, and $\Sigma_i x_i y_i = 74.8$, determine:

(3) (a) the least squares line (or line of best fit) for this data.

(sol) $\bar{x} = 3.5, \bar{y} = 4.967, S_{xx} = 17.5, S_{xy} = -29.5, TSS = S_{yy} = 51.5133, SSR = S_{xy}^2/S_{xx} = 49.7286, SSE = 51.5133 - 49.7286 = 1.7847, MSE = 1.7847/(6-2)$.
$b = S_{xy}/S_{xx} = (-29.5)/17.5 = -1.686, a = \bar{y} - (-1.686)\bar{x} = 4.967 + 1.686(3.5) = 10.868$.
$\hat{y} = 10.868 - 1.686x$.

(2) (b) the correlation coefficient. (Hint: the correlation coefficient r is the square root of the coefficient of determination r^2.)

(sol) $r = \sqrt{r^2} = \sqrt{\frac{SSR}{TSS}} = \sqrt{49.7286/51.5133} = \sqrt{0.9654} = 0.9825$.

(3) (c) Does the data provided sufficient evidence to indicate that x and y are linearly related? Test at $\alpha = .05$.

(sol) $H_0 : \beta = 0$ vs. $H_a : \beta \neq 0$ The rejection region is $|t| > 2.776 = |\pm t_{0.025}|$. Since the observed test statistic $t = \frac{b-0}{\sqrt{MSE/S_{xx}}} = (-1.686)/\sqrt{0.4462/17.5} = -10.56$. Since $|t_{obs}| > 2.776$, we reject the null hypothesis.

(5) 7. Medical statistics show that deaths due to four major diseases – call them A, B, C, and D – account for 15%, 21%, 18%, and 14%, respectively, of all non-accidental deaths. A study of the causes 308 non-accidental deaths at an area hospital gave the following counts:

Disease	A	B	C	D	Other
Deaths	43	76	85	21	83

Do these counts provide sufficient evidence to indicate that the true proportions of people dying of diseases A, B, C, and D differ from the claimed percentages? Use $\alpha = 0.025$.

(sol) $H_0 : p_1 = 0.15, p_2 = 0.21, p_3 = 0.18, p_4 = 0.14, p_5 = 0.32$. The test statistic $\chi^2 = (43 - 46.2)^2/46.2 + \cdots + (83 - 98.6)^2/98.6 = 31.77$. $df = 5 - 1 = 4$. The critical value is $\chi^2_{0.025} = 11.14$. Since the observed statistic falls in the rejection region, the null hypothesis is rejected.