Statistics (II) Midterm Exam April 22, 2016

Student ID: __________________ Name: __________________ Department: __________________

II. Multiple Choices Questions: (30 points, 2 points for each question)

1. The quantity S_P^2 is called the pooled variance estimate of the common variance of two unknown but equal population variances. It is the weighted average of the two sample variances, where the weights represent the:

a. sample variances.
b. sample standard deviations.
c. degrees of freedom for each sample.
d. None of these choices.

2. In constructing a confidence interval estimate for the difference between the means of two independent normally distributed populations, we:

a. pool the sample variances when the unknown population variances are equal.
b. pool the sample variances when the population variances are known and equal.
c. pool the sample variances when the population means are equal.
d. never pool the sample variances.

3. When testing $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 \neq 0$ with the variances of both populations known, the observed value of the *z*-score was found to be -2.15. Then, the *p*-value for this test would be

a. .0158
b. .0316
c. .9842
d. .9684

4. Which of the following statement is correct regarding the percentile points of the *F*-distribution?

a. $F_{.05,10,20} = 1/F_{.95,10,20}$
b. $F_{.05,10,20} = 1/F_{.05,20,10}$
c. $F_{.95,10,20} = 1/F_{.05,20,10}$
d. $F_{.95,10,20} = 1/F_{.95,20,10}$

5. In testing for the equality of two population variances, when the populations are normally distributed, the 10% level of significance has been used. To determine the rejection region, it will be necessary to refer to the *F* table corresponding to an upper-tail area of:

a. .90
b. .20
c. .10
d. .05

6. For testing the difference between two population proportions, the pooled proportion estimate should be used to compute the value of the test statistic when the:

a. populations are normally distributed.
b. sample sizes are small.
c. null hypothesis states that the two population proportions are equal.
d. samples are independently drawn from the populations.

7. In one-way ANOVA, the amount of total variation that is unexplained is measured by the:

a. sum of squares for treatments.
b. degrees of freedom.
c. total sum of squares.
d. sum of squares for error.

8. In a completely randomized design for ANOVA, the numerator and denominator degrees of freedom are 4 and 25, respectively. The total number of observations must equal:

a. 24
c. 25
c. 29
d. 30

9. The value of the test statistic in a completely randomized design for ANOVA is $F = 6.29$. The degrees of freedom for the numerator and denominator are 5 and 10, respectively. Using an F table, the most accurate statements to be made about the p-value is that it is:

a. greater than 0.05

b. between 0.001 and 0.010.

c. between 0.010 and 0.025

d. between 0.025 and 0.050

10. How does conducting multiple t-tests compare to conducting a single F-test?

a. Multiple t-tests increases the chance of a Type I error.

b. Multiple t-tests decreases the chance of a Type I error.

c. Multiple t-tests does not affect the chance of a Type I error.

d. This comparison cannot be made without knowing the number of populations.

11. In one-way analysis of variance, if all the sample means are equal, then the:

a. total sum of squares is zero.

b. sum of squares for treatments is zero.

c. sum of squares for error is zero.

d. sum of squares for error equals sum of squares for treatments.

12. Which of the following statement about multiple comparison methods is false?

a. They are to be use once the F-test in ANOVA has been rejected.

b. They are used to determine which particular population means differ.

c. There are many different multiple comparison methods but all yield the same conclusions.

d. All of these choices are true.

13. The primary interest of designing a randomized block experiment is to:

a. reduce the within-treatments variation to more easily detect differences among the treatment means.

b. increase the between-treatments variation to more easily detect differences among the treatment means.

c. reduce the variation among blocks.

d. None of these choices.

14. How do you calculate the expected frequency for one cell in a goodness-of-fit test?

a. The expected frequency is equal to the proportion specified in H_0 for that cell.

b. Use the total number of observations divided by the number of categories.

c. Multiply the specified proportion for that cell (found in H_0) by the total sample size.

d. None of these choices.

15. A chi-squared test of a contingency table with 6 degrees of freedom results in a test statistic $\chi^2 = 13.58$. Using the χ^2 tables, the most accurate statement that can be made about the p-value for this test is that:

a. p-value $> .10$

b. p-value $> .05$

c. $.05 < p$-value $< .10$

d. $.025 < p$-value $< .05$

Ans: cabcd cddba bcacd

III. Short answer questions by hand calculation: (50 points)

1. Suppose that a random sample of 60 observations was drawn from a population. After calculating the mean and standard deviation, each observation was standardized and the number of observations in each of the intervals below was counted. Can we infer at the 10% significance level that the data were drawn from a normal population? (6 points)

Intervals	Frequency
$Z \leq -1$	8
$-1 < Z \leq 0$	30
$0 < Z \leq 1$	17
$Z > 1$	5

2. A telemarketer makes five calls per day. A sample of 200 days gives the frequencies of sales volumes listed below:

Number of Sales	Observed Frequency (days)
0	10
1	38
2	69
3	63
4	18
5	2

Assume the population is binomial distribution with a probability of purchase p equal to .50.

(a) Compute the expected frequencies for $x =$ 0, 1, 2, 3, 4, and 5 by using the binomial probability function or the binomial tables. Combine categories if necessary to satisfy the rule of five. (6 points)

(b) Should the assumption of a binomial distribution be rejected at the 5% significance level? (4 points)

3. In a completely randomized design, 12 experimental units were assigned to the first treatment, 15 units to the second treatment, and 18 units to the third treatment. A partial ANOVA table is shown below:

Source of Variation	*SS*	*df*	*MS*	*F*
Treatments	*	*	*	9
Error	*	*	35	
Total	*	*		

(a) Fill in the blanks (identified by asterisks) in the above ANOVA table. (7 points)

(b) Test at the 5% significance level to determine if differences exist among the three treatment means. (3 points)

4. In order to examine the differences in ages of health inspectors among five counties, a Health Department statistician look random samples of six inspectors' ages in each county. The data are listed below. An F-test using ANOVA showed that average age differs for at least two counties.

Ages of Inspectors among Five School Districts

1	2	3	4	5
41	39	36	45	53
53	48	28	37	55
28	41	29	46	49
45	51	33	48	56
40	49	27	51	48
59	50	26	49	61

(a) Use Turkey's multiple comparison method to determine which means differ. (5 points)
(b) Use Fisher's LSD procedure with $a = .05$ to determine which population means differ. (5 points)

5. 35 undergraduate student who completed two years of college were asked to take a basic mathematics test. The mean and standard deviation of their scores were 75.1 and 12.8, respectively. In a random sample of 50 students who only completed high school, the mean and standard deviation of the test scores were 72.1 and 14.6, respectively. Can we infer at the 10% significance level that a difference exists between the two groups? (8 points)

6. A food processor wants to compare two antioxidants for their effects on retarding spoilage. Suppose 16 cuts of fresh meat are treated with antioxidant A and 16 are treated with antioxidant B, and the number of hours until spoilage begins is recorded for each of the 32 cuts of meat. The results are summarized in the table below.

	Antioxidant A	Antioxidant B
Sample Mean	108.7 hours	98.7 hours
Sample Standard Deviation	10.5 hours	13.6 hours

(a) Perform the test for determining if the population variances differ for Antioxidants A and B at $\alpha = .05$. (4 points)

(b) Develop the 95% confidence interval estimate of the ratio of the two population variances. (2 points)