

Math 2200 Final Exam Spring 2014

May 1, 2014

Student No.

Name.

Instructions:

1. There are 20 multiple choice problems, each worth 4 points.
2. There are 10 True/False problems, each worth 2 points. These problems are labelled **True/False**.
3. PRINT your name and BUBBLE IN your student number on the SCANTRON.
4. BUBBLE IN your answers on the SCANTRON.
5. You may use a graphic calculator with statistics functions, such as TI83, TI84, TI89 or TI-NSpire. You are allowed one double-side letter size 8.5x11 notecard.
6. When you look for the correct answer, keep it in mind that it may have been rounded off.
7. There are two blank pages at the end.
8. Good luck!

1. When a sum of squares is divided by its degrees of freedom, the result is called a(n) ...
 - A. F-ratio
 - B. R^2
 - C. p-value
 - D. t-ratio
 - E. mean square
2. **True/False** A professor runs a regression to see how students' exam scores (Y) are related to their homework grades (X). The R^2 of the regression is 69%. What does R^2 tell us?
 - A. 69% of the variation in the exam scores is explained by the regression analysis.
 - B. 69% of each student's exam grade will be determined by their homework grade.

Problem 3-4 Blood Donor

The American Red Cross says that about 11% of the U.S. population has Type B blood. A blood drive is being held at your school.

3. What is the probability that the tenth blood donor is the first one with Type B blood?
 - A. < 0.0001
 - B. 0.0385
 - C. 0.3028
 - D. 0.3853
 - E. 0.6882

4. What is the probability that at least 2 of the first 10 blood donors has Type B blood?
 - A. < 0.0001
 - B. 0.0385
 - C. 0.3028
 - D. 0.3853
 - E. 0.6882

Problem 5-6 Homelessness

Homelessness is a problem in many large U.S. cities. To better understand the problem, a multiple regression was used to model the rate of homelessness based on several explanatory variables. The following data were collected for 50 large U.S. cities. The regression results appear below.

Homeless number of homeless people per 10,000 in a city
Poverty percent of residents with income under the poverty line
Unemployment percent of residents unemployed
Temperature average yearly temperature (in °F.)
Vacancy percent of housing that is unoccupied
Rent Control indicator variable, 1 = city has rent control, 0 = no rent control

Dependent variable is Homeless

$$R^2 = 38.4\% \quad R^2(\text{adjusted}) = 31.5\%$$

$s = 2.861$ with 44 degrees of freedom

Variables	Coeff	SE(Coeff)	t-ratio	p-value
Constant	-4.275	3.465	-1.23	0.2239
Poverty	0.0823	0.0823	1.00	0.3228
Unemployment	0.159	0.218	0.73	0.4699
Temperature	0.135	0.0587	–	–
Vacancy	-0.247	0.138	-1.79	0.0809
Rent Control	2.944	1.37	2.15	0.0373

5. What is the p-value for Temperature?
- A. 0.5131
 - B. 0.5107
 - C. 0.0257
 - D. 0.0262
 - E. 0.0214
6. **True/False** Using the significance level $\alpha = 0.05$, is Poverty associated with Homelessness?
- A. Yes
 - B. No

Problem 7-9 Batteries

Three brands of AAA batteries are compared to see which last longest. Each brand of battery is tested in four different devices (a TV remote control, a hand-held game, a miniature flashlight, and a digital camera). The experiment is run once for each combination of brand and device. The twelve runs are ordered randomly. The time that the each battery lasts (in minutes) under continuous usage is recorded.

The two-way ANOVA table for response variable Time and factors Brand and Device is given below, albeit with some statistics omitted.

Source	df	Sums of Squares	Mean Square	F-ratio	p-value
Brand	2	2973.5	–	2.957	–
Device	3	2496534	832178	1654.9	< 0.00001
Error	–	3017.17	–		
Total	–	2502525			

7. What is the degree of freedom for residual/error?
- A. 2
 - B. 3
 - C. 6
 - D. 11
 - E. 12
8. What is the p-value for Brand?
- A. 0
 - B. 0.0489
 - C. 0.0564
 - D. 0.1278
 - E. 0.1792
9. **True/False** Using the significance level $\alpha = 0.05$, what is the conclusion for the brand effect?
- A. There is not enough evidence to conclude that the brand has an effect on how long the battery lasts.
 - B. There is strong evidence to conclude that the brand has an effect on how long the battery lasts.

Problem 10-12 Exams

To discourage cheating, a professor makes three different versions of an exam. For the 105 students in her class, she makes 35 copies of each version. The 105 exams are randomly scrambled, and one copy is given to each student. After the exam, the professor is concerned that one version might have been easier than the others. She uses a one-way ANOVA to test whether the average score was different for the three versions. The ANOVA table of the results, with some statistics omitted, are shown below.

Source	df	Sums of Squares	Mean Square	F-ratio	p-value
Version	–	771.943	–	–	–
Error	102	8883.49	87.093		
Total	–	2502525			

10. What is the degree of freedom for Version?

- A. 2
- B. 3
- C. 102
- D. 104
- E. 105

11. What is the p-value for Version?

- A. 0.9857
- B. 0.9763
- C. 0.7080
- D. 0.0237
- E. 0.0143

12. **True/False** Using the significance level $\alpha = 0.05$, what is the conclusion?

- A. At least one version of the exam has a different mean score than the others.
- B. The mean scores of all three versions of the exam are different.

Problem 13-16 Rent

A random sample of 76 apartments is collected near a university. All of the apartments in the sample have between 1 and 6 bedrooms. The variables recorded for each apartment are Rent (in dollars) and the number of Bedrooms. The regression output, with some statistics omitted, is:

Dependent variable is Rent

$$R^2 = 62.0\% \quad R^2(\text{adjusted}) = 61.5\%$$

$$s = 364.4$$

Variables	Coeff	SE(Coeff)	t-ratio	p-value
Constant	357.795	111.6	–	–
Bedrooms	400.554	36.42	–	–

13. What is the regression equation?

- A. $\hat{y} = 400.554x + 357.795$
- B. $\hat{y} = 400.554x + 364.4$
- C. $\hat{y} = 357.795x + 400.554$
- D. $\hat{y} = 400.554x + 64.42$
- E. $\hat{y} = 357.795x + 111.6$

14. What is the degree of freedom for Residual/Error?

- A. 2
- B. 5
- C. 74
- D. 75
- E. 76

15. Create a 95% confidence interval for the slope.

A. (327.99, 473.12)

B. (243.85, 556.26)

C. (329.17, 471.94)

D. (364.13, 436.97)

E. It cannot be determined.

16. **True/False** Based on the confidence interval for the slope, which of the following two statements is true?

A. For each additional bedroom, the rent increases by \$400.55.

B. There is strong evidence that the number of bedrooms is positively associated with the amount of rent charged.

Problem 17-19 Paper airplane

A student wants to build a paper airplane that gets maximum flight distance. She tries three ways of bending the wing (down, flat, and up) and two levels of nose weight (no and yes - a paper clip). She randomizes the 12 runs (each condition replicated twice). The analysis of variance for the 12 runs, with some statistics omitted, is shown in the table below.

Source	df	Sums of Squares	Mean Square	F-ratio	p-value
Wing Bend	–	13565.2	–	–	–
Weight	–	6768.75	–	–	–
Interaction	–	186.5	–	–	–
Error	–	266.5	–		
Total	11	20786.9			

17. What is the degree of freedom for residual/error?
- A. 1
 - B. 2
 - C. 3
 - D. 6
 - E. 11
18. **True/False** Calculate the p-value for the interaction term. Using the significance level $\alpha = 0.05$, is it reasonable to work with an additive model? That is, is it reasonable omit the interaction term?
- A. Yes.
 - B. No.
19. Compute the F-ratios for Wing Bend and Weight. Which of the following statements is true?
- A. Both the wing bend and the nose weight are associated with the distance the plans flies.
 - B. Only the wing bend is associated with the distance the plane flies.
 - C. Only the nose weight is associated with the distance the plane flies.
 - D. Neither the wing bend nor the nose weight are associated with the distance the plans flies.
 - E. The association cannot be determined.

Problem 20-24 Athletics

Of the 23 first year male students at State U. admitted from Jim Thorpe High School, 8 were offered baseball scholarships and 7 were offered football scholarships. The University admissions committee looked at the students' composite ACT scores (shown in the table), wondering if the University was lowering their standards for athletes. Assuming that this group of students is representative of all admitted students. Here is their composite ACT score.

Baseball	Football	Non-athletes
25	22	21
22	21	27
19	24	29
25	27	26
24	19	30
25	23	27
24	17	26
23		23

20. Find the sample standard deviation of the Football group.
- A. 2.0658
 - B. 2.7324
 - C. 2.9489
 - D. 3.0438
 - E. 3.2877
21. Perform an appropriate hypothesis test, assuming the conditions are satisfied. What is the p-value?
- A. < 0.00001
 - B. 0.0080
 - C. 0.016
 - D. 0.023
 - E. 0.286
22. **True/False** Using the significance level $\alpha = 0.05$, what is your conclusion?
- A. The average composite ACT score for non-athletes is higher than the average composite ACT score for the athletes.
 - B. There is evidence that the average composite ACT scores for the three groups are not the same.

We re-organize the data in two groups instead of three.

Athletics	Non-athletes
25	21
22	27
19	29
25	26
24	30
25	27
24	26
23	23
22	
21	
24	
27	
19	
23	
17	

23. Perform an appropriate hypothesis test. What is the p-value?

- A. < 0.00001
- B. 0.0080
- C. 0.016
- D. 0.023
- E. 0.286

24. **True/False** Using the significance level $\alpha = 0.05$, what is your conclusion?

- A. The average composite ACT score for non-athletes is higher than the average composite ACT score for the athletes.
- B. There is evidence that the average composite ACT scores for the three groups are not the same.

Problem 23-26 Diamonds

An article in the *Journal of Statistics Education* reported the price of diamonds of different sizes in Singapore dollars (SGD). The following table contains a data set that is consistent with this data, adjusted to US dollars in 2004:

2004 US \$	Carat
494.82	0.12
688.24	0.15
748.10	0.16
768.03	0.17
944.90	0.18
1076.18	0.19
1105.03	0.20
1071.75	0.21
1289.20	0.23
1508.88	0.25
1504.44	0.26
1597.63	0.27
1826.18	0.28
1908.28	0.29
2038.09	0.32
2096.89	0.33
2409.76	0.35

25. Create a linear regression model to predict diamond costs from the size of the diamond.

A. $\hat{y} = -603.29 + 8225.1x$

B. $\hat{y} = 8225.1 - 558.52x$

C. $\hat{y} = -603.29 + 7892.7x$

D. $\hat{y} = 7892.7 - 603.29x$

E. $\hat{y} = -558.52 + 8225.1x$

26. Let b_1 be the parameter for the slope. The standard error of the slope is

$$SE(b_1) = 239.1$$

We also have

$$\text{invT}(0.975, 15) = 2.13145, \quad \text{invT}(0.975, 16) = 2.1199$$

Create a 95% confidence interval for the slope.

- A. (7300.3, 8564.9)
 - B. (7715.4, 8734.7)
 - C. (7718.2, 8732.0)
 - D. (7383.1, 8402.3)
 - E. (7385.8, 8399.6)
27. **True/False** Writing the 95% confidence interval for the slope as (x_1, x_2) . How do we interpret it?
- A. We are 95% confident that for each additional carat in the diamond weight, the average increase in the cost of the diamond will be between x_1 dollars and x_2 dollars.
 - B. We are 95% confident that the average cost of the diamond will increase by $\frac{x_1+x_2}{2}$ dollars for each additional carat.
28. What is the correlation coefficient for the model?
- A. -0.987
 - B. 0.987
 - C. -0.994
 - D. 0.994
 - E. 1

Problem 29-30 Anxiety about Statistics

Before you took this course, you may have heard many stories about Statistics courses. Oftentimes parents of students have had bad experiences with Statistics courses and pass on their anxieties to their children. To test whether taking Statistics REDUCES students' anxieties about Statistics, a Statistics instructor gave a test to rate student anxiety at the beginning and at the end of his course. Anxiety levels were measured on a scale of 0-10, 10 being very high anxiety and 0 being no anxiety. Here are the data for 16 randomly chosen students from a class of 180 students:

Pre-course anxiety level	7	6	9	5	6	7	5	7	6	4	3	2	1	3	4	2
Post-course anxiety level	4	3	7	3	4	5	4	6	5	3	2	2	1	3	4	3

29. Perform an appropriate hypothesis test. What is the p-value?

- A. < 0.00001
- B. 0.00068
- C. 0.00136
- D. 0.054
- E. 0.172

30. **True/False** Based on the p-value, what is your conclusion?

- A. There is strong evidence that taking the statistics course reduces the anxiety level of the students about statistics.
- B. It is a great fun to take this statistics course!

This page is intentionally left blank.

This page is intentionally left blank.