

K300 (4392) Statistical Techniques (Fall 2007)**Final Exam Sample**

Instructor: Hun Myoung Park

kucc625@indiana.edu, (317) 274-0573

This is a sample of the final exam on December 14. The actual final exam will look like this sample but not the same as this sample.

If you have no idea, check powerpoint slides first and then go over seven assignments (ignore assignment 1). All questions here except one for the least squares method are included in 7 assignments. Definitely, the Lecture Note 1 is the key reading because most questions are about hypothesis testing using the p-value approach. Without knowing the approach correctly, you will NEVER get a high score in the final exam.

1. (5 points) Determine the level of measurement of the following variables. You may pick two arbitrary values, and then check “meaningful order”, “meaningful difference”, “discrete/continuous”, and proper operators (e.g., +, -, *, /). See Q1 of assignment 2, Q2 and Q3 of the midterm exam.

- 1) Your letter grade this semester
- 2) Your point you obtained from eight assignments
- 3) Mascot name of a college or university
- 4) IndyGo bus lines (e.g, 4, 8, ...)
- 5) The standing of a NFL team

2. (3 points) What is “statistical independence?” See Q5 of the midterm exam.

3. (4 points) The probability that a college student is a single is $P(S)$, the probability that a college student is married is $P(M)$, and the probability that a college student is a single AND get married at the same time is $P(S \cap M)$. Let us suppose $P(S)=.8$ and $P(M)=.2$. Are event S (being single) and M (being married) statistically independent? See Q13 and Q14 of assignment 3, Q5 of the midterm exam.

4. (3 points) What is the Central Limit Theorem? See Q6 of the midterm exam.

5. (5 points) What is the p-value? What does a small p-value, say .003, mean? See Q1.8 and Q2.8 of assignment 5 and slides.

6. (10 points) Your boss wants to know if taking a special vitamin is effective in strength of athletes (Example 9-12 on page 493). The following is SPSS output of the paired sample t-test. The null hypothesis is $\bar{\delta} = 0$, where δ_i is the difference of before and after measures of observation i . See Q2 of assignment 5.

Paired Samples Test

	Paired Differences	t	df	Sig. (2-tailed)

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
		Lower	Upper	Lower	Upper	Lower			
Pair 1	before - after	-2.37500	4.83846	1.71065	-6.42006	1.67006	-1.388	7	.208

- 6.1 Show how the t statistic is computed. Do not simply copy the formula.
- 6.2 Test the null hypothesis at the .05 level. Of course, you just need to take the p-value approach. DO NOT skip any one of five steps.
- 6.3 How do you explain the result to your boss? Is the special vitamin effective? Do not simply say “Yes” or “No.” Write down a complete sentence clearly so that your boss can fully understand the conclusion. The format is substantive interpretation with ($p < \dots$) at the end of the sentence (see assignments for examples). Of course, your boss knows what the p-value is.

7. (10 points) Your boss wants to know if two types of microwave oven have different average price. You went out and surveyed to get sample data. The following is a part of SPSS output for the independent same t-test. See Q3 of assignment 5.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
price	Equal variances assumed	.401	.537	2.774	14	.015	47.500	17.123
	Equal variances not assumed			2.774	12.690	.016	47.500	17.123

- 7.1 Your boss asked you, “What is F and what does that mean?” You need to interpret the meaning of F by testing the null hypothesis of $\sigma_1^2 = \sigma_2^2$ at the .05 level. Do not skip any of five steps. Of course, you just need to take the p-value approach. Do not try to look up the F distribution table to get the critical value.
- 7.2 By chance, you have the same t statistic. Your boss wants to know how t is computed. Show him the computation.
- 7.3 Your boss became wondering why there are two p-values. He knows p-values well but does not know t-test at all. How would you explain to your boss? Which p-value would you report as a correct one?
- 7.4 Test the null hypothesis of $\mu_1 = \mu_2$ using the p-value approach.
- 7.5 Your boss wants to hear the final conclusion from you. What is your “professional suggestion” on this issue? You need to write down a **complete sentence** to summarize the conclusion clearly. Do not simply say “reject H_0 ” or “reject H_0 ”; otherwise, your boss will get upset and probably you will be fired.

8. (5 points) Your boss wants to know if there is a significant difference in the average prices of different types of microwave ovens (Question 17 on page 601). The following is the SPSS output with some statistics removed. See Q1 of assignment 6.

ANOVA

price

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(1)	(2)	10864.867	(4)	.001
Within Groups	20402.083	19	(3)		
Total	42131.818	21			

- 8.1 Fill the all four blank cells from (1) through (4). Report N and T (# groups).
- 8.2 Conduct the hypothesis testing using the p-value approach and draw a conclusion. Use μ (e.x., $\mu_1, \mu_2...$) in a hypothesis. The test size is .05. YOU MUST follow all five steps.
- 8.3 Is there any significant different in prices? Write down a conclusion in a complete sencesce so that your boss can fully understand. Again your boss knows what the p-value is.

9. (5 points) Your boss wants to know if medication (new drug=1 and placebo=2) is independent of effectiveness (effective=1 and not effective=2). This is Question 20 on page 587. The following is the SPSS output with some unnecessary statistics taken out. See Q3 of assignment 6.

drug * effect Crosstabulation

			effect		Total
			1	2	
drug 1	Count		32	9	41
	Expected Count		25.4	15.6	41.0
2	Count		12	18	30
	Expected Count		18.6	11.4	30.0
Total	Count		44	27	71
	Expected Count		44.0	27.0	71.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.6421	1	.001		

- 9.1 State the null and alternative hypotheses clearly (Step 1)
- 9.2 Is this chi-square test valid? Check the expected frequency (count).
- 9.3 Show me how the degree of freedom is computed. Do not simply copy the formula.
- 9.4 Conduct the hypothesis testing at the .05 level. You may skip step 1.

- 9.5 Is the new drug effective? (Do not simply say “Yes” or “No”). Write down a conclusion in a complete sentence so that your boss can fully understand.

10. (5 points) Your boss wants to know if the number of fires (X) is related to the number of acres burned (Y). See Q2 of assignment 7.

Correlations

		X	Y
X	Pearson Correlation	1	.769(*)
	Sig. (2-tailed)		.026
	N	8	8
Y	Pearson Correlation	.769(*)	1
	Sig. (2-tailed)	.026	
	N	8	8

* Correlation is significant at the 0.05 level (2-tailed).

- 10.1 Report r . How would you interpret the (Karl Pearson) bivariate correlation coefficient r so that your boss can understand?
- 10.2 Is the correlation coefficient statistically significant? In order to answer, you should test the null hypothesis of ρ (rho)=0. Follow all five steps and use the .05 significance level.
- 10.3. Is your interpretation in 10.1 reliable? If $\rho=0$, it is not reliable.

11. (5 points) What is the least squares method? Describe the method to get ordinary least squares estimators. One full sentence will do. See slides.

12. (20 points) Your boss has a regression model in his mind. It is $Y = \beta_0 + \beta_1 X + \beta_2 D + \varepsilon$, where D is a dummy variable. Let us call the baseline ($D=0$) group 0. You collected data and regressed Y on X and D to get the following output. Unfortunately, your brother spilled coffee on the output and thus you cannot read two statistics as follows. See Q4 of assignment 7 and Q1 of assignment 8.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-26.832	21.587		-1.243	.269
	X	1.038	.338	.775	(1)	.028
	D	(2)	8.236	-.299	-1.186	.289

a Dependent Variable: Y

- 12.1 Fill the two blank cells (1) and (2). Your boss wants to see all statistics.
- 12.2 Interpret the coefficient of X . Your boss knows the concept of p-values but does not know the meaning of regression coefficients. How could you explain the coefficient to your boss?

- 12.3 Test the null hypothesis of $\beta_1 = 0$. Follow all five steps. Is your interpretation in 10.2 reliable? Of course, use the .05 significance level.
- 12.4 Report regression equation for group 0 and another equation for group 1.
- 12.5 Interpret the coefficient of dummy variable D. Your boss wants to know if there is any big (statistically significant) difference in intercept between group 0 and 1? You just need to read the p-value; do not follow all five steps here.

13. (10 points) Looking at the SPSS output below (ANOVA table of the regression model described in Q12 above), your boss asked, “Can you tell me what the F is?” Conduct the hypothesis testing using the p-value approach at the .05 level. Again, follow all five steps. Be careful when writing the alternative hypothesis. See Q4.11 of assignment 7 and Q1.3 of assignment 8.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1449.831	2	724.915	5.346	.057(a)
	Residual	678.044	5	135.609		
	Total	2127.875	7			

a Predictors: (Constant), D, X

b Dependent Variable: Y

14. (10 points) Finally, your boss wants to know the meaning of R^2 . How could you explain the following R^2 to your boss? You may show how R^2 is computed and describe its meaning.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.825(a)	.681	.554	11.645

a Predictors: (Constant), D, X

Your boss told you, “You are saying t, F, R^2 ... humm ... I am sorry. I’m lost. Please make it short. Is it a good or bad model?” As a public manager, you should provide your boss with your “professional opinion.” What is your answer to the question? See Q1.2 of assignment 8.