

## BIOS 600.001 Final Exam

December 11, 2012

**Honor Statement.** I pledge that I have not used any reference materials (including electronic materials) during this examination. I pledge that I have neither given nor received any aid from any other person during this examination, and that the work presented here is entirely my own. I furthermore pledge that I will not reveal any of the material on this examination, either in the form of the exact question or the topics covered, to any person for any reason until the end of fall term. I pledge that I will report all Honor Code violations observed by me. I understand that if I have committed any of the above, I have violated the UNC Honor Code.

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

INSTRUCTIONS: No electronic devices are permitted. Full credit will be given for correct answers that are unsimplified, e.g.  $1 + 2(3) + 4(5)$  is an acceptable form. There are 100 points on this exam not counting the bonus. Please ask if you have any questions. Note that this exam will be videorecorded.

EXAMS ARE DUE AT 10:45. At 10:45am you should be signing the honor statement and turning in the exam. Exams turned in after 10:45 am will be penalized one point for the first minute delay (10:46) and 5 points per minute thereafter (starting at 10:47).

1. (15 points total) Researchers are designing a study with  $\alpha = 0.05$  and power  $1 - \beta = 0.80$ . Compute the following conditional probabilities concerning the outcomes of this future study. If it is not possible to compute the probability, explain what additional information is needed.

(a) (3 points)  $Pr(\text{Do not reject } H_0 \mid H_0 \text{ true})$

0.95

(b) (3 points)  $Pr(\text{Reject } H_0 \mid H_0 \text{ true})$

0.05

(c) (3 points)  $Pr(\text{Do not reject } H_0 \mid H_0 \text{ false})$

0.20

(d) (3 points)  $Pr(\text{Reject } H_0 \mid H_0 \text{ false})$

0.80

(e) (3 points)  $Pr(H_0 \text{ false} \mid \text{Reject } H_0)$

We do not have enough information to calculate this probability. Using Bayes's rule, we would need to have the probability that  $H_0$  is false or a related probability.

2. (12 points) A researcher is interested in the potential association between artificial colorings in foods and child hyperactivity. A group of children will be randomized in a 1:1 ratio to receive either a typical American diet or a diet free of all artificial food colorings. After the study period, hyperactivity response will be measured on a continuous scale. The power calculation indicates that a sample size of 350 is sufficient to detect a difference of 0.3 standard deviations in the hyperactivity level between the two groups, assuming a type I error rate of 5% and 80% power for a two-sided hypothesis test.

- (a) (4 points) Researchers want to increase the power to 90%. How will increasing the power affect the required sample size if the other factors remain fixed?

To increase the power to 90%, we will need to increase the sample size.

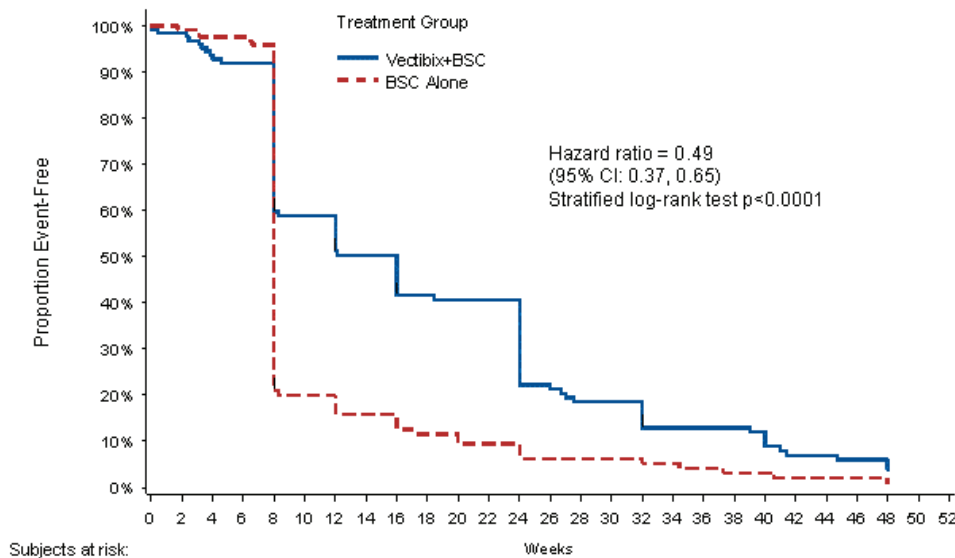
- (b) (4 points) Researchers decide instead they would like to detect a difference of 0.5 standard deviations. How will changing this minimum detectable difference affect the power if the other factors remain fixed?

If the minimum detectable difference is larger, we will have greater power to detect it. (Note: common mistake was to assume the standard deviation itself was increased, rather than the minimum detectable difference.)

- (c) (4 points) The investigators get a new grant and can increase the sample size to 500. How will this affect the power if the other factors remain fixed?

With increased sample size we will have increased power.

3. (12 points) A randomized, controlled trial was conducted to evaluate a new treatment (Vectibix) and best supportive care (BSC) for colorectal cancer versus BSC alone. The Kaplan-Meier plot and result of a log-rank test are provided below. The outcome of the study was the event of death or disease progression, so the desired outcome for a patient is to remain event-free.



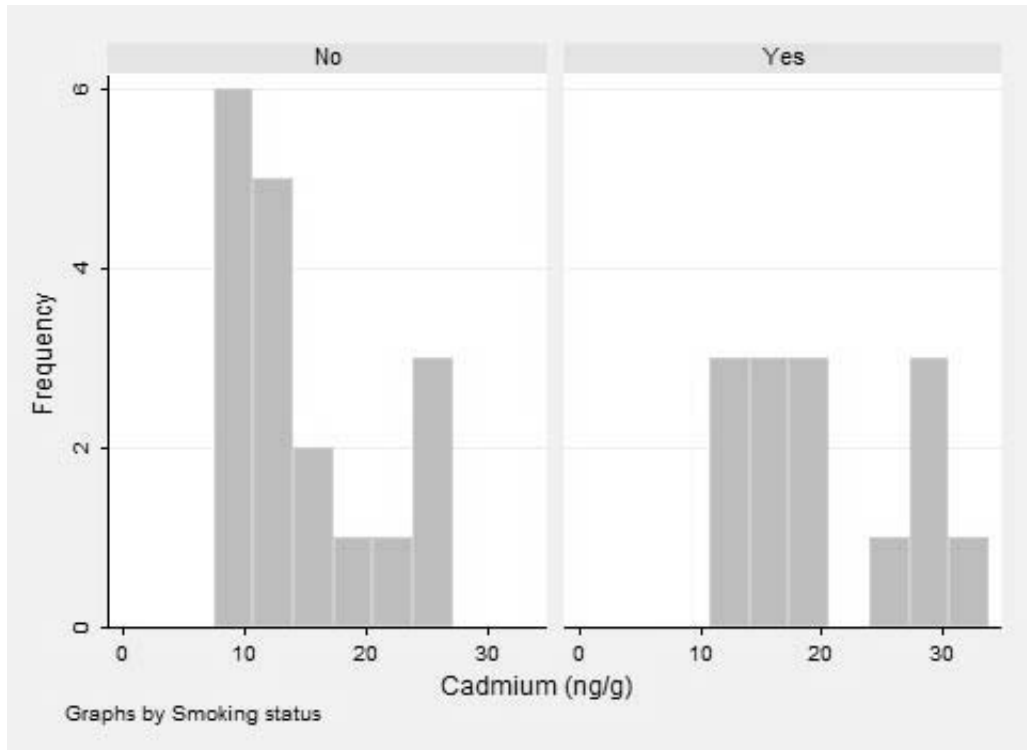
- (a) (6 points) What are the estimated median survival times in each of the 2 groups?

The median survival on BSC alone is 8 weeks. The median survival on the Vectibix+BSC group depends on the technical definition of the survival function one uses, and I accepted answers from 12-16 weeks.

- (b) (6 points) Which treatment is superior: Vectibix+BSC, BSC alone, or neither? Explain your answer.

Vectibix+BSC is the superior treatment and is associated with delayed time to event. Those on Vectibix+BSC had 0.49 (0.37, 0.65) the hazard of the event as those on the BSC alone group (logrank test  $p < 0.0001$ ).

4. (14 points total) In a study of factors thought to be responsible for adverse effects of smoking on human reproduction, cadmium level determinations (ng/g) were made on placental tissue of a sample of 14 smoking mothers (smoking status=yes) and an independent random sample of 18 nonsmoking mothers (smoking status=no). The data are summarized in the figure below.



- (a) (2 points) Approximately what is the median cadmium level in placental tissue of nonsmokers?

Around 12 ng/g

- (b) (2 points) Approximately what is the 75th %ile of cadmium levels in placental tissue of smokers?

Around 27 ng/g

- (c) (5 points) Which statistical test is most appropriate for evaluating whether cadmium levels were different on placentas from smoking and nonsmoking mothers? Explain your choice and provide the null and alternative hypotheses corresponding to this test.

With such a small sample, it's difficult to evaluate the assumptions behind any test. To be safe, I would use the Wilcoxon rank sum test, as it has more relaxed assumptions than the two sample t-test. The null hypothesis of the Wilcoxon rank sum test is that the median cadmium levels are the same among smokers and nonsmokers, and the alternative is that the medians are different.

- (d) (5 points) Suppose you carry out the test and obtain  $p = 0.008$ . Provide a short conclusion and explanation of the results suitable for publication in a scientific journal. Median cadmium levels differed by smoking status ( $p=0.008$ ), with estimated median levels of around 18 ng/g in smokers and only around 12 ng/g in nonsmokers.

5. (15 points total) Researchers used age (in years) and education level (years of schooling) to predict the capacity to direct attention (CDA) in elderly subjects. CDA refers to neural inhibitory mechanisms that focus the mind on what is meaningful, while blocking out distractions. The study collected information on 71 older women with normal mental status. The CDA measurement is continuous, with scores ranging from -7.7 to 9.6 and higher scores corresponding with better attentional functioning.

Researchers fit the linear regression model

$$CDA_i = \beta_0 + \beta_1 age_i + \beta_2 education_i + \varepsilon_i$$

and obtained the following parameter estimates.

Parameter	Estimate	95% CI
$\beta_0$	5.5	(-3.2, 14.2)
$\beta_1$	-0.2	(-0.3,-0.1)
$\beta_2$	0.6	(0.3, 0.8)

- (a) (5 points) Is the association between age and CDA statistically significant, controlling for education? Explain why or why not and describe how one additional year in age is expected to affect the CDA score for a fixed level of education.

We know the association between age and CDA is statistically significant ( $p < 0.05$ ) because the 95% confidence interval does not contain the null value of 0. Based on these data, elderly subjects who are one year older have CDA scores that are 0.2 (0.1, 0.3) units smaller on average than their counterparts with the same educational level. (NOTE on common misconception: The point estimate should ALWAYS be inside the confidence interval, whether or not the result is statistically significant.)

- (b) (5 points) What is the expected CDA score for a woman who is 70 years old and who has 20 years of education?

$$5.5 - 0.2(70) + 20(.6) = 5.5 - 14 + 12 = 3.5$$

- (c) (5 points) Suppose two women both have 20 years of education, but one is 80 years old and the other is 70 years old. Based on the model, how should you expect the 80 year old woman's CDA score to differ from the 70 year old woman's CDA score? Be as specific as possible.

The expected score for an 80 year old woman with 20 years of education is  $5.5 - 0.2(80) + 20(.6) = 5.5 - 16 + 12 = 1.5$ . Thus we expect an 80 year old woman with the same level of education to have a CDA score 2 points lower on average than her 70 year old counterpart.



6. (20 points total) In 2007, the Finnish National Board of Education and the National Public Health Institute recommended to schools that they quit selling candies and soft drinks. Finnish researchers wished to determine whether the proportion of schools selling candy and soft drinks changed from 2007 to 2008 (after the national recommendation). They provided surveys to all upper comprehensive schools in Finland in 2007 (before the recommendation) and again to the same schools in 2008 (after the recommendation).

(a) (4 points) What are  $H_0$  and  $H_A$ ?

$H_0$ : The proportion of schools selling soda and candy is the same in 2007 and 2008.

$H_A$ : The proportion of schools selling soda and candy is different in 2007 and 2008.

(b) (4 points) Which test should you use to test  $H_0$ ?

Because we visit the same schools both years, the data are matched by school. We would use McNemar's test to evaluate this hypothesis.

- (c) (4 points) Among the 150 schools that sold candy and soft drinks in both 2007 and 2008, researchers compared the number of items sold per week in 2007 and 2008 and wished to test whether average sales levels were the same in the two years. What are  $H_0$  and  $H_A$  in this setting, and which test should you use to evaluate  $H_0$ ?

$H_0$ : The mean number of items sold per week is the same in 2007 and 2008.

$H_A$ : The mean number of items sold per week is not the same in 2007 and 2008.

We can evaluate this hypothesis using the paired t-test or a nonparametric procedure for paired data, if you prefer.

- (d) (4 points) Suppose the estimated difference in the mean number of items sold per week among the 150 schools who sold candy and soft drinks in both years was 100 items/week, with significantly more items/week sold in 2007 than in 2008 ( $p < 0.05$ ). Which 95% confidence interval for the population mean difference is correct?

i. ~~X~~(-101.6, -98.4) Only 1 point off if you chose this one – the point estimate was 100 and should be inside the confidence interval.

ii. (98.4, 101.6)

iii. ~~X~~(98.4, 103.2)





iv. ~~X~~(-98.4, 103.2)

- (e) (4 points) Suppose the p-value in (d) above is exactly  $p = 0.02$ . Provide an appropriate interpretation of this p-value. Specifically, this p-value is the probability of what event?

The probability of getting a difference of 100 items/week or more in either direction in our sample, if the true population difference is 0, is 0.02. If we are conducting our test at the 0.05 level, we reject  $H_0$ . We estimate there were 100 (98.4, 101.6) fewer items per school per week sold in 2008 than in 2007.

7. (12 points) The National Audubon Society has sponsored a “Christmas Bird Count” for over 100 years. This event was founded by ornithologist Frank Chapman in 1900 as a new tradition to replace the usual Christmas “Side Hunt,” in which people formed teams to see which one could kill the greatest number of birds and small game animals.

You are testing whether the mean number of birds of four species is the same, corresponding to  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ . Researchers collected the data by counting the number of birds in seven different geographical areas for each species. They averaged the seven counts for each species to obtain  $\bar{y}_i$ ,  $i = 1, 2, 3, 4$ , and reported the estimated means and the estimated standard deviations  $s_i$  for each species. You will use the ANOVA model  $y_{ij} = \mu_i + \varepsilon_{ij}$ ,  $i = 1, 2, 3, 4$ ,  $j = 1, \dots, 7$  to analyze these data.

Species	Parameter	$n_i$	$\bar{y}_i$	$s_i$
Partridges 	$\mu_1$	7	16.44	0.87
Turtle Doves 	$\mu_2$	7	16.51	0.44
French Hens 	$\mu_3$	7	16.07	0.40
Colly Birds 	$\mu_4$	7	14.96	1.07

The F statistic corresponding to the null hypothesis  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$  is 6.45 (3 ndf, 24 ddf) with a corresponding p-value of 0.002.

- (a) (4 points) Do you reject or fail to reject  $H_0$ ? Does the p-value provide evidence of any difference between the groups, or not?

We reject  $H_0$  and conclude that the mean of at least one of the groups is different. The p-value tells us that if there is really no difference among the groups, then getting a test statistic as large as we saw would be extremely unlikely.

In order to investigate differences between species, the investigators tested differences between each pair of species and have provided the following table.

Parameter	Estimate	95% CI	p-value
$\mu_1 - \mu_2$	-0.07	(-0.89, 0.75)	0.860
$\mu_1 - \mu_3$	0.37	(-0.45, 1.19)	0.363
$\mu_1 - \mu_4$	1.49	(0.67, 2.31)	0.001
$\mu_2 - \mu_3$	0.44	(-0.38, 1.26)	0.280
$\mu_2 - \mu_4$	1.56	(0.74, 2.38)	0.001
$\mu_3 - \mu_4$	1.11	(0.29, 1.93)	0.010

- (b) (4 points) Using this table, describe differences in mean bird counts by species using language suitable for journal publication.

We conducted one-way ANOVA to evaluate the null hypothesis that the counts of the four bird species were equal. Because the overall test was significant ( $p=0.002$ ), we conducted step-down tests to evaluate pairwise differences among species. While there were no significant differences in the counts of partridges, turtle doves, and French hens, there were fewer colly birds than birds of each of the other species. In particular, we estimated that there were 1.49 (0.67, 2.31) more partridges per area than colly birds, 1.56 (0.74, 2.38) more turtle doves per area than colly birds, and 1.11 (0.29, 1.93) more French hens than colly birds per area. Perhaps this phenomenon is due to another old tradition of baking pies of four-and-twenty blackbirds (colly birds). (NOTE: OK also if you apply a Bonferroni correction.)

- (c) (4 points) Circle all of the assumptions below that are needed for ANOVA to be valid.

- i. Equal variances of the counts of each bird species.
- ii. ☒ Equal counts of each bird species.
- iii. ☒ Dependent observations
- iv. ☐ Independent observations
- v. ☒ Each bird must be capable of laying AN OVA. (OK, technically an ovum for the Latin scholars out there.) As several of you pointed out, we would incur considerable bias by excluding infertile or male birds!
- vi. ☒ A NOVA must not take out Earth before you can run the analysis in Stata. Though certainly the exercise would be pointless if this occurred!

- (d) 5 POINT BONUS: Famed researchers Drs. Donder and Blitzzen wish to test a firmly established historical null hypothesis instead of the null hypothesis  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ . In particular, the true *status quo* (null hypothesis) is that for every partridge seen, there should be on average two turtle doves, three French hens, and four colly birds. They think the hypothesis test should evaluate this more established null hypothesis. Express the new null hypothesis in terms of the  $\mu_i$ ,  $i = 1, 2, 3, 4$ .

$$H_0 : \mu_1 = \frac{1}{2}\mu_2 = \frac{1}{3}\mu_3 = \frac{1}{4}\mu_4$$

OR

$$H_0 : 12\mu_1 = 6\mu_2 = 4\mu_3 = 3\mu_4$$

OR

$$H_0 : \begin{pmatrix} \mu_2 = 2\mu_1 \\ \mu_3 = 3\mu_1 \\ \mu_4 = 4\mu_1 \end{pmatrix}$$