Name: _____

Biostatistics 1st year Comprehensive Examination: Applied Take-Home Exam

Due June 1st, 2018 by 5pm. Late exams will not be accepted.

Instructions:

1. ***This is exam is to be completed independently. Do not discuss your work with anyone else.***
2. There are 2 questions and 4 pages.
3. Answer each question to the best of your ability. Read the exam carefully.
4. Be as specific as possible and type up or Latex your answers.
5. *This is a take-home examination. You may consult books, notes, and papers. You may use the Internet as a research resource. However, you may not consult or discuss this exam with another human being, directly or indirectly, nor may you seek help from another individual on the Internet (e.g., no posting questions to chat rooms or message boards).*
6. **If you have any questions, please email Professors Greevy and Blume, AND also text Prof. Greevy at 615.403.8463 to let him know you just sent the email.** This will help us reply quickly. Do not worry about being polite. Contact Professor Greevy as needed; call for emergencies.
7. Turn in your exam by emailing it to Prof. Blume at j.blume@vanderbilt.edu **AND** Amanda Harding at amanda.harding@vanderbilt.edu **AND** Prof. Greevy at robert.greevy@vanderbilt.edu (all three). Your exam is not submitted until one of these three has confirmed your exam was received. Leave yourself time to handle any email problems when submitting. *If you do not receive confirmation you should assume that your exam has not been received.*
8. ***Vanderbilt's academic honor code applies; adhere to the spirit of this code.***

Exam Link: https://www.dropbox.com/s/cgjhhs58tewfzq8/TakeHome2018yr1.pdf?dl=0
Data Link: https://www.dropbox.com/s/u4gtvj3wjkf89nz/downer.dta?dl=0

Question	Points	Score	Comments
1	100		
2	100		
Total			

1. **Background:** A pharmaceutical company has been commissioned to rapidly develop a new antibiotic that will be more effective against a specific bacterial infection that is highly resistant to standard therapy. To speed up development the research team has decided to simultaneously test 10 new candidate medications against standard therapy in a randomized controlled trial. This has created methodological challenges the biostatistical team is working to resolve.

   They have agreed on a study design that will collect the study outcome (infection resolved or not resolved within 14 days) on 400-500 patients for the standard therapy (control arm) and 50-100 patients for each of the 10 candidate therapies (intervention arms). For sake of estimating the operational characteristics of the proposed analysis approaches, they assume each of the sample sizes will be uniformly distributed within those ranges. They will also assume all patients' outcomes will be independent of each other and *the probability of resolution under standard therapy is 0.10*.

   A group of biostatisticians (ONE) is concerned about controlling for the familywise Type I error (FWER) for testing the 10 treatments against the control group all at once. They propose first performing a single chi-square test on all 11 groups to test for any differences. This test would be evaluated at a 5% significance level. If it was not significant, they would conclude there were no differences in resolution rates. If it was significant, they would then individually test each of the treatments against the control with a chi-square test each at a 5% significance level. Any of those tests that returned p-values < 0.05 would be deemed statistically significant.

   Another group of biostatisticians (TWO) is concerned about controlling for the FWER without losing too much Power. They propose going straight to individually testing each of the treatments against the control with chi-square tests, but doing a Bonferroni correction to maintain a 5% familywise error rate. Any of those tests that returned p-values less than this threshold would be deemed statistically significant.

   A third group of biostatisticians (THREE) is concerned about not focusing on what is clinically meaningful and meaningless. They ask the researchers what a clinically meaningful improvement would be. The researchers reply around a 15% absolute gain, e.g. if the control had a 10% resolution rate, a treatment with a 25% rate would be truly meaningful. Then the biostatisticians ask what would be a clinically trivial change from the standard therapy. The team replies anything within a 2% absolute difference, e.g. going from 10% to 12%, would be trivial -- essentially meaningless. Group THREE proposes calculating the 95% confidence intervals for the risk difference for each of the 10 interventions vs the control, and deeming any interval that falls above a 2% absolute improvement to be clinically nontrivial, i.e. when the CI has a lower bound > 0.02.

   For the tests, all the biostatisticians agree on using the default chi-square test function in R, chisq.test(), with the default function settings. For the risk difference confidence intervals, they agree to use the Agresti-Caffo interval as implemented by the wald2ci() function in the R package PropCIs.

Questions:

a. Write a brief paragraph comparing the familywise Type I error rates (FWER) of the three analysis methods (Proposals ONE, TWO, and THREE) for testing the 10 new therapies versus the control therapy. Report your error rates to three decimal places and design your methods so you have a high degree of certainty on the first two decimal places.

Write a second paragraph describing your methods for estimating these error rates in detail and provide supporting code and figures as appropriate. – Note: Clearly explaining what you were trying to do is very important if it turns out you have an error in your code. It helps the graders give you partial credit.

b. Write a brief paragraph offering intuition on why each the three methods behaved the way they did in part a. Attempt to *explain why* each method either 1) achieved a FWER of 5% exactly, 2) was conservative by achieving a FWER less than 5%, or 3) failed to keep FWER under 5%.

c. Write a brief paragraph comparing the *Power* of the three analysis methods (Proposals ONE, TWO, and THREE) for detecting that one new treatment, call it treatment A, is different than the control therapy *assuming treatment A has a resolution rate of 26%* and all the other therapies retain the control rate of 10%. Here *Power* is referring to the probability of detecting the effect of treatment A specifically. Missing the effect of A and finding a false positive effect in a different therapy doesn't count as a successful study. Report your *Power* estimates to three decimal places and design your methods so you have a high degree of certainty on the first two decimal places.

Write a second paragraph describing your methods for estimating the *Power* in detail and provide supporting code and figures as appropriate.

d. Write a brief paragraph commenting on which method had the best *Power* out of the methods that controlled FWER at 5% or less. Attempt to offer insight into why the best performing method outperformed the others.

e. Assume that the settings for parts a and c were the only two scenarios that could occur with non-negligible probability and that they were equally likely to happen. Write a brief paragraph commenting on the *False Discovery* and *False Non-Discovery Rates* of the three approaches.

Write a second paragraph describing your methods for estimating the *FDR* and *FNR* in detail and provide supporting code and figures as appropriate. After the study was done and it was being analyzed, analysts wouldn't omnisciently know there is at most one real effect like we are assuming here. So it is important to allow the methods to potentially find more than one significant result in a given study.

f. Explain why the *FDR* and *FNR* are of interest to the statisticians designing the study. Attempt to offer insight into why the best performing method outperformed the others.

2. **Background:** For unknown reasons, dairy cows sometimes become recumbent—they lay down. Called *downers*, these cows may have a serious illness that may lead to their death. A study of blood samples of over 400 downer cows studied at the Ruakura New Zealand Animal Health Laboratory was conducted. A variety of blood tests were performed and the outcome (survived, died) was determined. The goal is to see how the risk of death is associated with key measures. The provided data were collected from veterinary records, and not all variables were recorded for all cows.

   Directions:
   Write a brief summary of your findings (two paragraphs or less) for each question or subpart. Put your code and raw output in an appendix. Only include code and raw output in the summary when explicitly requested in a question.

   Data:
   STATA file: https://www.dropbox.com/s/u4gtvj3wjkf89nz/downer.dta?dl=0
   View the variable labels and *codebook* output for definitions.

   Questions:

   a. What is myopathy and how might it relate to recumbency in cows?

   b. What is serum creatine phosphokinase - or just creatine kinase, *CK* - and how it is used in relation to myopathy?

   c. There are 213 observations with missing data on myopathy. What are some possible reasons that veterinarians may not record myopathic status?

   d. Treat "myopathy missing" as a separate category and describe the association between *Myopathy* and *Outcome*. [Suggestion: Recode so that 999 represents a missing value. The value-label associated with *Myopathy* will accommodate this missing value code.]

   e. Use the 3-category *Myopathy* as a predictor of death. Use *absence* as the reference category and discuss results in terms of odds ratios and their interval estimates.

   f. Does it make sense to use a ROC curve to assess the fitted model? Explain.

   g. Is it reasonable to log-transform *CK*? Explain.

   h. Use *log-CK* as a predictor of death. Discuss results in terms of odds ratios and their interval estimates.

   i. What does a ROC analysis say about the discriminative ability of the fitted model? Explain.

j. Does using both *Myopathy* and *log-CK* in the same model improve prediction? Does adding the square of *log-CK* help? Explain.

k. How would you advise a veterinarian to predict death in such a way that she/he does not need to make any calculations? Suggest a method that, conditioned on the presence or absence of *myopathy*, allows the veterinarian to predict death based on the *CK* measure. Use myopathy, *log-CK* and the square of log-CK in your model. Assume no prediction is made if *CK* is not measured. [**Note:** For this part, assume that a missing value on Myopathy means that myopathy was absent. With this change, there are 95 downer cows with myopathy present and 340 with myopathy absent.]

l. There are several other measures in the set of data. Use these to find the "best" model for predicting *Outcome*. Provide a profile of missing values for the retained regressors. How might this missingness affect your analysis?