Biost 517
Applied Biostatistics I

Midterm Examination Key

Instructions: Please provide concise answers to all questions. Rambling answers touching on topics not directly relevant to the question will tend to count against you. Nearly telegraphic writing style is permissible.

The examination is closed book and closed notes. If you come to a problem that you believe cannot be answered without making additional assumptions, clearly state the reasonable assumptions that you make, and proceed.

1. Suppose we are interested in studying whether higher values of Prostate Specific Antigen (PSA) are useful in the diagnosis of prostate cancer. For this problem we will regard that there is some "gold standard" for the diagnosis of prostate cancer (perhaps digital rectal exam and transrectal ultrasound directed needle biopsy) that is used when classifying patients as to their prostate cancer. Consider the following study designs for hypothetical studies done in King County:

    A. We randomly sample 200 men aged 50 - 80 years who have been recently diagnosed with prostate cancer and 400 men of similar ages who do not have prostate cancer. Each man has blood levels of PSA measured.

    B. Using hospital records of recently measured blood levels of PSA of men aged 50 - 80 years, we randomly select 500 men with PSA levels less than 4.0 and 500 men with PSA levels greater than 4.0. Each man is then examined for prostate cancer using the "gold standard" for diagnosis.

    C. We sample 1,000 men drawn randomly from the population of 50 - 80 year old men in King county. Each man has blood levels of PSA measured and is examined for prostate cancer via the "gold standard".

    a. (3 points) Which of the above study designs can provide an estimate of the prevalence of prostate cancer in the population of 50 - 80 year old men in King County?

Ans: Only study design C.

Study design A fixes the proportion of subjects in the study that have prostate cancer. Clearly, fixing that proportion precludes having any information about the prevalence in the general population. Study design B samples the population in a systematic manner based on PSA levels. To the extent that prevalence of cancer differs across groups defined by PSA level, the total number of cases of prostate cancer will not be an accurate reflection of the true distribution of the population.

    b. (3 points) Which of the above study designs can provide an estimate of the prevalence of PSA values less than 4.0 in the population of 50 - 80 year old men in King County?

Ans: Only study design C.

Study design B fixes the proportion of subjects in the study that have low PSA. Clearly, fixing that proportion precludes having any information about the prevalence in the general population. Study design A samples the population in a systematic manner based on prevalence of prostate cancer. To the extent that prevalence of cancer differs across groups defined by PSA level, the total number of subjects in the sample with low PSA will not be an accurate reflection of the true distribution of the population.

c. (3 points) Which of the above study designs can provide an estimate of the proportion of men with prostate cancer who have a PSA value greater than 4.0?

Ans: Study designs A and C.

Study design B fixes the proportion of subjects in the study that have low PSA. Clearly, fixing that proportion precludes having any information about the prevalence in either the general population or a population defined by presence of prostate cancer.

This measure is termed a "sensitivity" of PSA to detect prostate cancer. The threshold of 4 was of course arbitrary. Setting the threshold higher or lower would likely change the sensitivity. And we could use these study designs to explore different thresholds, because we are getting samples that are representative of the distribution of PSA values within diagnostic groups.

d. (3 points) Which of the above study designs can provide an estimate of the proportion of men without prostate cancer who have a PSA value less than 4.0?

Ans: Study designs A and C.

Same reasons as part c.

This measure is termed a "specificity" of PSA to detect prostate cancer. The threshold of 4 was of course arbitrary. Setting the threshold higher or lower would likely change the specificity. And we could use these study designs to explore different thresholds, because we are getting samples that are representative of the distribution of PSA values within diagnostic groups.

e. (3 points) Which of the above study designs can provide an estimate of the proportion of men with PSA values above 4.0 who actually have prostate cancer?

Ans: Study designs B and C.

Study design A fixes the proportion of subjects in the study that have prostate cancer. Clearly, fixing that proportion precludes having any information about the prevalence in either the general population or a population defined by PSA values.

This measure is termed a "predictive value of the positive" of PSA to detect prostate cancer. The threshold of 4 was of course arbitrary. Setting the threshold higher or lower would likely change the predictive value of the positive. However, as the study sampled systematically according to the threshold of PSA values, we cannot as easily explore the predictive value of other thresholds.

f. (3 points) Which of the above study designs can provide an estimate of the proportion of men with PSA values below 4.0 who do not have prostate cancer?

Ans: Study designs B and C.

Same reasons as part e.

This measure is termed a "predictive value of the negative" of PSA to detect prostate cancer. The threshold of 4 was of course arbitrary. Setting the threshold higher or lower would likely change the predictive value of the negativetive. However, as the study sample systematically according to the threshold of PSA values, we cannot as easily explore the predictive value of other thresholds.

g. (3 points) Which of the above study designs can provide information regarding an association between elevated PSA levels and prostate cancer? Justify your answer.

Ans: All of the studies. Lack of an association means the distribution of PSA is the same for each diagnostic group <u>and</u> the prevalence of cancer does not depend on PSA values. In every study we can examine at least one of these questions.

h. (3 points) Which of the above study designs can tell whether prostate cancer causes an elevation of prostate specific antigen in the blood?

Ans: None of them.

Causation can only be inferred from study design. None of these are interventional studies.

i. (3 points) Which of the above study designs would be the easiest to perform logistically?

Ans: Study A.

It is usually easier to identify patients with and without prostate cancer and get a single blood test, than to screen a population for prostate cancer using more costly procedures. Of course you would have to consider whether you could really believe that the subjects in the noncancer group were truly free of cancer. It will also depend upon the prevalence of cancer in the general population.

Actually, I accepted any choice here, so long as it was moderately well justified. It is easy to imagine situations where each of the study designs was advantageous.

Appendix A contains descriptive statistics from an early phase clinical trial of beta carotene in cancer prevention. Fifty (50) patients were randomized to receive either 0, 15, 30, 45, or 60 mg/day of beta-carotene supplementation. The goal of the study was to assess how beta-carotene supplementation affected levels of beta-carotene and vitamin E in plasma. Measurements of plasma beta-carotene and vitamin E levels were to be obtained at randomization and after 3 and 9 months of treatment. An additional measurement was to be made 3 months after treatment stopped (12 months after randomization). For each observation in the dataset, the following data are available The following variables are available:

$PTID$	patient's identification number
AGE	patient's age in years
$RACE$	patient's race: 1= Asian, 2= Black, 3= White, 4= Other
$MALE$	patient's sex: 0= female, 1= male
$WEIGHT$	patient's weight in pounds
$DOSE$	treatment group (0, 15, 30, 45, or 60)
$MONTH$	time since randomization in months
$CAROT$	plasma beta-carotene level
$VITE$	plasma vitamin E level

2. (10 points) Based on the descriptive statistics presented in Appendix A, do any of the variables appear prone to outlying values? Explain your answer briefly.

Ans: The plasma beta-carotene values show a mean and median that is somewhat removed from the midpoint of the range. The standard deviation is also relatively large compared to the mean of these positive measurements. Hence I might suspect that there was some tendency to outlying variables for that variable. None of the others seemed too bothersome to me. I note that dose has a large standard deviation relative to its mean, but as subjects were randomized to dose, I don't worry about it based on *a priori* **information. The same is true about month, though there could be more of a worry about loss to followup. Patient ID is really a nominal variable, so I am not sure that "outlying" has any meaning at all.**

3. 15 points How would you use the descriptive statistics presented in Appendix A? That is, how would such statistics aid you in the analysis of these data to answer the scientific question? Briefly explain.

Ans: These data represent repeated measurements on the same subjects. There were 50 patients who were to be measured four times, but only 184 measurements. Clearly, then, we do not have equal measurements on each subject, and the descriptive statistics would not be suitable description of the sample. Thus, the only use I would make of these descriptive statistics is looking for errors in the data. I would prefer subject specific statistics to describe the demographics (age, race, sex, etc.), and I would prefer to isolate the measurements by time for the plasma levels of beta-carotene and vitamin E.

I note that even if there had been no missing data (i.e., I had 200 measurements total on the 50 subjects, I would still not be too thrilled with these descriptive statistics, because baseline measurements of the beta-carotene and vitamin E measurements mean something very different to me than the later measurements.

Appendix B contains descriptive statistics from a study of sexual discrimination in salaries paid to university faculty. All nonadministrative faculty currently employed in 1995 were measured for the following variables

$EMPLID$	employee's identification number
$FEMALE$	employee's sex: 0= male, 1= female
$YRDEG$	year employee obtained highest degree
$FIELD$	employee's field (1= fine arts, 2= general, 3= professional)
$RANK95$	employee's rank (1= assistant prof, 2= assoc prof, 3= full prof)
$SLRY95$	employee's monthly salary in 1995
$TASSOC$	time spent as associate professor at the university (Assistant professors will have 0 for this variable, associate professors will have the years since they were promoted to associate, and full professors will have the time it took them to get promoted.)

4. 3 pts each part For each of the following descriptive statistics presented in Appendix B, identify the variables for which the specified statistic provides no scientifically meaningful descriptions of the sample. For each such variable, very briefly explain why not (just a few words should suffice to justify your entire answer).

   a. Mean

Ans: Employee ID and field are nominal variables, and the rank is ordered categorical. Time spent as associate professor includes censored observations. I would not use the sample mean for any of those variables.

   b. Standard deviation

Ans: Same as for mean. (In addition, the standard deviation of the binary variable sex produces no additional information than is given in the mean.)

   c. minimum

Ans: Employee ID and field are nominal variables, and time spent as associate professor includes censored observations. I would not use the minimum (or any statistic based on ordering) for these variables. (In addition, this statistic contains no information about the binary variable sex that is not given in the mean.)

(The minimum and maximum are heavily dependent on sample size, so they do not provide particularly good inference.)

   d. maximum

Ans: Same as for minimum.

   e. 25th percentile

Ans: Employee ID and field are nominal variables, and time spent as associate professor includes censored observations. I would not use the minimum (or any statistic based on ordering) for these variables. (In addition, this statistic contains no information about the binary variable sex that is not given in the mean.)

   f. median

Ans: Same as for 25th percentile.

   g. 75th percentile

Ans: Same as for 25th percentile.

5. (10 points) How would any of your answers to question 4 change if we were focusing on comparing distributions across sexes? Explain. (Ignore $FEMALE$ for this question.)

Ans: The only change would be that I might consider the mean to compare any ordered variable (including ordered categorical variables) when trying to answer questions about shifts to higher values. Hence my answer about using the mean for rank would change.

6. Suppose we are interested in comparing the time to promotion from Associate Professor to Full Professor for females and males.

   a. (10 points) What descriptive statistics would you want to see? How would they be computed?

Ans: As time spent as associate professor was censored, I would want to see statistics based on the Kaplan-Meier curves. Promotion probabilities at specific times and, say, median time to promotion would be of interest.

I note that it would not be appropriate just to discard the data for assistant and associate professors, as that would not detect blatant discrimination against women by the university. If they promoted a very few women in the same time frame as all men, the vast majority of women languishing at the lower ranks would not be detected in such an analysis. Kaplan-Meier type estimates are absolutely necessary here.

   b. (BONUS: 10 points) How might time trends in promotion policies affect the analysis you suggested in part a? That is, what key assumptions for the analysis of these data might be invalidated by time trends in the policies regarding criteria for promotion to full professor?

Ans: If there were time trends in the promotion to full professor, then people who had censored observations (e.g., current associate professors) are not equally likely to be like any of the measurements still at risk for promotion. This then suggests that we do not have noninformative censoring.

For instance, suppose it is now easier to be promoted to full professor than it used to be. Then an associate professor who has been at that rank for only 4 years is more likely to have a time to promotion of, say, 2 years than a time to promotion of 10 years that might have been seen in the past.

7. (10 points each part) Appendix B also contains mean salary levels by faculty rank and faculty field.

   a. Is there evidence of an association between faculty rank and faculty salary? Briefly explain the evidence.

Ans: Yes. The average salaries are markedly different across the ranks with associates tending to have lower salaries than assistants, who are lower than full.

An association exists if the distribution of salaries is different across the groups. Clearly, if the means are different, then the distributions must be different. I could answer this question either by looking within fields or by averaging across fields. I note that the U-shaped trend does not mean that there is not an association, but it is clearly somewhat surprising. There will be times that we might choose a method of analysis that is not very sensitive to finding such U-shaped trends. In fact, the U-shaped trend in that table is not the way the salaries really are in the dataset we will analyze next quarter.

   b. Is there evidence of an association between faculty field and faculty salary? Briefly explain the evidence.

Ans: Yes. The average salaries are markedly different across the fields with fine arts tending to be less than general, which is less than professional.

Again, I could answer this question either by looking within ranks or by averaging across ranks.

c. Is there evidence of that faculty field modifies the association between faculty rank and faculty salary? Briefly explain the evidence.

Ans: Yes. The difference between the ranks is not the same for each field. For instance, full professors average $2240 more than associates in fine arts, but $3185 more than associates in professional fields.

In judging whether there is effect modification, we will have to decide whether any differences in effect are something we care about. There are two bases we might use for such a decision: science and statistics. That is, scientifically we would have to decide whether the £945 per month difference in effect due to rank is scientifically important (it would matter to me in my paycheck). We might also want to consider whether we had enough information to be sure that such an observed difference were not just due to sampling error. This latter part we will judge according to statistical significance.

Of course, we could have judged these as ratios instead, and it is possible that something that looked like an effect modification when considering differences did not look like effect modification when considering ratios. My own opinion is that there was still effect modification when I looked at ratios: assistants averaged 19%, 14%, and 16% higher than associates, respectively, for fine arts, general, and professional. Full professors averaged 31%, 42%, and 37% higher than assistants, respectively, for those same fields. To me, the differences in percentage increase in salary with promotions across the fields is scientifically meaningful (I rarely turn down a 5% raise).

8. (5 points each part) Suppose we study the association between mental functioning and age. Two independent researchers have performed studies and found the same estimated slope in a linear regression of mental function tests on age. However, researcher A found a correlation of -.17 between age and mental function tests, and researcher B found a correlation of -.42 between age and mental function tests. For each of the following scenarios, specify whether the study conditions might explain why such different correlations would be observed in the presence of similar slope estimates. Briefly justify your answer.

a. Researcher A: 100 subjects of ages 50-100; all educational levels sampled

Researcher B: 100 subjects of ages 50-100; only considered college graduates

Ans: Yes, this could explain the difference. We were given that similar slopes were obtained. The variability of ages was similar for the two studies. The difference in educational background for the two studies would likely lead to a greater variance in mental function test scores for each age group in the study that sampled all educational levels. Thus, we would expect the correlation in that study to be closer to zero.

b. Researcher A: 100 subjects of ages 50-100; all educational levels sampled

Researcher B: 100 subjects of ages 70-80; all educational levels sampled

Ans: No, this could not explain the difference. We were given that similar slopes were obtained. Given that similar levels of educational background were sampled for both studies, we have no reason to suspect a greater variation in mental function test scores for either of the studies. Furthermore, because study B sampled a narrower range of ages, we would expect the lower variation in the predictor to lead to a correlation closer to zero, rather than farther from zero.

c. Researcher A: 100 subjects of ages 50-100; all educational levels sampled

Researcher B: 50 subjects of ages 50-100; all educational levels sampled

Ans: Yes and no. If sample size is all that is different between the two studies, then the study with the smaller sample size would have less precise estimates of correlation. But sample size does not systematically affect the correlation seen between two variables.