Biost 514
Biostatistics I

Second 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.

Much of the exam deals with selected analyses of the PSA dataset which we have discussed in class. Appendix A provides a brief description of this dataset. It should be noted that three additional variables have been computed from the variables originally provided in the dataset.

All confidence intervals may be computed using the critical values appropriate for normally distributed statistics. The .95, .975, .99, and .995 quantiles of the standard normal distribution are 1.645, 1.96, 2.326, and 2.576, respectively.

As always, if you do not have a calculator handy, you may just clearly state the formula you would use to answer the question.

1. (15 points) Appendix B presents descriptive statistics for all of the variables. Of the descriptive statistics presented in Appendix B, which are not of scientific interest?

   Ans: *ps*, *bss*, **and** *grade* **were all ordered categorical data and thus means and standard deviations are difficult to interpret for them (though as noted in class, comparisons based on the means are still useful in hypothesis tests).** *obstime* **is a censored random variable, and thus none of the descriptive statistics (none of mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum) are of scientific interest.** *inrem* **is measured over variable lengths of time, so again none of the descriptive statistics are of scientific interest. (There is no problem with** *inrem*24**, because everyone was followed for at least 24 months, as stated in the problem description.)**

2. (15 points) Suppose we are interested in estimating the average change in PSA values from pre-treatment to nadir. Consider two different approaches to constructing a confidence interval for the mean difference

   A. The point estimate is taken as the difference between the sample mean for NADIRPSA and the sample mean for PRETXPSA, and the estimated standard error is

   $$\sqrt{\frac{1287.64^2}{43} + \frac{39.25^2}{50}}$$

   B. The point estimate is taken as the sample mean for DIFFPSA, and the estimated standard error is

   $$\frac{1269.31}{\sqrt{43}}$$

   In either of the above methods, the CI is constructed as

   $$PointEstimate \pm CriticalValue \times StdError$$

What are the relative merits of the two methods? Which would you prefer, and why?

Ans: Method A assumes independence between the samples used to measure $PRETXPSA$ and $NADIRPSA$. This is not the case, so Method A is inappropriate. Method B merely assumes independence among the subjects in the sample. Thus Method B is preferred. (We can tell that independence between the $PRETXPSA$ and $NADIRPSA$ measurements is not true, because the variance of the difference is not the sum of the variances for the measurements.)

3. (15 points) Construct a 95% confidence interval for the mean of LOGNADIR. Recall that the geometric mean of a distribution is the exponentiation of the mean of log transformed data. Use that fact to construct a 95% confidence interval for the geometric mean of NADIRPSA.

Ans: 95% CI for mean $LOGNADIR$

$$.57 \pm 1.96 \times \frac{2.15}{\sqrt{50}} = (-.0259, 1.166)$$

95% CI for geometric mean of $NADIRPSA$ is found by exponentiating the above CI

$$(e^{-.0259}, e^{1.166}) = (.974, 3.209)$$

4. (10 points) Using the results in Appendix C, along with the fact that Kaplan-Meier estimates are asymptotically normally distributed, to construct a 99% confidence interval for the probability of remaining in remission for 30 months.

Ans: Using the standard error (which is not the standard deviation) given in the appendix

$$.498 \pm 2.576 \times .071 = (.315, .681)$$

5. (15 points) Using the results in Appendix C, compute a 95% confidence interval for the difference in 30 month probability of remaining in remission for those subjects who had initial performance status below 90 compared to those subjects whose performance status was 90 or above. Provide a concise interpretation of this confidence interval. From that confidence interval, what can you say about the possibility that performance status was associated with the probability of remaining in remission for 30 months?

Ans: We want to take the difference of the point estimates as the point estimate of the difference, and due to independent samples, the standard error of the difference will be the square root of the sums of the squared standard errors:

$$(.415 - .647) \pm 1.96 \times \sqrt{.089^2 + .116^2} = (-.519, .055)$$

The interpretation: Based on the 95% confidence interval, the observed difference in 30 month remission probabilities of -.232 is reasonably typical of what might be expected if the true difference in remission probabilities were anywhere between -.519 and .055.

Because the confidence interval suggests that these data might reasonably be obtained when the true difference is 0 (that is, 0 is in the CI), we cannot from this analysis state with high confidence that performance status is truly associated with time in remission.

6. (10 points each part)Recall that the primary scientific question to be addressed by this dataset was whether NADIRPSA was predictive of time in remission, and whether any such association was merely reflective of differences in performance status.

a. From the results in Appendix D, is there evidence of an association between the nadir PSA value and remission status at 24 months? Document and explain your reasoning.

Ans: Using as my measure of association the average difference in nadir PSA values across groups defined by remission status at 24 months, the 95% CI for that average difference is

$$(30.56 - 4.26) \pm 1.96 \times \sqrt{\frac{51.72^2}{23} + \frac{17.59^2}{27}} = (4.146, 48.454)$$

(note the adjustment of the sample sizes to account for missing data when computing my standard errors). Because that CI does not include 0, I can with 95% confidence state that there is an association between time in remission and nadir PSA value.

b. Does it appear that your answer to part a is confounded by performance status? Explain your reasoning.

Ans: Necessary conditions for confounding include an association between the predictor of interest and the potential confounder in the sample, and an association between the response and the potential confounder. These two conditions are clearly met. There is obviously an association between performance status and remission status at 24 months: 14/31 subjects with $ps < 90$ are still in remission at 24 months, while 13/17 subjects with $ps \geq 90$ are still in remission. There is similarly an association between performance status and nadir PSA: The average nadir PSA was 23.93 in subjects with $ps < 90$ and was 3.96 in patients with $ps \geq 90$. (Pointing out these relationships was judged an adequate answer.)

Before we adjusted our analyses for ps, however, we would need to consider whether such adjustment made scientific sense. We would do this by considering the possible causal mechanisms as well as the scientific question that needed to be answered. It is difficult to imagine that performance status is much of a cause of either the time in remission or the nadir PSA. However, insofar as performance status was a surrogate variable for severity of cancer, then we might question how the severity of cancer was related to time in remission and the nadir PSA. It seems clear that severity of cancer could quite likely be a common cause of both events, and thus severity of cancer is clearly a confounder. The scientific question is then whether nadir PSA might be a better surrogate for the severity of cancer than performance status, and adjustment for performance status is how we would answer that question.

(Of course, when we look at the association between remission time and nadir PSA within strata defined by performance status, we see that the issue of confounding is moot: The association does not look to be the same in the two strata, and that argues for an interaction.)

c. Does it appear that the association between nadir PSA and remission status at 24 months is modified by performance status? Discuss this both as trends and statistically significant effect modification. Explain your reasoning.

Ans: In the low performance status group, the association between remission time and nadir PSA is estimated by a difference in average nadir PSA levels of 37.68 - 7.24 = 30.49 between those who relapsed and those still in remission. For the high performance status group, the corresponding measure of association is 13.42 - 1.05 = 12.37. This difference in the estimates of association suggests that there is indeed the possibility of an interaction.

In order to test the statistical evidence of that difference, I constructed a 95% CI for the true difference in the measures of association. (Again note the adjustment of the denominators to account for missing data.)

$$((37.68 - 7.24) - (13.42 - 1.05)) \pm \sqrt{\frac{58.35^2}{17} + \frac{24.41^2}{14} + \frac{17.52^2}{4} + \frac{1.87^2}{13}} = (-16.983, 53.123)$$

Because the CI includes 0, I cannot with 95% confidence state that an interaction truly exists.

7. (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: It is reasonable to suspect that education level might be a predictor of performance on mental function tests. Thus a researcher who sampled only a single educational level would expect to have lower variance of measurements in each age group. When everything else remains the same, a lower variance within groups defined by the predictor age will tend to a more extreme correlation, and thus the more restrictive sampling in study B with respect to the third variable of educational level might be a reasonable explanation of the more extreme correlation in that study.

   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: When everything else remains the same, a larger variance in the predictor age will be expected to lead to a more extreme correlation. In this problem, study A has the greater variance of the predictor, but study B has the more extreme correlation. Thus this cannot be the explanation for the discrepancy in the results.

   (A couple people invoked possible nonlinearities in the relationship between age and mental function, and they suggested that over the wider interval those nonlinearities would cause the problem. In fact, I suggested that the estimated slope was the same for the two studies, but I guess the nonlinearities could cause different error variances, so I gave credit for such an answer.)

   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: Sample size does not affect the tendency for a particular correlation (though it does affect the sampling variability of the correlation). Thus the difference in sample sizes is not likely to explain the difference in correlations, unless it were just sampling variability.

Grade distribution: (115 Possible)

Mean (Std Dev): 69.2 (15.3)

Median: 68

Highest: 102

IQ range: 61, 78