Biost 517
Applied Biostatistics I

Midterm Examination Key
November 8, 2010

Name: _____ Disc Sect: M W F

Instructions: Please provide concise answers to all questions. The exam is worth a total of 150 points.

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. You may use calculators, but you may not use any special programs written for programmable calculators.

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.

Please adhere to and sign the following pledge. Should you be unable to truthfully sign the pledge for any reason, turn in your paper unsigned and discuss the circumstances with the instructor on Wednesday.

PLEDGE:

On my honor, I have neither given nor received unauthorized aid on this examination:

Signed:_____

Problems 1 – 2 relate to a clinical trial of experimental drug TFD725 plus docetaxel versus placebo plus docetaxel as second line treatment of non small cell lung cancer in 188 patients. The following variables are available:

Name	Description
ptid	Patient ID
age	Patient age at randomization (years)
male	Indicator of patient's sex (0= female, 1= male)
advdis	Indicator of advanced stage at initial diagnosis (0= stage IIIb without malignant pleural effusion, 1= malignant pleural effusion or stage IV)
LDH	Serum LDH level at time of randomization (U/l). An elevated LDH level has been reported to be highly predictive of poor patient outcomes.
alkphos	Serum alkaline phoshatase level at time of randomization (U/l). An abnormal alkaline phosphatase level has been reported to be highly predictive of poor patient outcomes.
ECOG	Patient's performance status on ECOG scale (0= best, 1, 2)
tx	Indicator of study treatment assignment (0= placebo, 1= TFD725)
obstime	Observation time from randomization to death or data analysis, whichever came first (days)
death	Status at last follow-up (0= alive, 1= dead)

The following table contains descriptive statistics on the sample.

Variable	N	Mean	SD	Min	25th %ile	Median	75th %ile	Max
ptid	188	58043	25956	10081	38063	58209	80269	99800
age	188	60.4	5.11	46.0	57.0	60.5	64.0	75.0
male	188	0.553	0.498	0.000	0.000	1.000	1.000	1.000
advdis	188	0.628	0.485	0.000	0.000	1.000	1.000	1.000
LDH	188	301	194	103	169	248	375	1283
alkphos	188	338	295	66	142	251	407	1775
ECOG	188	0.745	0.536	0.000	0.000	1.000	1.000	2.000
tx	188	0.521	0.501	0.000	0.000	1.000	1.000	1.000
obstime	188	389	127	56	309	396	491	615
death	188	0.745	0.437	0.000	0.000	1.000	1.000	1.000

1. (3 points each part) For each of the following variables, circle the descriptive statistics that are **NOT** of use to provide a scientifically meaningful description of the sample. Very briefly explain your reasons (just a few words should suffice to justify your entire answer).
   a. Consider **ptid**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      ~~*Mean*~~ ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

      Ans: ***ptid*** **is a nominal variable.**

   b. Consider **age**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      Mean *Std Dev* *Minimum* *25th Pctile* *Median* *75th Pctile* *Maximum*

      Ans: ***age*** **is a quantitative, continuous variable.**

   c. Consider **male**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      Mean ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

      Ans: ***male*** **is a binary variable. All but the mean (proportion) are boring.**

   d. Consider **advdis**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      Mean ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

      Ans: ***advdis*** **is a binary variable. All but the mean (proportion) are boring.**

   e. Consider **LDH**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      Mean *Std Dev* *Minimum* *25th Pctile* *Median* *75th Pctile* *Maximum*

      Ans: ***LDH*** **is a quantitative, continuous variable.**

   f. Consider **alkphos**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

      Mean *Std Dev* *Minimum* *25th Pctile* *Median* *75th Pctile* *Maximum*

      Ans: ***alkphos*** **is a quantitative, continuous variable.**

g. Consider **ECOG**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

~~*Mean*~~ ~~*Std Dev*~~ *Minimum* *25th Pctile* *Median* *75th Pctile* *Maximum*

Ans: ***ECOG*** **is an ordered categorical variable. Thus quantiles and extrema make sense.**

h. Consider **tx**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

Mean ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

Ans: ***tx*** **is a binary variable. All but the mean (proportion) are boring.**

i. Consider **obstime**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

~~*Mean*~~ ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

Ans: ***obstime*** **is a right censored, continuous variable. (Need to use Kaplan-Meier)**

j. Consider **death**. Circle the descriptive statistics that are NOT useful. Briefly explain why.

~~*Mean*~~ ~~*Std Dev*~~ ~~*Minimum*~~ ~~*25th Pctile*~~ ~~*Median*~~ ~~*75th Pctile*~~ ~~*Maximum*~~

Ans: ***death*** **is an indicator of observed events for the right censored continuous variable** ***obstime*****. (Need to use Kaplan-Meier)**

2. (5 points) How would your answers to problem 1 change if we were trying to compare distributions across populations, instead of just describing the sample. Briefly explain your reasons.

Ans: When comparing distributions across groups, I could use the mean for the ordered categorical variable *ECOG*, in which case the standard deviation is also of interest for assessing precision of inference.

Minima and maxima are not of much use for comparing across populations due to the heavy dependence of their sampling distributions on the sample size.

3. (10 points) For which of the variables is it of scientific interest to assess skewness? Do any of those variables appear to be skewed? Briefly explain your reasons.

Ans: Any of the quantitative continuous variables measured without censoring: *age, LDH, alkphos.*

Of these, both *alkphos* and *LDH* show a standard deviation that is very large relative to the mean for a positive random variable, a mean markedly different from the median, a median that is not midway between the 25th and 75th percentiles, and a maximum that is much further from the median than is the minimum. Hence, *alkphos* and *LDH* both appear to be skewed, and may well have some large outliers.

age **appears to have a remarkably symmetric distribution.**

4. Using the above descriptive statistics,
   a. (5 points) Can you tell how many patients have advanced disease? If so, do so. If not, briefly explain why not.

Ans: The mean of a binary 0-1 variable is the proportion having the value 1, so 0.628 x 188 = 118 have advanced disease.

b. (**Biost 514 or Bonus**: 10 points) Can you tell how many patients have ECOG performance status 0 (the healthiest category)? If so, do so. If not, briefly explain why not.

Ans: Yes, because there were only three categories, we can do this using both the mean and the standard deviation. It does, however, involve aspects of the formulas for these quantities that have not been stressed in this class (hence it was a problem for Biost 514):

Let p_0, p_1, p_2 be the proportion of subjects in each group. Then the sample mean will be

$$E[X] = 0 \times p_0 + 1 \times p_1 + 2 \times p_2 = 0.745$$

The sample variance of a random variable can be most easily computed using the computational forumula for a variance

$$Var(X) = n \{ E[X^2] - (E[X])^2 \} / (n-1)$$

from which we find $E[X^2] = (187 \times 0.536^2) / 188 + 0.745^2 = 0.84079$

Now, we can also express

$$E[X^2] = 0 \times p_0 + 1 \times p_1 + 4 \times p_2 = 0.84079$$

and solving these two equations in two unknowns simultaneously, we find

$p_1 = 2 \times 0.745 - 0.84079 = 0.64921$

$p_2 = (0.745 - 0.64921) / 2 = 0.047895$

$p_0 = 1 - p_1 - p_2 = 1 - 0.64921 - 0.047895 = 0.302895$

so there were 188 x 0.302895 = 57 subjects with ECOG = 0.

Note that this problem could also have been solved without using the computational formula by solving the following three equations in three unknowns simultaneously:

$E[X] = 0 \times p_0 + 1 \times p_1 + 2 \times p_2 = 0.745$

$Var[X] = 188 \times \{ (0 - .745)^2 \times p_0 + (1 - .745)^2 \times p_1 + (2 - 0.745)^2 \times p_2 \} / 187 = 0.536^2$

$p_0 + p_1 + p_2 = 1$

5. The following table presents descriptive statistics for selected variables according to whether the patient was in the experimental treatment group or the placebo group.

Variable	N	Mean	SD	Min	25th %ile	Median	75th %ile	Max
Control Arm (placebo)								
advdis	90	0.656	0.478	0.000	0.000	1.000	1.000	1.000
LDH	90	281	171	103	157	227	364	1009
obstime	90	376	124	56	307	372	482	607
death	90	0.800	0.402	0.000	1.000	1.000	1.000	1.000
Experimental Arm (TFD725)								
advdis	98	0.602	0.492	0.000	0.000	1.000	1.000	1.000
LDH	98	320	212	108	182	253	384	1283
obstime	98	401	130	59	312	414	496	615
death	98	0.694	0.463	0.000	0.000	1.000	1.000	1.000

a. (10 points) How would you use the above descriptive statistics to assess whether LDH might confound our ability to assess a beneficial treatment effect of TFD725 compared to placebo? Does it?

Ans: *(Because this is a randomized study, we can in some sense say that there is no confounding on average. Nevertheless, it is generally of interest to examine the treatment arms for any imbalance that occurred by chance.)*

We can use the above descriptives to assess whether the distribution of LDH appears to differ between the treatment groups. As my interest is in confounding, I would be primarily interested in

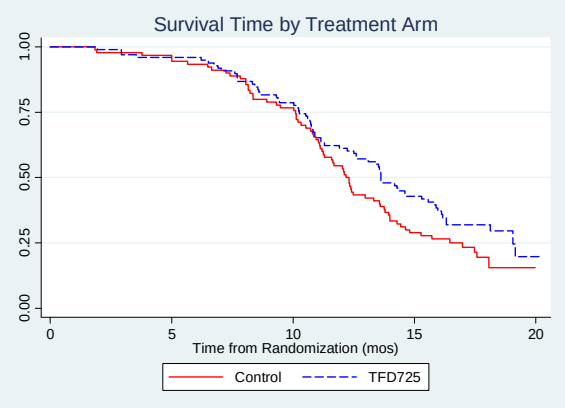
whether the mean LDH were different between the groups when I would adjust for LDH linearly and in whether the geometric mean LDH were different between the groups when I would adjust for a log transformed LDH. I note that the mean LDH does appear a bit higher in the Experimental Arm, though it is a judgment call as to whether this is enough to worry about.

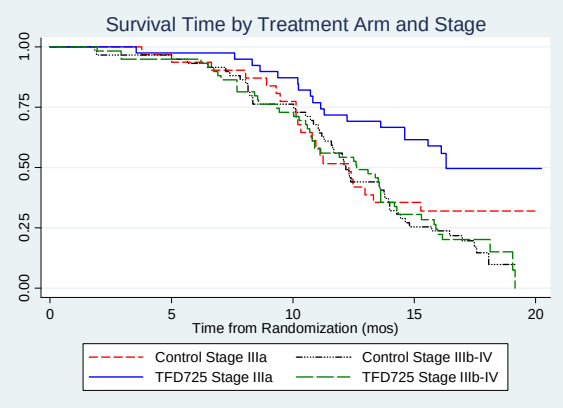
I cannot use these data to assess whether there is an association between LDH and survival time, because the survival times are censored. However, the data description notes that LDH is believed to be a marker of worse disease, and hence that presumed relationship is enough to go on—we would not need to see whether the relationship existed in this sample.

b. (10 points) How would you use the above descriptive statistics to assess whether the use of TFD725 is associated with prolonged survival?

Ans: I would not and could not: The survival times are subject to censoring. I need to use KM.

6. (30 points) The following are results from a Kaplan-Meier analyses of the time to death within strata defined by the presence of treatment (left panel) and strata defined by both treatment and advanced disease (right panel).





	Control Arm			TFD725 Arm		
		Survival Probability			Survival Probability	
	N	9month	18 month	N	9month	18 month
All Patients	90	0.789	0.195	98	0.816	0.320
Stage IIIa	31	0.839	0.319	39	0.897	0.497
Stage IIIb or IV	59	0.763	0.147	59	0.763	0.201

a. (10 points) Based on the above statistics, would you conclude that there is overall an association between the experimental treatment and the probability of survival? Provide statistics to quantify your answer.

Ans: There seems little difference in the 9 month survival probability (81.6% on TFD725 vs 78.9% on control), but more evidence of an association is evident at 18 months (32.0% probability of survival on TFD725, but only 19.5% probability on the control arm.) This is also seen in the plots of the Kaplan-Meier estimates, where there is a "late difference" between the survival curves.

(The above does not address the question of whether the observed differences are beyond that that might be due to chance in the absence of a true association. This possibility of a chance association is greatly increased when we look at more than one summary measure, such as I did here by looking at both the 9 month and 18 month survival probabilities. The subject of statistical inference on such observations is the subject of the rest of the course.)

b. (10 points) Based on the above statistics, would you conclude that the stage of disease modifies any association between the experimental treatment and survival? Provide statistics to quantify your answer.

Ans: Yes. From the plots we see that there does not appear to be substantial difference between the KM estimates for the two treatment groups within the advanced disease subgroup, while in the subjects with less advanced disease the KM estimates appear different. This is futher corroborated by looking at the 18 month survival probabilities: In the less advanced disease group (stage IIIa), the 18 month survival probability is 49.7% for TFD725 and 31.9% for control (an absolute difference of 17.8%), while in the advanced disease group (stage IIIab or IV), the 18 month survival probability is 20.1% for TFD725 and 14.7% for control (an absolute difference of 5.4%),.

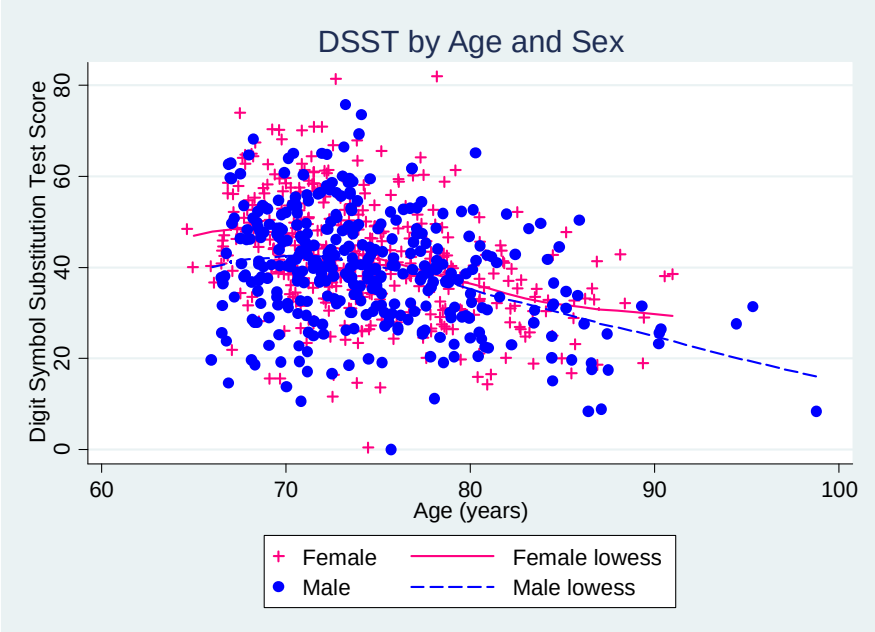
c. (10 points) Based on the above statistics, would you conclude that stage of disease confounds any association between the experimental treatment and survival? Provide statistics to quantify your answer.

Ans: Approximately 39.8% of subjects in the treatment arm are in the less advanced disease subgroup, while approximately 34.4% of subjects in the control arm have less advanced disease. To the extent that more advanced disease is (by definition) associated with worse survival, and to the extent that this absolute difference of about 5.4% is therefore clinically important, there will be some confounding of the treatment effect by the presence of more advanced disease. We might worry that this pattern might tend to make the treatment look more beneficial than it truly is.

Problems 7 and 8 pertain to a study of dementia in 735 elderly adults. Available data include

- **age** = Patient age (years)
- **male** = Patient sex (0= female, 1= male)
- **atrophy** = Cerebral atrophy as measured on a scale from 0 (best) to 100 (worst) from magnetic resonance imaging of the brain.
- **dsst** = Cognitive function as measured by the Digit Symbol Substitution Test (DSST) on a scale from 0 (worst) to 100 (best)

7. The following graph displays a scatterplot of DSST versus patient age. Data is stratified by age, with superimposed lowess curves.



a. (5 points) Briefly summarize the observations you would make from this graph.

Ans: I would comment on

- **No striking outliers.**
- **First order trend suggestive of a tendency for lower average DSST in older ages.**
- **The trend in mean DSST is well approximated by a straight line**
- **Some suggestion of greater variability in DSST in lower age groups than there is in the higher age groups. (Hence, greater variability of DSSTs in the groups that tend to have the higher mean DSSTs)** *(I note that in age groups that might have averaged even higher scores, say 80 – 90, there would likely be lower variance as well owing to the nature of the scores having to be between 0 and 100. I have never analyzed such data, however.)*

b. (5 points) What would you estimate the correlation of these two variables to be in the combined sample?

Ans: A negative correlation about -0.40. *(I gave you credit so long as you suggested a negative correlation. As noted below, the true correlation is -0.36.)*

c. (5 points) How might you expect the correlation within each sex to differ from the correlation in the combined sample?

Ans: I note

- **The slopes of the lowess lines are approximately the same for the two sexes,**
- **There is a slightly greater range of ages for males in that the oldest individuals are male, but I doubt that makes too much difference on the variance of ages.**
- **Given the heteroscedasticity, there might tend to be a lower average variance within age groups for males, but again I doubt this makes too much difference.**

Overall, I would expect the correlations to be comparable in the combined groups and in the two sexes separately.

(I wrote all of the above before I did the analysis. Truth is the correlation overall was -0.36, with a correlation of -0.39 in females and -0.33 in males. Turns out, the standard deviations of age within each sex was nearly identical: 5,25 for females, 5.45 for males The root mean squared error of DSST within age groups was also nearly identical for the two sexes: 11.78 for females, 11.77 for males. The least squares slopes showed a slightly more extreme slope for females than for males: -0.96 for females, -0.76 for males.)

8. Suppose we are interested in using the DSST to predict severe cerebral atrophy. I created an indicator of a low DSST score (below 30) and an indicator of a high atrophy score (above 50). The following table presents a cross tabulation of of these indicator functions.

. tabulate lowdsst highatrophy, row col

Key		
frequency		
row percentage		
column percentage		
lowdsst	highatrophy 0	1	Total
0	508	72	580
	87.59	12.41	100.00
	80.13	80.90	80.22
1	126	17	143
	88.11	11.89	100.00
	19.87	19.10	19.78
Total	634	89	723
	87.69	12.31	100.00
	100.00	100.00	100.00

a. (5 points) Suppose we had gathered our sample by randomly sampling 723 subjects and then assessing their DSST and atrophy scores, and we want to explore the ability of a low DSST score (DSST < 30) to predict severe atrophy (atrophy score > 50). Where they can be estimated from this study design, provide estimates of the prevalence of severe atrophy, sensitivity, specificity, predictive value of the positive, and predictive value of the negative. If they cannot be estimated from this study design, very briefly indicate why not.

Ans: From this cross-sectional sampling we can estimate

- **Prevalence of severe atrophy = 12.31%**
- **Sensitivity = Pr (low DSST | severe atrophy) = 19.10%**
- **Specificity = Pr (high DSST | low atrophy) = 80.13%**
- **Predictive Value of positive = Pr (severe atrophy | low DSST) = 11.89%**
- **Predictive Value of negative = Pr (low atrophy | high DSST) = 87.59%**

b. (5 points) Suppose we had gathered our sample by randomly sampling 143 subjects with low DSST and then separately found 580 subjects with high DSST. We then measured atrophy by MRI on each of these samples in order to explore the ability of a low DSST score (DSST < 30) to predict severe atrophy (atrophy score > 50). Where they can be estimated from this study design, provide estimates of the prevalence of severe atrophy, sensitivity, specificity, predictive value of the positive, and predictive value of the negative. If they cannot be estimated from this study design, very briefly indicate why not.

Ans: From this sampling based on test result, we find

- **Prevalence of severe atrophy = cannot be estimated**
- **Sensitivity = Pr (low DSST | severe atrophy) = cannot be estimated**
- **Specificity = Pr (high DSST | low atrophy) = cannot be estimated**
- **Predictive Value of positive = Pr (severe atrophy | low DSST) = 11.89%**

- **Predictive Value of negative = Pr (low atrophy | high DSST) = 87.59%**

c. (5 points) Suppose we had gathered our sample by randomly sampling 89 subjects with severe atrophy and then separately found 634 subjects without severe atrophy. We then measured DSST on each of these samples in order to explore the ability of a low DSST score (DSST < 30) to predict severe atrophy (atrophy score > 50). Where they can be estimated from this study design, provide estimates of the prevalence of severe atrophy, sensitivity, specificity, predictive value of the positive, and predictive value of the negative. If they cannot be estimated from this study design, very briefly indicate why not.

Ans: From this sampling based on disease status, we find

- **Prevalence of severe atrophy = cannot be estimated**
- **Sensitivity = Pr (low DSST | severe atrophy) = 19.10%**
- **Specificity = Pr (high DSST | low atrophy) = 80.13%**
- **Predictive Value of positive = Pr (severe atrophy | low DSST) = cannot be estimated**
- **Predictive Value of negative = Pr (low atrophy | high DSST) = cannot be estimated**

d. (10 points) Assuming the study design in part c, what would be the predictive value of the positive if the true prevalence of severe atrophy were 10%? Show how you obtained your answer.

Ans: Using Bayes' rule:

$$PV+ = \frac{Sens \times \Pr ev}{Sens \times \Pr ev + (1 - Spec) \times (1 - \Pr ev)}$$

$$PV+ = \frac{0.1910 \times 0.10}{0.1910 \times 0.10 + (1 - 0.8013) \times (1 - 0.10)} = 0.0965$$

(All in all, DSST does not seem very useful in diagnosing severe cerebral atrophy.)

Grades:

Maximum Possible: **150**
Highest Achieved: **140**
Mean (SD) **116 (17.1)**

Percentile	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
Grade	**90**	**100**	**110**	**115**	**119**	**124**	**127**	**130**	**135**