Biost 517
Applied Biostatistics I

Midterm Examination Key
November 15, 2005

Name: _____ Disc Sect: M W F

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. You may use calculators, but you may not use any special programs written for programmable calculators. Should you not have a calculator available, write down the equation that you would plug into a calculator.

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:_____

1. (20 points) A randomized clinical trial was conducted of a new treatment (idarubicin) versus an existing treatment (doxorubicin) in the treatment of acute myelogenous leukemia. For each of the patients, the following data is available.
   - *ptidno*= patient identification number uniquely identifying each patient
   - *date*= date of randomization in MMDDYY format
   - *tx*= treatment ("D"= doxorubicin, "I"= idarubicin)
   - *male*= sex of patient in coded format (0= female, 1= male)
   - *age*= age of patient in years
   - *fab*= French-American-British code for classification of type of AML (based on type of cell, appearance of cell under light microscopy, cytogenetics)
   - *karn*= patient's performance status on the Karnofsky scale (0 = dead, 100= perfect health)
   - *wbc*= patient's white blood cell count at randomization (1000 cells per cu mm)
   - *plt*= patient's platelet count at randomization (1000 cells per cu mm)
   - *hgb*= patient's hemoglobin levels at randomization (mg/dl)
   - *obstime*= days from randomization until death or last follow-up if still alive
   - *death*= indicator that patient was observed to die while under observation

The following table presents descriptive statistics for the dataset.

	N	msng	mean	std dev	min	25%-ile	median	75%-ile	maximum
ptid	130	0	96.3	59.7	1.0	33.3	115.5	147.8	180.0
date	130	0	64412	35206	10287	32387	62487	92487	121385
male	130	0	0.50	0.50	0.0	0.0	0.5	1.0	1.0
age	130	0	38.9	12.9	17.0	27.0	36.5	52.8	61.0
fab	130	12	3.07	1.49	0	2	3	4	6
karn	130	0	79.5	12.1	30	80	80	90	100
wbc	130	1	36.1	46.9	0.4	2.9	13.8	56.9	215.0
plt	130	1	80.0	77.8	11	33	57	95	457
hgb	130	1	9.43	1.67	2.8	8.3	9.3	10.2	13.9
obstime	130	0	497	395	3	262	394	598	1848
death	130	0	0.67	0.47	0	0	1	1	1

a. For each of the variables given above, indicate the descriptive statistics that are not of scientific use to answer <u>any</u> scientific question.

<u>Ans:</u> The descriptive statistics that are not of use:

- *ptidno*= **mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum** *(this is a nominal (unordered) variable that just happens to use numerals.)*
- *date*= **mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum** *(in its given format, this is a nominal (unordered) variable that just happens to use numerals. The absurdity of the mean and standard deviation should be readily apparent.)*
- *male*= **(all are OK)** *(all but the mean are pretty boring, however, so I accepted an answer that listed the others)*
- *age*= **(all are OK)** *(this is a quantitative, ratio variable.)*
- *fab*= **mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum** *(this is a nominal (unordered) variable that just happens to use numerals. I recognize that this was perhaps a more difficult thing to determine from the description, so if you told me your assumption about the type of variable, I gave credit)*
- *karn*=**(all are OK)** *(this is an ordered variable, but the scale is not really well-defined, so the mean and standard deviation are not as much of interest descriptively. I note, however, that when comparing groups the mean and standard deviation <u>are</u> generally of interest. I gave credit if you listed the mean and standard deviation, but mentioned your reasoning.)*
- *wbc*= **(all are OK)** *(this is a quantitative, ratio variable.)*
- *plt*=**(all are OK)** *(this is a quantitative, ratio variable.)*
- *hgb*= **(all are OK)** *(this is a quantitative, ratio variable.)*
- *obstime*= **mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum** *(this is a right censored variable. We would want to use Kaplan-Meier estimates.)*

- *death*= **mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum** *(this is a binary variable measured over varying time frames. I note that for the purposes of describing the statistical information, we do sometimes want to know the number of events, so the mean could really be of use here, but that is not really a scientific question.)*

b. Using those descriptive statistics that are relevant, do any of the variables appear to be prone to outliers? Briefly explain your reasoning.

Ans: Among the quantitative variables:

- *wbc* **had a mean very different from the median and the SD of this positive variable was larger than the mean. The maximum was markedly further from the mean and median than the minimum.**
- *plt* **had a mean very different from the median and the SD of this positive variable was about as large as the mean. The maximum was markedly further from the mean and median than the minimum.**
- *hgb* **had a minimum value much further from the mean/median than was its maximum. However, there were not any other striking signs of outliers.**

***(Note:** Some students tried to use as evidence of outliers that the maximum was more than 2 SD from the mean. You need to be careful here. Given a large enough sample size, a distribution without outliers (e.g., the normal distribution) will have a maximum that is many, many SD away from the mean. The distribution of the sample maximum depends heavily on the sample size, so you would have to talk about how the maximum was a large number of SD away from the mean relative to the sample size. For instance, with a normally distributed random variable, we should not tend to see a maximum that was more than 4SD from the mean until the sample size was 10000 or so. A far better criterion is the asymmetry in number of SD between the min and the mean and between the mean and the max.)*

2. (10 points) The following analysis was performed to examine survival time by treatment group.

Confidence intervals for mean *obstime* by *tx:*

```
. bysort tx: ci obstime

-------------------------------------------------------------------------------
-> tx = D

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
     obstime |         65    420.8923    39.96411        341.0548    500.7298

-------------------------------------------------------------------------------
-> tx = I

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
     obstime |         65    574.0154    55.21764        463.7055    684.3253
```

a. How would you use the analysis to address the question of whether Idarubicin provides survival advantage over Doxorubicin. (Include appropriate scientific interpretation of confidence intervals.)

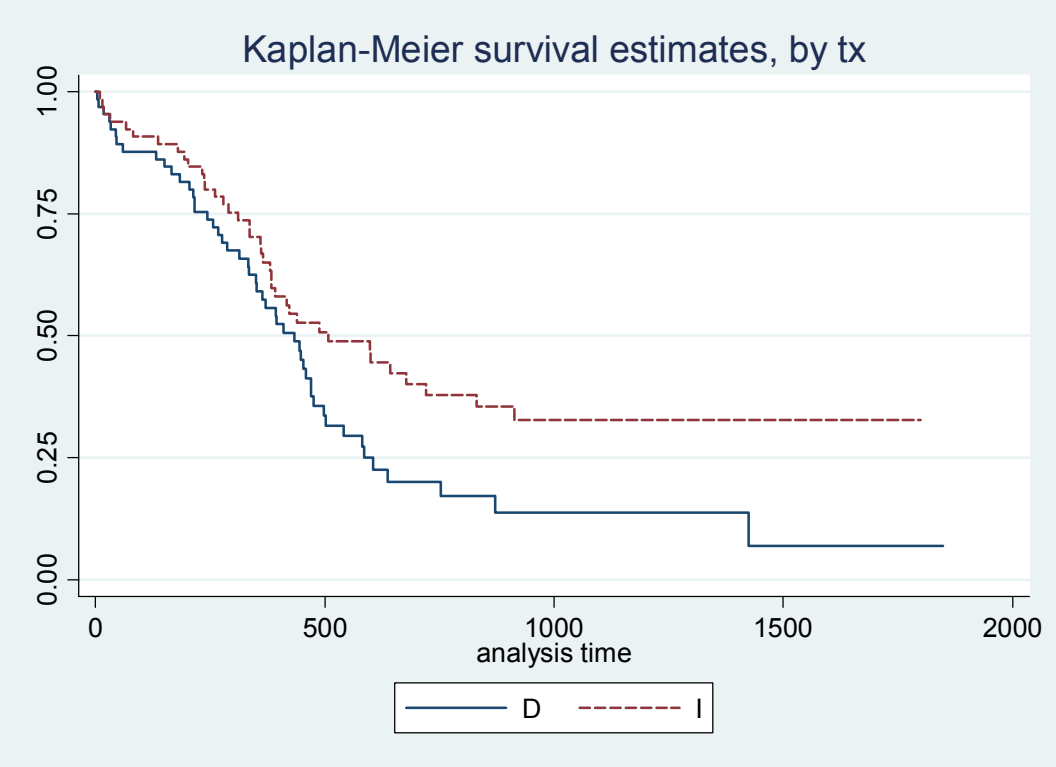
Ans: I would not, because the variable *obstime* is right censored. The confidence intervals are not relevant to the question of survival advantage.

3. (15 points) An additional analysis was performed to examine survival time by treatment group using Kaplan-Meier estimates.

```
. stset obstime death
. sts list, at(182, 365, 547, 730) by(tx)

         failure _d:  death
   analysis time _t:  obstime

              Beg.                        Survivor      Std.
    Time     Total     Fail              Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
D
     182        55       11               0.8308    0.0465     0.7153    0.9025
     365        35       16               0.5742    0.0624     0.4427    0.6853
     547        15       15               0.2950    0.0614     0.1818    0.4173
     730         8        4               0.1997    0.0573     0.1019    0.3210
I
     182        58        8               0.8769    0.0407     0.7689    0.9365
     365        38       14               0.6501    0.0606     0.5178    0.7544
     547        25        9               0.4880    0.0654     0.3554    0.6081
     730        18        5               0.3788    0.0666     0.2504    0.5062
-------------------------------------------------------------------------------
. sts graph, by(tx)
```



a. How would you use the analysis to address the question of whether Idarubicin provides survival advantage over Doxorubicin. (Include appropriate scientific interpretation of confidence intervals.)

<u>Ans:</u> The two year survival probability with Doxorubicin is estimated to be 20.0%, and the 95% CI suggests that these results would not be unusual if the true two year survival probability were anywhere between 10.2% and 32.1%. The two year survival probability with Idarubicin is estimated to be 37.9%, and the 95% CI suggests that these results would not be unusual if the true two year survival probability were anywhere between 25.0% and 50.6%. Hence, we estimate that Idarubicin has a 17.9% higher (absolute difference) two year survival probability, but we would need to perform a hypothesis test to determine whether the observed difference was beyond that which might happen by random chance when the two year survival was actually equal between the two treatments.

(I chose two year survival because it seemed to me to be a more clinically relevant time period and there was still reasonable precision. It would in general be wrong to look for the time period with the most significant differences, because then we would need to adjust for multiple comparisons.

I did not expect you to perform a hypothesis test. All I wanted was you to choose some summary measure of the distribution and compare the two treatment groups. But we have talked a little in class about how to go about doing a hypothesis test in this situation: The SE of the estimated difference in survival probabilities will be the square root of the sum of the squared SEs of the survival probabilities. Hence, the SE of the difference is sqrt($.0666^2 + .0573^2$)= 0.0879, and the Z score would be 0.1791 / 0.0879 = 2.036. As this is larger than the 97.5^{th} percentile of the standard

normal distribution (which is 1.96), we can reject the null hypothesis of equality of the two year survival probability for the two treatments.)

b. Would a level 0.05 hypothesis test reject the null hypothesis that the two year survival probability is 25% with Doxorubicin. Explain your reasoning.

Ans: No. The 95% confidence interval for the two year survival probability is 0.1019 to 0.3210, which contains 0.25.

c. Would a level 0.05 hypothesis test reject the null hypothesis that the two year survival probability is 25% with Idarubicin. Explain your reasoning.

Ans: Yes. The 95% confidence interval for the two year survival probability is 0.2504 to 0.5062, which does not contain 0.25.

4. (30 points) A clinical trial of beta carotene was performed in 46 healthy subjects. Of interest was the plasma beta carotene levels after 9 months of treatment. Subjects were randomized to placebo (dose=0), 15, 30, 45, or 60 mg/kg/day. The variable *carot0* is a measure of the plasma beta carotene level at time of randomization, and the variable *carot3* is a measure of the plasma beta carotene level at the end of the treatment period. The following analyses consider the placebo group and the highest dose group.

Mean plasma beta carotene levels at baseline, 9 months for placebo

```
. ttest carot0=carot3 if dose==0
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  carot0 |       7    287.8786     51.7768    136.9885    161.1853    414.5718
  carot3 |       7    186.3214     33.1844    87.79767    105.1221    267.5207
---------+--------------------------------------------------------------------
    diff |       7    101.5571    25.24236    66.78499    39.79132     163.323
------------------------------------------------------------------------------
     mean(diff) = mean(carot0 - carot3)                           t =   4.0233
 Ho: mean(diff) = 0                           degrees of freedom =        6

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9965         Pr(|T| > |t|) = 0.0069          Pr(T > t) = 0.0035
```

Mean plasma beta carotene levels at baseline, 9 months for dose 60 group

```
. ttest carot0=carot3 if dose==60
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  carot0 |       9    235.4815    38.47693    115.4308    146.7535    324.2094
  carot3 |       9     1877.63    143.2934    429.8801    1547.195    2208.065
---------+--------------------------------------------------------------------
    diff |       9   -1642.148    130.9412    392.8236   -1944.099   -1340.197
------------------------------------------------------------------------------
     mean(diff) = mean(carot0 - carot3)                           t = -12.5411
 Ho: mean(diff) = 0                           degrees of freedom =        8

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

a. From the above analyses, what conclusions do you reach about the trend in plasma beta carotene levels over the course of treatment in the placebo group? Provide point estimates, confidence intervals, and P values, and scientific interpretation.

Ans: Over the nine month period of treatment, the observed plasma beta carotene level decreased on average by 102, and based on the 95% CI, such an observation would be unusual if the true average change in plasma levels were to decrease by less than 39.8 or decrease by greater than 163. Based on the P value of .0069, we can reject the null hypothesis of no change in plasma levels over that time period. ***Because this is the placebo group, I would guess that any such tendency toward decreased levels were due to changes in diet (perhaps season or secular trends in the tendency to ingest foods with beta carotene), effects due to aging (everyone was 9 months older—okay, I don't really think 9 months would make such a difference), or perhaps laboratory drift (if the samples were analyzed at different times).***

(Note that it was a VERY BIG mistake to just compare the 95% CI at baseline and after 9 months to see whether they overlapped. The data from those two time points are correlated, because they were made on the same subjects. So the "elevator statistics" approach looking for nonoverlapping CI does not work here: It is possible to have nonoverlapping CI and not have statistical significance. Furthermore, even if the data had been independent, overlapping CI as we have here should be regarded as noninformative, rather than evidence that you cannot reject the null. In fact, as we see from the statistics for the difference, the results are relatively highly statistically significant.)

b. From the above analyses, what conclusions do you reach about the trend in plasma beta carotene levels over the course of treatment in the high dose group? Provide point estimates, confidence intervals, and P values, and scientific interpretation.

Ans: Over the nine month period of treatment, the observed plasma beta carotene level increased on average by 16422, and based on the 95% CI, such an observation would be unusual if the true average change in plasma levels were to increase by less than 1340 or increase by greater than 1944. Based on the P value less than 0.0001, we can reject the null hypothesis of no change in plasma levels over that time period. ***Because this group received supplementation, we would like to believe that this increase is a direct result of the supplement.***

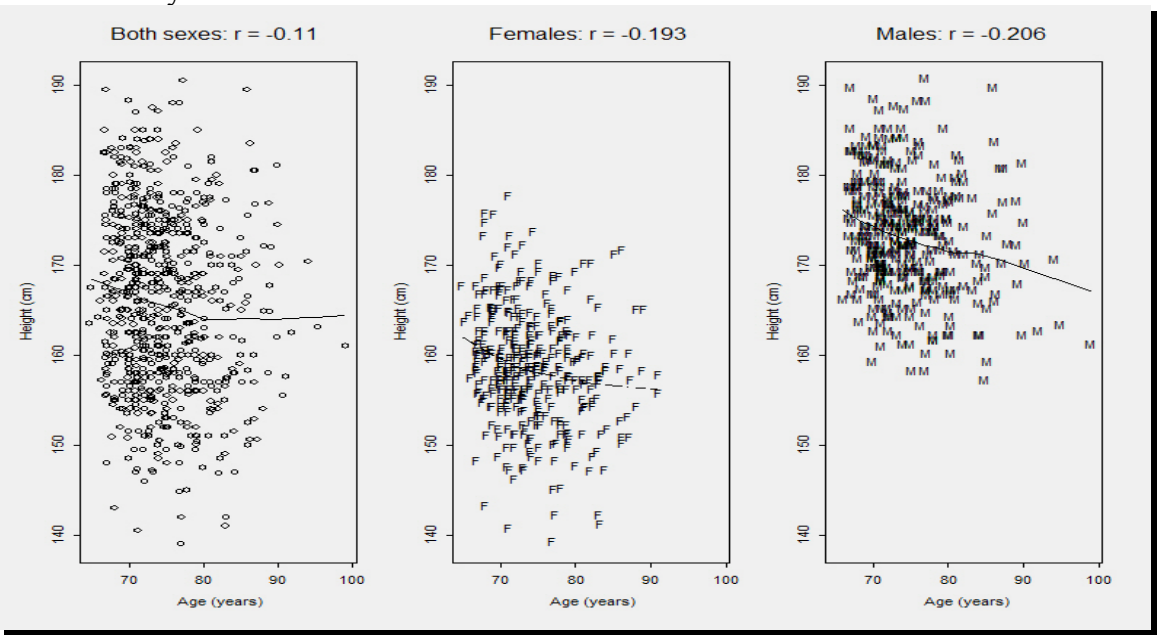
However, our observations for this group are confounded by time: The measurements made following supplementation were made when the subjects were 9 months older (so confounded with age) and 9 months later (so confounded with season and secular trends in calendar time). Ultimately, we will need to compare this group's results to the placebo group's results.

(Note again (and again and again and again) that it was a VERY BIG mistake to just compare the 95% CI at baseline and after 9 months to see whether they overlapped. The data from those two time points are correlated, because they were made on the same subjects. So the "elevator statistics" approach looking for nonoverlapping CI does not work here: It is possible to have nonoverlapping CI and not have statistical significance. Furthermore, even if the data had been independent, overlapping CI as we have here should be regarded as noninformative, rather than evidence that you cannot reject the null. In fact, as we see from the statistics for the difference, the results are relatively highly statistically significant.)

c. What conclusions do you reach about the effect of beta carotene supplementation on plasma beta carotene levels? Justify your answer, providing evidence for your conclusions.

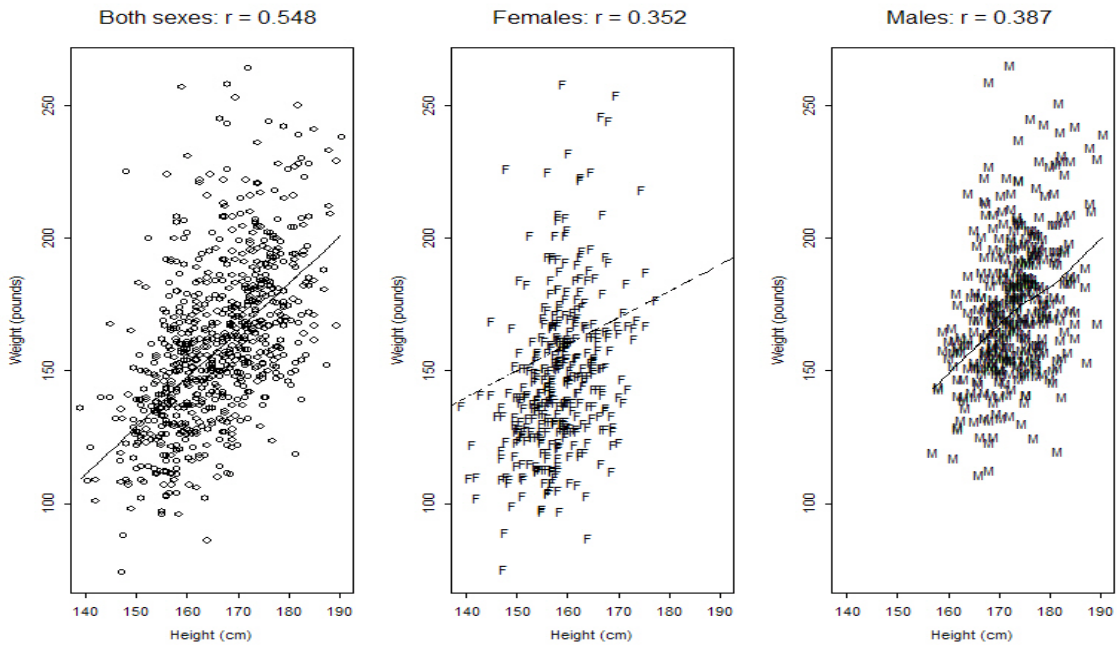
<u>Ans:</u> The placebo group had a statistically significant decrease, and the dose 60 group had a statistically significant increase. The two groups were independent of each other, so we can be confident that a comparison of the two groups would show a statistically significant effect of beta carotene supplementation on increasing plasma beta carotene levels.

5. (10 points) In an elderly population (age 65-97), the correlation between age and height was found to be r = -0.11 in the total sample, but r = -0.19 in females and r = -0.21 in males. The following figure shows scatterplots of the data along with lowess smooths. What can explain the difference in correlation observed in the total sample and the subgroups defined by sex?



Ans: The variance of the predictor Age and the downward trend in the slope are roughly comparable for each sex separately and the combined sample. On the other hand, the vertical spread of the data about the line in each sex is much less than that in the combined sample (this is evidence of a sex effect on height after adjusting for age), thus suggesting that the within predictor group variance of height is less in the individual sex strata than in the combined sample. This would make the correlation more extreme in the sex strata.

6. (10 points) In an elderly population (age 65-97), the correlation between height and weight was found to be r = 0.55 in the total sample, but r = 0.35 in females and r = 0.39 in males. The following figure shows scatterplots of the data along with lowess smooths. What can explain the difference in correlation observed in the total sample and the subgroups defined by sex?



Ans: The variance of the response Weight within groups defined by the predictor Height, as well as the upward trend in the slope, are roughly comparable for each sex separately and the combined sample. On the other hand, the variability of the predictor Height is much greater in the combined sample than in the individual sex strata. This would make the correlation more extreme in the combined sample.

7. (30 points) A study was performed to assess the utility of serum albumin (a blood protein) for prognosis after primary treatment for non small cell lung cancer. The goal was to determine if low albumin would predict who would die within 1 year following treatment. The study gathered data on 200 patients who survived more than 1 year and 300 patients

who died within 1 year. The following table provides the observed numbers of patients with albumin below 2.5 g/dl.

	Albumin < 2.5 g/dl	**Albumin ≥ 2.5g/dl**	**Total**
Survived > 1 year	30	170	200
Survived ≤ 1 year	90	210	300
Total	120	380	500

a. Can the above data be used to estimate the probability of surviving for 1 year past treatment? If so, provide the estimate. If not, briefly explain why not.

Ans: No. The proportion of patients in the sample surviving past 1 year was determined by study design.

b. Can the above data be used to estimate the sensitivity of a low albumin in predicting survival for 1 year past treatment? If so, provide the estimate. If not, briefly explain why not.

Ans: Yes. Of the 300 subjects who died within a year, 90 (or 30%) had low albumin.

c. Can the above data be used to estimate the specificity of a low albumin in predicting survival for 1 year past treatment? If so, provide the estimate. If not, briefly explain why not.

Ans: Yes. Of the 200 subjects who survived past a year, 170 (or 85%) had high albumin.

d. Can the above data be used to estimate the positive predictive value of a low albumin in predicting survival for 1 year past treatment? If so, provide the estimate. If not, briefly explain why not.

Ans: No. The proportion of patients in the sample surviving past 1 year was determined by study design.

e. Can the above data be used to estimate the negative predictive value of a low albumin in predicting survival for 1 year past treatment? If so, provide the estimate. If not, briefly explain why not.

Ans: No. The proportion of patients in the sample surviving past 1 year was determined by study design.

f. Suppose the one year probability of survival following treatment for non small cell lung cancer is 50%. What is the positive predictive value of an albumin less than 2.5 g/dl? Explain how you derived your answer.

Ans: Using Bayes' Rule:

$$PV_+ = \Pr(Death \mid lowAlb) = \frac{\Pr(lowAlb \mid Death) \times \Pr(Death)}{\Pr(lowAlb \mid Death) \times \Pr(Death) + \Pr(lowAlb \mid Survive) \times \Pr(Survive)}$$

$$= \frac{sens \times prev}{sens \times prev + (1 - spec) \times (1 - prev)} = \frac{0.30 \times 0.5}{0.30 \times 0.5 + 0.15 \times 0.5} = 0.667$$

8. (20 points) The following table contains descriptive statistics for FEV (l/sec) in a population aged 65 – 100. Descriptive statistics are presented for the combined sample, as well as within strata defined by smoking history (ever versus never). Also presented are descriptive statistics defined by smoking history within each sex.

	n	msng	mean	std dev	min	25%-ile	median	75%-ile	maximum
All	735	10	2.21	0.69	0.41	1.75	2.16	2.65	4.47
All Nonsmok	322	4	2.22	0.66	0.57	1.79	2.15	2.57	4.47
All Smokers	414	8	2.20	0.71	0.41	1.70	2.16	2.69	4.21
Nonsmok Females	200	4	1.94	0.43	0.57	1.62	1.99	2.22	2.86
Smoking Females	171	4	1.77	0.45	0.57	1.52	1.79	2.07	2.93
Nonsmok Males	123	1	2.67	0.72	0.58	2.27	2.66	3.03	4.47
Smoking Males	244	5	2.49	0.71	0.41	2.08	2.55	2.95	4.21

a. Is there evidence of an association between FEV and smoking? Provide descriptive statistics in support of your answer.

Ans: No. The mean FEV for smokers was .02 l/sec lower than that for nonsmokers. Such a difference was not so great that I would consider this evidence for an association.

(You could have compared medians or some other quantile. However, you only got half credit if you adjusted for sex in this question.)

b. Is there evidence that the association between FEV and smoking is confounded by sex? Provide descriptive statistics in support of your answer.

Ans: Yes. There is strong evidence that sex is associated with FEV: Among nonsmokers, males average an FEV of 2.67 while females average 1.94. A similar difference exists between the sexes among smokers. This, of course, fits in well with our understanding about the relationship between sex and body size and body size and FEV.

There is also an association between sex and smoking in the sample: 171 / 371 females smoke, while 244 / 367 males smoke.

Sex is most certainly not in any causal pathway of interest between smoking and differences in FEV, hence the above observations are sufficient for me to conclude that sex confounds the detection of an FEV – smoking association.

(A common symptom of confounding is that an unadjusted analysis differs substantially from an analysis adjusted for the confounder. In this example, we see that the difference between average FEV in nonsmokers and that in smokers is 0.18 l/sec for males and 0.17 l/sec for females. These stratified estimates of association are remarkably similar to each other, but they are quite different from the unadjusted analysis reported in part a. This would be enough to make me suspect confounding when using any summary measure as the basis for measuring association. When using the difference of means, this is actually diagnostic for confounding, though it would not necessarily prove confounding if we were using odds ratios as our measure of association. In any case, for this exam, I gave full credit if you pointed out the difference between the stratified and unadjusted analysis results. Be forewarned, however, that were I to ever use the odds ratio as a measure of association in such a problem, I would not have given full credit for such an argument.)

c. Is there evidence that the association between FEV and smoking is modified by sex? Provide descriptive statistics in support of your answer.

Ans: No. The difference between average FEV in nonsmokers and that in smokers is 0.18 l/sec for males and 0.17 l/sec for females. These stratified estimates of association are remarkably similar to each other, thus the effect of smoking on FEV is not modified by sex.

(Note that I chose to measure the association between smoking and FEV by the difference in average FEV between smokers and nonsmokers. My finding of no association is of course dependent upon the measure of association chosen. Had I decided to use ratio of mean FEV, I would have found that male smokers' average FEV was 2.49 /2.67 = 0.933 that of male nonsmokers and that female smokers' average FEV was 1.77 / 1.94 = 0.912 that of female nonsmokers. The answer would then depend upon whether I thought that 0.933 was sufficiently different from 0.912 to regard it as effect modification.)

d. What statistic would you present to describe the association between FEV and smoking? Provide the sentence you would use to report the results of your analysis.

Ans: While we found that the average FEV of 2.20 l/sec in smokers was only negligibly less than the average FEV of 2.22 l/sec in nonsmokers, this seeming lack of association may have

been due to the fact that a disproportionate number of smokers were male and males, due to their tendency toward larger body size, naturally average higher FEV. When we compare smokers to nonsmokers of the same sex, we find that smokers average 0.175 l/sec lower FEV.