Biost 517
Applied Biostatistics I

Final Examination Key
December 12, 2007

Name: ____

<u>Instructions:</u> 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.

<u>NOTE</u>: When you need to make calculations, always use at least four significant digits in your intermediate calculations, and report at least three significant digits. (Example: 1.045 and 0.0001234 and 1234000 each have four significant digits.)

If you come to a problem that you believe cannot be answered without making additional assumptions, <u>clearly</u> state the <u>reasonable</u> 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.

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

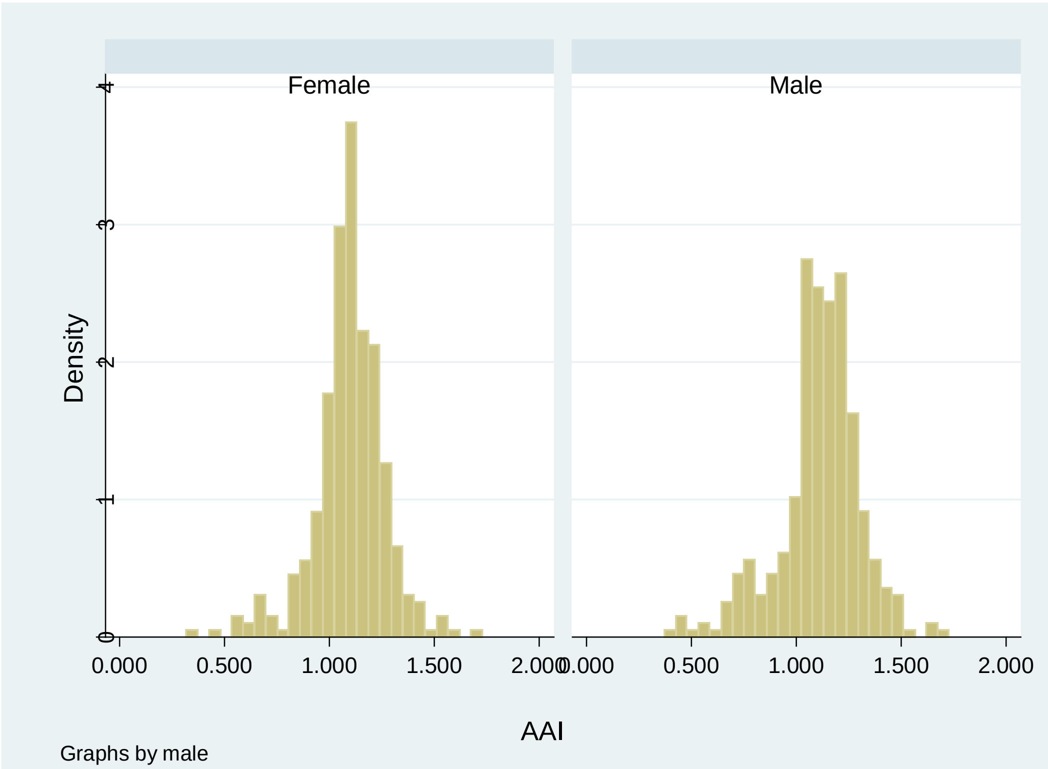
Signed:____

All problems consider investigations of peripheral vascular disease in the elderly. In particular, we are interested in how decreased blood flow in the legs might vary by sex and age. Our measure of peripheral vascular disease is the "ankle : arm index" (abbreviated *AAI*), which is the ratio of the systolic blood pressure measured in the ankle, to the systolic blood pressure measured in the arm. Ratios less than 1.0 suggest that the patient is getting less blood flow to his/her legs, which is generally regarded as a bad thing.

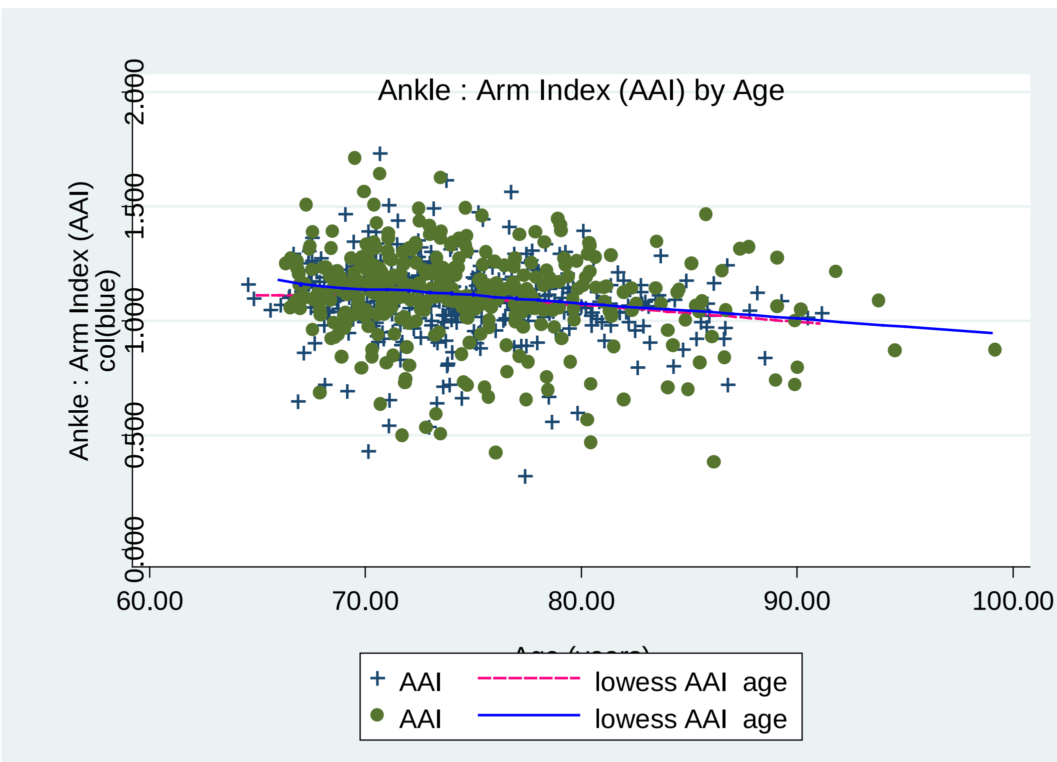
Descriptive statistics are provided on the following variables for data gathered on 735 patients:
age: the patient's age in years
male: an indicator that the patient is male (so 0= female, 1= male)
AAI: the ankle arm index (a dimensionless ratio, so having no units)

	N	**Mean**	**SD**	**Min**	**p25**	**Median**	**p75**	**Max**
Females (n= 369)								
Age	369	74.41	5.26	65	71	73	78	91
AAI	364	1.095	0.168	0.317	1.024	1.102	1.190	1.728
Males (n= 366)								
Age	366	74.73	5.64	66	71	74	78	99
AAI	362	1.111	0.196	0.421	1.033	1.126	1.226	1.711
All Subjects (n= 735)								
Age	735	74.57	5.45	65	71	74	78	99
AAI	726	1.103	0.183	0.317	1.027	1.112	1.207	1.728

Histogram of *AAI* by sex.



Scatterplot of AAI by Age and superimposed lowess smooths within sex strata (females denoted by "+" and dashed line, males denoted by "•" and solid line).



1. (15 points) The following is Stata output reporting the results of t-tests performed on the data.

```
. ttest AAI, by(male)
Two-sample t test with equal variances
   Group |  Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
  Female |  364    1.095299    .0088288    .1684419    1.077938    1.112661
    Male |  362    1.111236    .0103106    .1961721     1.09096    1.131513
combined |  726    1.103246    .0067859    .1828431    1.089923    1.116568
    diff |        -.0159367    .0135684               -.0425748    .0107014
    diff = mean(Female) - mean(Male)                             t =  -1.1745
Ho: diff = 0                                 degrees of freedom =      724
    Ha: diff < 0               Ha: diff != 0               Ha: diff > 0
 Pr(T < t) = 0.1203       Pr(|T| > |t|) = 0.2406       Pr(T > t) = 0.8797

. ttest AAI, by(male) unequal
Two-sample t test with unequal variances
   Group |  Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
  Female |  364    1.095299    .0088288    .1684419    1.077938    1.112661
    Male |  362    1.111236    .0103106    .1961721     1.09096    1.131513
combined |  726    1.103246    .0067859    .1828431    1.089923    1.116568
    diff |        -.0159367    .0135741                -.042587    .0107136
    diff = mean(Female) - mean(Male)                             t =  -1.1741
Ho: diff = 0                     Satterthwaite's degrees of freedom =  706.651
    Ha: diff < 0             Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1204      Pr(|T| > |t|) = 0.2408         Pr(T > t) = 0.8796
```

a. From the above analyses (state which one you use and why), what conclusion do you reach about an association between *AAI* and sex. Make clear the summary measure used to define an association, provide estimates of that summary measure if possible, and quantify the strength of statistical evidence used to justify your conclusion.

<u>Ans:</u> **The above t tests compare the groups with respect to their average (or arithmetic mean) AAI. I prefer using the t test that allows for the possibility of unequal variances, because I am most interested in inference about differences in a "central tendency" rather than just any difference in the distributions.** *(I do note that with approximately equal sample sizes, there is little difference between the results obtained in these two analyses, despite there being some suggestion of slightly different variances between men and women.)*

Men average AAI of 1.11, while women average AAI of 1.10. These observed results estimate an average AAI for men that is 0.0159 higher than that for women, with a 95% CI of 0.0426 higher to 0.0107 lower. Hence, these data are not unusual if the true difference in mean AAI were to be 0 (P = 0.2408), and thus we do not have sufficient evidence to state with high confidence that men and women differ in their average AAI.

2. (20 points) Suppose instead of the t-tests presented in problem 1, I had instead performed a classical linear regression (without "robust" standard errors) of *AAI* as the response variable and *male* as the predictor variable.

    a. What would be the estimate of the intercept from that regression?

<u>**Ans:**</u> **1.095** *(As the interpretation of the intercept would be the average AAI for women, and because classical linear regression with a binary predictor is exactly equivalent to the t test that presumes equal variances, the estimated intercept must be the sample mean for women.)*

    b. What would be the estimate of the slope from the linear regression?

<u>**Ans:**</u> **0.0159** *(As the interpretation of the slope would be the difference in average AAI associated with a one unit difference in the male variable, and because classical linear regression with a binary predictor is exactly equivalent to the t test that presumes equal variances, the estimated slope must be the difference in sample means, men minus women.)*

    c. What would have been the P value reported for the test that the slope parameter was 0?

<u>**Ans:**</u> **0.2406** *(As classical linear regression with a binary predictor is exactly equivalent to the t test that presumes equal variances, the P values will be exactly the same)*

    d. How would your answers to parts a – c change if I had used the robust standard error estimates?

<u>**Ans:**</u> **The estimates will not change, the p value could get either larger or smaller, depending on how the sample sizes vary across groups with different variances. When using robust standard error estimates, linear regression with a binary predictor approximates (but is not exactly equal to) the t test that allows for the possibility of unequal variances, so we cannot be sure what the p value would be, though it should not be much different from 0.2408.** *(In this case, it turned out to agree to 4 decimal places: It was 0.2408. And that was not very different from the t test that presumes equal variances because the sample sizes were nearly equal for the two groups.).*

3. (15 points) Suppose I created a variable *logAAI* that is the log transformation of the *AAI* measurement. The following is Stata output reporting the results of a t-test performed on the log transformed data.

```
. g logAAI= log(AAI)
. ttest logAAI, by(male) unequal
Two-sample t test with unequal variances
   Group |      Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
  Female |      364    .0772517    .0091816    .1751735    .0591959    .0953075
    Male |      362    .0870785    .0106321    .2022903    .0661698    .1079872
combined |      726    .0821516    .0070187    .1891143    .0683722    .0959309
    diff |            -.0098268    .0140479               -.0374073    .0177538
    diff = mean(Female) - mean(Male)                              t =  -0.6995
Ho: diff = 0                      Satterthwaite's degrees of freedom =  708.399
```

```
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2422         Pr(|T| > |t|) = 0.4845          Pr(T > t) = 0.7578
```

a. From the above analyses, what conclusion do you reach about an association between *AAI* and sex. Make clear the summary measure used to define an association, provide estimates of that summary measure if possible, and quantify the strength of statistical evidence used to justify your conclusion.

Ans: A t test performed on log transformed data is comparing the difference in log geometric means, and when that difference is exponentiated, we are considering the ratio of geometric means.

The geometric mean AAI for men is estimated to be $e^{0.0870785} = 1.091$, while the geometric mean AAI for women is estimated to be $e^{0.0772517} = 1.080$. These observed results therefore estimate that the geometric mean for women is only $e^{-0.0098268} = 0.990$ times as high as that for men, with a 95% CI of $e^{-0.0374073} = 0.963$ times as high to $e^{0.0177538} = 1.108$ times higher. Hence, these data are not unusual if the true ratio of geometric mean AAI were to be 1 (P = 0.4845), and thus we do not have sufficient evidence to state with high confidence that men and women differ in their geometric mean AAI.

(Alternative quantification with women as the reference group (i.e., just negating the estimated difference and the CI): The geometric mean AAI for men is estimated to be $e^{0.0870785} = 1.091$, while the geometric mean AAI for women is estimated to be $e^{0.0772517} = 1.080$. These observed results therefore estimate that the geometric mean for men is $e^{0.0098268} = 1.01$ times higher than that for women, with a 95% CI of from only $e^{-0.0177538} = 0.982$ times as high to $e^{0.0374073} = 1.038$ times higher.

(Alternative quantification based on percentages: The geometric mean AAI for men is estimated to be $e^{0.0870785} = 1.091$, while the geometric mean AAI for women is estimated to be $e^{0.0772517} = 1.080$. These observed results therefore estimate that the geometric mean for women is $100\,(1 - e^{-0.0098268}) = 0.978\%$ lower than that for men, with a 95% CI of $100\,(1 - e^{-0.0374073}) = 3.67\%$ lower to $100\,(e^{0.0177538} – 1) = 1.79\%$ higher.)

(Alternative quantification based on percentages with women as the reference group: The geometric mean AAI for men is estimated to be $e^{0.0870785} = 1.091$, while the geometric mean AAI for women is estimated to be $e^{0.0772517} = 1.080$. These observed results therefore estimate that the geometric mean for men is $100\,(e^{0.0098268} – 1) = 0.988\%$ higher than that for women, with a 95% CI of $100\,(1 - e^{-0.0374073}) = 3.67\%$ lower to $100(e^{0.0177538} – 1) = 1.79\%$ higher.)

4. (15 points) Suppose I created a variable *AAIlow* that indicates when the *AAI* is less than or equal to 0.9.The following is Stata output reporting the results of analyses performed on the indicator variable

```
. g AAIlow= AAI
. recode AAIlow 0/0.9=1 0.9/max=0
. ttest AAIlow, by(male) unequal
Two-sample t test with unequal variances
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
  Female |     364    .0961538    .0154731    .2952079    .0657257     .126582
    Male |     362     .121547     .017198    .3272141    .0877261    .1553678
combined |     726    .1088154    .0115654    .3116222    .0861098    .1315211
    diff |           -.0253931    .0231341                -.070812    .0200257
    diff = mean(Female) - mean(Male)                              t =  -1.0976
Ho: diff = 0                     Satterthwaite's degrees of freedom =  715.643
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1364         Pr(|T| > |t|) = 0.2727          Pr(T > t) = 0.8636

. cs AAIlow male
                 | male                   |
                 |   Exposed   Unexposed  |       Total
           Cases |        44          35  |          79
        Noncases |       318         329  |         647
           Total |       362         364  |         726
                 |                        |
            Risk |   .121547    .0961538  |    .1088154
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
Risk difference |         .0253931       |   -.0198864    .0706726
      Risk ratio |         1.264088       |    .8310635    1.922741
 Attr. frac. ex. |         .2089161       |   -.2032775     .479909
 Attr. frac. pop |         .1163583       |
                             chi2(1) =     1.21  Pr>chi2 = 0.2720
.
```

a. From the above analyses (state which one you use and why), what conclusion do you reach about an association between *AAI* and sex. Make clear the summary measure used to define an association, provide estimates of that summary measure if possible, and quantify the strength of statistical evidence used to justify your conclusion.

<u>Ans:</u> **I greatly prefer to summarize the distribution of a binary random variable within each group by the proportion, and to compare the groups with respect to the difference in proportions.** *(Alternatives would include using the proportions to summarize group, and then using the risk ratio to compare groups, or summarizing groups using the odds and then using the odds ratio to compare groups. Generally, the risk ratio is of greatest interpretability when investigating rare events, and in that setting, the odds ratio is a very good approximation to the risk ratio. Furthermore, the odds ratio computed from case-control sampling estimates the exact same quantity as the odds ratio computed from cohort study sampling. Hence, when studying a rare condition, we might use the case-control study for its greater ease of implementation, and summarize our data using the odds ratio in order to estimate the risk ratio that would have been observed in a cohort study. I note, however, that this is not the setting for this problem, so I will stick with inference based on the difference in proportions.)*

Traditionally, we would use the chi square test when comparing the distribution of binary random variables across two groups. *(As noted repeatedly in class, the statistical theory that allows us to use the t test to compare means would also apply to this setting which considers the mean of a binary random variable. When only considering the null hypothesis of equality of proportions across groups, it is entirely correct to use the t test that presumes equal variances, because if the proportions are equal, the variances must also be equal. When considering alternative hypotheses, the variances will tend to be unequal across groups. Hence, for CI it is probably better to use the t test that allows unequal variances. The t test uses the sample variance (i.e., dividing by n-1 instead of n) and uses the critical values from the t distribution, rather than the standard normal distribution. These two aspects will tend to make the t test just slightly more conservative than the chi square test, thus tending toward higher P values and wider confidence intervals. On the other hand, the way that the sample variances are combined across groups in the t test tends to lower values for the variance in the combined sample than does the chi square test, and that aspect would tend to cause the t test to have lower P values and narrower CI. Still, when this aspect is combined with the handling of the n vs n-1 and the critical value, the t test is generally slightly more conservative. Typically, it does not make a dime's worth of difference.)*

An AAI lower than 0.9 was observed in 12.2% of men and in 9.62% of women. We thus estimate that the proportion of men having low AAI is an absolute 2.54% higher than that for women, with a 95% CI for the absolute difference in proportions suggesting that the data are consistent with the men's proportion being anywhere from 1.99% lower to 7.07% higher. This observed difference is thus not atypical of the setting in which there was no true difference between men and women in the tendency to have an AAI less than 0.9 (P = 0.2720). *(Note my wording using "absolute difference". This is because it is not uncommon for people to speak of "relative difference" in proportions. I find it best to clarify my wording at all times. For instance, in this case, the risk ratio of men : women was 1.264. Hence, we could talk about the "relative difference" in proportions being 26.4%, because the proportion of men having low AAI was the women's proportion plus an additional 26.4% of the women's proportion. There are scientific reasons to sometimes prefer the relative difference. For instance, suppose we take a population of people who are HIV-, only proportion x of whom will be exposed to the virus. We administer either a placebo or a vaccine that is thought to reduce transmission of the virus by 2/3 (so if 100% of HIV exposed people will become infected when unvaccinated, we expect 33% of HIV exposed people to become infected after vaccination). Now, at the start of our study, we are unsure of the value of x. However, in the placebo group, we expect proportion x to become infected (the (1 – x) proportion representing unexposed people will not become infected), and in the vaccine group we expect proportion x/3 to become infected (neither the (1 – x) proportion representing unexposed people nor 2/3 of the x proportion of exposed people will become infected). If we were to take the difference in proportions, the difference x/3 - x depends heavily on the unknown value of x, and we do not know how to generalize our findings to the overall population of people at risk for infection. However, the ratio of the rates of infection consistently estimates the true risk ratio of 1/3 for the relevant population. It should be noted that this works because the "contamination" in our sample (i.e., inclusion of people not at risk for the event) corresponds to a group of people who would have no infection whether they were vaccinated or not. We could derive a similar approach when dealing with contamination by a population who would all have the event whether treated or not: We would just consider the rate of not having an event. This does not work, however, if the contaminating population has some event rate p that is strictly between 0 and 1 and that is unaffected by treatment. In that case, the risk ratio would be (p (1-x) + x/3) / (p (1-x) + x) and the* *risk difference would by x/3 – x. In both cases, the comparison across groups depends on the value of x.*

It should be noted that the relative difference is based on the risk ratio of 1.264, NOT the odds ratio. The estimated odds ratio for this data would be 1.301.)

5. (15 points) The following is Stata output reporting the results of Wilcoxon ranksum test performed on the data.

```
. ranksum AAI, by(male) porder

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        male |      obs    rank sum    expected
-------------+---------------------------------
      Female |      364    126008.5      132314
        Male |      362    137892.5      131587
-------------+---------------------------------
    combined |      726      263901      263901

unadjusted variance  7982944.67
adjustment for ties      -26.66
                     ----------
adjusted variance    7982918.01

Ho: AAI(male==Female) = AAI(male==Male)
             z =  -2.232
    Prob > |z| =   0.0256

P{AAI(male==Female) > AAI(male==Male)} = 0.452
```

a. From the above analyses, what conclusion do you reach about an association between *AAI* and sex. Make clear the summary measure used to define an association, provide estimates of that summary measure if possible, and quantify the strength of statistical evidence used to justify your conclusion.

<u>Ans:</u> **The Wilcoxon rank sum test compares the groups with respect to the probability that a randomly chosen female might have a higher AAI value than a randomly chosen male. Specifically, it considers the null hypothesis of exact equality of the two groups with respect to the entire distribution. Under this null hypothesis, the probability that a randomly chosen female would have a higher AAI value than a randomly chosen male should be 0.5.**

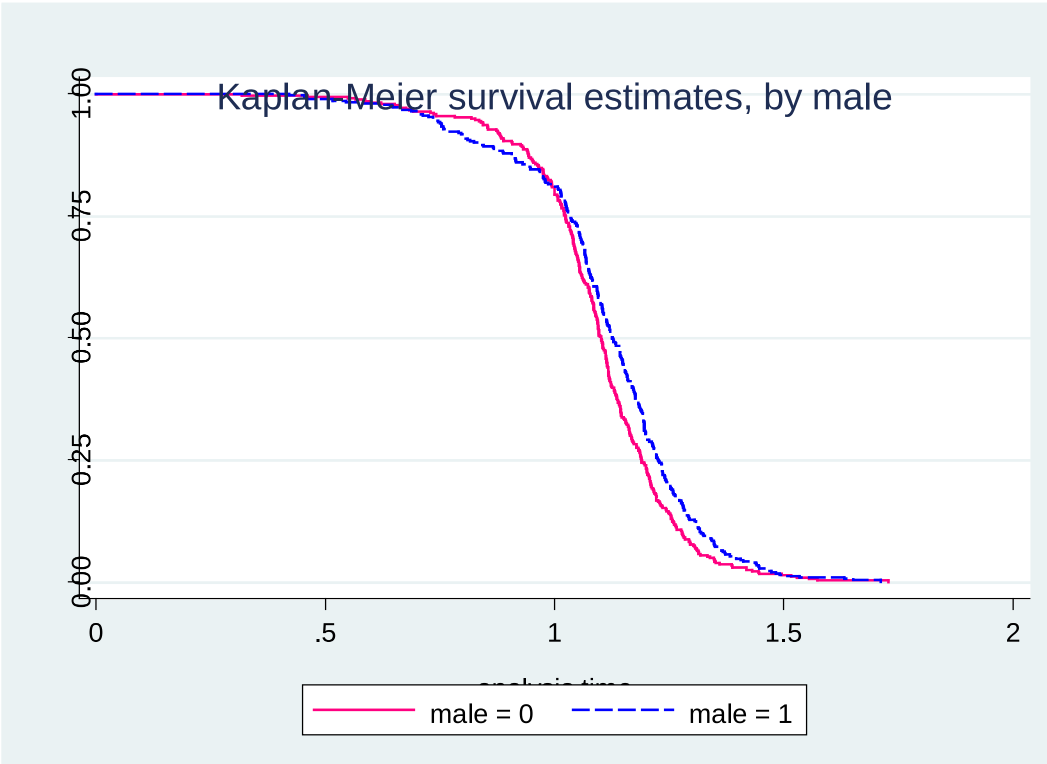
Based on a two-sided P value of 0.0256, we reject the null hypothesis of equality of AAI distributions for the two sexes. *(In this sample, there was a 45.2% probability that a randomly chosen woman's AAI would be higher than that for a man. However, we cannot use the Wilcoxon rank sum test to provide a CI for this probability in the general setting, because the Wilcoxon test presumes that when the probability is 50%, the two groups would have the exact same distribution. This need not be the case. For instance, I can construct two distributions that have the same mean, the same median, and that have a 50% probability that a randomly chosen individual from one group would have a larger value than a randomly chosen value from the other group. However, when using the Wilcoxon rank sum test to test the weak null that Pr (Y > X) = 0.5, it has the wrong type I error, incorrectly rejecting the null hypothesis about 11% of the time. But when viewed as a test of the strong null hypothesis, it is an "inconsistent test": Even with an infinite sample size, it will not always reject the null hypothesis in this case where the two distributions are in fact* *different. It can also be a "biased test", because sometimes when the strong null hypothesis is not true, the test rejects the null hypothesis with a probability less than the desired type I error- so the power in that situation is less than the type I error.)*

6. (15 points) The following is Stata output reporting the results of a proportional hazards analysis performed on the data.

```
. stset AAI
. stcox male, robust

         failure _d:  1 (meaning all fail)
   analysis time _t:  AAI

Cox regression -- Breslow method for ties

No. of subjects      =          726              Number of obs   =       726
No. of failures      =          726
Time at risk         =  800.9564995
                                                 Wald chi2(1)    =      6.23
Log pseudolikelihood =   -4058.1516              Prob > chi2     =    0.0125
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .8278939   .0626331    -2.50   0.013     .7138027    .9602208
------------------------------------------------------------------------------
```

a. Why is this analysis a bit unusual? (That is, when would this analysis more typically be used?) Is it valid in this instance?

<u>Ans:</u> **The proportional hazards model and the logrank test are most often used when the observations are right censored. There is no requirement that it only be used in that situation, however, and thus this is a valid analysis. It would be the most interpretable when the two groups exhibit "proportional hazards". Under the strong null hypothesis (exact equality of the entire distribution), proportional hazards does hold. So we can trust that the hypothesis test would have the correct type I error, regardless.** *(Inference under the alternative is best when we would have proportional hazards, though the use of the robust standard error protects us against an inflating the type I error when proportional hazards does not hold. Understanding what the hazard ratio means is very difficult when we do not have proportional hazards, however. For what it is worth, these data do not at all exhibit proportional hazards. By careful inspection of the histogram and the tabulated descriptive statistics, you can see that the females seem to be prone to more extreme low values <u>and</u> more extreme high values than the males, but then the bulk of the females appear to be more concentrated in the center of the distribution. This would lead to "crossing survival curves" as shown below. A setting with proportional hazards would not tend to exhibit such a pattern—though we do need to consider whether this is just random sampling error. In this case, I do have a larger data set with which I can convincingly demonstrate that the proportional hazards assumption does not hold between males and females for AAI in the CHS data.)*



b. From the above analyses, what conclusion do you reach about an association between *AAI* and sex. Make clear the summary measure used to define an association, provide estimates of that summary measure if possible, and quantify the strength of statistical evidence used to justify your conclusion.

Ans: The above analysis summarizes the difference between males and females using the hazard ratio. *(The hazard is typically described as the instantaneous rate of observing an event, but that wording does not seem particularly relevant in this setting. Suffice it to say that the hazard at any particular value a is the probability that the AAI=a, given that the AAI is at least as large as a. In a setting in which we are unsure that the hazard ratio between the groups is constant (that is, in a setting where there might not be proportional hazards), the estimated hazard ratio from a proportional hazards model is some time average of the hazard ratio.)*

From the proportional hazards analysis, we observed that the hazard for males is only 0.828 times as large as the hazard for females, with a 95% CI from 0.714 to 0.960. Such an observation is unusual were the two groups to have the same distribution of AAI (P = 0.013). Because the hazard ratio is less than 1, the males seem in some sense to have a tendency toward larger AAI over the range of the distributions.

7. (10 points) How would you resolve any discrepancies among the above analyses? That is, provide possible explanations for making different conclusions about the association between *AAI* and sex.

Ans: *(For this problem, I wanted at a minimum a discussion of the possibility that one or more of the tests might have been a statistical error, as well as some mention of the difference in* *interpretations based on the weak or strong null hypotheses. I provide here a far more in depth discussion that I would have expected during the examination.)*

In the above analyses, we were unable to demonstrate with high confidence that the distributions of AAI differed between males and females with respect to their means, their geometric means, or the probability of having a measurement less than 0.9. Possible explanations for these "negative" results include:

- **The distribution of AAI might be exactly the same for males and females.** *(In this case, the strong null is true.)*

- **The distribution of AAI might be different for males and females, but those distributions have the same mean (or geometric mean or probability of a value below 0.9).** *(In this case the respective weak nulls would be true.)*

- **The distribution of AAI might be different for males and females in such a way that they do not have the same mean (or geometric mean or probability of a value below 0.9), but we did not have enough precision with this sample size to be able to statistically demonstrate that difference: a type II statistical error.** *(In this setting, the weak nulls are false. And if the weak nulls are false, the strong nulls also have to be false.)*

We were able to state with high confidence that the distributions differed with respect to the hazard ratio (males tended toward a lower hazard in some sense) and the probability that a randomly chosen male would have a larger value than a randomly chosen female (this probability was greater than 0.5). Possible explanations for these "positive" results include:

- **The distribution of AAI might be different for males and females in such a way that the average hazard ratio is not 1 (or the probability is not 0.5 that a randomly chosen male would have a larger value than a randomly chosen female).** *(In this setting, the analysis would have been correct in telling us to reject the weak null hypothesis.)*

- *(This one is only relevant for the Wilcoxon rank sum test among the tests we performed, because that test used a permutation-type variance estimate which presumes that when the weak null is true, the strong null also has to be true. Had we obtained a statistically significant result for the t test that presumes equal variances, we would have had to worry about this issue there, too, because it presumes that when the weak null is true, the variances also have to be equal.)* **The distribution of AAI might be different for males and females in such a way that would cause the Wilcoxon rank sum test to be statistically significant with probability greater than 0.05, even though the probability is 0.5 that a randomly chosen male would have a larger value than a randomly chosen female.** *(This is merely saying that we can correctly reject the strong null hypothesis, but it was a type I error when rejecting the weak null hypothesis. Furthermore, it is possible that our type I error for rejecting the weak null hypothesis was potentially greater than our desired level of 0.05. As noted above, the Wilcoxon test is an "inconsistent" test in the sense that it is not very powerful unless Pr(Y > X) is not 0.5.)*

- **The distribution of AAI might be different for males and females, but not in a way that would cause the average hazard ratio to be different from 1. Nevertheless, by random chance we observed results that were spuriously suggestive of a difference: a type I statistical error.** *(In this setting, the weak null hypothesis is true. By using the robust standard errors in the proportional hazards analysis, we were focusing only on the weak null hypothesis. The strong null might or might not be true in this instance.)*

There is a multiple comparison issue when performing all these different tests: If we make a type I error of 0.05 on each of five different tests, there is the possibility that the probability of falsely rejecting the null hypothesis for one of them might be as high as 0.25. A "Bonferroni correction" in this setting would multiply each P value by 5 to account for the multiple tests. Were we to do that, we would not be able to reject the null hypothesis for any of the tests: Even the proportional hazards analysis would report a P value of 0.065. *(There are, however, better approaches that could have been used in this instance.)*

(For what it is worth, the 735 subjects used in this analysis are just a subset of the 5,000 some subjects in the Cardiovascular Health Study. Were we to use the larger sample size available in the Inflammatory Markers data set, we would find:

- *Men have a statistically significant higher mean AAI than do women.*
- *Men have a statistically significant higher geometric mean AAI than do women.*
- *Men have a statistically significant tendency toward lower AAI in the sense that the probability of a value lower than 0.9 is higher for men than it is for women.*
- *The probability that a randomly chosen male has a higher value than a randomly chosen female is greater than 0.5. (This was checked by using a Wilcoxon test with something akin to the "robust standard errors".)*
- *The average hazard ratio comparing males to females is significantly less than 1.*

From these results, I would thus leap to the conclusion that in problems 1, 3, and 4, we made a type II statistical error when we failed to reject the weak null hypotheses, and in problems 5 and 6 we could correctly reject the weak null hypotheses. It is of course very much of interest that according to all these analyses in the larger data set (which I am pretending is the truth), men "tend to have larger AAI" than women as judged by the mean, the geometric mean, the hazard ratio, and Pr(Y > X), but women "tend to have larger AAI" than men when judged by the proportion with an AAI less than 0.9. (I have no solution for this scientific dilemma: Which is the most relevant summary measure to define "tends to". It is undoubtedly a function of the uses we intend to make of our analysis. This would certainly have to be worked out in collaboration with people who are experts in the field of cardiovascular disease.)

8. (40 points) The following output provides a linear regression analysis of *AAI* regressed on age.

```
. regress AAI age, robust
Linear regression                                  Number of obs =     726
```

```
                                                   F(  1,   724) =   24.83
                                                   Prob > F      =  0.0000
                                                   R-squared     =  0.0289
                                                   Root MSE      =   .1803
------------------------------------------------------------------------------
             |               Robust
         AAI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0056911    .001142    -4.98   0.000    -.0079331   -.0034491
       _cons |   1.527554   .0850613    17.96   0.000     1.360558    1.694551
------------------------------------------------------------------------------
```

a. Provide an interpretation of the intercept in the above model. What scientific use would you make of this estimate?

<u>Ans:</u> **From the model, we estimate that newborns would average an AAI of 1.53. This is extrapolating way outside our data, and I would make no use of this estimate.**

b. Provide an interpretation of the slope in the above model? What scientific use would you make of this estimate?

<u>Ans:</u> **From the model, we estimate that when comparing different age groups, the difference in mean AAI would be 0.00569 per year difference in age, with the older age group tending toward lower AAI. I would use this estimate to quantify an association between age and AAI.**

c. Based on the above analysis, is there evidence of an association between age and the ankle : arm index? Explain your reasoning.

<u>Ans:</u> **From the above analysis, we find that the slope parameter is statistically significantly different from 0 ($P < 0.0005$), and thus we can conclude with high confidence that there is an association between age and AAI.**

d. Based on the above analysis, can you estimate the correlation between *AAI* and age? If so, do so.

<u>Ans:</u> **The correlation between age and AAI would be the square root of the R^2 and it would have the same sign as the slope, so $r = -0.17$.**

e. Is there evidence of a statistically significant correlation between age and the *AAI*? Explain your reasoning.

<u>Ans:</u> **A test for nonzero correlation is exactly equivalent to a test for nonzero slope in linear regression. Hence, we can conclude that there is a statistically significant correlation.**

f. According to the above model, what would be your estimate of the average *AAI* in a population of 70 year olds? Do you feel that this estimate is trustworthy? Explain your reasoning.

<u>Ans:</u> **From the linear model, we would estimate that the average AAI in 70 year olds would be $1.5276 - 0.005691 \times 70 = 1.129$. For this estimate to be reliable, we would need to have the relationship among the age specific means to be truly linear. Examining the scatterplot in a *post hoc* fashion suggests that a straight line is a reasonably good fit to the data.**

g. According to the above model, what would be your estimate of the standard deviation of *AAI* in a population of 70 year olds? Do you feel that this estimate is trustworthy? Explain your reasoning.

Ans: **From the root MSE, we estimate that the square root of the average variance across age groups is 0.1803. If a straight line relationship holds across group means and the variance is the same in each group (so homoscedastic), then this estimated standard deviation could be expected to apply to every age group, and hence to 70 year olds. As noted above, a straight line would seem to approximate the mean relationship well. It is a little difficult to judge whether the data are homoscedastic, but my overall impression would likely be to trust this estimate as a good approximation.**

h. Do you believe the above analysis results might be confounded by sex? Explain your reasoning.

Ans: **From the table of descriptive statistics, the distribution of ages is remarkably similar for the two sexes. Hence, there would not be confounding.** *(Note that I do not even have to consider whether the AAI distribution might differ across sexes after adjusting for age, because I already know that there is no association between the potential confounder (sex) and my predictor of interest (age) in the sample.)*

9. (Bonus: 20 points) The following output provides a linear regressions of *AAI* on age separately for each sex..

```
. regress AAI age if male==0, robust
Linear regression                                      Number of obs =     364
                                                       F(  1,   362) =   11.99
                                                       Prob > F      =  0.0006
                                                       R-squared     =  0.0222
                                                       Root MSE      =  .16679
------------------------------------------------------------------------------
             |               Robust
         AAI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0047668   .0013765    -3.46   0.001    -.0074737   -.0020599
       _cons |    1.44989   .1037099    13.98   0.000     1.245941     1.65384
------------------------------------------------------------------------------

. regress AAI age if male==1, robust
Linear regression                                      Number of obs =     362
                                                       F(  1,   360) =   14.09
                                                       Prob > F      =  0.0002
                                                       R-squared     =  0.0361
                                                       Root MSE      =  .19286
------------------------------------------------------------------------------
             |               Robust
         AAI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    -.00659   .0017558    -3.75   0.000    -.0100428   -.0031372
       _cons |   1.603684   .1301177    12.32   0.000     1.347798    1.859571
------------------------------------------------------------------------------
```

a. Does the above analysis suggest evidence that sex modifies the effect of age on *AAI?* Justify your answer

Ans: **The first order trend in AAI by age in males has an estimated slope of – 0.00477, while the corresponding slope in females is – 0.00659. This 40% higher association might be suggestive of effect modification.**

b. Is any effect modification by sex statistically significant? Justify your answer.

Ans: **The difference in slopes between males and females is estimated to be 0.001823. The standard error for this estimated difference in slopes can be found from the square root of the sum of the squared standard errors for each slope: sqrt ($0.0013765^2 + 0.0017558^2$) = 0.002231.**

We thus find an approximate 95% CI as $0.001823 \pm 1.96 \times 0.002231$ = (- 0.00255, 0.00620). As this CI includes 0, we conclude that we cannot reject the null hypothesis that sex does not modify the association between AAI and age.