Biost 518 / Biost 515
Applied Biostatistics II / Biostatistics II

Midterm Examination Key
February 12, 2014

Name: _____

Instructions: This exam is closed book, closed notes. You have 50 minutes. You may not use any device that is capable of accessing the internet.

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.

NOTE: 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, 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.

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

Signed:_____

All problems deal with a subset of data from an observational study of serum cholesterol and mortality in an elderly population. The appendices contain results from selected analyses:

Appendix A : Description of the variables and descriptive statistics **(all problems)**
Appendix B : Linear regression analyses of cerebral atrophy by age **(problems 1 and 2)**
Appendix C : Linear regression analyses of cerebral atrophy by alcohol intake **(problem 3)**
Appendix D : Linear regression analyses of cerebral atrophy by alcohol intake within sex strata **(problem 4)**
Appendix E : Linear regression analysis of log transformed atrophy score by history of stroke **(problem 5)**
Appendix F : Proportional hazards regression analysis of time to death by atrophy score **(problem 7)**

1. Suppose we are interested in any association between cerebral atrophy score and age. **Appendix B** contains the results of linear regression analyses exploring this question.

    a. (5 points) Based on the regression model, what is the best estimate for the mean atrophy score in subjects who are 70 years old? (**In this and all problems, whenever the precision of the analysis reports will allow, use at least four significant digits in your calculations, and provide at least three significant digits in your answer, or you will receive no credit.)**

Ans: -16.06 + 70 × 0.6980 = 32.80

    b. (5 points) Based on the regression model, what is the best estimate for the mean atrophy score in subjects who are 71 years old?

Ans: -16.06 + 71 × 0.6980 = 33.50

    c. (5 points) Based on the regression model, what is the best estimate for the mean atrophy score in subjects who are 80 years old?

Ans: -16.06 + 80 × 0.6980 = 39.78

    d. (5 points) Based on the regression model, what is the best estimate for the difference in mean atrophy score between 71 year old subjects and 70 year old subjects?

Ans: 0.6980 *(This is just the slope for the atrophy covariate, but you could have subtracted if you wanted to do the extra work.)*

    e. (5 points) Based on the regression model, what is the best estimate for the difference in mean atrophy score between 80 year old subjects and 70 year old subjects?

Ans: 10 × 0.6980 = 6.980 (Just 10 times the slope.)

    f. (5 points) Which regression analysis presented in Appendix B would you have chosen *a priori* to make statistical inference about any associations between mean cerebral atrophy score and age? Why? (A very brief answer should suffice here.)

Ans: *A priori* using the robust standard errors as in Model B2 would protect against heteroscedasticity. *(Note that heteroscedasticity would be expected if there were any trend by age, because the atrophy score is limited in its range to be between 0 and 100. You received 0 points on this problem if you used the results of the analyses to make this decision.)*

    g. (5 points) Using the regression analysis you identified in part (f), provide a 95% confidence interval for the difference in mean atrophy score between two populations who differ in age by 10 years. (Just the numbers, no interpretation necessary here.)

Ans: 10 × 0.5213 = 5.213, 10 × 0.8747 = 8.747 *(Just multiply the 95% CI by 10.)*

h. (5 points) Provide an interpretation for the intercept in the regression model. What scientific use would you make of this estimate?

Ans: -16.06 is the estimated mean atrophy score in newborns. This group is way outside our sampling range, so it is not too surprising that an impossible value results. I would not use this at all except as a mathematical construct to fit the regression model.

i. (5 points) Provide an interpretation for the slope in the regression model. What scientific use would you make of this estimate?

Ans: When comparing two populations that differ in age, we estimate that the mean atrophy score will be 0.6980 points higher for each year difference in age between the groups, with the older group having the higher mean atrophy.

j. (5 points) Is there evidence that there is an association between mean cerebral atrophy and age? State your evidence.

Ans: Yes, the p value for the atrophy regression coefficient has P < 0.0005, so using level 0.05 statistical significance we would reject the null hypothesis of no association. *(I accepted the answer: "Yes, P < 0.0005 for the slope.")*

k. (5 points) Is there evidence that there is a statistically significant correlation between cerebral atrophy and age? State your evidence.

Ans: Yes, tests for statistically significant correlation are exactly equivalent to tests for a statistically significant slope in linear regression.

l. (5 points) Can you provide an estimate of the correlation between cerebral atrophy score and age in this sample? If so, do so. If not, explain why not.

Ans: The square root of R^2 is 0.2944, and the correlation should have the same sign as the slope, so r= 0.2944.

m. (5 points) Based on the regression model, what is the best estimate for the average standard deviation of cerebral atrophy scores within a group that is homogeneous with respect to age?

Ans: The average SD within age groups would be estimated by the RMSE of 12.36.

2. **Appendix A** contains a scatterplot of cerebral atrophy scores versus age.

a. (5 points) From that plot, comment on the reliability of your estimates of age group specific means in parts (a) through (c) of problem 1.

Ans: The lowess curve on the scatterplot seems reasonably close to a straight line, so the estimated means for each group would be a reasonable estimate. *(If you said you thought it was nonlinear, then I accepted you also saying that they would not be good estimates. But personally I think this is pretty close to what I would see had I simulated the data.)*

b. (5 points) From this plot, comment on the reliability of your answers to the statistical inference you provided in parts (g), (i), and (j) of problem 1.

Ans: I used robust SE from the Huber-White sandwich estimator, so the inference about linear trends are reliable whether or not there is heteroscedasticity. *(Had we used classical linear regression, we would want to know that we had homoscedasticity, which I think we did.)*

c. (5 points) From this plot, comment on the reliability of your answers to part (m) of problem 1.

Ans: The scatterplot shows reasonable homoscedasticity and reasonable linearity, so I regard the estimated SD as relatively good. *(Note that if we had heteroscedasticity, the RMSE is estimating some sort of weighted average. If we have nonlinearity, then the RMSE includes the "systematic error" of the distance from the true group means to the estimated line. I graded this problem primarily on comments about heteroscedasticity. I note that I thought it looked homoscedastic and more complicated analyses I performed (looking for linear trends in the squared residuals) did not find any problem. However, you cannot prove equivalence with a finite sample size. If you said you thought it was heteroscedastic, that was fine, so long as you also then said that we would have to interpret the RMSE more carefully.)*

3. (15 points) Now suppose we are interested in investigating any association between cerebral atrophy score and alcohol intake. Provide a formal report of the methods and results of the analysis provided in **Appendix C.**

Ans:

Methods: Associations between cerebral atrophy and alcohol intake were assessed by examining the evidence for a linear trend in average atrophy score across groups when modeling alcohol intake as a linear continuous predictor in linear regression. 95% confidence intervals and two-sided P values were computed using an approximate normal distribution for the regression slope coefficient with robust standard errors as calculated with the Huber-White sandwich estimator. *(On some papers I made notations about the use of the term "robust regression". Unfortunately, "robust regression" is often used for methods that downweight outliers. "Robust SE" would probably always be understood to be the Huber-White estimator, but it is even better to state that explicitly.)*

Results: From the linear regression model we estimate that the difference in mean atrophy score is 0.2423 points for each drink per week difference between two populations, with the group with higher alcohol intake having higher average atrophy score. This difference is beyond that that might be expected due to random variation when there was no true association between mean atrophy score and alcohol intake (two-sided P = 0.007). Based on the 95% CI, the observed difference would not be judged unusual if the true association were such that the mean atrophy score was anywhere between 0.065 to 0.419 points higher for every 1 drink per week difference between two groups.

4. **Appendix D** contains results of regression analyses investigating associations between cerebral atrophy score and alcohol intake within strata defined by sex. Use the results of **Appendices A, C,** and **D** to answer the following questions.

    a. (5 points) Do the results of the available analyses suggest that sex modifies any association between cerebral atrophy score and alcohol intake? Clearly explain your reasoning.

Ans: No. The estimated slopes of 0.151 for females and 0.142 for males seem sufficiently similar that I would not be interested in characterizing it as effect modification. *(If you said that was a marked difference, then you got full points if you also said there was effect modification.)*

    b. (5 points) Do the results of the available analyses suggest that sex confounds any association between cerebral atrophy score and alcohol intake? Clearly explain your reasoning.

Ans: Yes. From Appendix A, males have higher average alcohol intake (2.89 drinks / week for males, 1.34 drinks / week for females). ***A priori*** **we might expect that males average higher cerebral atrophy, and that is borne out in Appendix A as well: the average atrophy score is 39.1 for males and 32.9 for females. As sex is clearly not in the causal pathway of any effect of alcohol on cerebral atrophy, the assessment of an atrophy-alcohol association would seem confounded by sex.** *(Because we are using differences in means as our measure of association, I would also have accepted a comparison of the unadjusted slope (0.2423) and the stratum specific slopes (about 0.14 – 0.15). Such differences would be evidence of confounding when using linear regression. But note that more justification would be required for logistic or PH regression, where we would have to also invoke that the stratum specific estimates were closer to the null than was the unadjusted estimate before we could claim that this was evidence of confounding.)*

(I also note that the analysis of this observational data cannot take into account changes in alcohol intake that might have been related to onset of cerebral atrophy. And even though the stratum specific estimates were not significant, that does not prove that alcohol is harmless. But if I overinterpret this data, it certainly suggests that alcohol intake is not as harmful for cerebral atrophy as is being male.)

5. **Appendix E** presents results of regression analyses exploring an association between cerebral atrophy score and prior history of cerebrovascular disease (transient ischemic attack (TIA) or stroke).

    a. (15 points) Provide interpretations of each of the three regression parameters in the regression models. (You do not need to report or interpret the CI.)

Ans: $e^{3.49} = 32.85$ is estimated geometric mean atrophy score in a population having had neither prior TIA nor stroke.

$e^{0.1046} = 1.111$ is the estimated ratio of geometric means for atrophy score when comparing a population having had prior TIA but not stroke (in the numerator) to a population having had neither prior TIA nor stroke (in the denominator). *(So geometric mean is estimated to be 11.1% higher in population with prior TIA compared to a population who had neither prior TIA nor stroke.)*

$e^{0.1820} = 1.1996$ is the estimated ratio of geometric means for atrophy score when comparing a population having had prior stroke with or without TIA (in the numerator) to a population having had neither prior TIA nor stroke (in the denominator). *(So geometric mean is estimated to be 19.96% higher in population with prior stroke compared to a population who had neither prior TIA nor stroke.)*

    b. (5 points) What conclusions would you reach about an association between cerebral atrophy score and prior history of cerebrovascular disease from this analysis? Quantify (provide a number in support of) your conclusion. (I must be able to tell how you reached your conclusion.)

Ans: Using linear regression with standard errors estimated using the Huber-White sandwich estimator, the two-sided P = 0.0001 suggests that we can with high confidence reject the null hypothesis of no association. *(I took this from the overall F test, because all parameters in the model were related to cerebrovascular disease.)*

6. Use the results presented in **Appendix A** to answer the following questions about associations between prior history of cardiovascular disease and sex. In all regressions, the indicator of prior cardiovascular disease is the response variable, and the indicator of male sex is the predictor.

    a. (5 points) What is the prevalence of prior cardiovascular disease for each sex?

Ans: From Appendix A, the prevalence of prior CVD is 0.1355 (13.55%) in females and 0.1831 (18.31%) in males.

    b. (5 points) Suppose we perform **a linear regression** investigating the risk difference (RD) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **intercept** in the regression analysis' linear predictor?

Ans: The intercept in this saturated model would be the sample prevalence of prior CVD in females: 0.1355.

c. (5 points) Suppose we perform a **linear regression** investigating the risk difference (RD) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **slope** in the regression analysis' linear predictor?

Ans: The slope in this saturated model would be the difference of the sample prevalence of prior CVD in females minus the sample prevalence in males: 0.1831 – 0.1355 = 0.0476.

d. (5 points) Suppose we perform **a Poisson regression** investigating the risk ratio (RR) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **intercept** in the regression analysis' linear predictor?

Ans: The intercept in this saturated model would be the $\log_e$ sample prevalence of prior CVD in females: $\log_e(0.1355) = -1.9987$.

e. (5 points) Suppose we perform a **Poisson regression** investigating the risk ratio (RR) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **slope** in the regression analysis' linear predictor?

Ans: The slope in this saturated model would be the $\log_e$ ratio of the sample prevalence of prior CVD in females divided by the sample prevalence in males: $\log_e(0.1831 / 0.1355) = 0.3011$. *(We could have described this also as the difference in log sample prevalences: log(0.1831) – log(0.1355).)*

f. (5 points) Suppose we perform a **logistic regression** investigating the odds ratio (OR) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **intercept** in the regression analysis' linear predictor?

Ans: The intercept in this saturated model would be the $\log_e$ sample odds of prior CVD in females: $\log_e(0.1355 / (1 – 0.1355)) = -1.853$.

g. (5 points) Suppose we perform a **logistic regression** investigating the odds ratio (OR) for prior cardiovascular disease between males and females, using the indicator of prior cardiovascular disease as the response variable and the indicator of male sex as the predictor. What would be the **slope** in the regression analysis' linear predictor?

Ans: The slope in this saturated model would be the $\log_e$ ratio of the sample odds of prior CVD in females divided by the sample odds in males:

$$\log_e([0.1831 / (1 – 0.1831)] / [0.1355 / (1 – 0.1355)]) = \log_e (1.43) = 0.3577.$$

(We could have described this also as the difference in log sample odds:

$$log(0.1831 / (1 – 0.1831)) – log(0.1355 / (1 – 0.1355)).)$$

7. **Appendix F** contains results of regression analyses investigating associations between all cause mortality and cerebral atrophy.

    a. (5 points) Provide an interpretation of the regression slope parameter from **Model F1**. (You do not need to consider the confidence interval)

Ans: When comparing two groups that differ by 1 point in their atrophy scores, we estimate that the instantaneous risk of death is 100(1.03088 – 1)% = 3.088% higher in the group with the higher atrophy score.

    b. (5 points) Provide an interpretation of the regression slope parameter from **Model F2**. (You do not need to consider the confidence interval)

Ans: When comparing two groups in which one group has double the atrophy score of the other, we estimate that the instantaneous risk of death is $3.202^{\log(2)} = 2.24$ fold higher in the group with the higher atrophy score. *(A much less desirable answer to my mind was to report that the instantaneous risk of death was 3.202 fold higher for a group having an atrophy score 2.718 times higher. An unacceptable answer was to say "1 log difference", because the reader would have to know the base logarithm used. Even though statistical software presumes $\log_e$, the average reader (if they understood it at all) would presume base 10 logarithm.)*

    c. (5 points) Does **Model F3** provide evidence of an association between all cause mortality and cerebral atrophy? Clearly state the basis for your conclusions.

Ans: The overall chi square statistic testing both the linear and quadratic terms shows P < 0.0005, so we can with high confidence reject the null hypothesis of no association between atrophy and all cause mortality.

    d. (5 point) Using the results in Appendix F, do you have evidence that any association between all cause mortality and cerebral atrophy score is nonlinear? Clearly state your reasoning and the evidence you use.

Ans: The P value for the quadratic term of P = 0.734 suggests that we do not have sufficient evidence of a nonlinear effect that is well-described by a parabolic relationship between the log hazard and atrophy score.

GRADING:

Maximum possible: 180

Highest achieved: 178

Mean (SD) : 124 (29.7)

Percent:	**90%**	**80%**	**50%**	**20%**
Percentile:	**160**	**151**	**126**	**93**

APPENDIX A: **Description of variables and descriptive statistics**

These data come from an observational study of brain changes in the elderly as measured by magnetic resonance imaging (MRI) in a population of elderly U.S. residents. All subjects were followed for at least 5 years from the time of study enrolment. This exam considers the following variables (all measured at time of study enrolment) on a subset of 735 subjects from that study.

age:	Age in years of the subject at the time of study enrolment
male:	Indicator that the subject is male (**0**= female, **1**= male)
alcoh*:*	Average alcohol intake for the participant for the two weeks prior to MRI (drinks per week, where 1 drink = 1 oz. whiskey, 4 oz. wine, or 12 oz. beer).
stroke	Indicator of whether the participant had been diagnosed with a cerebrovascular event prior to MRI (0= no, 1= diagnosis of a transient ischemic attack, 2= diagnosis of a stroke).
priorCVD	Indicator that the participant had been previously diagnosed with cardiovascular disease: congestive heart failure, angina, myocardial infarction, transient ischemic attacks (or TIA), or stroke. (0= no, 1= yes)
atrophy	A measure of global brain atrophy detected on MRI. (A number between 0 and 100, with 0 indicating no shrinking and 100 indicating the most severe degree of brain shrinking.)
obstime	The total time (in years) that the participant was observed on study between the date of MRI and death or September 16, 1997, whichever came first.
death	An indicator that the participant was observed to die while on study. If death=1, the number of years recorded in obstime is the number of years between that participant's MRI and his/her death. If death=0, the number of years recorded in obstime is the number of years between that participant's MRI and September 16, 1997.

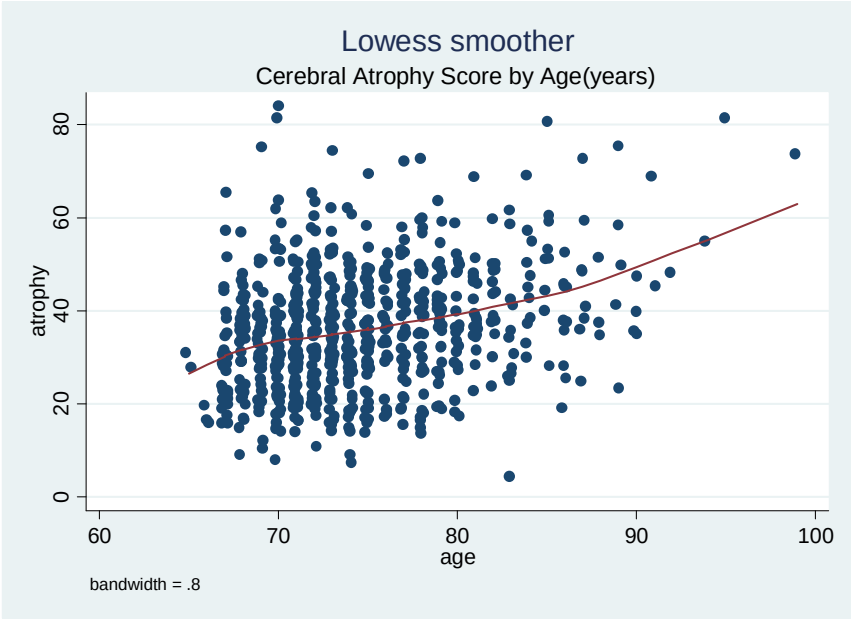
APPENDIX A (cont.): **Description of variables and descriptive statistics**

The following table presents descriptive statistics for the above variables within strata defined by subject sex, as well as for the entire sample. There is no missing data for any variable. Descriptive statistics include the sample size (N), sample mean, standard deviation (sd), minimum (min), 25th percentile (p25), median (p50), 75th percentile (p75) and maximum (max).:

```
. tabstat age alcoh stroke priorCVD atrophy obstime death, ///
>         stat(n mean sd min q max) col(stat) by(male) long
```

male	variable	N	mean	sd	min	p25	p50	p75	max
0	age	369	74.407	5.2578	65	71	73	78	91
	alcoh	369	1.33813	3.4547	0	0	0	.5192	21.5
	stroke	369	.170732	.52192	0	0	0	0	2
	priorCVD	369	.135501	.3427	0	0	0	0	1
	atrophy	369	32.9052	12.225	5	24	32	40	82
	obstime	369	5.07525	.79719	.1889	5.038	5.133	5.4949	5.908
	death	369	.127371	.33384	0	0	0	0	1
1	age	366	74.7268	5.642	66	71	74	78	99
	alcoh	366	2.88693	5.8394	0	0	.0385	2.25	35
	stroke	366	.303279	.70095	0	0	0	0	2
	priorCVD	366	.183060	.38725	0	0	0	0	1
	atrophy	366	39.0874	12.881	10	30	39	48	84
	obstime	366	4.80221	1.2811	.1862	5.016	5.1581	5.6372	5.911
	death	366	.234973	.42456	0	0	0	0	1
Total	age	735	74.5660	5.4514	65	71	74	78	99
	alcoh	735	2.10937	4.852	0	0	.0192	1.25	35
	stroke	735	.236735	.62072	0	0	0	0	2
	priorCVD	735	.159184	.36610	0	0	0	0	1
	atrophy	735	35.9837	12.923	5	27	35	44	84
	obstime	735	4.93929	1.0740	.1862	5.029	5.1444	5.5989	5.911
	death	735	.180952	.38524	0	0	0	0	1

Scatterplot of cerebral atrophy score versus age (with superimposed lowess smooth)



APPENDIX B: Linear regression analyses of cerebral atrophy score by age.

######## MODEL B1

```
. regress atrophy age

      Source |       SS       df       MS              Number of obs =     735
-------------+------------------------------           F(  1,   733) =   69.58
       Model |   10626.648     1   10626.648           Prob > F      =  0.0000
    Residual |  111953.156   733  152.732819           R-squared     =  0.0867
-------------+------------------------------           Adj R-squared =  0.0854
       Total |  122579.804   734  167.002458           Root MSE      =  12.359

------------------------------------------------------------------------------
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .6979831   .0836783     8.34   0.000     .5337054    .8622609
       _cons |  -16.06213   6.256186    -2.57   0.010    -28.34431   -3.779947
------------------------------------------------------------------------------
```

######## MODEL B2

```
. regress atrophy age, robust

Linear regression                               Number of obs =     735
                                                F(  1,   733) =   60.12
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0867
                                                Root MSE      =  12.359

------------------------------------------------------------------------------
             |               Robust
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .6979831   .0900192     7.75   0.000      .521257    .8747093
       _cons |  -16.06213   6.700595    -2.40   0.017    -29.21677   -2.907482
------------------------------------------------------------------------------
```

APPENDIX C: Linear regression analyses of cerebral atrophy score by alcohol intake.

######## MODEL C1

```
. regress atrophy alcoh

      Source |       SS       df       MS              Number of obs =     735
-------------+------------------------------           F(  1,   733) =    6.12
       Model |  1014.17023     1  1014.17023           Prob > F      =  0.0136
    Residual |  121565.634   733  165.846704           R-squared     =  0.0083
-------------+------------------------------           Adj R-squared =  0.0069
       Total |  122579.804   734  167.002458           Root MSE      =  12.878

------------------------------------------------------------------------------
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alcoh |   .2422712   .0979715     2.47   0.014      .049933    .4346095
       _cons |   35.47264   .5180243    68.48   0.000     34.45565    36.48962
------------------------------------------------------------------------------
```

######## MODEL C2

```
. regress atrophy alcoh, robust

Linear regression                               Number of obs =     735
                                                F(  1,   733) =    7.22
                                                Prob > F      =  0.0074
                                                R-squared     =  0.0083
                                                Root MSE      =  12.878

------------------------------------------------------------------------------
             |               Robust
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alcoh |   .2422712   .0901418     2.69   0.007     .0653043    .4192382
       _cons |   35.47264   .5204059    68.16   0.000     34.45097     36.4943
------------------------------------------------------------------------------
```

<u>APPENDIX D</u>: Linear regression analyses of cerebral atrophy score by alcohol intake within strata defined by sex.

######## MODEL D1 : FEMALES

. regress atrophy alcoh if male==0, robust

```
Linear regression                                       Number of obs =     369
                                                        F(  1,   367) =    1.06
                                                        Prob > F      =  0.3034
                                                        R-squared     =  0.0018
                                                        Root MSE      =   12.23

------------------------------------------------------------------------------
             |               Robust
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alcoh |   .1510547     .14658     1.03   0.303    -.1371873    .4392967
       _cons |   32.70302   .6814908    47.99   0.000      31.3629    34.04314
------------------------------------------------------------------------------
```

######## MODEL D2 : MALES

. regress atrophy alcoh if male==1, robust

```
Linear regression                                       Number of obs =     366
                                                        F(  1,   364) =    1.70
                                                        Prob > F      =  0.1927
                                                        R-squared     =  0.0041
                                                        Root MSE      =  12.872

------------------------------------------------------------------------------
             |               Robust
     atrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alcoh |   .1416756   .1085577     1.31   0.193    -.0718034    .3551547
       _cons |   38.67842   .7573296    51.07   0.000     37.18913    40.16772
------------------------------------------------------------------------------
```

<u>APPENDIX E</u>: Linear regression analyses of logarithmically transformed cerebral atrophy score by prior history of TIA and stroke.

```
######## MODEL E1

. g logatrophy= log(atrophy)
. regress logatrophy i.stroke

      Source |       SS       df       MS              Number of obs =     735
-------------+------------------------------           F(  2,   732) =    8.11
       Model |  2.39455073     2  1.19727537           Prob > F      =  0.0003
    Residual |  108.022481   732  .147571695           R-squared     =  0.0217
-------------+------------------------------           Adj R-squared =  0.0190
       Total |  110.417031   734  .150431923           Root MSE      =  .38415

------------------------------------------------------------------------------
  logatrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      stroke |
          1  |   .1055772   .0798802     1.32   0.187    -.0512445    .2623989
          2  |   .1819618   .0469005     3.88   0.000     .0898863    .2740373
             |
       _cons |   3.491568   .0152326   229.22   0.000     3.461664    3.521473
------------------------------------------------------------------------------


######## MODEL E2

. regress logatrophy i.stroke, robust

Linear regression                                       Number of obs =     735
                                                        F(  2,   732) =    9.34
                                                        Prob > F      =  0.0001
                                                        R-squared     =  0.0217
                                                        Root MSE      =  .38415

------------------------------------------------------------------------------
             |               Robust
  logatrophy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      stroke |
          1  |   .1055772    .071626     1.47   0.141    -.0350398    .2461942
          2  |   .1819618   .0436964     4.16   0.000     .0961767    .2677469
             |
       _cons |   3.491568   .0154253   226.35   0.000     3.461285    3.521851
------------------------------------------------------------------------------
```

APPENDIX F: Proportional hazards regression analyses of time to death by cerebral atrophy.

######## **MODEL F1**

```
. stcox atrophy, robust
        failure _d:  death
  analysis time _t:  obstime
Cox regression -- Breslow method for ties
No. of subjects      =          735                 Number of obs   =       735
No. of failures      =          133
Time at risk         =  3630.376453
                                                    Wald chi2(1)    =     23.49
Log pseudolikelihood =   -844.04581                 Prob > chi2     =    0.0000
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     atrophy |   1.030879   .0064684     4.85   0.000     1.018278    1.043635
------------------------------------------------------------------------------
```

######## **MODEL F2**

```
. stcox logatrophy
        failure _d:  death
  analysis time _t:  obstime
Cox regression -- Breslow method for ties
No. of subjects =          735                      Number of obs   =       735
No. of failures =          133
Time at risk    =  3630.376453
                                                    LR chi2(1)      =     22.25
Log likelihood  =   -844.09131                      Prob > chi2     =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  logatrophy |   3.202313    .822938     4.53   0.000     1.935174    5.299166
------------------------------------------------------------------------------
```

######## **MODEL F3**

```
. g atrophysqr= atrophy^2
. stcox atrophy atrophysqr, robust
        failure _d:  death
  analysis time _t:  obstime
Cox regression -- Breslow method for ties
No. of subjects      =          735                 Number of obs   =       735
No. of failures      =          133
Time at risk         =  3630.376453
                                                    Wald chi2(2)    =     22.72
Log pseudolikelihood =   -843.98762                 Prob > chi2     =    0.0000
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     atrophy |   1.040706   .0291416     1.42   0.154      .985129    1.099419
  atrophysqr |   .9998931    .000315    -0.34   0.734     .9992759    1.000511
------------------------------------------------------------------------------
```