Biostatistics 140.623
Third Term, 2015-2016
Final Examination
Answer Key
March 10, 2016

Instructions: You will have two hours for this examination. There are 20 problems. The formula page and Stata output are at the **back** of the exam for your use. Choose one best response for each question.

Questions 1-4 address general knowledge.

1. What is the difference between **Kaplan-Meier** estimates versus **life-table** estimates of a **survivor function**? (*Circle only one response*)
    a) Life-table estimates are the cumulative probability of survival beyond the beginning of a time bin whereas Kaplan-Meier estimates are the cumulative probability of survival before an event time.
    b) Life-table estimates are the cumulative probability of survival beyond the midpoint of a time bin whereas Kaplan-Meier estimates are (1-hazard) at a particular event time.
    c) Life-table estimates are the cumulative probability of survival beyond the end of a time bin(s) whereas Kaplan-Meier estimates are the cumulative probability of survival beyond a specific event time(s).
    For grouped data: $S_j = \prod (1-P_j)$ ***where*** P_j ***is the probability of having the event in bin j of those at risk at the beginning of the time bin.***
    For ungrouped data: $S_t = \prod (1-h_j)$ ***where*** h_j ***is the hazard of the event at time t.***

    d) Life-table estimates are (1-incidence) during a particular time bin whereas Kaplan-Meier estimates are (1-hazard) at a particular event time.
    e) There is no difference; both are calculated in the same way.

2. In order to assess whether the **proportional hazard assumption** holds in an analysis investigating treatment effect on time to an event using a Poisson regression model, the analysts would: *(Circle only one response).*
    a) Include a spline term for time in the Poisson regression model and test whether this term is needed in the model.
    b) Include an interaction term between a function of time and treatment in the Poisson regression model and test whether this term is statistically significant.
    If the proportional hazards assumption holds, then this term will not be statistically significant, suggesting that the hazard ratio for treatment remains constant over time.
    c) Include indicator variables for time bins in the Poisson regression model and test whether these terms are statistically significant.
    d) Determine which model provides the lowest AIC.
    e) Compare the extended versus null Poisson regression models to assess treatment effect.

3. The following is a **Poisson regression model** of lung cancer death incidence, for individuals followed for up to 10 years after diagnosis, as a function of baseline smoking and age defined as:
   smokecat = 1 none; 2 cigars/pipes only; 3 cigarettes only; 4 both cigars/pipes and cigarettes
   age in years

$$\log(\text{expected lung cancer deaths}) = \log(\text{person-years}) + \beta_0 + \beta_1 \text{smokcat2} + \beta_2 \text{smokecat3} + \beta_3 \text{smokecat4} + \beta_4 \text{age}$$

   This model assumes that: (*Circle only one response*)
   - **a) The log hazard of a lung cancer death is constant over time.**
   - **b) The log hazard of a lung cancer death changes linearly over time.**
     We accepted either of the above responses. Response a) assumes that age is baseline age and that there is no time variable in the model. Response b) assumes that age varies over follow-up time (and, thus, represents time).
   - c) The log hazard of a lung cancer death varies from year to year.
   - d) The relative hazard of lung cancer death for cigarette smokers versus non-smokers changes over time.
   - e) The relative hazard of lung cancer death for cigarette smokers versus non-smokers varies by age.

4. The output below provides the estimated coefficients for the Poisson regression model defined in question 3. What would one conclude about the lung cancer death and smoking after adjusting for age? (*Circle only one response*)
   - a) The lung cancer death incidence rate is significantly decreased in cigarette smokers (with or without cigar/pipes) as compared to non-smokers.
   - b) The lung cancer death incidence rate ratio for age is not significantly different from one.
   - c) The risk of lung cancer death for individuals who smoke both cigarettes and cigars/pipes is 23.6 deaths per 100 population.
   - d) The lung cancer death incidence rate is significantly increased in individuals who smoke cigars/pipes as compared to non-smokers.
   - **e) The lung cancer death incidence rate is significantly increased in cigarette smokers (with or without cigar/pipes) as compared to non-smokers.**
     The lowest category (non-smokers) serves as the reference group.

```
Poisson regression
------------------------------------------------------------------------------
      deaths |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    smokecat |
       cigars |   .0329268   .0468939     0.70   0.483    -.0589835    .1248371
  cigarettes |   .4379459   .0398025    11.00   0.000     .3599344    .5159574
        both |   .2363532   .0385969     6.12   0.000     .1607047    .3120018
             |
         age |   .3330061   .0055912    59.56   0.000     .3220475    .3439647
       _cons |  -3.738877   .0500089   -74.76   0.000    -3.836892   -3.640861
      ln(pop) |          1  (exposure)
------------------------------------------------------------------------------
```


Questions 5 – 9 refer to data from the most recent National Hospital Ambulatory Medical Care Survey regarding factors associated with waiting time (in minutes) for nearly 25,000 emergency department visits at U.S. Hospitals in 2011.

The primary outcome of interest is:
waittime = waiting time in minutes

Predictors of interest include:
age = age in years
agecat = four age categories (age quartiles):
1: 0- 20 years
2: > 20- 40 years
3: > 40 – 60 years
4: >= 60 years
white: 1 if white, 0 if non-white
private: 1 if private insurance, 0 if public

A series of linear regression models, **Models A-E** are given at the back of the exam.

5. Based on a comparison of **Models A and B,** which of the following can be concluded? (*Circle only one response*)
   a) Age categories contribute to the prediction of waiting times beyond that contributed by age in years.
   b) The relationship between average waiting times and age is linear because the p-value for the slope of age in Model A is less than 0.05.
   c) Model A is statistically significantly better than Model B because F(1, 24676) is less than F(3, 24674) .
   d) The relationship between average waiting times and age is not linear because the mean waiting time differences for two consecutive age categories is not consistent across age categories.
   Age is modeled as a categorical variable with 20-year age categories. The coefficients represent the difference in average waiting time between the category and the baseline category (0-20 years). The differences in the estimated coefficients are not constant between age categories.
   e) Age is not a statistically significant predictor of waiting times.

6. Which is the following conclusion can be made from the results for **Model C**? (*Circle only one response*)

   a) The relationship between waiting times and race (white vs. non-white) is confounded by age differences between the race groups.
   b) The relationship between waiting times and race (white vs. non-white) is not confounded by age differences between the race groups.
   c) The relationship between waiting times and race (white vs. non-white) is modified by age.

d) The relationship between waiting times and race (white vs. non-white) is not modified by age.

e) White subjects had wait times of 16.4 minutes less that of black subjects of comparable age.
Model A includes continuous race; Model B includes categorical race; Model C includes categorical age and dichotomous race. These models do not allow an assessment of whether age is a confounded of the relationship between waiting times and race. Since Model C does not include an interaction term, it cannot be used to assess effect modification. However, the estimated coefficient for race ("white") suggests that the average wait time for whites is 16.4 minutes less than that blacks after adjusting for age.

7. After fitting **Model C** in Stata, which of the following commands could be used to get an estimated mean waiting time, and 95% confidence interval, for 65 year old white subjects? (*Circle only one response*)

a) lincom agecat4 + white
b) test agecat4 white
c) lincom _cons+ agecat4 + white
From Model C, we can write:
$E[waittime] = \beta_0+\beta_1 agecat2+\beta_2 agecat3+\beta_3 agecat4 + \beta_4 white$
For a 65-year old white, this reduces to $\beta_0+\beta_3(1) + \beta_4(1)$

d) test _cons agecat4 white
e) No command is necessary. The estimated mean is 64.9+-16.4+-0.07 with 95% confidence interval (62.4+-18.6+-2.9, 67.3+-14.2+2.8)

8. Based on the results of **Model E**, what it the estimated mean difference in waiting time for **white subjects with private insurance** compared to **white subjects with public insurance** of the same age? (*Circle only one response*)

a) -8.8 (-13.0, -4,5)
b) -16.6 (-19.1, -14.0)
c) -13.0 (-17.3, -8.8)
d) -5.3 (-7.7, -2.8)
From Model E, we can write:
$E[waittime] = \beta_0+\beta_1 agecat2+\beta_2 agecat3+\beta_3 agecat4 + \beta_4 white+ \beta_5 private+ \beta_4 white{\cdot}private.$
For white subjects with private insurance, we can write $\beta_0+ \beta_4+ \beta_5+ \beta_6.$
For white subjects with public insurance, we can write $\beta_0+ \beta_4+ 0.$
By subtraction, we obtain $\beta_5+ \beta_6$ ***which we can estimate by using***
`.lincom private+ white_private`

e) Mean differences cannot be estimated with linear regression.

9. The R^2 value for **Model B** is 0.002. How can this be interpreted in conjunction with the rest of the information given in the output for Models A-E? (*Circle only one response*)

    a) Model B predicts waiting times well for observations not used in fitting Model B.
    b) Taken together race, and insurance type (private or public) are statistically significant predictors of waiting time after accounting for age.
    c) While the mean waiting times are statistically different across (at least some of) the age categories, there is still substantial variation in the individual waiting times within each age group.
    d) Model B does not fit the observed data well.
    e) The association between waiting time and age is not modified by other factors.

Questions 10 through 14 refer to a randomized controlled study[1] that was performed to assess the efficacy of financial incentives for quitting smoking. A total of 878 employees (who smoked) of a multinational company based in the United States were randomized to either a:

control group: received information about smoking-cessation programs (442 employees) **or intervention group:** received information about smoking-cessation programs plus financial incentives (436 employees)

The primary outcome for this study was quitting smoking within 12 months after randomization. (***referred to as "quitting smoking")***

The results from a simple logistic regression are as follows:

$$\ln(\text{odds of quitting smoking}) = -2.95 + 1.20x_1,$$

where $x_1 = 1$ for the intervention group, and 0 for the control group. The standard error of the intercept is 0.20, and the standard error of the slope for x_1 is 0.25

10. What is the odds ratio (and 95% CI) of quitting smoking for the intervention group compared to the control group? (*Circle only one response*)

    a) -2.95 (-3.34, 2.55)
    b) 0.50 (0.20, 0.80)
    c) 1.20 (0.70, 1.70)
    d) 3.3 (2.0, 5.5)
    From this model, the logOR of quitting smoking by treatment is estimate by b_1=1.20. Thus, the OR = exp(1.2) = 3.3
    The 95% CI for the true logOR = b1± 1.96 SE(b_1) = 1.2± 1.96 SE(b_1)
    = 1.2 ± 1.96 (0.25) = (0.71, 1.69). Exponentiating, we obtain (2.0, 5.4) which is the 95% CI for the true OR.
    e) Odds ratios cannot be estimated from logistic regression.

[1] Volpp K, et al. A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation (2009) New England Journal of Medicine; Vol 360 (7); pps 699-709.

11. What proportion of the persons randomized to the intervention group quit smoking? (i.e: what is the probability that a person randomized to the intervention group quit smoking?) (*Circle only one response*)

a) 5%
b) 15%
Since p = odds/(1+odds), we can solve for the estimated:
log(odds|intervention) = $-2.95+1.20(1)$ ***=-1.75***
Then, odds = exp(-1.75) 0.1738
Thus, p = 0.1738/(1+0.1738) = 0.148 or approximately 15%
c) 77%
d) 20%
e) Logistic regression can only be used to estimate odds and odds ratios.

12. How does randomization minimize the potential of the relationship between quitting smoking and the intervention group versus control group being confounded by person's age? (*Circle only one response*)

a) Randomization minimizes the potential for an association between quitting smoking and age.
b) Randomization minimizes the potential for an association between the intervention (versus control) and age.
Randomization aims to eliminate bias in treatment assignment. As a result, the intervention and control groups should be well-balanced with respect to covariates and possible confounders. Thus, we would expect no association between age and treatment group.
c) Randomization minimizes the potential for an association between quitting smoking and the intervention (versus control).
d) Randomization minimizes the potential that the association between quitting smoking and the intervention (versus control) differs by age.
e) Randomization minimizes the potential that the relationship between quitting smoking and person's age is statistically significant.

In order to investigate whether the relationship between quitting smoking and the financial incentives intervention is modified by sex, the researchers examined the results from the following logistic regression model:

$$\ln(\text{odds of quitting smoking}) = -3.1 + 1.40x_1 + 0.15x_2 + -0.30x_3$$

where $x_1 = 1$ for intervention group, and 0 for persons randomized to the control group
$x_2 = 1$ for males and 0 for females
$x_3 = x_1 * x_2$ (the interaction term)

13. Based on the estimated regression coefficients given above, what is the odds ratio of quitting smoking for males randomized to the intervention group compared to males randomized to the control group? (*Circle only one response*)

a) 3.0
log(odds|males, intervention) $= -3.1 + 1.40(1) + 0.15(1) + -0.30(1)$ ***and***
log(odds|males, control) $= -3.1 + 1.40(0) + 0.15(1) + -0.30(0)$
By subtraction, log OR = 1.4 – 0.30=1.1
Thus, OR = exp(1.1)= 3.0

b) 4.1
c) 0.86
d) 1.10
e) -0.15

14. The **standard error** for the estimated coefficient of the interaction term (x_3) is 0.55. Given this information, what conclusion can be made? (*Circle only one response*)

a) Because the 95% confidence interval for the coefficient (β_3) of the interaction term **does not include 1,** the relationship between quitting smoking and the intervention is (statistically significantly) confounded by sex.
b) Because the 95% confidence interval for the coefficient (β_3) of the interaction term **does not include 1**, the relationship between quitting smoking and the intervention is (statistically significantly) modified by sex.
c) Because the 95% confidence interval for the coefficient (β_3) of the interaction term **includes 0**, the relationship between quitting smoking and the intervention is not (statistically significantly) confounded by sex.
d) Because the 95% confidence interval for the coefficient (β_3) of the interaction term includes 0, the relationship between quitting smoking and the intervention is not (statistically significantly) modified by sex.
The 95% CI for β_3 is -0.30 ± 1.96 (0.55) = (-1.38, 078) which overlaps zero.
e) This cannot be answered without the results from a likelihood ratio test.

Questions 15 - 20 refer to an investigation of all-cause mortality (time to death) for a subset of participants enrolled in the longitudinal Framingham Heart Study.

15. **Figure 1** displays the estimated **cumulative survival function (Kaplan-Meier curve)** by baseline diabetes status (diabetes versus no diabetes) by time **in years.** The difference in the estimated survivor function at **10 years** between participants without diabetes versus with diabetes, S(10|without diabetes) – S(10|with diabetes) , is approximately: *(Circle only one response)*

    a) .90 – 0.70 ***by looking at 10 years and extrapolating to the Y-axis which shows the cumulative probability of survival.***
    b) .10 – 0.30
    c) 0.90/0.70
    d) 0.10/0.30
    e) 0.70 – 0.90

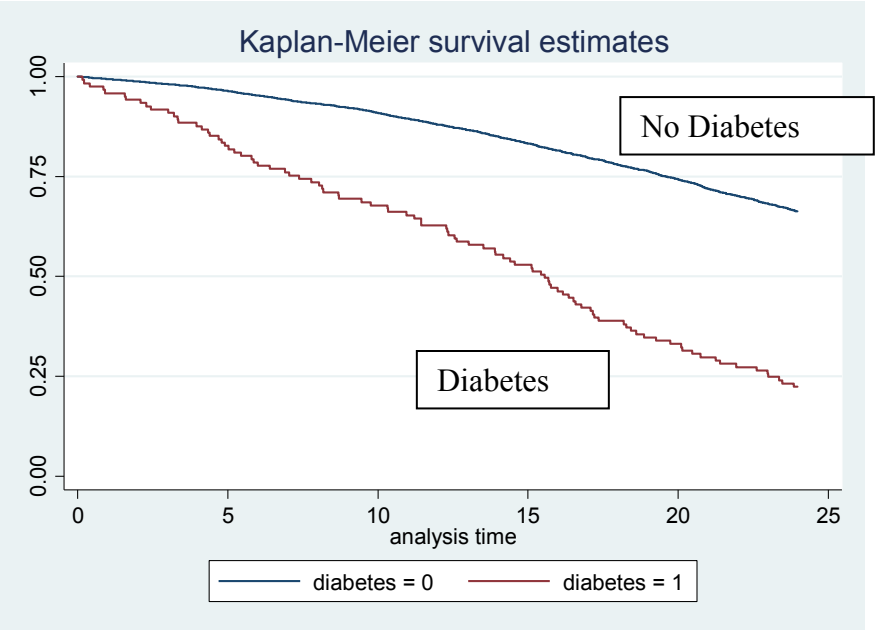

Figure 1

16. The **logrank test statistic** to accompany Figure 1 is calculated as 176.32 with an associated p-value of approximately zero. One could conclude that: *(Circle only one response)*

    a) There is a statistically significant difference in overall survival by diabetes status. ***The null hypothesis for a logrank test is no difference in overall survival by covariate level. The logrank statistic is distributed as a chi-squared statistic with one degree of freedom. Based on 176.32, p~0, we would reject the null hypothesis and conclude that there is a statistically significant difference by diabetes status.***
    b) There is no significant difference in overall survival by diabetes status.
    c) There is a statistically significant difference in survival at 10 years by diabetes status.
    d) There is no difference in survival at time zero by diabetes status.
    e) The hazard of death is constant in the two groups.

Table 1 shows the unadjusted and adjusted hazard ratios of all cause-mortality from Cox regression models.

	Unadjusted HR	**95% CI**	**Adjusted* HR**	**95% CI**
Covariate				
Diabetes	3.74	3.03, 4.61	2.50	2.01, 3.09
Sex	0.59	0.54, 0.65	-	-
BMI category				
< 25 kg/m^2	1.00		1.00	-
25- 29	1.32	1.18, 1.47	1.05	0.93, 1.17
> 29	1.73	1.50, 1.99	1.44	1.24, 1.67
Age (years)				
≤ 60	1.08	1.08, 1.09	1.09	1.08,1.10
> 60	1.14	1.11, 1.17	1.14	1.11, 1.17
Current Smoker	1.09	0.99, 1.20	1.50	1.34, 1.67
BP Medication	2.69	2.18, 3.31	-	-
Females				
No BP Meds	-	-	1.0	-
BP Meds	-	-	1.55	1.18, 2.02
Males				
No BP Meds	-	-	1.73	1.55, 1.92
BP Meds	-	-	?	

*Adjusted for all covariates listed plus the **interaction** of sex and BP medication.

17. Using the output from **Model Z** at the back of the exam, what is the **Adjusted HR** for the Males using BP medications compared to Females not using BP medications (fill in the cell with the ? in the last line of Table 1): (*Circle only one response*)

a) 1.64
b) 1.73
c) 3.37
d) 4.38

From Model Z

$$log[\lambda(t;X)] = log[\lambda_0(t;X)] + \beta_1 sex + \beta_2 bmicat2 + \beta_3 bmicat3 + \beta_4 age1 + \beta_5 age2 + \beta_6 diabetes + \beta_7 cursmoke + \beta_8 bpmeds + \beta_9 bpmeds*sex$$

For Males using BP medications :
$$\log[\lambda(t;X)] = \log[\lambda_0(t;X)] + \beta_1(1) + \beta_2 bmicat2 + \beta_3 bmicat3 + \beta_4 age1 + \beta_5 age2 + \beta_6 diabetes + \beta_7 cursmoke + \beta_8(1) + \beta_9(1)*(1)$$

For Females notusing BP medications :
$$\log[\lambda(t;X)] = \log[\lambda_0(t;X)] + \beta_1(0) + \beta_2 bmicat2 + \beta_3 bmicat3 + \beta_4 age1 + \beta_5 age2 + \beta_6 diabetes + \beta_7 cursmoke + \beta_8(1) + \beta_9(0)*(0)$$

Holding the other factors constant, we subtract and obtain : log (HR) = $\beta_1 + \beta_8 + \beta_9$
To obtain the HR, we use the lincom command to obtain the estimated HR of 4.38 :

```
. lincom sex +bpmeds +bpmeds_sex, hr
```

e) 6.02

18. An assessment of the multivariable-adjusted hazard ratios versus the unadjusted hazard ratios for all-cause mortality provided in **Table 1** and **Model Z** suggests: (*Circle only one response*)

a) No evidence of confounding of the relationship between age and the log hazard of all-cause-mortality by any of the other multiple adjustment variables.
We can see that there is no appreciable change in log HR for the age categories between the unadjusted and adjusted models.

b) No evidence of confounding of the relationship between diabetes and the log hazard of all-cause mortality by any of the other multiple adjustment variables.

c) Possible interaction between diabetes and sex on the log hazard of all-cause mortality, controlling for the other adjustment variables.

d) No evidence of interaction between sex and the use of blood pressure medications on the log hazard of all-cause mortality.

e) Evidence of a possible effect modification of the relationship between diabetes and the log hazard of all-cause mortality by time.

19. A test of the null hypothesis of no interaction between sex and use of blood pressure medications in **Model Z** can be performed using: (*Circle only one response*)

    a) An F-test of the null hypothesis that the regression coefficient for the interaction term equals zero.
    b) A Likelihood- Ratio Test of the null hypothesis that the regression coefficient for the interaction term equals zero.
    c) A log-rank test of the null hypothesis that the survival distribution is the same by sex.
    d) A comparison of the AIC values for the multivariable-adjusted versus unadjusted models.
    e) A Z-test resulting from the sum (linear combination) of the regression coefficients for sex, use of blood pressure medication, and the interaction term.

20. A correct interpretation of the **exponentiated interaction coefficient**, $\exp(\beta_9)$ in **Model Z** is: (*Circle only one response*)

    a) The difference, between males and females, in the log hazard ratio for death for those taking BP medication versus not taking BP medication.
    b) The difference, between those taking BP medication and those not taking BP medication, in the log hazard ratio for death for males versus females.
    c) The factor by which the hazard ratio for death for females versus males differs between those taking BP medication and those not taking BP medication.
    d) The factor by which the hazard ratio for death for those taking BP medication versus not taking BP medication differs for males versus females.

    The coefficient β_9 is the difference between log (HR for death for those taking BP medication in males) and log (HR for death for those taking BP medication in females). This equals the log (ratio of these two HRs). The exponentiated coefficient can be interpreted as a ratio of HRs.

    e) The hazard ratio of death for those taking BP medication versus those not taking BP medication, adjusting for sex.

Biostatistics 140.623
Final Exam Formula Sheet

Tabled chi-squared values: (α=0.05)

df=1, χ2= 3.84
df=2, χ2= 5.99
df=3, χ2= 7.81
df=200, χ2= 233.99

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\varepsilon$$

$$F_{s,\ n-p-s-1} = \frac{(RSS_{Null} - RSS_{Extended}) / s}{RSS_{Extended} / (n-p-s-1)}$$

$$AIC = RSS + 2(\text{model df})$$

$$\ln = \log_e$$

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$$

$$\frac{e^{a+b}}{e^{a}} = e^{b}$$

$$\log \text{odds} = \text{logit}[\Pr(Y=1)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_s X_s$$

$$\Pr(Y=1) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_s X_s}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_s X_s}} = \frac{\text{odds}}{1 + \text{odds}}$$

LRT (Likelihood Ratio Test) = $-2\ (LL_{Null} - LL_{Extended})$
where LL = log likelihood
AIC = - 2 LL + 2(model df)

Poisson Regression (LLR) Model:

$$\log(\mu_i) = \log N_i + \beta_1 X_1 + ... + \beta_p X_p$$

$$\log(\lambda_i) = \beta_1 X_1 + ... + \beta_p X_p$$

Proportional Hazards Model:

$$\log \lambda(t;X) = \log \lambda_0(t:X) + \beta_1 X_1 + ... + \beta_p X_p$$

$$\lambda(t;X) = \lambda_0(t;X) e^{\beta_1 X_1 + ... + \beta_p Xp}$$

$$S(t;X) = [S_0(t)]^{e^{X\beta}}$$

Questions 5- 9 pertain to **Models A – E** where:

The primary outcome of interest is: **waittime** = waiting time in minutes

Predictors of interest include:
age = age in years
agecat = four age categories (age quartiles):
1: 0- 20 years
2: > 20- 40 years
3: > 40 – 60 years
4: >= 60 years
white: 1 if white, 0 if non-white
private: 1 if private insurance, 0 if public

Model A

```
. regress waittime age

      Source |       SS           df       MS      Number of obs   =    24,678
-------------+----------------------------------   F(1, 24676)     =      9.93
       Model |  63197.0214         1  63197.0214   Prob > F        =    0.0016
    Residual |   157026495    24,676  6363.53118   R-squared       =    0.0004
-------------+----------------------------------   Adj R-squared   =    0.0004
       Total |   157089693    24,677  6365.83428   Root MSE        =    79.772

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0669103   .0212321    -3.15   0.002    -.1085265    -.025294
       _cons |   59.00096   .9478167    62.25   0.000     57.14318    60.85873
------------------------------------------------------------------------------
```

Model B

```
. regress waittime i.agecat

      Source |       SS           df       MS      Number of obs   =    24,678
-------------+----------------------------------   F(3, 24674)     =     18.10
       Model |  344932.847         3  114977.616   Prob > F        =    0.0000
    Residual |   156744760    24,674  6352.62866   R-squared       =    0.0022
-------------+----------------------------------   Adj R-squared   =    0.0021
       Total |   157089693    24,677  6365.83428   Root MSE        =    79.703

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
           2 |   6.846341   1.450474     4.72   0.000     4.003324    9.689358
           3 |   5.766656   1.432259     4.03   0.000     2.959342    8.573969
           4 |  -1.989739   1.437056    -1.38   0.166    -4.806456    .8269778
             |
       _cons |   53.82912   1.024621    52.54   0.000      51.8208    55.83744
------------------------------------------------------------------------------
```


Model C

```
. regress waittime i.agecat white

      Source |       SS           df       MS      Number of obs   =    24,678
-------------+----------------------------------   F(4, 24673)     =     68.01
       Model |  1713048.79         4  428262.198   Prob > F        =    0.0000
    Residual |   155376644    24,673  6297.43621   R-squared       =    0.0109
-------------+----------------------------------   Adj R-squared   =    0.0107
       Total |   157089693    24,677  6365.83428   Root MSE        =    79.356

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
           2 |   6.876619   1.444161     4.76   0.000     4.045976    9.707261
           3 |   5.965783   1.426088     4.18   0.000     3.170566    8.761001
           4 |  -.0672839   1.436733    -0.05   0.963    -2.883367    2.748799
             |
       white |  -16.41755   1.113856   -14.74   0.000    -18.60078   -14.23433
       _cons |   64.85284   1.264948    51.27   0.000     62.37346    67.33221
------------------------------------------------------------------------------
```

Model D

```
. regress waittime i.agecat white private

      Source |       SS           df       MS      Number of obs   =    24,678
-------------+----------------------------------   F(5, 24672)     =     60.72
       Model |  1909591.9         5   381918.38   Prob > F        =    0.0000
    Residual |   155180101    24,672  6289.72522   R-squared       =    0.0122
-------------+----------------------------------   Adj R-squared   =    0.0120
       Total |   157089693    24,677  6365.83428   Root MSE        =    79.308

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
           2 |   6.658733   1.443803     4.61   0.000     3.828792    9.488674
           3 |   6.125212     1.4255     4.30   0.000     3.331147    8.919277
           4 |  -.9506666   1.444523    -0.66   0.510    -3.782018    1.880685
             |
       white |  -15.60523   1.122618   -13.90   0.000    -17.80563   -13.40483
     private |  -6.138852   1.098181    -5.59   0.000    -8.291353    -3.98635
       _cons |   66.49571   1.297886    51.23   0.000     63.95178    69.03965
------------------------------------------------------------------------------
```


Model E

```
. gen white_private = white*private

. regress waittime i.agecat white private white_private

      Source |       SS           df       MS      Number of obs   =    24,678
-------------+----------------------------------   F(6, 24671)     =     50.93
       Model |  1921835.77         6  320305.962   Prob > F        =    0.0000
    Residual |   155167857    24,671  6289.48388   R-squared       =    0.0122
-------------+----------------------------------   Adj R-squared   =    0.0120
       Total |   157089693    24,677  6365.83428   Root MSE        =    79.306

------------------------------------------------------------------------------
     waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
       agecat |
            2  |   6.682855   1.443879     4.63   0.000     3.852766    9.512945
            3  |   6.148449   1.425569     4.31   0.000     3.354247    8.942651
            4  |  -.8586839   1.445999    -0.59   0.553    -3.692929    1.975561
               |
         white |   -16.5572   1.313675   -12.60   0.000    -19.13208   -13.98231
       private |  -8.765908   2.179704    -4.02   0.000    -13.03826   -4.493558
white_private |   3.511582   2.516813     1.40   0.163    -1.421522    8.444686
         _cons |   67.09411   1.366888    49.09   0.000     64.41493    69.77329
------------------------------------------------------------------------------


. lincom white + white_private
 ( 1)  white + white_private = 0

------------------------------------------------------------------------------
     waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
         (1) |  -13.04562   2.150743    -6.07   0.000     -17.2612   -8.830029
------------------------------------------------------------------------------

. test white white_private
 ( 1)  white = 0
 ( 2)  white_private = 0

       F(  2, 24671) =   97.59
            Prob > F =    0.0000

. lincom private+ white_private
 ( 1)  private + white_private = 0

------------------------------------------------------------------------------
     waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
         (1) |  -5.254326   1.268012    -4.14   0.000    -7.739706   -2.768945
------------------------------------------------------------------------------

. test private white_private
 ( 1)  private = 0
 ( 2)  white_private = 0

       F(  2, 24671) =   16.60
            Prob > F =    0.0000
```


Questions 17- 20 pertain to investigating the log hazard of death and the following covariates :

Model Z $\log[\lambda(t;X)] = \log[\lambda_0(t;X)] + \beta_1 sex + \beta_2 bmicat2 + \beta_3 bmicat3 + \beta_4 age1 + \beta_5 age2 + \beta_6 diabetes + \beta_7 cursmoke + \beta_8 bpmeds + \beta_9 bpmeds*sex$

sex= 0 if female ; 1 if male
bmicat= 1if < 25 ; 2 if 25-29 ; 3 if > 29 kg/m2
age1 = age-60**,**
and **age2**=0 if age ≤ 60 years or **age2** = (age-60) if age > 60 years.
diabetes=0 if no ; 1 if yes
cursmoke=0 if no ; 1 if yes
bpmeds= 0 if no ; 1 if yes

```
.stcox sex i.bmicat age1 age2 diabetes cursmoke bpmeds bpmeds_sex

     Cox regression -- Breslow method for ties

No. of subjects =        4,373                 Number of obs    =       4,373
No. of failures =        1,518
Time at risk    =     32858668
                                               LR chi2(9)       =     1156.18
Log likelihood  =   -11847.173                 Prob > chi2      =      0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   1.730151    .096061     9.87   0.000     1.551758    1.929053
             |
      bmicat |
           2 |   1.045331   .0607683     0.76   0.446     .9327622    1.171485
           3 |    1.43955   .1091144     4.81   0.000     1.240817    1.670113
             |
        age1 |    1.09016   .0047743    19.71   0.000     1.080842    1.099558
        age2 |   1.042639   .0173954     2.50   0.012     1.009096    1.077297
    diabetes |   2.495091   .2732532     8.35   0.000     2.013103     3.09248
    cursmoke |   1.496908   .0823015     7.34   0.000     1.343987    1.667228
      bpmeds |   1.546226   .2130261     3.16   0.002     1.180325    2.025557
  bpmeds_sex |   1.636712   .3607016     2.24   0.025     1.062632    2.520935
------------------------------------------------------------------------------

. lincom age1+age2, hr
 ( 1)  age1 + age2 = 0

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.136643   .0160758     9.06   0.000     1.105568    1.168592
------------------------------------------------------------------------------

. lincom sex +bpmeds +bpmeds_sex, hr
 ( 1)  sex + bpmeds + bpmeds_sex = 0

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   4.378541     .76663     8.43   0.000     3.106663    6.171128
------------------------------------------------------------------------------
```

