Name________________

I will adhere to the Hopkins code of academic ethics.

Signature_________________

Lecturer (please check one):

( ) Diener-West ( ) McGready

Biostatistics 140.623
Third Term, 2016-2017
Final Examination
March 16, 2017

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.

Questions 1 through 5 concern general knowledge.

1. Suppose that a Likelihood Ratio Test (LRT) was performed for a comparison on an extended model with p+s covariates versus a null model with p covariates. If the observed p-value for the LRT is p=0.22, one would conclude that: (*Circle only one response*)

   a) Taken together, the p covariates $X_1, \ldots, X_p$ do not contribute to the model.
   b) Taken together, the s covariates $X_{p+1}, \ldots, X_{p+s}$ do not contribute to the model.
   c) Taken together, the p+s covariates $X_1, \ldots, X_{p+s}$ do not contribute to the model.
   d) None of the individual s covariates $X_{p+1}, \ldots, X_{p+s}$ are statistically significantly associated with the outcome.
   e) None of the p+s covariates $X_1, \ldots, X_{p+s}$ are statistically significantly associated with the outcome.

2. Consider the log odds [obesity] = $\beta_0+\beta_1 age + \beta_2(age-40)^+ + \beta_3(age-65)^+ + \beta_4$ exercise $+ \beta_5 age{*}exercise + \beta_6(age-40)^+{*}exercise + \beta_7 (age-65)^+{*}exercise$

   where $(age-40)^+ = 0$ if age ≤ 40; or = (age-40) if age > 40
   where $(age-65)^+ = 0$ if age ≤ 65; or = (age-65) if age > 65
   and exercise=1 for daily; 0 for not daily

   In this model, the **log odds ratio** for obesity per unit increase in age among those who are over 65 years of age <u>and</u> exercise daily is: (*Circle only one response*)

   a) $\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\beta_6+\beta_7$
   b) $\beta_3+\beta_5+\beta_6+\beta_7$
   c) $\beta_1+\beta_2+\beta_3+\beta_5+\beta_6+\beta_7$
   d) $\beta_3+\beta_4+\beta_7$
   e) $\beta_0+\beta_3+\beta_4+\beta_7$

3. Let y be the number of events per interval of time for a given treatment group. Let N be the person-years per interval for a given treatment group (such that $\lambda = y/N$). Also let X=1 if treatment A and 0 if treatment B, and Interval=0 if the interval is 0-5 years, 1 if the interval is 5 -10 years, 2 if the interval is 10-15 years, and 3 if the interval is 15 -20 years. Indicator variables are defined as: Int1=1 if Interval=1, 0 otherwise; Int2=1 if Interval=2, 0 otherwise; and Int3=1 if Interval=3; 0 otherwise. Assuming the following model

$$\log\{E(y)\} = \log(N) + \beta_0 + \beta_1 X + \beta_2 \text{Int1} + \beta_3 \text{Int2} + \beta_4 \text{Int3}$$

what assumption(s) are made regarding the risk of an event per unit time? (*Circle only one response*)

a) The risk is constant over the entire time period.
b) The risk changes linearly over the entire time period.
c) The risk varies across time intervals but is constant within a time interval.
d) The risk varies across time intervals and is not constant within a time interval.
e) The risk varies within a time interval but is constant across time intervals.

4. The difference between a plot of Kaplan-Meier estimates for $S(t_j)$ for **ungrouped data** and a plot of life-table survival estimates for S_j, for **grouped data** is that: (*Circle only one response*)

a) $S(t_j)$ is the cumulative probability of survival beyond time t_j whereas S_j is the cumulative probability of survival beyond the end of bin j.
b) $S(t_j)$ is the cumulative probability of survival before time t_j whereas S_j is the cumulative probability of survival before the beginning of bin j.
c) $S(t_j)$ is the probability of survival at time t_j whereas S_j is the probability of survival during bin j.
d) $S(t_j)$ is the probability of survival beyond the end of bin j whereas S_j is the cumulative probability of survival beyond time j.
e) There is no difference; both estimate the same probabilities.

5. The generic formulation of a multiple regression model including age (x_1=age in years), smoking status (x_2 = 1 if smoker, and 0 if non-smoker) and sex (x_3= 1 if female, 0 if male) is as follows:

$$\text{LHS} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 * X_3$$

What comparison is being made by the coefficient, β_2? (*Circle only one response*)

a) Smokers to non-smokers.
b) Smokers to non-smokers of the same sex and age.
c) Smokers to non-smokers of the same age, but only among females.
d) Smokers to non-smokers of the same age, but only among males.
e) The difference, between females and males, between smokers to non-smokers.

Questions 6 through 9 refer to the results of a survey of married women aged 18 to 49 years who were interviewed at an outpatient service in a hospital in Uganda. Of interest was the association between HIV status and desire for future children. Women also reported their partner's desire for future children. (*Am J Public Health 2013; 103: 278-285*).

Table 1 shows the tabulation of desire for future children in 784 married partners by woman's HIV status:

	HIV status of Woman		
Future Children	HIV+	HIV-	Total
Both want	120	275	395
	34.29	63.36	50.38
Both do not want	162	111	273
	46.29	25.58	34.82
Only man wants	49	25	74
	14.00	5.76	9.44
Only woman wants	19	23	42
	5.43	5.30	5.36
Total	350	434	784
	100.00	100.00	100.00

Pearson chi2(3) = 70.3223 Pr = 0.000

6. Suppose that a model was specified in the following way to investigate this association:

$$\textbf{logit(P(Y=1))} = \beta_0 + \beta_1 \mathbf{X}$$

where Y= 1 if **both** partners want future children; 0 otherwise
and X = 1 if HIV+; 0 if HIV-

From the data in **Table 1,** what is the estimate of the regression coefficient, β_1 ? (*Circle only one response*)

a) [(120/230)/(275/159)]
b) $\log_e$[(120/230)/(275/159)]
c) [(120/350)/(275/434)]
d) $\log_e$(275/159)
e) [(275/159)/(120/230)]
f) $\log_e$[(275/159)/(120/230)]

The authors used **logistic regression models** to investigate the association between a woman's desire for future children and her HIV status, along with other characteristics of the individual.

The outcome was defined as: **desire for future children**; 1 for yes, 0 for no.
The covariates are defined in Table 2. In addition, the unadjusted and adjusted results are presented in **Table 2**: (95% CIs are given in parentheses next to each estimated odds ratio).

Table 2

Characteristics	**Unadjusted OR (95% CI)**	**Adjusted* OR (95% CI)**
HIV status		
HIV -	1.000	1.000
HIV +	0.295 (0.228, 0.382)	0.461 (0.326, 0.653)
Age, years		
≤ 24	1.000	1.000
25-29	0.559 (0.394, 0.793)	1.108 (0.772, 1.700)
30-34	0.287 (0.199, 0.414)	0.979 (0.609, 1.573)
35-39	0.214 (0.046, 0.204)	1.193 (0.622, 2.289)
≥ 40	0.071 (0.040, 0.125)	0.346 (0.160, 0.755)
Educational attainment		
Any primary	1.000	1.000
≥ Secondary	1.892 (1.453, 2.465)	1.004 (0.709, 1.420)
Parity	0.473 (0.425, 0.526)	0.505 (0.439, 0.581)
Foster child < 18 years		
0	1.000	1.000
≥1	0.525 (0.399, 0.691)	0.638 (0.450, 0.904)
Household income, UGX		
0 – 50,000	1.000	1.000
50,001-150,000	1.430 (1.039, 1.969)	1.306 (0.872, 1.958)
≥ 150,001	1.990 (1.461, 2.711)	2.006 (1.325, 3.036)
HIV + child in household		
No	1.000	1.000
Yes	0.320 (0.190, 0.539)	0.740 (0.387, 1.415)
Current marriage		
First marriage	1.000	1.000
Second marriage	0.602 (0.444,0.818)	0.805 (0.664,1.467)

***Adjusted for all variables in Table 2**

7. Based on the results in **Table 2**, what is the estimated **adjusted odds ratio** of desiring future children between HIV+ women aged 30-34 and HIV- women aged 25-29 (i.e. HIV+ women aged 30-34 as compared to HIV- women aged 25-29) who are otherwise similar with respect to the other characteristics? (*Circle only one response*)

   a) 0.15
   b) 0.41
   c) 0.88
   d) 1.51
   e) 2.46

8. What **assumption** is made about the relationship between parity (number of times the woman previously has given birth) and the desire for future children in the logistic regression models in **Table 2**? (*Circle only one response*)

   a) The odds ratio of the desire for future children is multiplied by the number of previous births.
   b) The change in the log odds of desire for future children for each additional previous birth is constant.
   c) The odds of desire for future children is not associated with number of previous births.
   d) The variability in the odds of desire for future children is constant.
   e) Parity is independent of the desire for future children.

9. What **conclusion** can be made about the adjusted relationships between each covariate and the desire for future children in the logistic regression models in **Table 2**? (*Circle only one response*)

   a) The adjusted odds of the desire for future children significantly **decreased** in women with positive HIV status, older age, higher parity, one or more foster child, HIV+ child in the household, and/or second marriage and significantly **increased** in women with secondary education and/or from households with > 50,000 UGX income.
   b) The adjusted odds of the desire for future children significantly **decreased** in women with positive HIV status, decreased in women having higher parity, and/or more than one foster child and significantly **increased** in women from households with > 150,000 UGX income.
   c) There is no confounding by any of the other covariates of the relationship between a woman's HIV status and her desire for future children
   d) There is no interaction between HIV status and any of the other covariates on a woman's desire for future children.
   e) There is no statistically significant association between HIV status and desire for future children, after adjusting for all other covariates.

Questions 10-14 pertain to data on men aged 39-40 years enrolled in the Western Collaborative Group Study which investigated the relationship between coronary heart disease (**CHD**) and smoking status and factors such as coronary-prone behavior type. In this data set, the number of CHD events per number at risk are **binned** within smoking status-behavior type groups.

The variables are defined such that:
chd is the number of CHD events in bin j

at risk is the total persons at risk ("exposure) in bin j

ab = 1 for Type A behavior, 0 for Type B behavior
smoke = 1 if never smoker; 2 if former smoker; 3 if light smoker (<20 cigarettes daily); 4 if heavy smoker (≥20 cigarettes daily).

Models A through E at the back of the exam, provide the results of several Poisson models investigating the risk of CHD.

10. What is the expected **overall** CHD incidence rate across all individuals? (*Circle only one response*)

    a) 2.51 fewer CHD events with each one year increase in age.
    b) 8.15 CHD events per 100 individuals.
    c) 79.7 CHD events per 100 person-years.
    d) The CHD incidence rate is over 2 times higher in those with Type A behavior as compared to those with Type B behavior.
    e) 0.08 total CHD events

11. The result of the Likelihood Ratio Test comparing **Models B and D** suggests that? (*Circle only one response*)

    a) Smoking contributes information about the CHD outcome beyond that contributed by behavioral type alone.
    b) Smoking modifies the relationship between CHD and behavioral type.
    c) Smoking and behavioral type interact on the outcome of CHD.
    d) Smoking substantially confounds the relationship between CHD and behavioral type.
    e) CHD substantially confounds the relationship between behavioral type and smoking.

12. The Stata output accompanying **Model D** suggests that, after adjusting for behavioral type: (*Circle only one response*)

    a) The log incidence of CHD is significantly increased in all smoking categories (former, light and heavy smokers) as compared to never smokers but there are no significant differences among smoking categories.
    b) The log incidence of CHD is significantly increased in all smoking categories (former, light and heavy smokers) as compared to never smokers and there are statistically significant differences among pairwise smoking categories.
    c) The log incidence of CHD statistically significantly increases in a linear relationship with the smoking categories.
    d) The log incidence of CHD statistically significantly decreases in a linear relationship with the smoking categories.
    e) There is no association between smoking category and log incidence of CHD.

13. From **Model E,** what is the estimated difference in log(CHD incidence rate) between individuals with <u>Type A behavior who are heavy smokers</u> and individuals with <u>Type B behavior who are former smokers</u>? (*Circle only one, response*)

    a) $b_1 + b_2 + b_3 + b_4 + b_5 + b_6 + b_7 - b_3$
    b) $b_1 + b_4 + b_7 - b_3$
    c) $b_2 + b_3 + b_4 + b_2 - b_5 - b_6 - b_7$_
    d) $b_1 + b_4 + b_7 - b_2$
    e) $b_1 + b_4 + b_7 - b_2 - b_5$

14. What is the null hypothesis tested by the **Likelihood Ratio Test** comparing **Model E** to **Model D**? (*Circle only one response*)

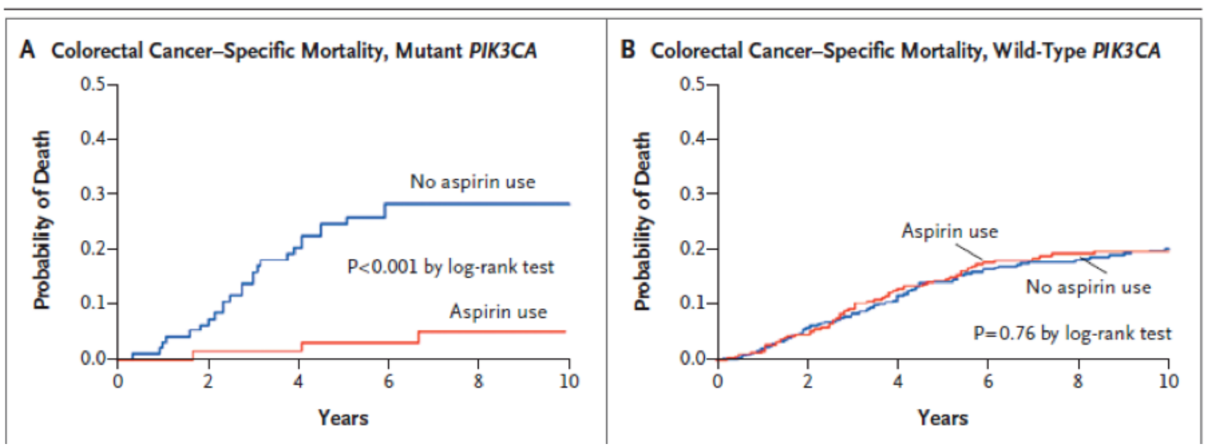
    a) Smoking category modifies the effect of behavioral type on CHD risk.
    b) There is an interaction between smoking category and behavior type on CHD risk.
    c) Individually, each interaction coefficient equals zero.
    d) Taken together, the 3 interaction terms do not contribute to the model of CHD risk above that contributed by smoking category and behavioral type.
    e) Model E provides a precise prediction of the outcome.

Questions 15 through 20 focus on the results of a study examining aspirin and survival among patients with colorectal cancer. (*NEJM* 367 (17): 1596-1606, 2012. As per the authors:

"We obtained data on 964 patients with rectal or colon cancer from the Nurses' Health Study and the Health Professionals Follow-up Study, including data on aspirin use after diagnosis and the presence or absence of PIK3CA mutation."

"We used data from two prospective cohort studies, the Nurses' Health Study (NHS, involving 121,700 women who were enrolled in 1976) and the Health Professionals Follow-up Study (HPFS, involving 51,500 men who were enrolled in 1986)."

The authors present the following **Figure A** and **Figure B** as part of the article. The starting point (time 0) for each patient was the year of colorectal cancer diagnosis. Each is a Kaplan-Meier estimate show the **proportion who had died** by the given follow-up time (i**.e. 1- S(t))**



15. What is the null hypothesis for the **log-rank test** with p-value <0.001 presented in Figure A? (*Circle only one response*)

a) The population level Survival Curves (and, hence, hazards of death over-time) are different for the aspirin and no-aspirin groups.
b) The population level Survival Curves (and, hence, hazards of death over-time) are the same for the aspirin and no-aspirin groups.
c) The hazard of death is constant over time in both the aspirin and non-aspirin groups.
d) The hazard of death is not constant over time in both the aspirin and non-aspirin groups.
e) The ratio of the hazard of death in the aspirin group versus the non-aspirin group changes over time.

16. Suppose the researchers had fit the following Cox regression model, using only the data on subjects with the PIK3CA mutation (the data used to create the curves in **Figure A**):

$$\log(\lambda(t,x_1)) = \log(\lambda_o(t)) + \beta_1 x_1 ,$$

where $x_1 = 1$ for the aspirin group, 0 for the non-aspirin group

What can be said about b_1, the estimate of β_1 for this model? (*Circle only one response*)

a) $b_1=0$
b) $b_1>0$
c) $b_1<0$
d) $b_1>1$
e) $b_1<1$

17. The validity of the **proportional hazards assumption** for the model depicted in **question 16** can be confirmed when: (*Circle only one response*)

a) Observing that the plot of the log(-log S(t)) versus log t results in approximately parallel straight lines for the aspirin and non-aspirin groups.
b) Observing that the plot of the log(-log S(t)) versus log t results in diverging straight lines for the aspirin and non-aspirin groups.
c) The p-value for the log-rank statistic is less than 0.05.
d) The AIC achieves the minimum value.
e) There is a statistically significant interaction between aspirin status (aspirin versus non-aspirin) and time.

18. Suppose the researchers had fit the following Cox regression model, using only the data on subjects with the PIK3CA mutation (the data used to create the curves in **Figure A**):

$$\log(\lambda(t,x_1)) = \log(\lambda_o(t)) + \beta_1 x_1 ,$$

where $x_1 = 1$ for the aspirin group, 0 for the non-aspirin group

What does the function $\lambda_o(t)$ quantify? (*Circle only one response*)

a) The hazard of death for the aspirin group at t=0.
b) The hazard of death for the aspirin group as a function of time over the follow-up period.
c) The hazard of death for the non-aspirin group at t=0.
d) The hazard of death for the non-aspirin group as a function of time over the follow-up period.
e) The hazard ratio of death the aspirin group compared to the non-aspirin group.

19. What assumption did the researchers have to make when fitting the model from **question 18**? (*Circle only one response*)

    a) The hazard of death is constant over time in both the aspirin and non-aspirin groups.
    b) The hazard of death is constant over time in only the non-aspirin group.
    c) The difference in log(hazard of death) between the aspirin and non-aspirin groups is constant over time.
    d) The difference in hazard of death between the aspirin and non-aspirin groups is constant over time.
    e) The log(hazard) of death is a linear function of time in both the aspirin and non-aspirin groups.

20. Which of the following Cox regression models, *based on data from all subjects* (with or without the PIK3CA mutation) would correspond to the results presented in **Figure A and Figure B**? (*Circle only one response*)

    For each of the following models, the variables are defined as:

    x_1 = 1 for the aspirin group, 0 for the non-aspirin group
    x_2 = 1 for those with the PIK3CA mutation , 0 for those with the Wild-Type PIK3CA (no mutation)
    t= follow-up time

    a) $\log(\lambda(t,x_1)) = \log(\lambda_o(t)) + \beta_1 x_1$
    b) $\log(\lambda(t,x_{1,}\ x_2)) = \log(\lambda_o(t)) + \beta_1 x_2$
    c) $\log(\lambda(t,x_1\ x_{2,})) = \log(\lambda_o(t)) + \beta_1 x_1 + \beta_2 x_2$
    d) $\log(\lambda(t,x_1\ x_2)) = \log(\lambda_o(t)) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 t + \beta_4 (x_1 * t)$
    e) $\log(\lambda(t,x_1,x_2,x_3)) = \log(\lambda_o(t)) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 * x_2)$

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

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

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

AIC = RSS + 2(model df)

$$\ln = \log_e$$

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

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

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

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

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:

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

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

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

Models A through D pertain to questions 10 - 14:

The variables are defined such that:
chd is the number of CHD events in bin j
at risk is the total persons at risk (“exposure) in bin j

ab = 1 for Type A behavior, 0 for Type B behavior
smoke = 1 if never smoker; 2 if former smoker; 3 if light smoker (<20 cigarettes daily); 4 if heavy smoker (≥20 cigarettes daily).

```
.gen sm2=0
.gen sm3=0
.gen sm4=0
.replace sm2=1 if smoke==2
.replace sm3=1 if smoke==3
.replace sm4=1 if smoke==4

.gen absm2=ab*sm2
.gen absm3=ab*sm3
.gen absm4=ab*sm4
```

Model A log(Expected Event Rate) = β_0

```
. poisson chd, exposure(atrisk)


Poisson regression                                Number of obs     =          8
                                                  LR chi2(0)        =       0.00
                                                  Prob > chi2       =          .
Log likelihood = -51.745809                       Pseudo R2         =     0.0000

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -2.507351   .0623783   -40.20   0.000     -2.62961   -2.385091
  ln(atrisk) |          1  (exposure)
------------------------------------------------------------------------------

. est store A
```

Model B log(Expected Event Rate) = $\beta_0+\beta_1$ab

```
. poisson chd ab, exposure(atrisk)

Poisson regression                                Number of obs     =          8
                                                  LR chi2(1)        =      37.65
                                                  Prob > chi2       =     0.0000
Log likelihood = -32.921588                       Pseudo R2         =     0.3638

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ab |   .7971166   .1351895     5.90   0.000     .5321501    1.062083
       _cons |  -2.986193   .1125088   -26.54   0.000    -3.206706    -2.76568
  ln(atrisk) |          1  (exposure)
------------------------------------------------------------------------------

. est store B
```

Model C $\log(\text{Expected Event Rate}) = \beta_0 + \beta_1 sm2 + \beta_2 sm3 + \beta_3 sm4$

```
. poisson chd sm2 sm3 sm4  , exposure(atrisk)

Iteration 0:   log likelihood = -38.256964
Iteration 1:   log likelihood = -38.256572
Iteration 2:   log likelihood = -38.256572

Poisson regression                              Number of obs     =          8
                                                LR chi2(3)        =      26.98
                                                Prob > chi2       =     0.0000
Log likelihood = -38.256572                     Pseudo R2         =     0.2607

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sm2 |   .3824274   .1564922     2.44   0.015     .0757084    .6891464
         sm3 |   .7329104   .1737932     4.22   0.000      .392282    1.073539
         sm4 |   .8115823    .189328     4.29   0.000     .4405062    1.182658
       _cons |  -2.824774   .1010153   -27.96   0.000    -3.022761   -2.626788
  ln(atrisk) |          1  (exposure)
------------------------------------------------------------------------------
.est store C
```

Model D $\log(\text{Expected Event Rate}) = \beta_0 + \beta_1 ab + \beta_2 sm2 + \beta_3 sm3 + \beta_4 sm4$

```
. poisson chd ab sm2 sm3 sm4, exposure(atrisk)

Poisson regression                              Number of obs     =          8
                                                LR chi2(4)        =      59.54
                                                Prob > chi2       =     0.0000
Log likelihood = -21.975714                     Pseudo R2         =     0.5753

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ab |   .7471917   .1358939     5.50   0.000     .4808446    1.013539
         sm2 |    .368102   .1565106     2.35   0.019     .0613468    .6748572
         sm3 |   .6858998   .1739665     3.94   0.000     .3449317    1.026868
         sm4 |    .695461   .1902331     3.66   0.000     .3226109    1.068311
       _cons |  -3.248266    .134724   -24.11   0.000     -3.51232   -2.984211
  ln(atrisk) |          1  (exposure)
------------------------------------------------------------------------------

. est store D


. lincom sm4-sm3

 ( 1)  - [chd]sm3 + [chd]sm4 = 0

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0095611   .2139089     0.04   0.964    -.4096927    .4288149
------------------------------------------------------------------------------
```

(continued on next page)

```
. lincom sm3-sm2

 ( 1)  - [chd]sm2 + [chd]sm3 = 0

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3177978   .1852416     1.72   0.086     -.045269    .6808646
------------------------------------------------------------------------------

. lincom sm4-sm2

 ( 1)  - [chd]sm2 + [chd]sm4 = 0

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3273589   .2004667     1.63   0.102    -.0655487    .7202665
------------------------------------------------------------------------------

. lrtest D B

Likelihood-ratio test                                 LR chi2(3)  =     21.89
(Assumption: B nested in D)                           Prob > chi2 =    0.0001
```

Model E **log(Expected Event Rate) = β_0+$\beta_1$ab+ $\beta_2$sm2+ $\beta_3$sm3+ $\beta_4$sm4 +$\beta_5$ab*sm2+ B$_6$ab*sm3+ $\beta_7$ab*sm4**

```
. poisson chd ab sm2 sm3 sm4 absm2 absm3 absm4 , exposure(atrisk)

Poisson regression                                Number of obs     =          8
                                                  LR chi2(7)        =      62.16
                                                  Prob > chi2       =     0.0000
Log likelihood = -20.667256                       Pseudo R2         =     0.6006

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ab |   1.018073   .2236068     4.55   0.000     .5798122    1.456335
         sm2 |   .6440498   .2751623     2.34   0.019     .1047417    1.183358
         sm3 |   .9808293   .3133916     3.13   0.002      .366593    1.595065
         sm4 |   1.092181   .3683942     2.96   0.003      .370142    1.814221
       absm2 |  -.4079256   .3349959    -1.22   0.223    -1.064505    .2486542
       absm3 |  -.4276267   .3767118    -1.14   0.256    -1.165968    .3107148
       absm4 |  -.5454138   .4295146    -1.27   0.204    -1.387247    .2964193
       _cons |  -3.433987   .1889822   -18.17   0.000    -3.804386   -3.063589
  ln(atrisk) |          1  (exposure)
------------------------------------------------------------------------------
. est store E

. lrtest E D

Likelihood-ratio test                                 LR chi2(3)  =      2.62
(Assumption: D nested in E)                           Prob > chi2 =    0.4545
```