Lab #9 Discussion

Question 1. *Note the estimated odds ratios for nonsmokers and smokers and their 95% confidence intervals, and the estimated summary (combined) odds ratio and its 95% confidence interval. What do the results of the test for homogeneity and inspection of the odds ratios allow you to conclude about the effect of smoking on the association between type A behavior and CHD? Is there any evidence for interaction? If we assume that no interaction is present, is there any evidence that smoking has a confounding influence on the measured association between* `chd69` *and* `dibpat`?

```
. cs chd69 dibpat, by(smoke) or

           smoke |       OR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
       nonsmoker |   2.941176     1.882394   4.594855      12.10169 (Cornfield)
          smoker |   1.962697     1.385228   2.780702      23.66644 (Cornfield)
-----------------+-------------------------------------------------
           Crude |   2.372929     1.804034   3.121147
    M-H combined |   2.293753     1.741568   3.021016
-----------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    1.938  Pr>chi2 = 0.1639

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     36.68
                                                Pr>chi2 =    0.0000
```

The estimated odds ratios (95% C.I.s) for the association between `chd69` and `dibpat` among non-smokers and smokers are 2.94 (1.88, 4.59) and 1.96 (1.39, 2.78), respectively. The odds ratio for the association between behavior pattern and CHD adjusted for the binary measure of smoking is 2.29 (1.74, 3.02). The *test of homogeneity* of odds ratios across the strata of the smoking variable is not significant at the 5% level. Visual inspection of the component odds ratios indicates that both represent positive associations between type-A behavior and CHD, and that the 95% confidence intervals overlap substantially. Recall that rejection of this test indicates that the effect of type A behavior on CHD differs in smokers and non-smokers, obligating us to consider smoking when investigating the association of the outcome with behavior. If we operate under the assumption that the indication of interaction is not strong enough to warrant further consideration, we can summarize the behavior-CHD association with the adjusted (MH-combined) odds ratio estimate. If this differs from the unadjusted (Crude) estimate, we can conclude that smoking has a confounding influence on the association of interest. In this case, the degree of confounding is very slight, with the adjusted measure of association differing from the crude estimate by 0.08.

The issue of what level of significance constitutes sufficient evidence for interaction is a delicate one and is somewhat context dependent. As mentioned in lecture, power to detect interactions in observational studies can be limited. This is related both to the fact that the test for homogeneity focuses on detecting all possible departures from the null hypothesis of equal odds ratios (rather than on detecting a particular departure) and that the additional stratification involved in evaluating interaction frequently leads to sparse

data in strata. Adopting a less stringent standard for rejection than the conventional 0.05 cut-off is one way to avoid missing potentially important interactions. However, uncritical adoption of this procedure may also lead to inappropriate identification of interaction. In this example, further investigation of this issue would include use of the underlying quantitative measure of smoking (`ncig`). We may also want to evaluate the impact of controlling for the interaction in further analyses involving additional predictors. As it stands, the relatively large P-value and the observation that both the component odds ratios reveal fairly strong positive associations indicate that this particular interaction is not likely to be important.

Question 2. *How do the estimated odds ratio and 95% confidence interval for* `dibpat` *compare to their counterparts from the results reported above?*

```
. logistic chd69 dibpat smoke
Logistic regression                               Number of obs   =       3154
                                                  LR chi2(2)      =      60.51
                                                  Prob > chi2     =     0.0000
Log likelihood = -860.36587                       Pseudo R2       =     0.0340

------------------------------------------------------------------------------
       chd69 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dibpat |   2.301319   .3238066     5.92   0.000     1.746661     3.03211
       smoke |     1.8002   .2422852     4.37   0.000       1.3828    2.343593
------------------------------------------------------------------------------
```

The estimated odds ratio and 95% C.I. for `dibpat` from the logistic regression fit is 2.30 (1.75, 3.03). These results differ very slightly from the results above (to two decimal places). In fact, there is no advantage to using logistic regression in this example where only two binary predictors are involved. This will not be the case when we consider age as a continuous measure (below). As noted in the lab handout: We don't expect an exact correspondence between the frequency table and logistic regression results when there are multiple risk factors involved. (This is related to the fact that the logistic model represents the effect of risk factors on the outcome as additive on the log odds scale.) Also, a complete analysis of the relationship between CHD, smoking and behavior type would involve assessing the effect of smoking using the quantitative measure `ncig`.

Question 3. *What do the above results allow you to conclude about the effect of age on the association between arcus and CHD risk? About the effect of arcus on the association between age and CHD risk?*

```
. cs chd69 arcus, by(dage) or

             dage |       OR      [95% Conf. Interval]    M-H Weight
------------------+------------------------------------------------
            39-49 |   1.911643     1.348286   2.710533      20.34964 (Cornfield)
            50-59 |     1.0575     .7086192   1.578275      23.00885 (Cornfield)
------------------+------------------------------------------------
            Crude |    1.63528     1.257732   2.126197
    M-H combined |   1.458379     1.118748   1.901115
```

```
------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    4.742  Pr>chi2 = 0.0294

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =      8.05
                                              Pr>chi2 =    0.0046
```

```
. cs chd69 dage, by(arcus) or
```

```
   arcus senilis |       OR      [95% Conf. Interval]    M-H Weight
-----------------+------------------------------------------------
          absent |     2.4431     1.745663   3.419325      18.83853 (Cornfield)
         present |   1.351497     .8957664    2.03919      18.99575 (Cornfield)
-----------------+------------------------------------------------
           Crude |    2.04938      1.58146   2.655782
    M-H combined |    1.89503     1.457833   2.463341
------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    4.747  Pr>chi2 = 0.0294

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =     23.99
                                              Pr>chi2 =    0.0000
```

The results presented above display the association between CHD and arcus, stratified by age, and CHD and age, stratified by arcus. Comparison of the stratum-specific odds ratio estimates and the (identical) tests of homogeneity reveal that a significant interaction between the two predictors is operating. This implies that any further conclusions about the association between either and CHD risk must consider the presence of the other to be valid. CHD risk appears to be elevated among younger participants with arcus relative to their older counterparts.

Question 4. *Compare these results with those obtained from the last two* `cs` *commands and comment on any correspondences. What do the results imply about the presence of interaction?*

```
. logistic chd69 arcus dage arcdage
Logistic regression                               Number of obs   =       3152
                                                  LR chi2(3)      =      40.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -865.43251                       Pseudo R2       =     0.0228
------------------------------------------------------------------------------
       chd69 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       arcus |   1.911643   .3419235     3.62   0.000     1.346349    2.714287
       dage |     2.4431   .4205159     5.19   0.000     1.743529    3.423366
     arcdage |   .5531892    .150593    -2.17   0.030     .3244544     .943178
       _cons |   .0559748   .0060971   -26.47   0.000     .0452142    .0692964
------------------------------------------------------------------------------
```

The odds ratios for `arcus` and `dage` correspond The fact that the 95% confidence interval for the product variable `arcdage` excludes 1, (and the associated P-value of 0.03 is smaller than 0.05) corresponds to the result of the test for homogeneity presented above, indicating a significant interaction between arcus and this binary representation of age in influencing CHD risk. As demonstrated in lab, identifying the component odds ratios from the `logistic` output involves a bit more work.

Question 5. *Interpret each of the above odds ratios using the definition of the logistic model given above*

Using the `lincom` procedure after fitting a logistic regression model allows estimation and inference about combinations of regression coefficients and particular values of predictors using the estimated regression equation. For this example, this has the form:

$$\log(\text{odds of CHD}) = \beta_0 + \beta_1 \texttt{arcus} + \beta_2 \texttt{dage} + \beta_3 \texttt{arcus} \times \texttt{dage}$$

As discussed in lecture, the component log odds ratios for the interaction between arcus and age can be obtained by differences between appropriate pairs of log-odds. For example, the log odds ratio comparing CHD is individuals with arcus to those without can be obtained by taking the difference in the log odds in these two groups. Because the interaction with age is operating (as represented by the presence of the product variable `arcus#dage`), we must specify an age group to perform this calculation. For the younger age group, `dage` =0 and `arcus#dage` = 0, so the log odds of CHD for individuals with arcus in this group is given by $\beta_0 + \beta_1$. The corresponding log odds for individuals in this group without arcus is then β_0. The difference between these log odds gives the log odds ratio associated with arcus in the younger individuals, i.e. β_1. The odds ratio and confidence interval for this are given directly in the logistic output (if we do not request coefficients to be printed). A similar calculation shows that the corresponding log odds ratio for arcus among older individuals is $\beta_1 + \beta_3$. Although we can calculate this from the coefficients of the logistic model, the standard error and 95% confidence interval are not provided. This is where `lincom` comes in. The command

```
lincom 1.arcus + 1.arcus#1.dage
```

requests that the coefficients for these two variables be summed and an associated 95% confidence interval estimated. Stata exponentiates `lincom` results when the preceding model is fitted by logistic regression. Note that even though the output is labeled "Odds Ratio", `lincom` can also be used to estimate odds. For example, the odds for CHD among older individuals with arcus is given by

```
lincom _cons + 1.arcus + 1.dage + 1.arcus#1.dage
```

The odds ratios and 95% confidence intervals produced using `lincom` in this lab correspond very closely to the results from the frequency table approach applied earlier. The latter approach would clearly suffice for this particular analysis. However, if we want to consider age as a continuous variable, the logistic model would clearly be preferred.

The tables used to aid in identifying the requested odds ratios are given below.

arcus	dage	_cons β_0	arcus β_1	dage β_2	arcus#dage β_3
Yes	Yes	1	1	1	1
Yes	No	1	1	0	0
	Difference	0	0	1	1

arcus	dage	_cons	arcus	dage	arcus#dage
Yes	No	1	1	0	0
No	No	1	0	0	0
	Difference	0	1	1	0

arcus	dage	_cons	arcus	dage	arcus#dage
No	Yes	1	0	1	0
No	No	1	0	0	0
	Difference	0	0	1	0

Question 6. *Do you see evidence for an additive interaction?*

Below are the results from fitting the additive risk (risk difference) regression model to look at the same interaction:

```
. binreg chd69 i.arcus##i.dage, rd

Generalized linear models                         No. of obs      =      3152
Optimization     : MQL Fisher scoring             Residual df     =      3148
                   (IRLS EIM)                     Scale parameter =         1
Deviance         =  1730.865012                   (1/df) Deviance =  .5498301
Pearson          =         3152                   (1/df) Pearson  =  1.001271

Variance function: V(u) = u*(1-u)                 [Bernoulli]
Link function    : g(u) = u                       [Identity]

                                                  BIC             = -23628.77
------------------------------------------------------------------------------
             |                 EIM
      chd69 | Risk Diff.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      arcus |
   present  |   .0436531   .0135409     3.22   0.001     .0171134    .0701927
            |
       dage |
     50-59  |    .067293   .0151269     4.45   0.000     .0376448    .0969412
            |
 arcus#dage |
present#50-59|  -.0376097   .0260577    -1.44   0.149    -.0886819    .0134625
            |
      _cons |   .0530077   .0054679     9.69   0.000     .0422909    .0637246
------------------------------------------------------------------------------
```

Interestingly, the *p*-value for the coefficient of the product variable is not significant at the 5% level. Evidence for additive interaction is weak.

As discussed in lecture (and in the text), frequently we can't fit additive risk models directly because of numerical issues. In these cases, we can still evaluate evidence for the presence of additive interaction by computing the "relative excess risk for interaction"

($RERI_{RR}$) statistic based on estimated relative risks from a regression model. This statistic is defined for two binary predictors as

$$RERI_{RR} = RR_{11} - RR_{10} - RR_{01} + 1$$

Here, the first relative risk compares outcome risk in those with both predictors present, with the corresponding risk in those with both absent. In terms of a log relative risk model, this can be rewritten as follows:

$$\log(P) = \beta_0 + \beta_1 \text{arcus} + \beta_2 \text{dage} + \beta_3 \text{arcus} \times \text{dage}$$

$$RERI_{RR} = e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1} - e^{\beta_2} + 1$$

We can successfully fit the relative risk regression model for this example:

```
. binreg chd69 i.arcus##i.dage, rr

Generalized linear models                         No. of obs      =      3,152
Optimization     : MQL Fisher scoring             Residual df     =      3,148
                   (IRLS EIM)                     Scale parameter =          1
Deviance         =  1730.865012                   (1/df) Deviance =   .5498301
Pearson          =  3151.881069                   (1/df) Pearson  =   1.001233

Variance function: V(u) = u*(1-u)                 [Bernoulli]
Link function    : g(u) = ln(u)                   [Log]
                                                  BIC             =  -23628.77
------------------------------------------------------------------------------
             |                 EIM
       chd69 | Risk Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       arcus |
     present |   1.823522   .2999889     3.65   0.000     1.320929    2.517346
             |
        dage |
       50-59 |   2.269494   .3543959     5.25   0.000     1.671123     3.08212
             |
  arcus#dage |
present#50-59|   .5759376   .1403651    -2.26   0.024     .3572105    .9285957
             |
       _cons |   .0530077   .0054677   -28.48   0.000     .0433051    .0648843
------------------------------------------------------------------------------
```

The following `nlcom` statement computes $RERI_{RR}$ based on the coefficients from the model:

```
. nlcom exp(_b[1.arcus]+_b[1.dage]+_b[1.arcus#1.dage]) - exp(_b[1.arcus]) -
exp(_b[1.dage]) + 1

       _nl_1:  exp(_b[1.arcus]+_b[1.dage]+_b[1.arcus#1.dage]) - exp(_b[1.arcus]) -
exp(_b[1.dage]) + 1

------------------------------------------------------------------------------
       chd69 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.7095139   .5119629    -1.39   0.166    -1.712943     .293915
------------------------------------------------------------------------------
```

The conclusion about the lack of additive interaction is similar to what we obtained by fitting the additive model directly. Note that the outcome is relatively rare in categories defined by age and arcus for this dataset. The following table illustrates that the outcome prevalence doesn't exceed 13% in subgroups:

```
. tabulate arcus dage, summarize(chd69) means

                           Means of CHD event

     arcus |        dage
   senilis |     39-49      50-59 |     Total
-----------+----------------------+----------
    absent | .05300774  .12030075 | .06919946
   present | .09666081  .12634409 | .10839532
-----------+----------------------+----------
     Total | .06405694  .12278761 | .08090102
```

Repeating the above analysis with the logistic regression-based analog $RERI_{OR}$ yields a similar result:

```
. quietly logistic chd69 i.arcus##i.dage
. nlcom exp(_b[1.arcus]+_b[1.dage]+_b[1.arcus#1.dage]) - exp(_b[1.arcus]) -
exp(_b[1.dage]) + 1

_nl_1:  exp(_b[1.arcus]+_b[1.dage]+_b[1.arcus#1.dage]) - exp(_b[1.arcus]) -
exp(_b[1.dage]) + 1

------------------------------------------------------------------------------
       chd69 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.7711643   .6159101    -1.25   0.211    -1.978326    .4359974
------------------------------------------------------------------------------
```

Question 6. *What do the estimated coefficient for the product variable and its 95% confidence interval imply about the presence of a possible arcus-age interaction?*

The results (below) imply that a significant multipicative interaction is present. The *p*-value for the Wald test associated with the product term corresponds closely to the result of the test for homogeneity obtained above.

```
. logistic chd69 i.arcus##c.age, coef

Logistic regression                               Number of obs   =       3152
                                                  LR chi2(3)      =      53.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -858.93362                       Pseudo R2       =     0.0301

------------------------------------------------------------------------------
       chd69 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       arcus |
    present  |   2.754185   1.140118     2.42   0.016     .5195952    4.988774
         age |    .089647   .0148904     6.02   0.000     .0604623    .1188317
             |
 arcus#c.age |
    present  |  -.0498298   .0233431    -2.13   0.033    -.0955814   -.0040782
             |
       _cons |  -6.788086   .7179977    -9.45   0.000    -8.195335   -5.380836
------------------------------------------------------------------------------
```

Question 7. *Compute the above odds ratio for a 40 year old. Given what we know about the age range of subjects in the WCGS, does it make any sense to use the above model to*

estimate this odds ratio for an individual of age 20?

We can compute the odds ratio for the effect of arcus on CHD for a 40-year old individual by using the model coefficients (stored in internal variables after the last model fitted), or by using `lincom` again.

```
. lincom 1.arcus + 1.arcus#c.age*40

 ( 1)  [chd69]1.arcus + 40*[chd69]1.arcus#c.age = 0

------------------------------------------------------------------------------
       chd69 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.140398   .5140334     3.17   0.002     1.336816    3.427025
------------------------------------------------------------------------------
```

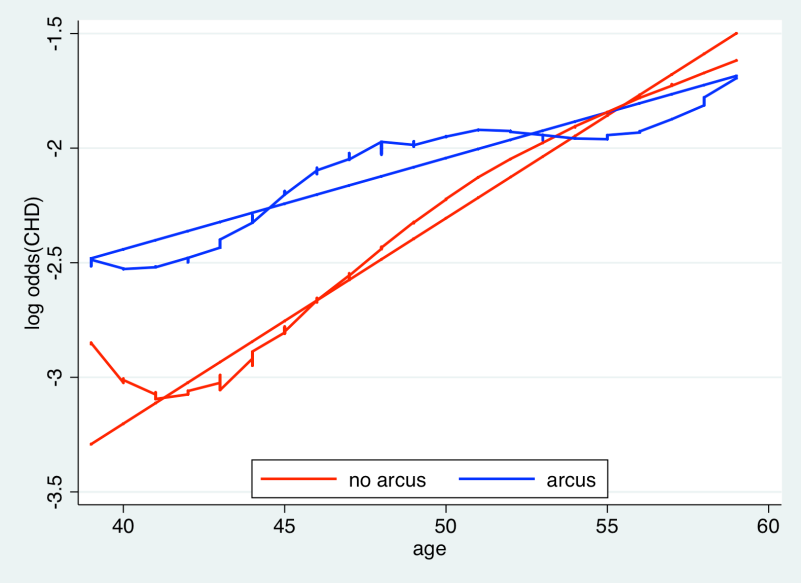
Recall that the age range of participants in this study is 35-60. We would not expect regression models to apply to ages well outside this range, especially considering the disease outcome under consideration.

Question 8. *What do the plotted lines reveal about the relationship between CHD risk and age in the two arcus groups?*

Although both groups exhibit increased CHD risk with increasing age, individuals with arcus appear to be at greater risk overall except among individuals over age 55. (This observation is consistent with the results obtained above for the categorical representation of age.) The slope of increase in risk with age is greater among individuals without arcus.

Question 9. *What do you conclude from the graph about the adequacy of the linear representation of age in the two groups?*

Since the smoothed estimates follow the observed data quite closely, differences with the corresponding linear estimates reveal possible non-linearity in the relationship between the log-odds of CHD risk and age in the two groups. The lines appear to represent the data quite well with the exception of the younger age range (35-45), where some increased risk is evident. Without further investigation of this phenomenon (possibly including formal comparison of a model allowing the log odds to increase with younger ages with the linear alternative), we do not have enough information to conclude that the linear representation is adequate. However, the linear models appear to capture the differences between the two arcus groups well and in this sense provide a reasonable summary of the interaction effect.



Question 10. *Is there evidence for additive interaction in this model?*

The *p*-value for the product term `arcus#c.age` in model output (below) is 0.484, which is very different from what we saw in the logistic model. We conclude that we don't require an interaction to model the relationship between CHD risk and these variables on this scale. So this is a case where a multiplicative interaction is required to model the relationship between these variable on the log odds scale, but not on the risk/probability scale (i.e. no convincing evidence of additive interaction).

```
. binreg chd69 i.arcus##c.age, rd

Generalized linear models                         No. of obs      =      3152
Optimization     : MQL Fisher scoring             Residual df     =      3148
                   (IRLS EIM)                     Scale parameter =         1
Deviance         =   1720.92654                   (1/df) Deviance =   .546673
Pearson          =  3151.956739                   (1/df) Pearson  =  1.001257

Variance function: V(u) = u*(1-u)                 [Bernoulli]
Link function    : g(u) = u                       [Identity]

                                                  BIC             = -23638.71

------------------------------------------------------------------------------
             |                 EIM
       chd69 | Risk Diff.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       arcus |
    present  |   .0951563   .0945146     1.01   0.314    -.0900888    .2804015
         age |    .005478   .0010767     5.09   0.000     .0033677    .0075883
             |
 arcus#c.age |
    present  |  -.0014309   .0020431    -0.70   0.484    -.0054354    .0025735
             |
       _cons |   -.180477   .0474075    -3.81   0.000     -.273394   -.0875599
------------------------------------------------------------------------------
```

Definition of the $RERI_{RR}$ or $RERI_{OR}$ statistics when one of the predictors in the model is continuous requires us to specify the change in the continuous variable for which we wish

to evaluate interaction (i.e. by specifying the actual values of the continuous predictor that we wish to compare). Let's assume that the outcome is sufficiently rare in this case that we can use $RERI_{OR}$ to approximate $RERI_{RR}$. Because we are considering age as continuous, the modification below is based on an increase in age from age a_1 to age a_2:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \texttt{arcus} + \beta_2 \texttt{age} + \beta_3 \texttt{arcus*age}$$

$$RERI_{OR} = e^{\beta_1 + (a_2 - a_1)\beta_2 + a_2\beta_3} - e^{\beta_1} - e^{(a_2 - a_1)\beta_2} + 1$$

(Note that this formula depends on the particular ages specified, not just the change in age.) For a ten-year age increase from 50 to 60 (i.e. a_2=60, a_1=50):

```
. quietly logistic chd69 i.arcus##c.age
. nlcom exp(_b[1.arcus]+10*_b[c.age]+60*_b[1.arcus#c.age]) - exp(_b[1.arcus]) -
exp(10*_b[c.age]) + 1

       _nl_1:  exp(_b[1.arcus]+10*_b[c.age]+60*_b[1.arcus#c.age]) - exp(_b[1.arcus]) -
exp(10*_b[c.age]) + 1

------------------------------------------------------------------------------
      chd69 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
      _nl_1 |   -15.2227   18.44542    -0.83   0.409    -51.37505    20.92966
------------------------------------------------------------------------------
```

The results from the accompanying test that the effect is zero (p=0.41) are quite similar to the p-value from the product term in the additive model fitted using `binreg`.

Question 11. *What do you conclude about the contribution of behavior pattern in our model for CHD risk?*

```
. logistic chd69 chol sbp age smoke bmi i.behpat2

Logistic regression                               Number of obs   =       3142
                                                  LR chi2(8)      =     189.40
                                                  Prob > chi2     =     0.0000
Log likelihood = -794.89999                       Pseudo R2       =     0.1065

------------------------------------------------------------------------------
      chd69 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
       chol |   1.010798   .0015221     7.13   0.000     1.007819    1.013785
        sbp |   1.018203   .0042049     4.37   0.000     1.009995    1.026478
        age |   1.062438   .0127373     5.05   0.000     1.037764    1.087698
      smoke |   1.828109    .257801     4.28   0.000     1.386645    2.410122
        bmi |   1.057105   .0280768     2.09   0.037     1.003483    1.113592
            |
    behpat2 |
          2 |    .898615    .252463    -0.38   0.704     .5181212    1.558533
          3 |   1.868554   .4926845     2.37   0.018     1.114469    3.132877
          4 |   1.747332   .5578522     1.75   0.080     .9345907    3.266853
            |
      _cons |   4.89e-06   4.89e-06   -12.23   0.000     6.89e-07    .0000347
------------------------------------------------------------------------------

. lrtest A1
Likelihood-ratio test                                 LR chi2(3)  =     24.81
(Assumption: . nested in A1)                          Prob > chi2 =    0.0000
```

A significant result in the LR test indicates that the increase in likelihood associated with including `behpat2` (or `behpat`) in a model containing the other predictors is substantial, and exceeds the value required for significance at the 5% level with the chi-squared distribution. (The minimum value required to achieve significance at the 5% level turns out to be 7.815.) We conclude that behavior pattern makes a significant contribution to the model already containing the other variables. We could also say that the estimated effect of behavior pattern on CHD risk is significant when adjusting for the other included predictors.

Question 12. *What does the LR test just performed allow you to conclude about the relative contributions of the two different representations of behavior pattern in our model for CHD risk?*

```
. logistic chd69 chol sbp age smoke bmi dibpat

Logistic regression                               Number of obs   =       3142
                                                  LR chi2(6)      =     189.16
                                                  Prob > chi2     =     0.0000
Log likelihood = -795.01767                       Pseudo R2       =     0.1063

------------------------------------------------------------------------------
       chd69 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        chol |   1.010779   .0015212     7.12   0.000     1.007802    1.013765
         sbp |   1.018241   .0041957     4.39   0.000     1.010051    1.026498
         age |    1.06223    .012712     5.04   0.000     1.037605     1.08744
       smoke |   1.826545   .2575497     4.27   0.000     1.385504    2.407981
         bmi |   1.056673   .0280322     2.08   0.038     1.003135    1.113069
      dibpat |   2.008081   .2899056     4.83   0.000     1.513191    2.664826
       _cons |   4.59e-06   4.50e-06   -12.54   0.000     6.72e-07    .0000313
------------------------------------------------------------------------------

. lrtest A1

Likelihood-ratio test                                 LR chi2(2)  =      0.24
(Assumption: . nested in A1)                          Prob > chi2 =    0.8890
```

The first LR test clearly showed that the four-level version of behavior pattern made a significant contribution to a model already containing the other predictors. The non-significant results for the second test agree with the numerical observation that the likelihood for the model including the four-level version was not much larger than the two-level version. Considering that the larger model uses two additional parameters to represent the behavior effect, a LR statistic of 0.24 is not very impressive (and does not even come close the that value of 6 required for significance of the chi-squared test with two degrees of freedom). Clearly, controlling for behavior with the binary indicator of "type A" is sufficient here.