Self Evaluation Problems
Class 3

Questions 1 through 4 concern the 2007 National Hospital Ambulatory Medical Care Survey (NHAMCS) which is a national (United States) sample survey of visits to hospital outpatient and emergency departments. Of interest is the association between waiting time (in minutes) of persons admitted to the Emergency Departments (EDs) and patient demographic and other characteristics.

waittime = waiting time for patient to be seen in ED (in minutes)
race = 1 if non-white; 0 if white
and **age defined with two spline terms:**
age1 = age (in years)**,**
where **age2**=0 if age ≤ 18 years or **age2** = (age-18) if age > 18 years.
where **age3**=0 if age ≤ 65 years or **age3** = (age-65) if age > 65 years.

The following Models are defined:

Model A E[waittime] = $\beta_0 + \beta_1$**race**

```
. regress waittime   race

      Source |       SS       df       MS              Number of obs =   27928
-------------+------------------------------           F(  1, 27926) =  170.44
       Model |  1051120.55     1  1051120.55           Prob > F      =  0.0000
    Residual |   172219095 27926   6166.9804           R-squared     =  0.0061
-------------+------------------------------           Adj R-squared =  0.0060
       Total |   173270215 27927  6204.39772           Root MSE      =   78.53

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        race |   12.65684   .9694724    13.06   0.000     10.75663    14.55706
       _cons |   51.50819   .5954894    86.50   0.000     50.34101    52.67538
------------------------------------------------------------------------------
```

Model B E[waittime] = $\beta_0 + \beta_1(\mathbf{age}1) + \beta_2(\mathbf{age}2) + \beta_3(\mathbf{age}3)$

```
. mkspline age1 18 age 2 65 age3 = age, marginal
. regress waittime  age1 age2 age3

      Source |       SS       df       MS              Number of obs =   27928
-------------+------------------------------           F(  3, 27924) =   20.30
       Model |   377098.83     3   125699.61           Prob > F      =  0.0000
    Residual |   172893116 27924  6191.55982           R-squared     =  0.0022
-------------+------------------------------           Adj R-squared =  0.0021
       Total |   173270215 27927  6204.39772           Root MSE      =  78.686

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |   .4333112    .104551     4.14   0.000     .2283862    .6382362
        age2 |  -.4920716   .1277496    -3.85   0.000    -.742467   -.2416761
        age3 |  -.4606402   .1253335    -3.68   0.000      -.7063   -.2149803
       _cons |   51.68964   1.475652    35.03   0.000     48.79729    54.5819
```


Model C $E[\text{waittime}] = \beta_0 + \beta_1(\textbf{age1}) + \beta_2(\textbf{age2}) + \beta_3(\textbf{age3}) + \beta_4(\textbf{race})$

```
. regress waittime  age1 age2 age3 race

      Source |       SS       df       MS              Number of obs =   27928
-------------+------------------------------           F(  4, 27923) =   54.97
       Model |  1353714.37     4  338428.592           Prob > F      =  0.0000
    Residual |   171916501 27923  6156.80624           R-squared     =  0.0078
-------------+------------------------------           Adj R-squared =  0.0077
       Total |   173270215 27927  6204.39772           Root MSE      =  78.465

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |   .4559665   .1042726     4.37   0.000      .251587     .660346
        age2 |  -.4956842   .1273909    -3.89   0.000    -.7453766   -.2459919
        age3 |  -.4074942   .1250525    -3.26   0.001    -.6526033   -.1623852
        race |   12.28505   .9754226    12.59   0.000     10.37317    14.19692
       _cons |   46.18539   1.535032    30.09   0.000     43.17665    49.19413
------------------------------------------------------------------------------
```

Model D $E[\text{waittime}] =$
$\beta_0 + \beta_1(\textbf{age1}) + \beta_2(\textbf{age2}) + \beta_3(\textbf{age3}) + \beta_4(\textbf{race}) + \beta_5(\textbf{age1race}) + \beta_6(\textbf{age2race}) + \beta_7(\textbf{age3race})$

```
. gen age1race=age1*race
. gen age2race=age2*race
. gen age3race=age3*race
. regress waittime  age1 age2 age3 race age1race age2race age3race

      Source |       SS       df       MS              Number of obs =   27928
-------------+------------------------------           F(  7, 27920) =   33.83
       Model |  1457375.01     7   208196.43           Prob > F      =  0.0000
    Residual |   171812840 27920  6153.75502           R-squared     =  0.0084
-------------+------------------------------           Adj R-squared =  0.0082
       Total |   173270215 27927  6204.39772           Root MSE      =  78.446

------------------------------------------------------------------------------
    waittime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |   .1871963   .1386911     1.35   0.177    -.0846451    .4590376
        age2 |  -.2549328   .1671775    -1.52   0.127    -.5826089    .0727433
        age3 |  -.3033843   .1486893    -2.04   0.041    -.5948226    -.011946
        race |   1.949157   2.957607     0.66   0.510    -3.847896    7.746211
    age1race |   .5807609   .2106753     2.76   0.006      .167827    .9936947
    age2race |  -.4890108   .2592729    -1.89   0.059    -.9971984    .0191768
    age3race |  -.3233276   .2812286    -1.15   0.250    -.8745493    .2278941
       _cons |   50.85607   1.982651    25.65   0.000     46.96998    54.74216
------------------------------------------------------------------------------
```

1. Based on a comparison of **Model A** and **Model C**, which of the following statements is correct? (*Circle only one response*)

   a) The unadjusted association between ED waiting times and race is completely explained by age differences in the race categories.
   b) The association between ED waiting times and race is modified by age.
   c) Non-whites have longer average waiting times than whites, and this difference is statistically significant after adjusting for age.
   d) When taken together, neither race nor age are statistically significant predictors of ED waiting times.
   e) Age substantially confounds the association between ED waiting times and race.

2. Suppose the researchers were interested in testing whether the relationship between ED waiting times and age is modified by race. Which of the following steps would be appropriate? (*Circle only one response*).

   a) Use the results from the overall F-test for Model D.
   b) Compare the R^2 values between Model D and Model A.
   c) Use a F-test to compare Model C to Model B.
   d) Use the t-test to test the significance of the coefficient of *age2race* in Model D.
   e) Use a F-test to compare Model D to Model C.

3. Based on the results from **Model D**, what is the average ED wait time (in minutes) for **white** individuals who are **50 years** of age? (*Circle only one response*)

   a) 50.85+0.19(50)+-0.25(50-18)
   b) 50.85 +0.19(50) +-0.25(50)
   c) 50. 85+ 0.19(50)+ -0.25(50-18) + -0.30(50-65)
   d) 50.85+0.19(50)
   e) 50.84

4. The best interpretation of the R^2 value from **Model D** is: (*Circle only one response*)

   a) Taken together, age and race precisely predict waiting times for future ED patients.
   b) In this sample, age and race explain nearly all of the variability in ED waiting times.
   c) Neither age or race is statistically significantly associated with waiting times.
   d) Even after adjusting for their relationship to age and race, there is still substantial unexplained variability in the waiting times.
   e) There is a statistically significant interaction between race and age.

Questions 5 – 10 concern a study investigating the associations between smoking, systolic blood pressure and myocardial infarction (MI) in 117 individuals.

Variables are defined as:

MI = 1 if myocardial infarction; 0 if not
SMK= 1 for smoker; 0 for non-smoker
SBP (systolic blood pressure category) =1 if SBP < 140 mm Hg;
=2 if SBP 140-160 mm Hg;
=3 if SBP > 160 mm Hg

with dummy variables defined as **sbp2**=1 if SBP 140-160 mm Hg; 0 otherwise
and **sbp3**=1 if SBP > 160 mm Hg; 0 otherwise

5. From the following **Model 1**: log(odds of MI) = $\beta_0 + \beta_1 \mathbf{SMK} + \beta_2 \mathbf{sbp}2 + \beta_3 \mathbf{sbp}3$

```
. logit MI SMK sbp2 sbp3

Log likelihood = -53.373051

------------------------------------------------------------------------------
          MI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         SMK |   .5546278   .5219739     1.06   0.288    -.4684223    1.577678
        sbp2 |  -1.380778   .6850522    -2.02   0.044    -2.723455   -.0381001
        sbp3 |   2.370212   .5638804     4.20   0.000     1.265026    3.475397
       _cons |  -1.215697    .367703    -3.31   0.001    -1.936381   -.4950122
------------------------------------------------------------------------------
```

what can you conclude about the relationship between MI for individuals with SBP, **adjusting for smoking**? (*Circle only one response*)

a) The odds of MI are statistically significantly decreased in individuals with SBP ≥ 140 mm Hg as compared to SBP < 140 mm Hg, holding smoking status fixed.
b) **The odds of MI are statistically significantly decreased in individuals with SBP 140-160 mm Hg as compared to SBP < 140 mm Hg, holding smoking status fixed.**
c) There is no significant association between SBP and the odds of MI after controlling for smoking status.
d) The odds of MI are statistically significantly decreased in individuals with SBP > 160 mm hg as compared to SBP < 140 mm Hg, holding smoking status fixed.
e) There is a statistically significant increase in the change in log odds of MI each mm Hg increase in SBP beyond the cutpoint of 160 mm Hg as compared to before the cutpoint, after adjusting for smoking status.

6. Based on **Model 1** from question 1, what is the **predicted probability** of MI for a smoker with SBP > 160 mm Hg? *(Circle only one response)*

   a) 0.15
   b) 0.37
   c) 0.63
   d) 0.75
   e) **0.85 since log(odds) = log (p/(1-p) = $b_0+b_1+b_2$**
   And, p=odds/(1+odds)
   = exp(-1.216+0.555+2.370)/(1+exp(-1.216+0.555+2.370))= 0.85

7. Suppose you also investigated the following **Model 2**:

$$\text{log(odds of MI)} = \beta_0 + \beta_1\mathbf{SMK} + \beta_2\mathbf{sbp}2 + \beta_3\mathbf{sbp}3 + \beta_4\mathbf{smksbp}2 + \beta_5\mathbf{smksbp}3$$

```
. gen smkbp2=SMK*sbp2
. gen smkbp3=SMK*sbp3
. logit MI SMK sbp2 sbp3 smkbp2 smkbp3

Log likelihood = -53.306321

------------------------------------------------------------------------------
          MI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         SMK |   .6649763   .6868955     0.97   0.333    -.6813141    2.011267
        sbp2 |  -1.386294    .834523    -1.66   0.097    -3.021929    .2493406
        sbp3 |   2.505526   .6943651     3.61   0.000     1.144595    3.866456
      smkbp2 |   .0281709   1.466423     0.02   0.985    -2.845966    2.902307
      smkbp3 |  -.4136619   1.185059    -0.35   0.727    -2.736335    1.909011
       _cons |  -1.252763   .4008919    -3.12   0.002    -2.038497   -.4670294
------------------------------------------------------------------------------
```

The Likelihood Ratio Test which compares the **extended Model 2** with the **null Model 1** suggests that: (*Circle only one response*)

a) **Fail to reject the null hypothesis and conclude that there is no substantial added contribution in predicting MI provided by the addition of the smoking-SBP interactions above that contributed by SBP and smoking alone.**
   The LRT for nested models = -2(LL_{Null} - $LL_{Extended}$) = -2(-53.37-(53.31)) =0.12 ~ χ^2 with 2 degrees of freedom, p > 0.05

b) Fail to reject the null hypothesis and conclude that SBP is not needed in a model predicting MI.
c) Reject the null hypothesis and conclude that there is a statistically significant interaction between smoking and SBP on the log odds of MI.
d) There is a statistically significant association between the log odds of MI and smoking.
e) Reject the null hypothesis and conclude that there is a substantial additional contribution in predicting MI provided by smoking above that contributed by SBP alone

8. The interpretation of the logistic regression coefficient, $ß_0$, in **Model 2** is: *(Circle only one response*).

   a) The log odds ratio for MI in smokers versus non-smokers by level of SBP.
   b) The difference in the log odds of MI in smokers versus non-smokers after controlling for SBP categories.
   c) The log odds of MI across all 117 individuals in the data set.
   d) The log odds of MI in non-smokers with SBP < 140 mm Hg.
   e) The odds of MI in non-smokers after adjusting for SBP.

9. Suppose you are interested in relating MI to a continuous predictor (age) via a regression model. Which of the following graphical displays use the data to describe the approximate nature of the relationship between MI and age? (*Circle only one response*).

   a. A residuals versus fitted values plot from a simple linear regression of MI on age.
   b. A scatterplot of the values of MI versus the values of age.
   c. A lowess plot showing the estimated smoothed relationship between the log odds of MI and age.
   d. A lowess plot showing the estimated smoothed relationship between the fitted values from a linear regression and age, including two spline terms.
   e. A residuals versus predicted values plot from a simple linear regression of MI on age.

10. Which sentence correctly describes the exponentiated **smkbp3** coefficient in **Model 2**? (*Circle only one response*)

   a) The odds ratio for MI in smokers versus non-smokers having SBP > 160 mm Hg.
   b) The ratio of the odds ratio for MI in smokers versus non-smokers among individuals with SBP > 160 mm Hg to the odds ratio for MI in smokers versus non-smokers among individuals with SBP < 140 mm Hg.
   c) The difference in the log odds of MI for smokers versus non-smokers for individuals with SBP > 160 mm Hg versus > 140 mm Hg.
   d) The ratio of the odds ratio for MI in non-smokers having SBP > 160 mm Hg versus < 140 mm Hg in non-smokers to the odds ratio for MI in smokers having SBP < 140 mm Hg versus > 160 mm Hg.
   e) The difference between smokers and non-smokers in the log odds ratio for MI in individuals having SBP > 160 mm Hg versus < 140 mm Hg.

