Sociology 63993
Exam 2
March 30, 2012

I. True-False. (20 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A researcher has constructed scales that measure health, diet, exercise, income and education. When she regresses health on the other four variables, she finds that the effect of education is 0. This means that, at least when it comes to health, it makes no difference whether you are well educated or poorly educated.

2. A key advantage of centering independent variables is that centering often eliminates the need to include interaction terms in the model.

3. A researcher runs the following:

```
. ovtest

Ramsey RESET test using powers of the fitted values of health
     Ho:  model has no omitted variables
            F(3, 10330) =      5.62
                Prob > F =      0.0008
```

This indicates that our model has too many extraneous variables.

4. A researcher hypothesizes that income positively affects the self-image of men but has a negative effect on the self-image of women. She gets

$$\hat{\beta}_{Income} = 12$$
$$\hat{\beta}_{Female} = 0$$
$$\hat{\beta}_{Income \ * \ Female} = -4$$

Female = 1 if female, 0 if male. The T values for Income and for the interaction term are both highly significant. The evidence supports the researcher's hypothesis.

5. A researcher has regressed Y on X1. However, X2 should also be in the model. As a result of this omission, the coefficient for X1 will definitely be biased.

II.   Path Analysis/Model specification (25 pts). A sociologist believes that the following model describes the relationship between X1, X2, X3, and X4. All her variables are in standardized form. The estimated value of each path in her model is included in the diagram.
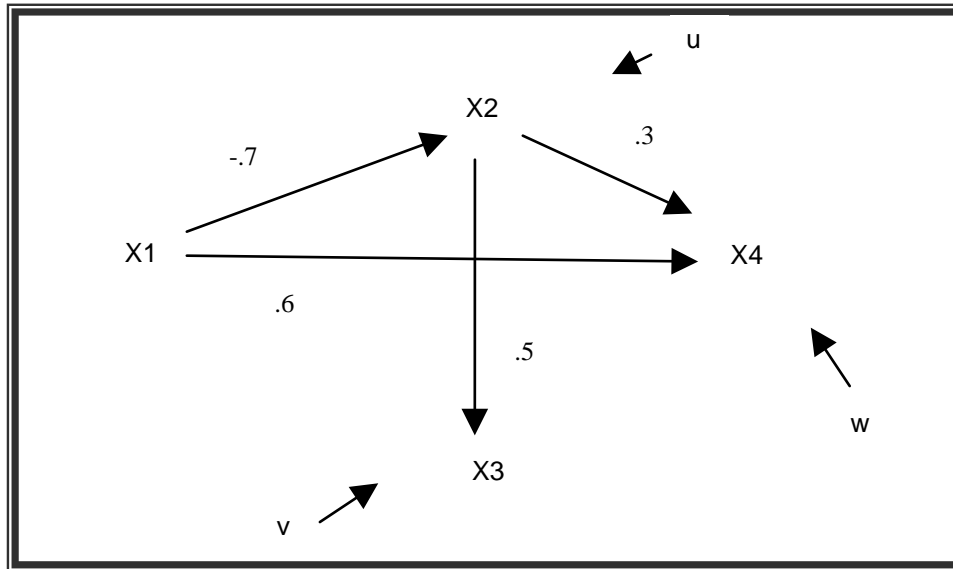
a.   (5 pts) Write out the structural equation for each endogenous variable, using both the names for the paths (e.g. β_{42}) and the estimated value of the path coefficient.

b.   (10 pts) Part of the correlation matrix is shown below. Determine the complete correlation matrix. (Remember, variables are standardized. You can use either normal equations or Sewell Wright, but you might want to use both as a double-check.)

```
             |      x1         x2         x3         x4
-------------+-----------------------------------------
          x1 |   1.0000
          x2 |  -0.7000     1.0000
          x3 |      ?           ?      1.0000
          x4 |     .?           ?          ?      1.0000
```

c.   (5 pts) Decompose the correlation between X1 and X4 into

- Correlation due to direct effects
- Correlation due to indirect effects
- Correlation due to common causes

d.   (5 pts) Suppose the above model is correct, but instead the researcher believed in and estimated the following model:

X2 ⟶ **X4** ⟵ **w**
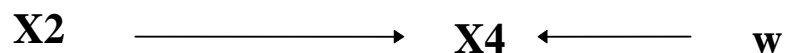
What conclusions would the researcher likely draw? In particular, what would the researcher conclude about the effect of changes in X2 on X4? Discuss the consequences of this mis-specification, and in what ways, if any, the results would be misleading. Why would she make these mistakes?

III.   Group comparisons (25 points). The Catholic Bishops are adamantly opposed to proposals that they feel would mandate that Catholic Institutions pay for contraceptive health care. However, they want to know more about how the Catholic laity feel about these issues. In particular, how do Catholic men and women differ in their beliefs? They have therefore collected information from almost 2,300 US Catholics on the following variables.

Variable	Description
contracept	Scale that measures support for government-mandated coverage of contraceptive health care. Scale ranges from 0 to 125, where a higher score indicates greater support.
female	Coded 1 if female, 0 otherwise
religious	Scale that measures how religious the respondent is. Original scale ranges from 0 (not religious at all) to 75 (extremely religious). *The scale used here has been centered to have a mean of zero.*
femrel	female * religious

The results of the analysis are as follows:

```
. * Estimate Models
. nestreg: reg  contracept  religious female femrel

Block  1: religious

      Source |       SS       df       MS              Number of obs =    2270
-------------+------------------------------           F( 1,  2268) =  540.90
       Model | 488051.725      1  488051.725           Prob > F      =  0.0000
    Residual | 2046391.94   2268  902.289213           R-squared     =  0.1926
-------------+------------------------------           Adj R-squared =  0.1922
       Total | 2534443.66   2269  1116.98707           Root MSE      =  30.038


------------------------------------------------------------------------------
  contracept |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   religious | -3.689038   .1586182    -23.26   0.000    -4.000091   -3.377986
       _cons |  61.28458   .6304635     97.21   0.000     60.04824    62.52093
------------------------------------------------------------------------------


Block  2: female

      Source |       SS       df       MS              Number of obs =    2270
-------------+------------------------------           F( 2,  2267) = 1317.39
       Model | 1362299.24      2  681149.622           Prob > F      =  0.0000
    Residual | 1172144.42   2267    517.0465           R-squared     =  0.5375
-------------+------------------------------           Adj R-squared =  0.5371
       Total | 2534443.66   2269  1116.98707           Root MSE      =  22.739


------------------------------------------------------------------------------
  contracept |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   religious | -1.359155   .1327702    -10.24   0.000    -1.619519   -1.098791
      female |  43.49509   1.057762     41.12   0.000     41.42081    45.56937
       _cons |  38.09997   .7386994     51.58   0.000     36.65138    39.54857
------------------------------------------------------------------------------
```

Block 3: femrel

```
      Source |       SS       df       MS              Number of obs =    2270
-------------+------------------------------           F(  3,  2266) =  879.57
       Model | 1363521.85        3  454507.284         Prob > F      =  0.0000
    Residual | 1170921.81     2266  516.735131         R-squared     =  0.5380
-------------+------------------------------           Adj R-squared =  0.5374
       Total | 2534443.66     2269  1116.98707         Root MSE      =  22.732


------------------------------------------------------------------------------
   contracept |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    religious | -1.133977   .1976051    -5.74   0.000    -1.521483   -.7464716
       female | 43.60929    1.060046    41.14   0.000     41.53052    45.68805
       femrel | -.4102893   .2667354    -1.54   0.124    -.9333604    .1127818
        _cons | 37.69189    .7846873    48.03   0.000     36.15311    39.23067
------------------------------------------------------------------------------


  +------------------------------------------------------------------+
  |       |         Block  Residual                         Change   |
  | Block |       F     df        df    Pr > F        R2     in R2    |
  |-------+----------------------------------------------------------|
  |     1 |  540.90      1      2268    0.0000    0.1926              |
  |     2 | 1690.85      1      2267    0.0000    0.5375    0.3449    |
  |     3 |    2.37      1      2266    0.1241    0.5380    0.0005    |
  +------------------------------------------------------------------+
```

. * Differences by gender
. ttest contracept, by(female)

Two-sample t test with equal variances
```
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |    1060    35.63679    .6628397     21.5805    34.33616    36.93742
  Female |    1210    83.75289     .707922    24.62511     82.364     85.14178
---------+--------------------------------------------------------------------
combined |    2270    61.28458    .7014733    33.42136    59.90899    62.66018
---------+--------------------------------------------------------------------
    diff |            -48.1161    .9782482               -50.03446   -46.19775
------------------------------------------------------------------------------
    diff = mean(Male) - mean(Female)                            t = -49.1860
Ho: diff = 0                                   degrees of freedom =     2268

    Ha: diff < 0                  Ha: diff != 0                  Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

. ttest religious, by(female)

Two-sample t test with equal variances
```
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |    1060     1.81229   .1085763    3.534988    1.599241    2.025339
  Female |    1210   -1.587626      .1049    3.648953   -1.793432   -1.381819
---------+--------------------------------------------------------------------
combined |    2270    2.45e-08   .0834429    3.975598   -.1636324    .1636325
---------+--------------------------------------------------------------------
    diff |            3.399915   .1512899                3.103234    3.696596
------------------------------------------------------------------------------
    diff = mean(Male) - mean(Female)                            t =  22.4729
Ho: diff = 0                                   degrees of freedom =     2268

    Ha: diff < 0                  Ha: diff != 0                  Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Based on the above results, advise the Bishops on the following. When thinking about your answers, keep in mind the various reasons that two groups can differ on some outcome measure.

a) (5 pts) In block 3, the coefficient for female is 43.60929. Assuming the model is correct, the analyst for the Bishops thinks this means that, when a man and woman are equally religious, the woman is expected to score 43.6 points higher than the man on the contraception measure. Indicate whether you agree or disagree, and explain why.

b) (10 pts) The researchers begin by estimating a series of models. Which of the models do you think is best, and why? What do these models tell us about how religiosity affects support for mandatory contraceptive care? What ways (if any) do the determinants of support for contraceptive care differ by gender?

c) (10 pts) The researchers then run a series of t-tests. What do these t-tests tell us about how attitudes differ by gender? What additional insights, if any, do these tests give us as to why support for mandatory contraceptive care differs by gender?

IV.   Short answer. Answer *both* of the following questions. (15 points each, 30 points total.) Each of the following describes a nonlinear or nonadditive relationship between variables. Draw a scatterplot that illustrates the relationship. Describe the harms that might result if you simply regressed Y on X, e.g. would values be over-estimated, under-estimated, or what? Indicate the model you think should be estimated, e.g. $E(Y) = \alpha + \beta_1 X + \beta_2 X^2$ and/or give the Stata commands that would estimate the model. Explain what variables you would need to compute in order to actually estimate the model, e.g. logs of variables, interaction terms. Finally, indicate how you would actually test whether or not nonlinearity or nonadditivity actually was a problem. If you find it helpful, you are welcome to present the Stata commands you would use, but the statistical rationale behind the command still needs to be clear.

a.   Mitt Romney is worried that issue positions that are helping him to win Republican votes in the primaries may hurt him in the general election. Specifically, he believes that, the more he opposes Obama's health care plan, the more Republicans tend to support him. But, at the same time, he fears that the more he opposes Obama's health care plan, the less support he will get from Independents and Democrats.

b.   Obama too is wondering how much to emphasize his health care program. Based on data from past electoral campaigns, his advisors believe that each additional dollar he spends on ads promoting his health care program, up to $10 million, will steadily increase his support among voters. However, their research suggests that any dollars spent above that on health care advertising will have no effect on his support.