1. In their classic 1982 paper, “Beyond Wives' Family Sociology: A Method for Analyzing Couple Data,” Thomson and Williams examined the relationship between the subjective expected utility of children and childbearing expectations. Husbands and wives were presented with several possible consequences of having another child within the next 20 months. Products of their subjective probability of each consequence (0 = no chance...10 = certain) and their evaluation of the consequence (-3 = extremely bad...+3 = extremely good) were constructed to form “subjective expected utilities” of another child. For the analyses presented below, the subjective expected utilities “a fulfilled family life” (X1 for wives, X3 for husbands) and “watching another child grow and develop” (X2 for wives, X4 for husbands) were used as indicators of child utility. The dependent variable, Y, is wife's childbearing expectations. Wives and husbands were asked to estimate the likelihood that they would have another child within 20 months (1 = extremely unlikely...7 = extremely likely; only the wife's measure is presented here though). In the spss printout presented below, only the wife variables (X1, X2, and Y) are used in the regression analysis.

```
            Mean  Std Dev  Label

X1        11.360   11.450  W's SEU for fulfilled family life
X2        22.340   10.890  W's SEU for watching another child grow
Y          3.640    2.660  W's Childbearing expectations

N of Cases =   340

Correlation:

                  X1         X2          Y

X1            1.000       .470       .628
X2             .470      1.000       .421
Y              .628       .421      1.000

Equation Number 1    Dependent Variable..   Y   W's Childbearing expectations

Multiple R           .64398
R Square             .41471          R Square Change    .41471
Adjusted R Square    .41124          F Change        119.39130
Standard Error      2.04104          Signif F Change    .0000

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2          994.73432        497.36716
Residual           337         1403.89408          4.16586

F =     119.39130       Signif F =  .0000

---------------------------------- Variables in the Equation -----------------------------------

Variable             B        SE B     95% Confdnce Intrvl B       Beta    SE Beta   Correl Part Cor

X1             .128257     .010969      .106682      .149833    .552086    .047214  .628000  .487308
X2             .039453     .011533      .016768      .062138    .161520    .047214  .421000  .142568
(Constant)    1.301617     .252942      .804072     1.799161


--------------- Variables in the Equation ----------------

Variable    Partial  Tolerance        VIF         T  Sig T

X1          .537238    .779100      1.284    11.693  .0000
X2          .183199    .779100      1.284     3.421  .0007
(Constant)                                    5.146  .0000
```

```
------------------------ Variables not in the Equation ------------------------

Variable     Beta In  Partial  Tolerance        VIF  Min Toler         T  Sig T

X3           .255428  .295754    .784685      1.274    .659420     5.675  .0000
X4           .192958  .238633    .895172      1.117    .733926     4.504  .0000
```

a. According to the above analysis, which form of subjective expected utility is more important for determining childbearing expectations? Justify your answer by citing at least 2 or 3 pieces of evidence from the printout.

b. What is the squared semipartial for X1? How much would R^2 decline if X1 were dropped from the model?

c. If you were doing forward stepwise regression, what variable, if any, would be added next? Why?

2. The data is the same as above, except that the corresponding SEU measures for the husbands (X3 and X4) have been added to the model.

```
            Mean  Std Dev  Label

X1        11.360   11.450  W's SEU for fulfilled family life
X2        22.340   10.890  W's SEU for watching another child grow
X3         9.750   10.730  H's SEU for fulfilled family life
X4        18.500   10.300  H's SEU for watching another child grow
Y          3.640    2.660  W's Childbearing expectations

N of Cases =   340

Correlation:

                 X1         X2         X3         X4          Y

X1            1.000
X2             .470      1.000
X3             .460       .270      1.000
X4             .312       .223       .495      1.000
Y              .628       .421       .498       .381      1.000

Equation Number 1    Dependent Variable..   Y   W's Childbearing expectations

Multiple R           .68955
R Square             .47548
Adjusted R Square    .46921
Standard Error      1.93794

Analysis of Variance
                   DF      Sum of Squares      Mean Square
Regression          4          1140.49257        285.12314
Residual          335          1258.13583          3.75563

F =      75.91887       Signif F =  .0000

---------------------------------- Variables in the Equation ----------------------------------

Variable             B        SE B     95% Confdnce Intrvl B       Beta    SE Beta   Correl Part Cor

X1                 (1)        (2)       .078521      .123175    .434102    .048857  .628000  .351579
X2             .033271     .011002          (3)          (4)    .136212    .045042  .421000  .119661
X3                 (5)        (6)          (7)      .074821    .205330    .049050  .498000  .165643
X4             .029324     .011860      .005995      .052653    .113546    .045923     (8)   .097836
(Constant)     .712303     .283045      .155533     1.269074       (9)
```

```
-------------- Variables in the Equation ----------------

Variable    Partial  Tolerance        VIF         T  Sig T

X1          .436709    .655938      1.525     8.885  .0000
X2          .163014    .771751      1.296     3.024  .0027
X3          .222956       (10)      1.537     4.186  .0000
X4          .133872    .742424        (11)    2.473  .0139
(Constant)                                    2.517  .0123
```

a. Fill in the missing information (1) - (11). HINT: There are many ways to double-check your answers. Some approaches are far easier than others.

b. Do an F-test of the hypothesis H_0: $ß_3 = ß_4 = 0$.

c. Do an F-test of the hypothesis H_0: $ß_1 = 0$.

d. Suppose the null hypothesis H_0: $ß_4 = 0$ is true. If we drew 10,000 samples, in how many of them would we expect to find a value of b_4 that is as big or bigger in magnitude as what we found here (i.e. how many samples would likely contain a value of b_4 that is bigger than .029324 or smaller than -.029324)?

e. Suppose we did two-tailed tests for each b coefficient separately (i.e., first we tested H_0: $ß_1 = 0$, then we tested H_0: $ß_2 = 0$, etc). If we are using the .01 level of significance, which coefficients will we conclude are statistically significant?

f. Suppose X2 was completely uncorrelated with the other three X variables, i.e. its tolerance equaled 1. What would the standard error of b_2 be? What would its semipartial correlation be?

g. Suppose the tolerance of X3 equaled 0. What would the standard error of b_3 be?

h. Using the standardized coefficients, confirm that the value of R^2 given in the printout is correct.

i. Suppose the following two variables were added to the model: X5 = 0 if black, 1 otherwise. X6 = 1 if black, 0 otherwise. Drawing on your knowledge of sociological and political theory and your mastery of statistics, do you think the effects of X5 and X6 would be statistically significant? Why or why not?

j. Interpret these results. Among other things, your discussion should mention the following:

1. Who tends to have a higher subjective expected utility of children - husbands or their wives?
2. Who tends to be more influential in determining childbearing expectations - the husband or the wife? What do you think is the simplest explanation for this finding?
3. Does the average wife think it is likely or unlikely that she will have another child in the next 20 months?

4. How strongly are childbearing expectations determined by the subjective expected utility of children?

k. Suppose you now did backwards stepwise regression using the .05 level of significance. What variable, if any, would be dropped from the model? Explain your answer. How would your answer change if you used the .01 level of significance?

3. In all of the following problems, Y = income (in thousands of dollars). Assume each sample only includes members of the groups mentioned, and that the sample size for each group is the same (e.g. there are 50 blacks and 50 whites; or, 100 Catholics, 100 Protestants, 100 Jews.)

a. X = 0 if black, 1 if white. If a = 10 and b = 5, what is the average white income? Average black income? Average income for the entire sample? (HINT: The average income for the entire sample is not $10,000.)

b. X1 = 1 if Catholic, 0 if Protestant, -1 if Jewish.
X2 = 0 if Catholic, 1 if Protestant, -1 if Jewish.
a = 10, b1 = -1, b2 = -2.
Find the average income for each religious group and for the entire population.

c. The average male makes $30,000, the average female $20,000. X = 1 if male, 0 if female. Compute a and b.

d. X1 = 1 if Sociologist, 0 if Political Scientist, -1 if Chemist.
X2 = 0 if Sociologist, 1 if Political Scientist, -1 if Chemist.
The average Sociologist makes $35,000 a year, the average Political Scientist makes $40,000, the average Chemist makes $60,000. Compute a, b_1, and b_2.

4. (This problem is adapted from Education, Occupation, & Earnings, by William H. Sewell and Robert M. Hauser. 1975, Academic Press. See especially pages 72 and 79.) Focusing on earnings as the end product of the achievement procedure, Sewell and Hauser did extensive research on the post-secondary schooling, occupational achievements, and earnings during the first 10 years after high school graduation of a large sample of young men who graduated from Wisconsin high schools in 1957. A portion of their analysis can be replicated using Hw10.sps. The variables examined are:

V Father's educational attainment (i.e. years of education)
M Mother's educational attainment
X Status of father's occupation when son graduated from high school (Duncan SEI)
I Parent's average income, 1957-1960 (in thousands of dollars)
Q Son's score on Henmon-Nelson Test of Mental Ability
U Son's educational attainment
W Son's 1964 occupation (Duncan SEI)
Y Canonically weighted average of son's 1965-1967 earnings (in thousands of dollars)

[Note: The Duncan Socio-Economic Index is a measure of occupational prestige.]

In Hw10.sps, three models are estimated using these data:

I. W is regressed on V and M
II. W is regressed on X, Q, and U
III. W is regressed on X, Q, U, V, and M.

Copy and run Hw10.sps. Use the printout to answer the following questions: [HINT: No calculations are required to answer any of these questions, the necessary numbers are in the printout]

a. On average, who is the best educated - the mother, the father, or the son? What does this suggest about educational mobility across generations?

b. According to Model I, should we accept or reject H_0: $ß_M = ß_V = 0$? Use the .05 level of significance.

c. According to Model III, should we accept or reject H_0: $ß_M = ß_V = 0$? Use the .05 level of significance.

d. Are your results from (b) and (c) consistent? If not, offer a possible sociological explanation for the discrepancy.

e. What would you say is the most important determinant of W, the son's 1964 occupational prestige? What is the second most important? Offer a brief sociological explanation as to why these two variables are more important than the rest.

NOTE: As an added bonus, Hw10.sps also contains the cards needed to double-check your answers to problem 2. You just need to un-comment the regression card. But of course, don't do this until AFTER you have worked the problems by hand.

5. Use Stata to replicate the results from problem 2. You can either use hw10-2.dta or create a pseudo-replication of the data using the `corr2data` command (which is what I did, since I do not have the original data). Specifically, use Stata to compute

a. The means, correlations and standard deviations

b. The metric coefficients and their 95% confidence intervals

c. The standardized coefficients

d. The tolerances and VIFs

e. The semipartial and partial correlations

Note that each of the above will require the use of a different Stata command.