1. Consider OLS estimation of $y$ on $X$ where an intercept is included. Then

$$R^2 = \frac{\hat{y}'A\hat{y}}{y'y}$$

where $A = I - \frac{1}{N}ll'$ and $l$ is an $N \times 1$ vector of ones.

(a) Show that $\hat{y}'A\hat{y} = \hat{y}'Ay$.

(b) Hence show that

$$R^2 = \frac{(\hat{y}'Ay)^2}{(y'Ay)(\hat{y}'Ay)}.$$  

(c) Hence show that $R^2$ equals the squared correlation coefficient between $\hat{y}$ and $y$ (i.e. between the fitted and actual values of $y$).

2. Consider the linear regression model $y = X\beta + u$, where $y$ is an $N \times 1$ vector, $X$ is an $N \times k$ matrix of constants, $\beta$ is a $k \times 1$ vector and $u$ is an $N \times 1$ vector, and it is assumed that $u$ has mean $0$ and variance $\sigma^2 I$. Consider the estimator

$$\tilde{\beta} = (Z'X)^{-1}Z'y,$$

where $Z$ is an $N \times k$ matrix of constants. Note that $\tilde{\beta}$ is not the OLS estimator. Also assume plim $\frac{1}{N}Z'X$ exists, plim $\frac{1}{N}Z'u = 0$ and $\frac{1}{\sqrt{N}}Z'u \overset{d}{\rightarrow} N(0, \Omega)$. Be clear stating any other assumptions you use to derive your results.

(a) Show whether or not $\tilde{\beta}$ is consistent. Then show whether or not $\tilde{\beta}$ is unbiased.

(b) Show that $\tilde{\beta}$ has variance matrix $\sigma^2 (Z'X)^{-1}Z'Z(X'Z)^{-1}$, stating clearly any assumptions.

(c) Derive the limiting distribution of $\sqrt{N}(\tilde{\beta} - \beta)$. Then show that $\tilde{\beta}$ is normally distributed, stating clearly any assumptions needed for this result to hold.

(d) Let $A = V[\tilde{\beta}]$ and $B = V[^{\text{OLS}}\tilde{\beta}] = \sigma^2 (X'X)^{-1}$. Show that $B^{-1} - A^{-1}$ is a positive semidefinite.

(e) Part (d) implies that $B - A$ is a negative semidefinite matrix. State clearly what you conclude about the efficiency of $\tilde{\beta}$ compared to that of the OLS estimator.

3. Consider the linear regression model $y = X\beta + u$ where $u \sim [0, \sigma^2 I]$, and estimator

$$\tilde{\beta} = (X'AX)^{-1}X'Ay.$$  

(a) Show that $\tilde{\beta} - \beta = (X'AX)^{-1}X'Au$.

(b) Show that we can write $\tilde{u} = y - X\tilde{\beta}$ as $\tilde{u} = Mu$, where

$$M = I - X(X'AX)^{-1}X'A.$$  

(c) Show that $E[\tilde{u}'\tilde{u}] = \sigma^2 \text{trace}[P]$ where $P = M'M$. 

1
(d) Suppose that $A$ is symmetric idempotent. Will $\text{trace}[P] = N - k$ where $X$ is $N \times k$?


**Dependent Variable:**

\[ \text{lnwage} = \text{Natural logarithm of hourly wage rate in dollars} \]

**Regressors:**

- jirate = job injury rate per 100 workers per year for the industry that the individual works in
- exper = years of work experience
- expersq = experience-squared
- educ = years of schooling
- union = 1 if union member and 0 otherwise
- nonwhite = 1 if nonwhite and 0 otherwise

The data set herschdataextract.csv and associated program ps3.do are at the course website. You will need to adapt ps3.do to answer the questions below.

You are to compute Wald F-tests using Stata matrix commands and the formula

\[ W = (R\hat{\beta} - r)'[RV[\hat{\beta}]R']^{-1}(R\hat{\beta} - r)/q. \]

Compare your results to those given in the OLS output (where appropriate).

(a) Test at level 0.05 that the coefficient of exper equals zero. State clearly your conclusion.

(b) Test at level 0.05 that all coefficients aside from the intercept equal zero. State clearly your conclusion.

(c) Test at level 0.05 that the coefficient of exper and expersq jointly equal zero. State clearly your conclusion.

(d) Test at level 0.05 that being the member of a union has the same impact as one more year of education. State clearly your conclusion.

(e) Redo parts (a)-(d) using the Stata command test. [For details give Stata command help test].

5. The following question uses the same data as question 4. You will also find ps3.do helpful.

(a) The restricted OLS estimator is

\[ \tilde{\beta} = \tilde{\beta} - (XX)^{-1}R'[R(XX)^{-1}R']^{-1}(R\tilde{\beta} - r), \]

where $\tilde{\beta}$ is the OLS estimator.

Suppose we wish to impose the restriction that the coefficients of exper and expersq equal zero. Use Stata matrix commands to compute $\tilde{\beta}$ given $\hat{\beta}$ obtained in question 4.

[Hint: Specify the appropriate matrix $R$, and note that $V[\tilde{\beta}] = s^2(XX)^{-1}$ and the $s^2$ terms cancel out in the above formula. Also, you may want to type help matrix and help mkmat in STATA to see how to construct the matrix $X$ from the data].

(b) Compare your answer in part (a) with that obtained by OLS regression of lnwage on jirate, educ, union, nonwhite, and an intercept.

(c) Consider testing that exper and expersq jointly equal zero using the formula

\[ F = \frac{(SSR_r - SSR_u)/q}{SSR_u/(N - k)}, \]
where $SSR$ denotes the sum of squared residuals, and subscripts $u$ and $r$ denote restricted and unrestricted residuals. Show that this yields the same value test statistic as that in the previous question 4 part(b).

6. The following questions use the same data as question 4. You will also find ps3.do helpful. Consider prediction of wage (now in levels) based on regression on jirate, educ, union, nonwhite, and an intercept for a white unionized worker with 12 years of education, 10 years of experience working in an industry with job injury rate of 3.

(a) Run the regression and obtain the predicted value of the wage.

(b) Find the standard error of the expected wage for a white unionized worker with 12 years of education, 10 years of experience working in an industry with job injury rate of 3.

[Hint: In general $\tilde{V}[c'\beta] = c'\tilde{V}[\beta]c$.]

(c) Hence give a 95 percent confidence interval for the expected wage for a white unionized worker with 12 years of education, 10 years of experience working in an industry with job injury rate of 3.

(d) Now give the standard error of the actual wage for a white unionized worker with 12 years of education, 10 years of experience working in an industry with job injury rate of 3.