Economics 522A, Homework 6 Suggested Solutions

1. (a) First, note

$$\begin{aligned} E[y_i - x_i'\beta \mid x_i] &= E[y_i \mid x_i] - x_i'\beta \\ &= x_i'\beta - x_i'\beta \\ &= 0 \end{aligned}$$

Therefore, using the law of iterated expectations,

$$\begin{aligned} E[x_i(y_i - x_i'\beta)] &= E\left[E\left[x_i(y_i - x_i'\beta) \mid x_i\right]\right] \\ &= E\left[x_i E\left[y_i - x_i'\beta \mid x_i\right]\right] \\ &= E\left[x_i 0\right] = 0. \end{aligned}$$

(b) $\hat{\beta}$ solves:

$$\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) = 0,$$

$$\Rightarrow \quad \sum_i x_i(y_i x_i'\hat{\beta}) = 0,$$

$$\Rightarrow \quad \sum_i x_i y_i - \left(\sum_i x_i x_i'\right)\hat{\beta} = 0,$$

$$\Rightarrow \quad \hat{\beta} = \left(\sum_i x_i x_i'\right)^{-1} \sum_i x_i y_i.$$

Note that:

$$\begin{aligned} X'X &= \sum_i x_i x_i' \\ X'y &= \sum_i x_i y_i \end{aligned}$$

So

$$\hat{\beta} = (X'X)^{-1}X'y.$$

(c) The question did not make this clear, but we can assume that the first element of x_i is a constant.

First, we will show that

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = 0.$$

Since the first element of x_i is 1, the fact that $\frac{1}{n}\sum_i x_i(y_i - x_i'\hat{\beta}) = 0$ implies that

$$\frac{1}{n}\sum_i 1 \cdot (y_i - x_i'\hat{\beta}) = \frac{1}{n}\sum_i e_i = 0.$$

Next, consider the sample covariance between x_{ij} and e_i. Let

$$\bar{x}_j = \frac{1}{n}\sum_i x_{ij}.$$

Then

$$\begin{aligned}
\widehat{Cov}(x_{ij}, e_i) &= \frac{1}{n}\sum_i (x_{ij} - \bar{x}_j)(e_i - \bar{e}) \\
&= \frac{1}{n}\sum_i (x_{ij} - \bar{x}_j)e_i \\
&= \frac{1}{n}\sum_i x_{ij}e_i - \bar{x}_j\frac{1}{n}\sum_i e_i \\
&= \frac{1}{n}\sum_i x_{ij}e_i - 0 \\
&= \frac{1}{n}\sum_i x_{ij}(y_i - x_i'\hat{\beta}) = 0
\end{aligned}$$

2. Form lagrangean:

$$\mathcal{L} = (y - X\beta)'(y - X\beta) + 2\lambda'(R\beta - r).$$

FOC:

$$\begin{aligned}
-2X'y + 2X'X\beta + 2R'\lambda &= 0 \\
2R\beta - r &= 0
\end{aligned}$$

This implies:

$$\begin{aligned}
X'X\hat{\beta}_R + R'\lambda &= X'y \\
R\hat{\beta}_R &= r
\end{aligned}$$

From the first equation:

$$\hat{\beta}_R = (X'X)^{-1}X'y - (X'X)^{-1}R'\lambda$$

and using the second equation,

$$R\hat{\beta}_R = R(X'X)^{-1}X'y - R(X'X)^{-1}R'\lambda = r$$

So

$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}\left[R\hat{\beta} - r\right].$$

Plugging this back into the first equation,

$$\hat{\beta}_R = (X'X)^{-1}X'y - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)$$

which proves the result.

3. Since the treatment variable is binary, the assumption that

$$y_i|x_i \sim N(\beta_1 + \beta_2 \texttt{treat}_i, \sigma^2)$$

is equivalent to assuming that

$$y_i|\texttt{treat}_i = 0 \sim N(\beta_1, \sigma^2),$$

and

$$y_i|\texttt{treat}_i = 1 \sim N(\beta_1 + \beta_2, \sigma^2).$$

We can therefore interpret β_1 as $E[y_i|\texttt{treat}_i = 0]$, and we can interpret β_2 as

$$\beta_2 = E[y_i|\texttt{treat}_i = 1] - E[y_i|\texttt{treat}_i = 0].$$

In words, β_2 is the difference in expected earnings between those who received the program and those who did not.

You might tempted to say that β_2 is the "effect" of the job training on annual earnings, but this does not necessarily hold, because the $\texttt{treat}_i = 1$ group consists of different individuals than the $\texttt{treat}_i = 0$ group.

(In fact, in this data, the treatment was randomly assigned, so this interpretation might be reasonable, but this information was not given to you in the statement of the homework.)

The M-file `hw6.m` carries out the calculations to analyze this data set. Here are the results:

```
b =
  4554.8
  1794.3
s2 =  4.3290e+07
SE =
  408.05
  632.85
R2 = 0.017823
tstat = 2.8353
Fstat = 8.0390
CI_lower = 553.95
CI_upper = 3034.7
```

The point estimates are $\hat{\beta}_1 = 4554.8$, meaning that average real earnings among untreated workers in 1978 is \$4,554.80, and $\hat{\beta}_2 = 1794.3$, meaning that treated workers earn \$1,794.30 more on average than untreated workers.

The t statistic of 2.835 is greater than 1.96, so we would reject the hypothesis that $\beta_2 = 0$ (the hypothesis that treated and untreated workers have the same expected earnings) at the 5% level. The F statistic is the square of the t statistic, as we showed in class, and it also leads to the same conclusion.

The 95% confidence interval for β_2 is $[554, 3035]$, so our interval estimate is that the difference in expected earnings is somewhere between \$554 and \$3035. This is a fairly wide interval, reflecting a relatively large standard error on the estimate of β_2, but would suggest that treated workers earn an economically meaningful amount more than untreated workers.

```
% hw6.m - suggested solutions to HW6, Q3
load hw6_nohead.dat  % this contains data, with variable names removed
y = hw6_nohead(:,10);
n = length(y);
X = [ones(n,1) hw6_nohead(:,11)];

% OLS estimates
b = inv(X'*X)*X'*y

% form s^2
e = y-X*b;
s2 = (e'*e)/(n-2)

% standard errors
Vb = s2*inv(X'*X);
SE = sqrt(diag(Vb))   % make sure you understand logic here!

% form R^2
ybar = mean(y);
R2 = 1 - (e'*e )/((y-ybar)'*(y-ybar))

% t-test for beta_2 = 0
tstat = abs( b(2)/SE(2) )
% reject if tstat > 1.96

% F-test for beta_2 = 0
Fstat = tstat^2
% reject if Fstat > 3.84

% 95 percent confidence interval for beta_2
CI_lower = b(2) - 1.96*SE(2)
CI_upper = b(2) + 1.96*SE(2)

% end hw6.m
```