Cornell University
Department of Economics

Econon 620
Instructor: Prof. Kiefer

Solution to Problem set # 1

1) We will use the following fact: $log_e\ w = b \Leftrightarrow e^b = w$. By taking log to base "10" to the last expression, we get that $b\ log_{10}\ e = \log_{10} w$. Denote by "ln" the log to base "e", and by log, the log to base 10.

Therefore, log X = log e ln X, for any variable X. Let $\overline{\log x} = \frac{1}{n}\sum_{i=1}^{n}\log x_i$ and $\overline{\log y} = \frac{1}{n}\sum_{i=1}^{n}\log y_i$ and similarly, $\overline{\ln x} = \frac{1}{n}\sum_{i=1}^{n}\ln x_i$ and $\overline{\ln y} = \frac{1}{n}\sum_{i=1}^{n}\ln y_i$

Hence,

$\hat{\beta}_{10} = \frac{\sum_{i=1}^{n}(\log x_i - \overline{\log x})\log y_i}{\sum_{i=1}^{n}(\log x_i - \overline{\log x})^2} = \frac{\sum_{i=1}^{n}(\log e)^2(\ln x_i - \overline{\ln x})\ln y_i}{\sum_{i=1}^{n}(\log e)^2(\ln x_i - \overline{\ln x})^2} = \frac{(\log e)^2\sum_{i=1}^{n}(\ln x_i - \overline{\ln x})\ln y_i}{(\log e)^2\sum_{i=1}^{n}(\ln x_i - \overline{\ln x})^2} = \frac{\sum_{i=1}^{n}(\ln x_i - \overline{\ln x})\ln y_i}{\sum_{i=1}^{n}(\ln x_i - \overline{\ln x})^2} = \hat{\beta}_e$

However, it is not true that $\hat{\alpha}_{10} = \hat{\alpha}_e$.To see this, note that

$\hat{\alpha}_{10} = \overline{\log y} - \hat{\beta}_{10}\overline{\log x} = \overline{\log y} - \hat{\beta}_e\overline{\log x} = \log e\,\overline{\ln y} - \hat{\beta}_e \log e\,\overline{\ln x} = \log e(\overline{\ln y} - \hat{\beta}_e\overline{\ln x}) = \log e\ \hat{\alpha}_e.$

Also, if the model is $logY_t = \alpha_{10} + \beta_{10}t + \varepsilon_t$, *then* $\hat{\alpha}$ and $\hat{\beta}$ will be different if we take log to base 10 or log to base e, simply because $logY_t = \log e \ln Y_t$, and hence

$logY_t = \alpha_{10} + \beta_{10}t + \varepsilon_t$ is equivalent to $\ln Y_t = \frac{\alpha_{10}}{\log e} + \frac{\beta_{10}t}{\log e} + \frac{\varepsilon_t}{\log e}$. To see this, let $\overline{t} = \frac{1}{n}\sum_{t=1}^{n} t$

$\hat{\beta}_{10} = \frac{\sum_{t=1}^{n}(t-\overline{t})\log y_i}{\sum_{t=1}^{n}(t-\overline{t})^2} = \frac{\log e\sum_{t=1}^{n}(t-\overline{t})\ln y_i}{\sum_{t=1}^{n}(t-\overline{t})^2} = \log e\hat{\beta}_e$ and

$\hat{\alpha}_{10} = \overline{\log y} - \hat{\beta}_{10}\overline{t} = \log e\ \overline{\ln y} - \log e\ \hat{\beta}_e\ \overline{t} = \log e(\ \overline{\ln y} - \hat{\beta}_e\ \overline{t}) = \log e\hat{\alpha}_e.$

2)

The statement is false. Here is a counterexample: let the joint density of (X,Y) be

$$g(x, y) = 2zf(x)f(y),$$

where f is univariate standard normal pdf and z is a function of x and y taking value 1 if $xy > 0$ and taking value 0 if $xy \leq 0$. *Clearly*, the support of (X,Y) are the northeast and southwest quadrants (i.e., both x and y are positive or both x and y are negative), so (X,Y) is not bivariate normal (since the support of a bivariate normal is R^2).

The marginal density (pdf) of X is $\int_{-\infty}^{+\infty} g(x,y)dy = \int_{-\infty}^{+\infty} 2zf(x)f(y)dy = 2f(x)\int_{-\infty}^{+\infty} zf(y)dy.$

Now, if $x > 0$, then $\int_{-\infty}^{+\infty} zf(y)dy = \int_{0}^{+\infty} f(y)dy = \frac{1}{2}$.

And if $x \leq 0, then\ \int_{-\infty}^{+\infty} zf(y)dy = \int_{-\infty}^{0} f(y)dy = \frac{1}{2}$.

Therefore, the marginal pdf of X if f(x). Similarly for Y.

This exercise shows you that if (X,Y) is a bivariate normal, that is stronger than just saying that the univariate distribution of X and of Y is normal.

3)

a) Yes. The model is linear, $E(\varepsilon_i) = 0$ for all i, X is full rank (has rank one) and $Var(\varepsilon_i) = 1$ for all i (and the errors are uncorrelated since they are independent random variables).

b) The OLS estimator for β is

$\hat{\beta} = \frac{\sum_{i=1}^{2} x_i y_i}{\sum_{i=1}^{2}} = \beta + \frac{\sum_{i=1}^{2} x_i \varepsilon_i}{\sum_{i=1}^{2}} = 1 + \frac{\varepsilon_1 + 2\varepsilon_2}{5}$.

This comes from minimizing the sum of squares residuals. Note that we do not demeaned x and y in the formula for $\hat{\beta}$ as in the case where there is an intercept in the model.

So the exact distribution of $\hat{\beta}$ is given by its probability mass function (pmf), which is: $\frac{1}{4}$ if $\hat{\beta} = \frac{2}{5}, \frac{4}{5}, \frac{6}{5}$ *or* $\frac{8}{5}$ and 0 otherwise.

c) $\beta^* = \frac{\sum y}{\sum x} = \beta + \frac{\varepsilon_1 + \varepsilon_2}{3}$. It is unbiased, and its pmf is : $\frac{1}{2}$ if $\beta^* = 1, \frac{1}{4}$ if $\beta^* = \frac{1}{3}$ or $\frac{5}{3}$ and 0 otherwise.

d) Var ($\hat{\beta}$) $= \frac{1}{5}$ and Var (β^*) $= \frac{2}{9}$. So Var (β^*) >Var ($\hat{\beta}$).

4)

(a) Recall that

$$\widehat{\beta} = \frac{\sum_i (y_i - \overline{y})(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}$$

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$$

The information given in the question is not directly usable. However,

$$\sum_i (y_i - \overline{y})(x_i - \overline{x}) = \sum_i (x_i - \overline{x}) y_i = \sum_i x_i y_i - \overline{x} \sum_i y_i = \sum_i x_i y_i - n\overline{xy}$$
$$= 4430 - 22 \times \frac{220}{22} \times \frac{440}{22} = 30$$
$$\sum_i (x_i - \overline{x})^2 = \sum_i (x_i - \overline{x})(x_i - \overline{x}) = \sum_i (x_i - \overline{x}) x_i = \sum_i x_i^2 - n\overline{x}^2$$
$$= 2260 - 22 \times \left(\frac{220}{22}\right)^2 = 60$$

Hence,

$$\widehat{\beta} = \frac{30}{60} = 0.5$$
$$\widehat{\alpha} = \frac{440}{22} - 0.5 \times \frac{220}{22} = 15$$

(b) R^2 is defined as the ratio of the explained sum of squares(ESS) to total sum of squares(TSS).

$$R^2 = \frac{\widehat{\beta}^2 \sum_i (x_i - \overline{x})^2}{\sum_i (y_i - \overline{y})^2} = \widehat{\beta}^2 \frac{\sum_i x_i^2 - n\overline{x}^2}{\sum_i y_i^2 - n\overline{y}^2} = 0.5^2 \times \frac{60}{8900 - 22 \times 20^2} = 0.15$$

(c) By the normality assumption, we know that

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_i (x_i - \overline{x})^2}\right)$$

Moreover,

$$\frac{(n-2)\, s^2}{\sigma^2} \sim \chi^2 (n-2)$$

where $s^2 = \frac{1}{(n-2)} \sum_i e_i^2$. We can also show that $\widehat{\beta}$ and s^2 are independent each other. Then,

$$\frac{\frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_i (x_i - \overline{x})^2}}}}{\sqrt{\frac{\frac{(n-2)s^2}{\sigma^2}}{(n-2)}}} = \frac{\widehat{\beta} - \beta}{\frac{s}{\sqrt{\sum_i (x_i - \overline{x})^2}}} = \frac{\widehat{\beta} - \beta}{\widehat{\sigma}_\beta} \sim t\,(n-2)$$

where $\widehat{\sigma}_\beta = \sqrt{\frac{s^2}{\sum_i (x_i - \overline{x})^2}}$. We want to reject the null hypothesis if

$$T = \left| \frac{\widehat{\beta} - \beta}{\widehat{\sigma}_\beta} \right| > t_{0.975}\,(20)$$

under the null hypothesis. On the other hand,

$$
\begin{aligned}
s^2 &= \frac{1}{(n-2)} \sum_i e_i^2 = \frac{1}{(n-2)} \sum_i \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i \right)^2 \\
&= \frac{1}{(n-2)} \sum_i \left(y_i^2 + \widehat{\alpha}^2 + \widehat{\beta}^2 x_i^2 - 2\widehat{\alpha} y_i + 2\widehat{\alpha}\widehat{\beta} x_i - 2\widehat{\beta} x_i y_i \right) \\
&= \frac{1}{20} \left[\begin{array}{c} 8900 + 22 \times 15^2 + 0.5^2 \times 2260 - 2 \times 15 \times 440 \\ +2 \times 15 \times 0.5 \times 220 - 2 \times 0.5 \times 4430 \end{array} \right] \\
&= 4.25
\end{aligned}
$$

Hence, the test statistic is given by

$$
T = \left| \frac{0.5 - 0}{\sqrt{\frac{4.25}{60}}} \right| = 1.8787
$$

Since $t_{0.975}(20) = 2.086$, we do not reject the null hypothesis.

1. (d) The distribution of $\widehat{\alpha}$ is given by

$$
\widehat{\alpha} \sim N\left(\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_i (x_i - \overline{x})^2} \right] \right)
$$

Therefore, $\widehat{\tau} \equiv \widehat{\alpha} - \widehat{\beta}$ is distributed as

$$
\widehat{\tau} \sim N\left(\tau, \sigma_\tau^2 \right)
$$

where $\tau = \alpha - \beta$. By Gauss-Markov theorem $\widehat{\tau}$ is the BLUE of τ. The variance of $\widehat{\tau}$ is given by

$$
\begin{aligned}
\sigma_\tau^2 &= Var(\widehat{\tau}) = Var\left(\widehat{\alpha} - \widehat{\beta} \right) = Var(\widehat{\alpha}) + Var\left(\widehat{\beta} \right) - 2Cov\left(\widehat{\alpha}, \widehat{\beta} \right) \\
&= \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_i (x_i - \overline{x})^2} \right] + \sigma^2 \frac{1}{\sum_i (x_i - \overline{x})^2} + \sigma^2 \frac{2\overline{x}}{\sum_i (x_i - \overline{x})^2} \\
&= \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2 + 1 + 2\overline{x}}{\sum_i (x_i - \overline{x})^2} \right]
\end{aligned}
$$

We can estimate the variance of $\widehat{\tau}$ as

$$
\widehat{\sigma}_\tau^2 = s^2 \left[\frac{1}{n} + \frac{\overline{x}^2 + 1 + 2\overline{x}}{\sum_i (x_i - \overline{x})^2} \right]
$$

where $s^2 = \frac{1}{n-2} \sum_i e_i^2$.

What do we know? We know that

$$
\frac{\widehat{\tau} - \tau}{\sigma_\tau} \sim N(0, 1)
$$

and

$$\frac{(n-2)\, s^2}{\sigma^2} \sim \chi^2\,(n-2)$$

and $\widehat{\tau}$ and s^2 are independent. Then,

$$\frac{\frac{\widehat{\tau}-\tau}{\sigma_\tau}}{\sqrt{\frac{(n-2)s^2}{\sigma^2}/\,(n-2)}} = \frac{\widehat{\tau}-\tau}{\widehat{\sigma}_\tau} \sim t\,(n-2)$$

We want to reject the null hypothesis if

$$T = \left|\frac{\widehat{\tau}-\tau}{\widehat{\sigma}_\tau}\right| > t_{0.975}\,(n-2)$$

under the null hypothesis. Note that $\widehat{\tau} = 15 - 0.5 = 14.5$ and $\tau = 10$ under the null. Moreover,

$$\widehat{\sigma}_\tau^2 = s^2\left[\frac{1}{n} + \frac{\overline{x}^2 + 1 + 2\overline{x}}{\sum_i (x_i - \overline{x})^2}\right] = 4.25\left[\frac{1}{22} + \frac{10^2 + 1 + 2\times 10}{60}\right] = 8.764$$

The test statistic is now

$$T = \left|\frac{14.5 - 10}{\sqrt{8.764}}\right| = 1.5201$$

Since $t_{0.975}\,(20) = 1.725$, again, we do not reject the null hypothesis.

5)

1. You can write

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = \frac{1}{n}\sum_i y_i - \overline{x}\frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} = \frac{1}{n}\sum_i y_i - \overline{x}\frac{\sum_i (x_i - \overline{x})\, y_i}{\sum_i (x_i - \overline{x})^2}$$
$$= \sum_i \left[\frac{1}{n} - \overline{x}\frac{(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right] y_i = \sum_i m_i y_i$$

where $m_i = \left[\frac{1}{n} - \overline{x}\frac{(x_i-\overline{x})}{\sum_i (x_i-\overline{x})^2}\right] = \left[\frac{1}{n} - \overline{x}w_i\right]$ with $w_i = \frac{(x_i-\overline{x})}{\sum_i (x_i-\overline{x})^2}$. Then, $Var\,(\widehat{\alpha}) = \sigma^2 \sum_i m_i^2$. Now, consider an alternative linear estimator such that

$$\widetilde{\alpha} = \sum_i h_i y_i = \sum_i h_i\,(\alpha + \beta x_i + \varepsilon_i) = \alpha\sum_i h_i + \beta\sum_i h_i x_i + \sum_i h_i \varepsilon_i$$

Then,

$$E\,(\widetilde{\alpha}) = \alpha\sum_i h_i + \beta\sum_i h_i x_i$$

Therefore, unbiasedness requires that $\sum_i h_i = 1$ and $\sum_i h_i x_i = 0$. Introduce a new expression for h_i;

$$h_i = m_i + g_i$$

We can always do this! - g_i may be negative-. Now,

$$\begin{aligned}
Var\left(\widetilde{\alpha}\right) &= E\left[\left(\sum_i h_i \varepsilon_i\right)^2\right] = \sum_i h_i^2 E\left(\varepsilon_i^2\right) = \sigma^2 \sum_i h_i^2 \\
&= \sigma^2 \sum_i \left(m_i + g_i\right)^2 = \sigma^2 \sum_i m_i^2 + \sigma^2 \sum_i g_i^2 + 2\sigma^2 \sum_i m_i g_i \\
&= \sigma^2 \sum_i m_i^2 + \sigma^2 \sum_i g_i^2 \geq \sigma^2 \sum_i m_i^2 = Var\left(\widehat{\alpha}\right)
\end{aligned}$$

since

$$\begin{aligned}
\sum_i m_i g_i &= \sum_i m_i \left(h_i - m_i\right) = \sum_i m_i h_i - \sum_i m_i^2 = \sum_i \left[\frac{1}{n} - \overline{x} w_i\right] h_i - \sum_i \left[\frac{1}{n} - \overline{x} w_i\right]^2 \\
&= \frac{1}{n} \sum_i h_i - \overline{x} \sum_i w_i h_i - \sum_i \left(\frac{1}{n}\right)^2 + \frac{2\overline{x}}{n} \sum_i w_i - \overline{x}^2 \sum_i w_i^2 \\
&= \frac{1}{n} - \overline{x} \sum_i \left(\frac{\left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2}\right) h_i - \frac{1}{n} - \overline{x}^2 \frac{1}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \frac{1}{n} - \overline{x} \sum_i \frac{x_i h_i}{\sum_i \left(x_i - \overline{x}\right)^2} + \overline{x}^2 \sum_i \frac{h_i}{\sum_i \left(x_i - \overline{x}\right)^2} - \frac{1}{n} - \overline{x}^2 \frac{1}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \frac{1}{n} + \overline{x}^2 \frac{1}{\sum_i \left(x_i - \overline{x}\right)^2} - \frac{1}{n} - \overline{x}^2 \frac{1}{\sum_i \left(x_i - \overline{x}\right)^2} = 0
\end{aligned}$$

The third row follows from $\sum_i h_i = 1$ and $\sum_i w_i^2 = 1$. The last row follows from $\sum_i x_i h_i = 0$ and $\sum_i h_i = 1$.