Assignment Eight

① (a) $y_0 \sim N(x_0'\beta, \sigma^2)$ & $x_0'b \sim N(x_0'\beta, \sigma^2 x_0'(X'X)^{-1}x_0)$
are independent because b is a function of
$y_1, \ldots, y_n$, which are independent of y_0

So $y_0 - x_0'b \sim N(0, \sigma^2 + \sigma^2 x_0'(X'X)^{-1}x_0)$.

Standardizing,

$$z = \frac{y_0 - x_0'b}{\sqrt{\sigma^2(1 + x_0'(X'X)^{-1}x_0)}} \sim N(0,1)$$

And since $w = \frac{e'e}{\sigma^2} \sim \chi^2(n-k-1)$,

$$t = \frac{z}{\sqrt{w/\nu}} = \frac{\frac{y_0 - x_0'b}{\sigma\sqrt{1 + x_0'(X'X)^{-1}x_0}}}{\sqrt{\frac{e'e}{\sigma^2}/(n-k-1)}} \sim t(n-k-1)$$

Numerator & denominator are independent because
- e is a function of $y_1, \ldots y_n$, independent of y_0
- e is independent of b
- Functions of independent random vectors are ind.

$$t = \frac{\frac{y_0 - x_0'b}{\sigma\sqrt{1 + x_0(X'X)^{-1}x_0}}}{\frac{1}{\sigma}\sqrt{s^2}} = \frac{y_0 - x_0'b}{s\sqrt{1 + x_0'(X'X)^{-1}x_0}}$$

which is the expression on the formula sheet

(1b) $1-\alpha = P\{-t_{\alpha/2} < t < t_{\alpha/2}\}$

$$= P\left\{-t_{\alpha/2} < \frac{y_0 - x_0'b}{s\sqrt{1+x_0'(X'X)^{-1}x_0}} < t_{\alpha/2}\right\}$$

$$= P\{-t_{\alpha/2}\, s\sqrt{1+x_0'(X'X)^{-1}x_0} < y_0 - x_0'b$$
$$< t_{\alpha/2}\, s\sqrt{1+x_0'(X'X)^{-1}x_0}\}$$

$$= P\{x_0'b - t_{\alpha/2}\, s\sqrt{x_0(X'X)^{-1}x_0} < y_0$$
$$< x_0'b + t_{\alpha/2}\, s\sqrt{x_0'(X'X)^{-1}x_0}\}$$

or, $x_0'b \pm t_{\alpha/2}\, s\sqrt{x_0'(X'X)^{-1}x_0}$

(2) (a) $w \sim N\left(\sum_{j=n+1}^{n+m} x_j'\beta,\ m\sigma^2\right)$

(b) $\hat{w} = \sum_{j=n+1}^{n+m} \hat{y}_j = \sum_{j=n+1}^{n+m} x_j' b = v'b,$

where $v = \sum_{j=n+1}^{n+m} x_j$ This notation will save a lot of writing.

(c) $\hat{w} \sim N(v'\beta,\ \sigma^2 v'(X'X)^{-1}v)$

(d) w & $\hat{w}$ are independent since w is based on a new set of independent observations.

$E(w) = v'\beta$, so

$w - \hat{w} \sim N(0,\ \sigma^2(m + v'(X'X)^{-1}v))$

(e)

$$z = \frac{w - \hat{w}}{\sigma\sqrt{m + v'(X'X)^{-1}v}} \sim N(0,1)$$

(f)

$$t = \frac{\dfrac{w-\hat{w}}{\cancel{\sigma}\sqrt{m+v'(X'X)^{-1}v}}}{\sqrt{\dfrac{e'e}{\cancel{\sigma^2}}/(n-k-1)}} = \frac{w-\hat{w}}{s\sqrt{m+v'(X'X)^{-1}v}}$$

(g) Numerator & denominator are independent because $\hat{w}$ is a function of b, independent of e, and w is a function of $y_{n+1}, \ldots, y_{n+m}$. The denominator is a function of $y_1, \ldots, y_n$ through s, & all the y_j are independent.

(2h) $1-\alpha = P\left\{ -t_{\alpha/2} < t < t_{\alpha/2} \right\}$

$$= P\left\{ -t_{\alpha/2} < \frac{w-\hat{w}}{s\sqrt{m + v'(X'X)^{-1}v}} < t_{\alpha/2} \right\}$$

$$= P\left\{ -t_{\alpha/2}\, s\sqrt{m + v'(X'X)^{-1}v} < w - \hat{w} < t_{\alpha/2}\, s\sqrt{m + v'(X'X)^{-1}v} \right\}$$

$$= P\left\{ \hat{w} - t_{\alpha/2}\, s\sqrt{m + v'(X'X)^{-1}v} < w < \hat{w} + t_{\alpha/2}\, s\sqrt{m + v'(X'X)^{-1}v} \right\}$$

Where again, $w = \sum_{j=n+1}^{n+m} y_j$

$$\hat{w} = \sum_{j=n+1}^{n+m} x_j'\, b \quad , \text{ and}$$

$$v = \sum_{j=n+1}^{n+m} x_j$$

(3) Computer

(4) Simple random sampling is regression with $k=0$ independent variables. The X matrix is a column of ones, and $(X'X)^{-1} = \frac{1}{n}$. $b = \bar{y}$, $x_0 = 1$, and from the formula sheet, the prediction interval is

$$x_0' b \pm t_{\alpha/2}\, s \sqrt{1 + x_0'(X'X)^{-1} x_0}$$

$$= \bar{y} \pm t_{\alpha/2}\, s\sqrt{1 + 1/n}, \text{ where}$$

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

Deriving it, use $t = \dfrac{z}{\sqrt{w/\nu}}$

with $w = \dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

$\bar{y} \sim N(\mu, \frac{\sigma^2}{n})$, $y_0 \sim N(\mu, \sigma^2)$

Independent, so $y_0 - \bar{y} \sim N(0, \sigma^2 + \frac{\sigma^2}{n})$,

and $z = \dfrac{y_0 - \bar{y}}{\sqrt{\sigma^2(1+\frac{1}{n})}}$, so

$$t = \frac{\dfrac{y_0 - \bar{y}}{\cancel{\sigma}\sqrt{1+1/n}}}{\sqrt{\dfrac{\cancel{(n-1)}s^2}{\cancel{\sigma^2}} \Big/ \cancel{(n-1)}}} = \frac{y_0 - \bar{y}}{s\sqrt{1+1/n}}$$

$\sim t(n-1)$

4 continued

Then, $1-\alpha = P\left\{-t_{\alpha/2} < \dfrac{y_0 - \bar{y}}{s\sqrt{1+1/n}} < t_{\alpha/2}\right\}$

$= P\{\bar{y} - t_{\alpha/2}\, s\sqrt{1+1/n} < y_0$

$< \bar{y} + t_{\alpha/2}\, s\sqrt{1+1/n}\}$, or

$$\bar{y} \pm t_{\alpha/2}\, s\sqrt{1+1/n},$$

which is what we got from the formula sheet.

For $\bar{y} = 7.5$, $s^2 = 3.82$ and $n = 14$, R gives a critical value of $qt(0.975, 13) = 2.16$

The 95% margin of error is

$t_{\alpha/2}\sqrt{s^2(1+1/n)} = 2.16\sqrt{3.82(1+1/14)}$

$= 4.37$, and the prediction interval is

$(7.5 - 4.37,\ 7.5 + 4.37)$

$= (3.13,\ 11.87)$

⑤ Since there is an intercept, $\sum_{i=1}^{n} e_i = 0 \Rightarrow \bar{y} = \bar{\hat{y}}$

$$r^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}$$

$$= \frac{\left(\sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$$

$$= \frac{\left(\sum_{i=1}^{n}(e_i + \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{SST \cdot SSR}$$

$$= \left(\sum_{i=1}^{n}\left(e_i(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2\right)\right)^2 / (SST \cdot SSR)$$

$$= \left(\sum_{i=1}^{n}\hat{y}_i e_i - \bar{y}\underbrace{\sum_{i=1}^{n} e_i}_{0} + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2\right)^2 / (SST \cdot SSR)$$

$$= \left((Xb)'e - 0 + SSR\right)^2 / (SST \cdot SSR$$

$$= \frac{SSR^2}{SST \cdot SSR} = \frac{SSR}{SST} = R^2$$

(6) Numerator of correlation coefficient is

$$\sum_{i=1}^{n} (e_i - \overset{\swarrow 0}{\bar{e}})(x_{ij} - \bar{x}_j) = \sum_{i=1}^{n} e_i x_{ij} - \bar{x}_j \sum_{i=1}^{n} e_i$$

$$= \underbrace{\sum_{i=1}^{n} x_{ij} e_i} + 0 = 0$$

Because this is the inner product of a column of X and e, , and $X'e = 0$

(7) (a) Using $\mathrm{cov}(Ay, By) = A\,\mathrm{cov}(y)\,B'$,

$$\mathrm{cov}(\hat{y}, e) = \mathrm{cov}(Hy, (I-H)y)$$

$$= H\sigma^2 I (I-H)' = \sigma^2 H (I-H)$$

$$= \sigma^2 (H - H^2) = \sigma^2 (H - H) = 0_{n \times n}$$

Since zero covariance implies independence for the multivariate normal, $\hat{y}$ & e are independent.

(7b) Numerator of correlation coefficient is

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})(e_i - \bar{e}) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}) e_i$$

$$= \sum_{i=1}^{n} \hat{y}_i e_i - \bar{y} \underbrace{\sum_{i=1}^{n} e_i}_{0} = \sum_{i=1}^{n} \hat{y}_i e_i$$

$$= (Xb)' e = b' \underbrace{X' e}_{0} = b' 0 = 0_{1\times 1}$$

(8) Computer