1.

Let Y_1, Y_2, Y_3 be a random sample from $N(\beta, \sigma^2)$ population. An alternative to the least squares estimator is the weighted estimator, which weighs the observations unequally,

$$b_w = \frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3$$

(a)

Is b_w a linear estimator?

Yes, Y_1, Y_2, and Y_3 enter the function in a linear manner.

(b)

Show that b_w is an unbiased estimator.

For b_w to be an unbiased estimator, the expected value of b_w has to equal to β.

$$\begin{aligned}
b_w &= \frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3 && (1)\\
E(b_w) &= E(\frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3) && (2)\\
&= E(\frac{1}{2}Y_1) + E(\frac{1}{3}Y_2) + E(\frac{1}{6}Y_3) && (3)\\
&= \frac{1}{2}E(Y_1) + \frac{1}{3}E(Y_2) + \frac{1}{6}E(Y_3) && (4)\\
&= \frac{1}{2}\frac{\sum Y_{1t}}{T} + \frac{1}{3}\frac{\sum Y_{2t}}{T} + \frac{1}{6}\frac{\sum Y_{3t}}{T} && (5)\\
&= \frac{1}{2}\beta + \frac{1}{3}\beta + \frac{1}{6}\beta && (6)\\
&= \frac{3}{6}\beta + \frac{2}{6}\beta + \frac{1}{6}\beta && (7)\\
E(b_w) &= \beta && (8)
\end{aligned}$$

(1) Given equation
(2) Find the expected value of b_w, take the expected values of both sides
(3) Since, Y_1, Y_2, and Y_3 are independent, their expected values can be summed separately
(4) Constants can be taken outside of the expected values, since the expected value of a constant is the constant
(5) The expected value of each sample is just the mean of each sample
(6) The mean of each sample is given: $N(\beta, \sigma^2)$, which would be β
(7) Create a common denominator in order to sum up
(8) $1 * \beta = \beta$, so we see the expected value of b_w is indeed β

(c)

Find the variance of b_w.

Since Y_1, Y_2, and Y_3 are independent, we do not need to calculate each pair of covariances, so the general equation for a weighted variance is simply:

$$var(c_1X_1 + c_2X_2 + c_3X_3) = c_1^2 var(X_1) + c_2^2 var(X_2) + c_3^2 var(X_3)$$

$$var(\frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3) = \frac{1}{2}\sigma^2 + \frac{1}{3}\sigma^2 + \frac{1}{6}\sigma^2 \quad (1)$$
$$= \frac{1}{4}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 \quad (2)$$
$$= \frac{9}{36}\sigma^2 + \frac{4}{36}\sigma^2 + \frac{1}{36}\sigma^2 \quad (3)$$
$$= \frac{14}{36}\sigma^2 \quad (4)$$
$$= \frac{7}{18}\sigma^2 \quad (5)$$

(d)

What is the variance of the least squares estimator b of β?

$$b = \sum_{i=1}^{T} \frac{Y_i}{T} \quad (1)$$
$$var(b) = var(\sum_{i=1}^{T} \frac{Y_i}{T}) \quad (2)$$
$$= var(\frac{1}{T}Y_1 + \frac{1}{T}Y_2 + \frac{1}{T}Y_3) \quad (3)$$
$$= (\frac{1}{T})^2 var(Y_1) + (\frac{1}{T})^2 var(Y_2) + (\frac{1}{T})^2 var(Y_3) \quad (4)$$
$$= \frac{1}{T^2}\sigma^2 + \frac{1}{T^2}\sigma^2 + \frac{1}{T^2}\sigma^2 \quad (5)$$
$$= \frac{3}{T^2}\sigma^2 \quad (6)$$
$$= \frac{3}{3^2}\sigma^2 \quad (7)$$
$$= \frac{1}{3}\sigma^2 \quad (8)$$

(e)

Is b_w as good of an estimator as b?

Although b_w is unbiased ($E(b_w = \beta)$), b_w variance is still greater than $var(b)$:

$$var(b) < var(b_w) \rightarrow \frac{1}{3}\sigma^2 < \frac{7}{18}\sigma^2.$$

As a result, b_w is still *not* as good of an estimator as b.

(f)

If $\sigma^2 = 9$, calculate the probability that each estimator is within 1 unit (on either side) of β.

The probability of an estimator between a lower and upper limit can be written generally as:

$$P(-z_c \leq z \leq z_c)$$

Since we want the probability that each estimator is within 1 unit on either side of β, we can find it for b first and rewrite the general formula as:

$$P(-1 \leq b - \beta \leq 1) \quad (1)$$
$$P(\beta - 1 \leq b \leq \beta + 1) \quad (2)$$

For the **lower** limit:

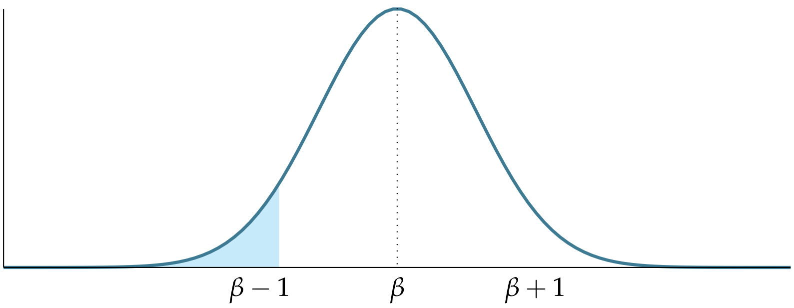
$$
\begin{aligned}
z_c &= \frac{b-\beta}{\sqrt{var(b)}} && (1)\\
&= \frac{(\beta-1)-\beta}{\sqrt{var(b)}} && (2)\\
&= \frac{-1}{\sqrt{\frac{1}{3}\sigma^2}} && (3)\\
&= \frac{-1}{\sqrt{\frac{1}{3}(9)}} && (4)\\
&= \frac{-1}{\sqrt{3}} && (5)\\
&= -0.577 && (6)
\end{aligned}
$$

For the **upper** limit:

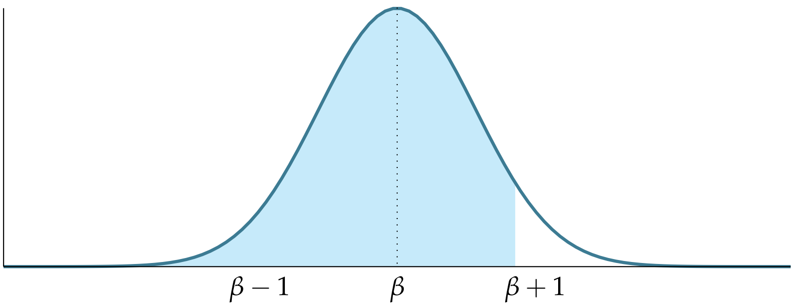
$$
\begin{aligned}
z_c &= \frac{b-\beta}{\sqrt{var(b)}} && (1)\\
&= \frac{(\beta+1)-\beta}{\sqrt{var(b)}} && (2)\\
&= \frac{1}{\sqrt{\frac{1}{3}\sigma^2}} && (3)\\
&= \frac{1}{\sqrt{\frac{1}{3}(9)}} && (4)\\
&= \frac{1}{\sqrt{3}} && (5)\\
&= 0.577 && (6)
\end{aligned}
$$

Graphically, we can represent the probabilities of our lower and upper limit as such, with the probability of b being in the region shaded cyan:

$P((\beta-1) \leq b)$

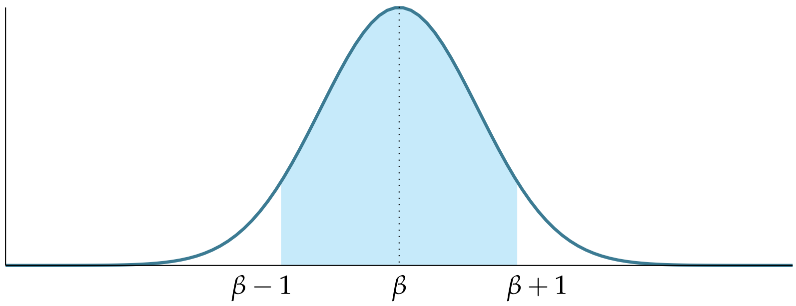


$P(b \leq (\beta+1))$



Thus, for the probability our estimator b is within 1 unit on either side of β would be the probability of the upper limit minus the probability of the lower limit.

$$P(b \leq (\beta+1)) - P((\beta-1) \leq b)$$



Looking up the probabilities from the z-table, we get:

$$\begin{aligned}
P(\beta - 1 \leq b \leq \beta + 1) &= P(b \leq (\beta+1)) - P((\beta-1) \leq b) && (1)\\
&= P(z_c = -0.577) - P(z_c = 0.577) && (2)\\
&= 0.719 - 0.281 && (3)\\
&= 0.438 && (4)
\end{aligned}$$

Now, for the probability of b_w being within 1 unit on either side of β, we can do the same following steps:

$$\begin{aligned}
&P(-1 \leq b_w - \beta \leq 1) && (1)\\
&P(\beta - 1 \leq b_w \leq \beta + 1) && (2)
\end{aligned}$$

For the **lower** limit:

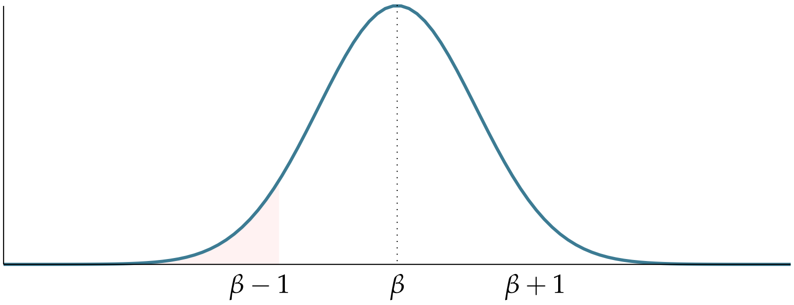
$$\begin{aligned}
z_c &= \frac{b_w - \beta}{\sqrt{var(b_w)}} && (1)\\
&= \frac{(\beta - 1) - \beta}{\sqrt{var(b_w)}} && (2)\\
&= \frac{-1}{\sqrt{\frac{7}{18}\sigma^2}} && (3)\\
&= \frac{-1}{\sqrt{\frac{7}{18}(9)}} && (4)\\
&= \frac{-1}{\sqrt{\frac{7}{2}}} && (5)\\
&= -0.534 && (6)
\end{aligned}$$

For the **upper** limit:

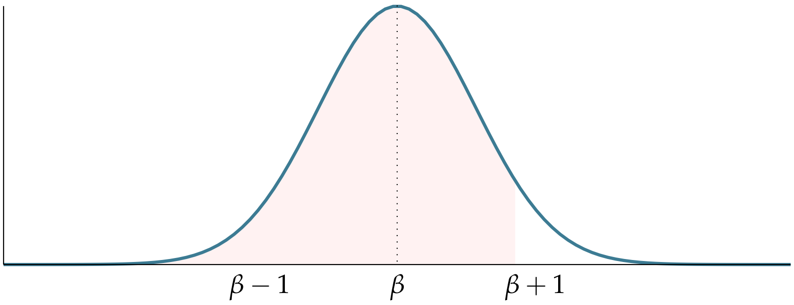
$$\begin{aligned}
z_c &= \frac{b_w - \beta}{\sqrt{var(b_w)}} && (1)\\
&= \frac{(\beta - 1) - \beta}{\sqrt{var(b_w)}} && (2)\\
&= \frac{1}{\sqrt{\frac{7}{18}\sigma^2}} && (3)\\
&= \frac{1}{\sqrt{\frac{7}{18}(9)}} && (4)\\
&= \frac{1}{\sqrt{\frac{7}{2}}} && (5)\\
&= 0.534 && (6)
\end{aligned}$$

Graphically, we can represent the probabilities of our lower and upper limit as such, with the probability of b_w being in the region shaded pink:

$P((\beta - 1) \leq b_w)$

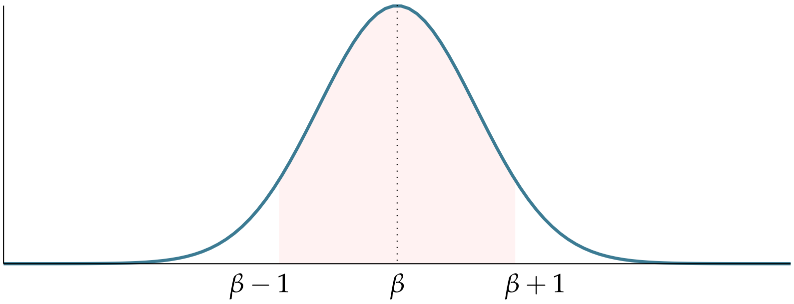


$P(b_w \leq (\beta + 1))$



Thus, for the probability our estimator b_w is within 1 unit on either side of β would be the probability of the upper limit minus the probability of the lower limit.

$P(b \leq (\beta + 1)) - P((\beta - 1) \leq b_w)$



Looking up the probabilities from the z-table, we get:

$$
\begin{aligned}
P(\beta - 1 \leq b_w \leq \beta + 1) &= P(b_w \leq (\beta + 1)) - P((\beta - 1) \leq b_w) && (1)\\
&= P(z_c = -0.534) - P(z_c = 0.534) && (2)\\
&= 0.7019 - 0.2981 && (3)\\
&= 0.4038 && (4)
\end{aligned}
$$

Thus, we see our less than optimal estimator b_w has a smaller probability compared to b of being within 1 unit on either side of β, which makes sense since $var(b_w) > var(b)$.

2.

Consider the statistical model $y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$ and

$$X = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} 5 \\ 2 \\ 3 \\ 2 \\ -2 \end{bmatrix}$$

(a)

Find $X'X$, $X'Y$ and $(X'X)^{-1}$.

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & -1 & 0 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \quad (1)$$

$$= \begin{bmatrix} (1+1+1+1+1) & 3+2+1-1+0 \\ 3+2+1-1+0 & (3)(3)+(2)(2)+(1)(1)+(-1)(-1)+0 \end{bmatrix} \quad (2)$$

$$X'X = \begin{bmatrix} 5 & 5 \\ 5 & 15 \end{bmatrix} \quad (3)$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \\ 2 \\ -2 \end{bmatrix} \quad (1)$$

$$= \begin{bmatrix} 5+2+3+2-2 \\ (3)(5)+(2)(2)+(1)(3)+(-1)(2)+0 \end{bmatrix} \quad (2)$$

$$X'Y = \begin{bmatrix} 10 \\ 20 \end{bmatrix} \quad (3)$$

$$(X'X)^{-1} = \frac{1}{|X'X|}\text{adj}(X'X)$$

Since we already found

$$X'X = \begin{bmatrix} 5 & 5 \\ 5 & 15 \end{bmatrix}$$

Then $|X'X| = 5(15) - 5(5) = 50$

And

$$\text{adj } X'X = \begin{bmatrix} 15 & -5 \\ -5 & 5 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{|X'X|}\text{adj}(X'X) \quad (1)$$
$$= \frac{1}{50}\begin{bmatrix} 15 & -5 \\ -5 & 5 \end{bmatrix} \quad (2)$$
$$(X'X)^{-1} = \begin{bmatrix} \frac{3}{10} & \frac{-1}{10} \\ \frac{-1}{10} & \frac{1}{10} \end{bmatrix} \quad (3)$$

(b)

Find the least squares estimate $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$.

$$\beta = (X'X)^{-1}X'Y \quad (1)$$
$$= \begin{bmatrix} \frac{3}{10} & \frac{-1}{10} \\ \frac{-1}{10} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \\ 2 \\ -2 \end{bmatrix} \quad (2)$$
$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (3)$$

(c)

Give an interpretation of the parameter estimate of the explanatory variable.

We expect the mean of Y to increase by 1 whenever X increases by 1, ceteris paribus. The y-intercept is equal to 1 when X is 0.

(d)

Find the variance covariance of the least squares estimator.

To find the variance:

$$var(\beta) = \sigma^2(X'X)^{-1}$$

Solving for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{Y'Y - b'X'Y}{T-K} \quad (1)$$
$$= \frac{\begin{bmatrix} 5 & 2 & 3 & 2 & -2 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \\ 2 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \end{bmatrix}}{5-2} \quad (2)$$
$$= \frac{\begin{bmatrix} 25+4+9+4+4 \end{bmatrix} - \begin{bmatrix} 10+20 \end{bmatrix}}{3} \quad (3)$$
$$= \frac{46-30}{3} \quad (4)$$
$$= \frac{16}{3} \quad (5)$$

Plugging this back to the equation for $var(\beta)$, we get:

$$\begin{aligned} var(\beta) &= \frac{16}{3}\begin{bmatrix} \frac{3}{10} & \frac{-1}{10} \\ \frac{-1}{10} & \frac{1}{10} \end{bmatrix} \\ &= \begin{bmatrix} \frac{8}{5} & \frac{-8}{15} \\ \frac{-8}{15} & \frac{8}{15} \end{bmatrix} \end{aligned}$$

So we get $var(\beta_1) = \frac{8}{5}$, $var(\beta_2) = \frac{8}{15}$, and $cov(\beta_1, \beta_2) = \frac{-8}{15}$

3.

Suppose an investigator is using least squares principle and get the following system of linear equation from the first order conditions of minimizing the sum of squared errors given by

$$SSE = (y - X\beta)'(y - X\beta)$$

$$\begin{cases} 3\beta_1 + 18\beta_2 = 24 \\ 18\beta_1 + 110\beta_2 = 140 \end{cases}$$

(a)

Provide the statistical model leading to the above normal equations. Be explicit, the answer $y = X\beta + e$, although correct, will not earn you any credit.

Since our statistical matrix model is given as:

$$y = X\beta + e$$

Where

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{21} & ... & x_{K1} \\ 1 & x_{22} & ... & x_{K2} \\ ... & ... & ... & ... \\ 1 & x_{2T} & ... & x_{KT} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, e = \begin{bmatrix} e_1 \\ e_2 \\ ... \\ e_n \end{bmatrix}$$

This can be written as:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + e_t,$$

which is our general statistical model.

Plugging in the β's we will find in part (b), $\beta_1 = 20$ and $\beta_2 = -2$, we get:

$$y_t = 20x_{1t} - 2x_{2t} + e_t$$

Thus, we see $y = x\beta \to \hat{y} = \hat{\beta_1}x_1 + \hat{\beta_2}x_2 \to \hat{y} = 20x_1 - 2x_2$

(b)

Estimate the β's using the least squares rule.

$$\begin{cases} 3\beta_1 + 18\beta_2 = 24 \\ 18\beta_1 + 110\beta_2 = 140 \end{cases}$$

In matrix form:

$$X'X = \begin{bmatrix} 3 & 18 \\ 18 & 110 \end{bmatrix} \tag{1}$$

$$X'Y = \begin{bmatrix} 24 \\ 140 \end{bmatrix} \tag{2}$$

$$(X'X)^{-1} = \frac{1}{(3)(110) - (18)(18)} \begin{bmatrix} 110 & -18 \\ -18 & 3 \end{bmatrix} \tag{3}$$

$$(X'X)^{-1} = \begin{bmatrix} \frac{55}{3} & -3 \\ -3 & \frac{1}{2} \end{bmatrix} \tag{4}$$

$$(X'X)^{-1}X'Y = \begin{bmatrix} \frac{55}{3} & -3 \\ -3 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 24 \\ 140 \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} 440 - 420 \\ -72 + 70 \end{bmatrix} \tag{6}$$

$$\beta = \begin{bmatrix} 20 \\ -2 \end{bmatrix} \tag{7}$$

So $\beta_1 = 20$ and $\beta_2 = -2$.

(c)

Find the variance-covariance matrix of the least squares estimates as a function of σ^2.

$$var(b) = \sigma^2 (X'X)^{-1} \tag{1}$$

$$= \sigma^2 \begin{bmatrix} \frac{55}{3} & -3 \\ -3 & \frac{1}{2} \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} \frac{55}{3}\sigma^2 & -3\sigma^2 \\ -3\sigma^2 & \frac{1}{2}\sigma^2 \end{bmatrix} \tag{3}$$

4. Exercise 5.6, page 197.

Consider the problem of estimating a production function that expresses the relationship between the level of output of a commodity and the level of input of a factor and assume that you have the input data given in the Table 5.3.

(a)

Assume that the data can be described by the statistical model, $y_1 = \beta_1 + x_t\beta_2 + e_t$, where the random $e_t \sim (0, \sigma^2)$. Use the least squares rule to estimate β_1 and β_2.

	X	Y	X^2	XY
	1	0.58	1	0.58
	2	1.1	4	2.2
	3	1.2	9	3.6
	4	1.3	16	5.2
	5	1.95	25	9.75
	6	2.55	36	15.3
	7	2.6	49	18.2
	8	2.9	64	23.2
	9	3.45	81	31.05
	10	3.5	100	35
	11	3.6	121	39.6
	12	4.1	144	49.2
	13	4.35	169	56.55
	14	4.4	196	61.6
	15	4.5	225	67.5
$\sum$	120	42.08	1240	418.53

$$\beta_2 = \frac{T\sum(x_t y_t) - \sum x_t \sum y_t}{T\sum x_t^2 - (\sum x_t)^2} \tag{1}$$

$$= \frac{15(418.53) - (120)(42.08)}{15(1240) - 120^2} \tag{2}$$

$$= \frac{6277.95 - 5049.6}{18600 - 14400} \tag{3}$$

$$= \frac{1228.35}{4200} \tag{4}$$

$$\beta_2 = 0.29246 \tag{5}$$

Plugging β_2 into the equation for β_1, we get:

$$\beta_1 = \bar{y} - \beta_2 \bar{x} \tag{1}$$

$$= \frac{42.08}{15} - 0.2924642857142857(\frac{120}{15}) \tag{2}$$

$$\beta_1 = 0.4656 \tag{3}$$

(b)

Give an economic interpretation of the estimated parameters.

$$\hat{y} = 0.4656 + 0.29246\hat{x} + \hat{e}$$

Every additional unit increase of input X (feed), increases the output of Y (poultry meat) by 0.29246, ceteris paribus. Although when input X is 0, the y-intercept, or β_1, is equal to 0.4656, in this case it would make no economic sense, since chickens cannot produce poultry meat with no feed. Unless, of course, the chickens consume something else in addition to the feed. In this case, it would still make economic sense that with no feed, there would still be some output of poultry meat.

(c)

Use the result of part (a) to plot the production function.

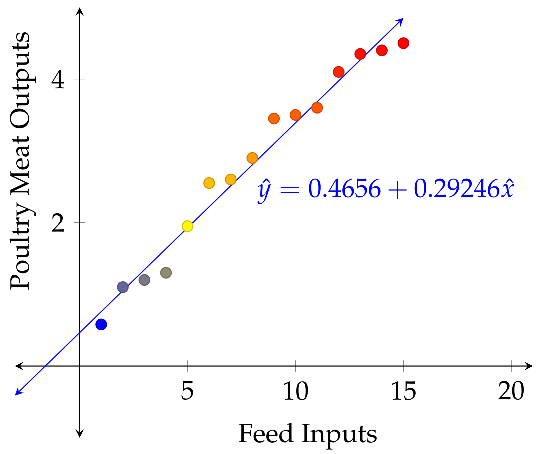


Figure 1: Production function of a sample of observations from a controlled experiment involving poultry meat outputs and feed inputs.

(d)

If the cost of feed is 6 cents per pound, derive the total cost and marginal cost functions.

$$TC = 0.06\hat{x} \quad (1)$$
$$\hat{y} = 0.4656 + 0.29246\hat{x} \quad (2)$$
$$\hat{x} = \frac{\hat{y} - 0.4656}{0.29246} \quad (3)$$
$$TC = 0.06(\frac{\hat{y} - 0.4656}{0.29246}) \quad (4)$$
$$TC = \frac{0.06\hat{y} - 0.027936}{0.29246} \quad (5)$$

For the marginal cost function:

$$MC = \frac{dTC}{dy} \quad (1)$$
$$= \frac{d}{dy}\frac{0.06\hat{y} - 0.027936}{0.29246} \quad (2)$$
$$= \frac{0.06(0.29246) - 0.06\hat{y}(0)}{0.29246^2} \quad (3)$$
$$= 0.205156 \quad (4)$$

(e)

Show how to use the total cost function (which relates total cost to output) to obtain estimates of the parameters of the production function connecting feed inputs and poultry meat outputs.

Given our general equation:

$$\hat{y} = \beta_1 + \beta_2\hat{x}$$

and our total cost function:

$$TC = 0.06\hat{x}$$

We can rearrange our general equation to solve for x_t:

$$\hat{x} = \frac{\hat{y} - \beta_1}{\beta_2}$$

Plugging this back into our total cost function for x_t, we get:

$$TC = 0.06(\frac{\hat{y} - \beta_1}{\beta_2})$$

And solving for each of the β's:

$$\beta_1 = \hat{y} - \frac{TC\beta_2}{0.06}$$
$$\beta_2 = \frac{0.06\hat{y} - 0.06\beta_1}{TC}$$

So we can see, TC and β_1 have an inversely proportional relationship (as TC increases, β_1 decreases, and vice versa, ceteris paribus), while TC and β_2 have a proportional relationship (as TC increases, β_2 increases, and vice versa, ceteris paribus).

Furthermore, taking the TC function:

$$TC = 0.06(\frac{\hat{y} - \beta_1}{\beta_2})$$

We can also rewrite the equation as:

$$TC = \frac{0.06\hat{y}}{\beta_2} - \frac{0.06\beta_1}{\beta 2}$$

If we set $\alpha_1 = \frac{0.06}{\beta_2}$ and $\alpha_2 = \frac{\beta_1}{\beta_2}$, we can rewrite the TC function as:

$$TC = \alpha_1\hat{y} - \alpha_2$$

Therefore, if we know the values of α_1 and α_2, we can estimate the values of β's.

5. Exercise 9.6, pages 316-317.

Data on per capita consumption of beef, the price of beef, the price of lamb, the price of pork, and per capita disposable income for Australia, for the period 1949 to 1965, are given in Table 9.8. All prices and income have been deflated with 1953 as the base year. Consider the log-linear demand curve:

$$\ln qb_t = \beta_1 + \beta_2 \ln pb_t + \beta_3 \ln pl_t + \beta_4 \ln pp_t + \beta_5 \ln y_t + e_t$$

where

qb_t is per capita consumption of beef in year t (pounds),
pb_t is the price of beef in year t (pence per pound),
pl_t is the price of lamb in year t (pence per pound),
pp_t is the price of pork in year t (pence per pound),
y_t is per capita disposable income in year t (Australian currency pounds).

(a)

What signs do you expect on each of the coefficients?

For the price of beef (pb_t), I expect the sign to be negative, because with increased prices, all else equal, people would buy other meat substitutes and consume less beef.

For the price of lamb (pl_t) and the price of pork (pp_t), I expect their signs to be positive. As those prices increase, Australian consumers will seek cheaper meat substitutes whose prices did not rise, or rise as much, thereby increasing the per capita consumption of beef.

For per capita disposable income (y_t), the sign should be positive, assuming beef is a normal good, as disposable income increases, per capita beef consumption should also increase. However, if beef is an inferior good, then the sign should be negative.

(b)

Estimate $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ using least squares. Interpret the results. Do they seem reasonable?

If we take the natural logs of each value, we obtain this table:

Year	ln qb	ln y	ln pb	ln pl	ln pp
1949	4.798267	5.872118	3.245712	2.996232	3.822973
1950	4.822698	5.940171	3.269569	2.93066	3.970292
1951	4.879767	6.054439	3.407511	3.111291	4.058717
1952	4.778283	5.866468	3.500439	3.299903	4.035302
1953	4.784989	5.869297	3.453474	3.02965	4.011868
1954	4.741448	5.888878	3.406185	2.938633	4.019801
1955	4.757891	5.940171	3.472898	2.952825	3.924347
1956	4.779963	5.968708	3.459466	3.00072	3.958143
1957	4.859037	5.940171	3.426215	3.025291	3.988243
1958	4.829113	5.908083	3.468233	2.968361	3.947004
1959	4.767289	5.953243	3.506758	2.884801	3.932022
1960	4.589041	5.998937	3.601686	2.882564	4.000766
1961	4.447346	6.006353	3.681603	2.95178	3.971046
1962	4.53582	6.018593	3.632045	2.892037	3.901366
1963	4.609162	6.056784	3.647536	2.93492	3.983413
1964	4.653008	6.12905	3.666122	2.946542	4.011325
1965	4.593098	6.150603	3.684871	3.026746	4.025887

Transforming the table into a matrix and ordering the columns in order of when the β's enter the equation: $\ln qb_t = \beta_1 + \beta_2 \ln pb_t + \beta_3 \ln pl_t + \beta_4 \ln pp_t + \beta_5 \ln y_t + e_t$, we get:

$$X = \begin{bmatrix} 1 & 3.245712 & 2.996232 & 3.822973 & 5.872118 \\ 1 & 3.269569 & 2.93066 & 3.970292 & 5.940171 \\ 1 & 3.407511 & 3.111291 & 4.058717 & 6.054439 \\ 1 & 3.500439 & 3.299903 & 4.035302 & 5.866468 \\ 1 & 3.453474 & 3.02965 & 4.011868 & 5.869297 \\ 1 & 3.406185 & 2.938633 & 4.019801 & 5.888878 \\ 1 & 3.472898 & 2.952825 & 3.924347 & 5.940171 \\ 1 & 3.459466 & 3.00072 & 3.958143 & 5.968708 \\ 1 & 3.426215 & 3.025291 & 3.988243 & 5.940171 \\ 1 & 3.468233 & 2.968361 & 3.947004 & 5.908083 \\ 1 & 3.506758 & 2.884801 & 3.932022 & 5.953243 \\ 1 & 3.601686 & 2.882564 & 4.000766 & 5.998937 \\ 1 & 3.681603 & 2.95178 & 3.971046 & 6.006353 \\ 1 & 3.632045 & 2.892037 & 3.901366 & 6.018593 \\ 1 & 3.647536 & 2.93492 & 3.983413 & 6.056784 \\ 1 & 3.666122 & 2.946542 & 4.011325 & 6.12905 \\ 1 & 3.684871 & 3.026746 & 4.025887 & 6.150603 \end{bmatrix} \quad Y = \begin{bmatrix} 4.798267 \\ 4.822698 \\ 4.879767 \\ 4.778283 \\ 4.784989 \\ 4.741448 \\ 4.757891 \\ 4.779963 \\ 4.859037 \\ 4.829113 \\ 4.767289 \\ 4.589041 \\ 4.447346 \\ 4.53582 \\ 4.609162 \\ 4.653008 \\ 4.593098 \end{bmatrix}$$

Since we know $b = (X'X)^{-1}X'Y$,

$$(X'X)^{-1} = \begin{bmatrix} 559.440717784217 & 22.0874733004413 & -14.9439358453672 & -47.8035484926007 & -67.3075192772525 \\ 22.0874733004413 & 6.77612399504596 & 0.803576126045965 & -2.17862729347991 & -6.62135294075760 \\ -14.9439358453672 & 0.803576126045965 & 9.19147918242113 & -8.92561037144844 & 3.37299052952592 \\ -47.8035484926007 & -2.17862729347991 & -8.92561037144844 & 29.9028114645170 & -6.15165551541797 \\ -67.3075192772525 & -6.62135294075760 & 3.37299052952592 & -6.15165551541797 & 17.5534407842293 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 80.2262195062685 \\ 280.716586001900 \\ 239.684148324574 \\ 318.843429805907 \\ 479.200389499748 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 4.6726 \\ -0.8266 \\ 0.1997 \\ 0.4371 \\ 0.1017 \end{bmatrix}$$

$$\beta' = \begin{bmatrix} 4.6726 & -0.8266 & 0.1997 & 0.4371 & 0.1017 \end{bmatrix}$$

Therefore, $\beta_1 = 4.6726$, $\beta_2 = -0.8266$, $\beta_3 = 0.1997$, $\beta_4 = 0.4371$, and $\beta_5 = 0.1017$.

As we predicted in (a), every coefficient should be positive except for the price of beef pb_t, which means as the price of beef increases, per capita beef consumption decreases, and as the prices of lamb and pork and as disposable income increases, per capita beef consumption increases. This also means, according to this data, ceteris paribus, beef in Australia during the period 1949-1965 was a normal good.

(See the last few pages to view the Matlab inputs and calculations for this part and the next following parts for this problem).

(c)

Compute the estimated covariance matrix for the least squares estimator and the standard errors. Also compute R^2.

$$\begin{aligned} var(b) &= \sigma^2 (X'X)^{-1} && (1) \\ \sigma^2 &= \frac{(Y'Y - b'X'Y)}{T-K} && (2) \\ \sigma^2 &= \frac{(Y'Y - b'X'Y)}{17-5} && (3) \\ &= 0.0049 && (4) \\ & && (5) \end{aligned}$$

So then,

$$var(b) = \begin{bmatrix} 2.75419540012069 & 0.108739345261294 & -0.0735708325415178 & -0.235342743533642 & -0.331363188437861 \\ 0.108739345261294 & 0.0333596911067197 & 0.00395610401539679 & -0.0107256498848448 & -0.0325977341875288 \\ -0.0735708325415178 & 0.00395610401539679 & 0.0452507815033473 & -0.0439418766758317 & 0.0166056468643638 \\ -0.235342743533642 & -0.0107256498848448 & -0.0439418766758317 & 0.147215215425230 & -0.0302853560441525 \\ -0.331363188437861 & -0.0325977341875288 & 0.0166056468643638 & -0.0302853560441525 & 0.0864177460226673 \end{bmatrix}$$

And

$$\begin{aligned} R^2 &= 1 - \frac{\hat{e}^2}{Y'Y - T\bar{Y}^2} && (1) \\ &= 0.7609 && (2) \end{aligned}$$

(d)

Reproduce each observation twice so that, instead of having a sample of size $T = 17$, you have a sample size of $T = 51$. Repeat parts (b) and (c) with this increased sample and note any changes in the results. Can you explain these changes? Do you think this procedure of reproducing the observations is a valid one?

By reproducing each observations twice, b and R^2 remains the same but the σ^2 has changed.

$$b = \begin{bmatrix} 4.6726 \\ -0.8266 \\ 0.1997 \\ 0.4371 \\ 0.1017 \end{bmatrix}$$

$$var(b) = \sigma^2(X'X)^{-1} \tag{1}$$
$$\sigma^2 = \frac{(Y'Y - b'X'Y)}{T-K} \tag{2}$$
$$\sigma^2 = \frac{(Y'Y - b'X'Y)}{51-5} \tag{3}$$
$$= 0.0039 \tag{4}$$
$$\tag{5}$$

$$R^2 = 1 - \frac{\hat{e}^2}{Y'Y - T\bar{Y}^2} \tag{1}$$
$$= 0.7609 \tag{2}$$

These results make sense. By doubling the observations, there are just 3 data points on top of each other rather than just 1 data point. The distance between each of the data points and the least squares line remains unchanged, therefore b and R^2 remain the same. However, since variance depends on the size of the sample, the larger the sample size, the smaller the variance (an inversely proportional relationship)—which we obtained from our results since as T increased from 17 to 51, σ^2 decreased from 0.0049 to 0.0039.

Such a procedure to reproduce observations would be invalid. This procedure does not obtain any additional real observations from the population, it merely replicates one observation sample. We do not know if additional sampling will result in a repeat of the original sample even once much less twice. In other words, it is unlikely or even impossible that sampling the population two more times would result in a repeat doubling of our original data. Thus, reproducing observations will not get us closer towards knowing the true β, σ^2, and R^2 of the population.

```
>> x = [ones(17,1) lnpb lnpl lnpp lny]

x =

    1.0000    3.2457    2.9962    3.8230    5.8721
    1.0000    3.2696    2.9307    3.9703    5.9402
    1.0000    3.4075    3.1113    4.0587    6.0544
    1.0000    3.5004    3.2999    4.0353    5.8665
    1.0000    3.4535    3.0297    4.0119    5.8693
    1.0000    3.4062    2.9386    4.0198    5.8889
    1.0000    3.4729    2.9528    3.9243    5.9402
    1.0000    3.4595    3.0007    3.9581    5.9687
    1.0000    3.4262    3.0253    3.9882    5.9402
    1.0000    3.4682    2.9684    3.9470    5.9081
    1.0000    3.5068    2.8848    3.9320    5.9532
    1.0000    3.6017    2.8826    4.0008    5.9989
    1.0000    3.6816    2.9518    3.9710    6.0064
    1.0000    3.6320    2.8920    3.9014    6.0186
    1.0000    3.6475    2.9349    3.9834    6.0568
    1.0000    3.6661    2.9465    4.0113    6.1291
    1.0000    3.6849    3.0267    4.0259    6.1506

>> y = lnqb

y =

    4.7983
    4.8227
    4.8798
    4.7783
    4.7850
    4.7414
    4.7579
    4.7800
    4.8590
    4.8291
    4.7673
    4.5890
    4.4473
    4.5358
    4.6092
    4.6530
    4.5931

>> b = inv(x'*x)*x'*y

b =

    4.6726
   -0.8266
```

```
    0.1997
    0.4371
    0.1017

>> sigma2 = (y'*y-b'*x'*y)/(17-5)

sigma2 =

    0.0049

>> varb = sigma2*inv(x'*x)

varb =

    2.7542    0.1087   -0.0736   -0.2353   -0.3314
    0.1087    0.0334    0.0040   -0.0107   -0.0326
   -0.0736    0.0040    0.0453   -0.0439    0.0166
   -0.2353   -0.0107   -0.0439    0.1472   -0.0303
   -0.3314   -0.0326    0.0166   -0.0303    0.0864

>> ehat = y - x*b

ehat =

   -0.0580
   -0.0721
    0.0127
   -0.0203
    0.0115
   -0.0584
    0.0469
    0.0306
    0.0670
    0.1045
    0.0931
   -0.0409
   -0.1181
   -0.0295
    0.0084
    0.0457
   -0.0233

>> averagey = mean(y)

averagey =

    4.7192

>>  r2 = 1 - (ehat'*ehat)/[(y'*y)-(17*averagey*averagey)]
```

```
r2 =

    0.7609

>>
```

```
>> x = [ones(51,1) lnpb lnpl lnpp lny]

x =

    1.0000    3.2457    2.9962    3.8230    5.8721
    1.0000    3.2696    2.9307    3.9703    5.9402
    1.0000    3.4075    3.1113    4.0587    6.0544
    1.0000    3.5004    3.2999    4.0353    5.8665
    1.0000    3.4535    3.0297    4.0119    5.8693
    1.0000    3.4062    2.9386    4.0198    5.8889
    1.0000    3.4729    2.9528    3.9243    5.9402
    1.0000    3.4595    3.0007    3.9581    5.9687
    1.0000    3.4262    3.0253    3.9882    5.9402
    1.0000    3.4682    2.9684    3.9470    5.9081
    1.0000    3.5068    2.8848    3.9320    5.9532
    1.0000    3.6017    2.8826    4.0008    5.9989
    1.0000    3.6816    2.9518    3.9710    6.0064
    1.0000    3.6320    2.8920    3.9014    6.0186
    1.0000    3.6475    2.9349    3.9834    6.0568
    1.0000    3.6661    2.9465    4.0113    6.1291
    1.0000    3.6849    3.0267    4.0259    6.1506
    1.0000    3.2457    2.9962    3.8230    5.8721
    1.0000    3.2696    2.9307    3.9703    5.9402
    1.0000    3.4075    3.1113    4.0587    6.0544
    1.0000    3.5004    3.2999    4.0353    5.8665
    1.0000    3.4535    3.0297    4.0119    5.8693
    1.0000    3.4062    2.9386    4.0198    5.8889
    1.0000    3.4729    2.9528    3.9243    5.9402
    1.0000    3.4595    3.0007    3.9581    5.9687
    1.0000    3.4262    3.0253    3.9882    5.9402
    1.0000    3.4682    2.9684    3.9470    5.9081
    1.0000    3.5068    2.8848    3.9320    5.9532
    1.0000    3.6017    2.8826    4.0008    5.9989
    1.0000    3.6816    2.9518    3.9710    6.0064
    1.0000    3.6320    2.8920    3.9014    6.0186
    1.0000    3.6475    2.9349    3.9834    6.0568
    1.0000    3.6661    2.9465    4.0113    6.1291
    1.0000    3.6849    3.0267    4.0259    6.1506
    1.0000    3.2457    2.9962    3.8230    5.8721
    1.0000    3.2696    2.9307    3.9703    5.9402
    1.0000    3.4075    3.1113    4.0587    6.0544
    1.0000    3.5004    3.2999    4.0353    5.8665
    1.0000    3.4535    3.0297    4.0119    5.8693
    1.0000    3.4062    2.9386    4.0198    5.8889
    1.0000    3.4729    2.9528    3.9243    5.9402
    1.0000    3.4595    3.0007    3.9581    5.9687
    1.0000    3.4262    3.0253    3.9882    5.9402
    1.0000    3.4682    2.9684    3.9470    5.9081
    1.0000    3.5068    2.8848    3.9320    5.9532
    1.0000    3.6017    2.8826    4.0008    5.9989
```

```
    1.0000    3.6816    2.9518    3.9710    6.0064
    1.0000    3.6320    2.8920    3.9014    6.0186
    1.0000    3.6475    2.9349    3.9834    6.0568
    1.0000    3.6661    2.9465    4.0113    6.1291
    1.0000    3.6849    3.0267    4.0259    6.1506

>> y = lnqb

y =

    4.7983
    4.8227
    4.8798
    4.7783
    4.7850
    4.7414
    4.7579
    4.7800
    4.8590
    4.8291
    4.7673
    4.5890
    4.4473
    4.5358
    4.6092
    4.6530
    4.5931
    4.7983
    4.8227
    4.8798
    4.7783
    4.7850
    4.7414
    4.7579
    4.7800
    4.8590
    4.8291
    4.7673
    4.5890
    4.4473
    4.5358
    4.6092
    4.6530
    4.5931
    4.7983
    4.8227
    4.8798
    4.7783
    4.7850
    4.7414
```

```
    4.7579
    4.7800
    4.8590
    4.8291
    4.7673
    4.5890
    4.4473
    4.5358
    4.6092
    4.6530
    4.5931

>> b = inv(x'*x)*x'*y

b =

    4.6726
   -0.8266
    0.1997
    0.4371
    0.1017

>> sigma2 = (y'*y-b'*x'*y)/(51-5)

sigma2 =

    0.0039

>> varb = sigma2*inv(x'*x)

varb =

    0.7185    0.0284   -0.0192   -0.0614   -0.0864
    0.0284    0.0087    0.0010   -0.0028   -0.0085
   -0.0192    0.0010    0.0118   -0.0115    0.0043
   -0.0614   -0.0028   -0.0115    0.0384   -0.0079
   -0.0864   -0.0085    0.0043   -0.0079    0.0225

>> ehat = y - x*b

ehat =

   -0.0580
   -0.0721
    0.0127
   -0.0203
    0.0115
   -0.0584
    0.0469
    0.0306
```

```
    0.0670
    0.1045
    0.0931
   -0.0409
   -0.1181
   -0.0295
    0.0084
    0.0457
   -0.0233
   -0.0580
   -0.0721
    0.0127
   -0.0203
    0.0115
   -0.0584
    0.0469
    0.0306
    0.0670
    0.1045
    0.0931
   -0.0409
   -0.1181
   -0.0295
    0.0084
    0.0457
   -0.0233
   -0.0580
   -0.0721
    0.0127
   -0.0203
    0.0115
   -0.0584
    0.0469
    0.0306
    0.0670
    0.1045
    0.0931
   -0.0409
   -0.1181
   -0.0295
    0.0084
    0.0457
   -0.0233

>> averagey = mean(y)

averagey =

    4.7192
```

```
>> r2 = 1 - (ehat'*ehat)/[(y'*y)-(51*averagey*averagey)]

r2 =

    0.7609

>>
```