Block 7
MLR

LAND ACKNOWLEDGEMENT

The School of Public Health at the University of Minnesota Twin Cities is built within the traditional homelands of the Dakota people. Minnesota comes from the Dakota name for this region, Mni Sóta Maḳoce, which loosely translates to the land where the waters reflect the skies.

It is important to acknowledge the peoples on whose land we live, learn, and work as we seek to improve and strengthen our relations with our tribal nations. We also acknowledge that words are not enough. We must ensure that our institution provides support, resources, and programs that increase access to all aspects of higher education for our American Indian students, staff, faculty, and community members.

Multiple Linear Regression

Recall simple linear regression

$$E(Y|X_1 = x_1) = \beta_0 + \beta_1 x_1$$

Now suppose we have a second variable X_2 Consider the model:

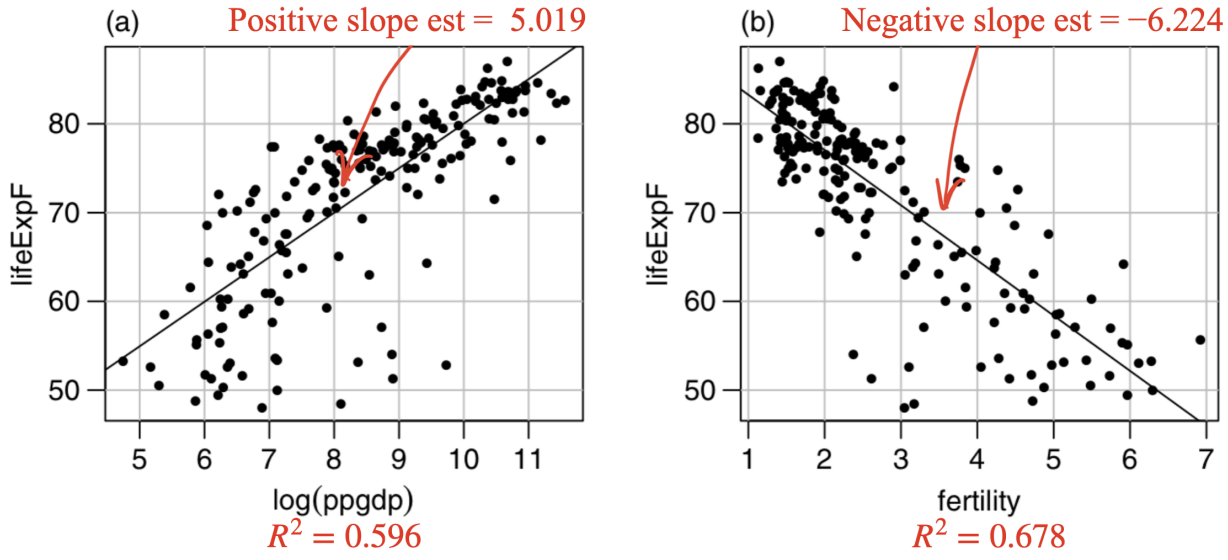
$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Idea: use X_2 to help explain part of (the variation in) Y not already explained by X_1.

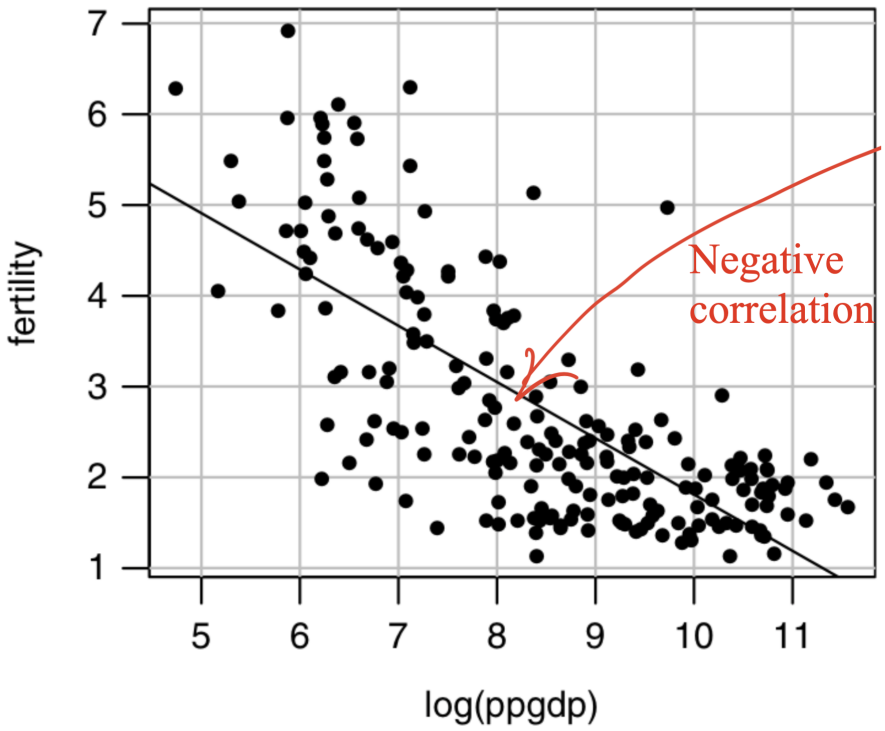
United Nations Fertility Data

lifeExpF = life expectancy (response)
fertility = birth rate per 1000 females from 2009
ppgdp = GDP per person in USD

Data from 199 UN member countries and a few others (e.g., Hong Kong)
[Data collected in 2011]



Fact: If log(ppgdp) and fertility were uncorrelated, above figure would provide a complete summary of the dependence of response on regressors.



$\Longrightarrow$ The regressors will in part be explaining the same variation in lifeExpF.

Intuition About Explained Variability

Total explained variation explained jointly by two variables X_1, X_2 can take different scenarios:

1) If population correlation $\rho(X_1, X_2) = 0$, then variation explained jointly = sum of variations explained individually.

2) If $\rho(X_1, X_2) \neq 0$, then the situation is more complicated

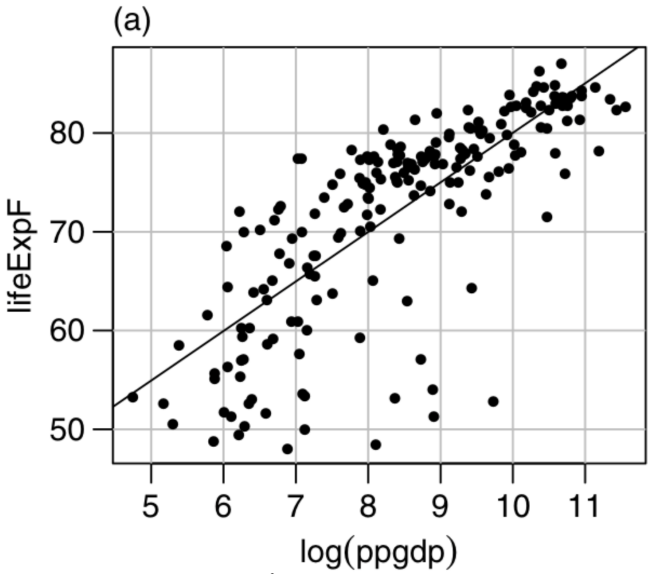
– total variation can exceed or be less than the sum of individual variation.

Added Variable Plots

Used to get the effect of adding in X_2 to a model that already includes X_1. Let's consider the fertility dataset again.

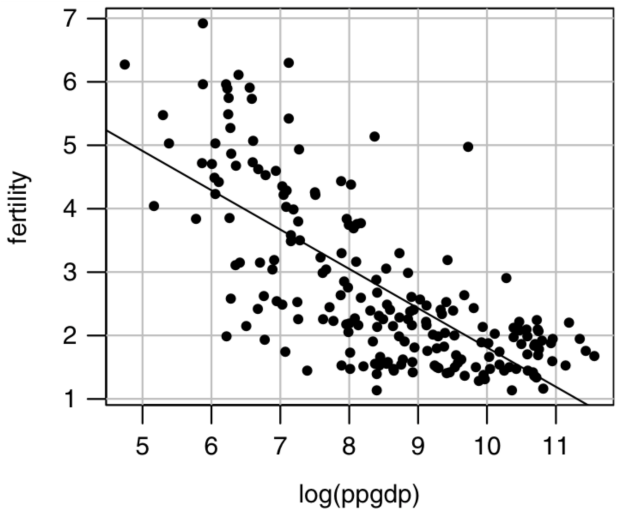
Steps

1. Compute regression of lifeExpF on log(ppgdp) (see Fig. below)



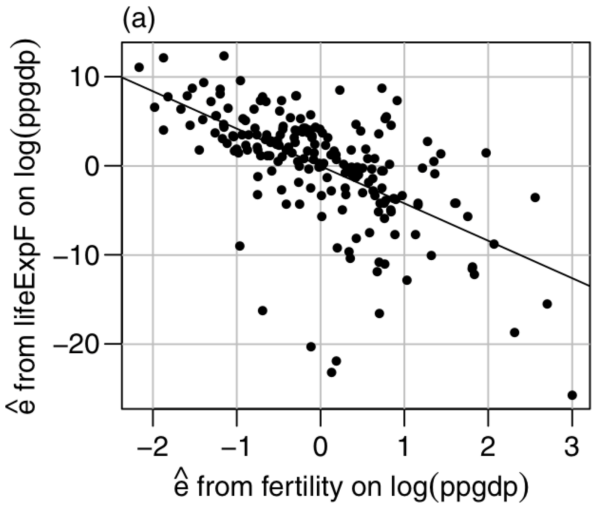
Extract residuals (part of lifeExpF not explained by log(ppgdp))

2. Compute regression of fertility on log(ppgdp) – see Fig. below



Extract residuals (part of fertility not explained by log(ppgdp))

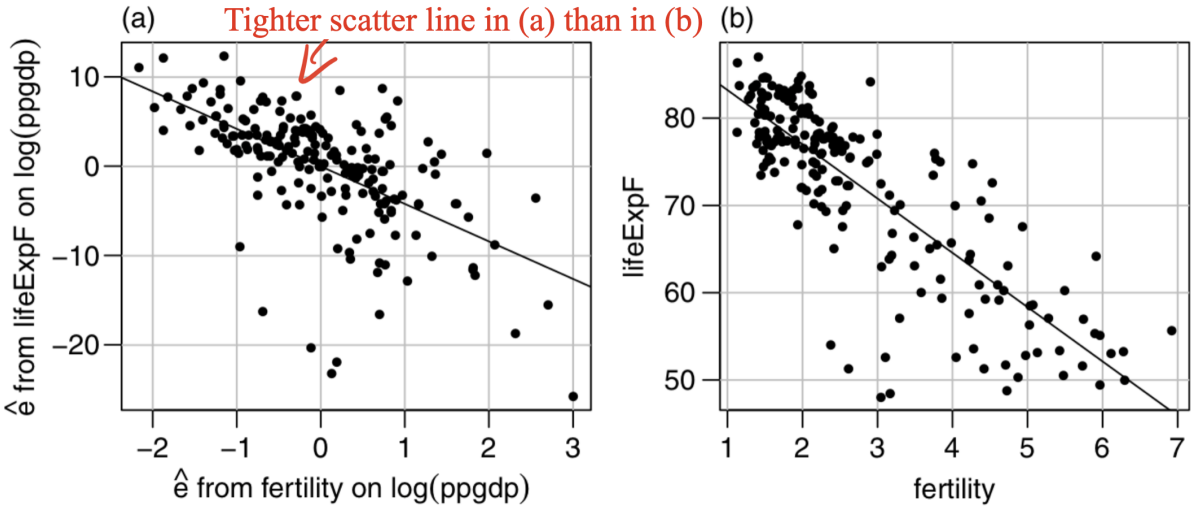
3. Plot residuals from (1) against those from (2)



Interpretation

AVP summarizes the relationship between lifeExpF and fertility adjusting for log(ppgdp).

If AVP plot shows a stronger relationship than the marginal plot on fertility, then the two variables act jointly to explain extra variation.



Note: From AVP, when $\hat{e}$ (x-axis) $= 0$, $\hat{e}$ (y-axis) $= 0$ *(intercept)*, estimated slope $= -4.199$

$\equiv$ Same as slope for fertility when fitting both predictors together *(which we learn about next)*

Multiple Linear Regression Model

Assume we have response r.v. Y and predictors $X_1, \ldots, X_p$. Write

$$E(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

(conditioning on all $X_1, \ldots, X_p$)

With specific values $(x_1, \ldots, x_p) = x$

$$E(Y|X = x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

$\beta_0 \ldots \beta_p$ unknown parameters to be estimated from sample of data

Fuel Consumption Data

Table 1.1 Variables in the Fuel Consumption Data[a] ←Across 50 states and DC

`Drivers`	Number of licensed drivers in the state
`FuelC`	Gasoline sold for road use, thousands of gallons
`Income`	Per person personal income for the year 2000, in thousands of dollars
`Miles`	Miles of Federal-aid highway miles in the state
`Pop`	2001 population age 16 and over
`Tax`	Gasoline state tax rate, cents per gallon
`Fuel`	1000 × `Fuelc/Pop` ← response of interest
`Dlic`	1000 × `Drivers/Pop`
log(`Miles`)	Natural logarithm of `Miles`

[a]All data are for 2001, unless otherwise noted. The last three variables do not appear in the data file, but are computed from the previous variables, as described in the text.

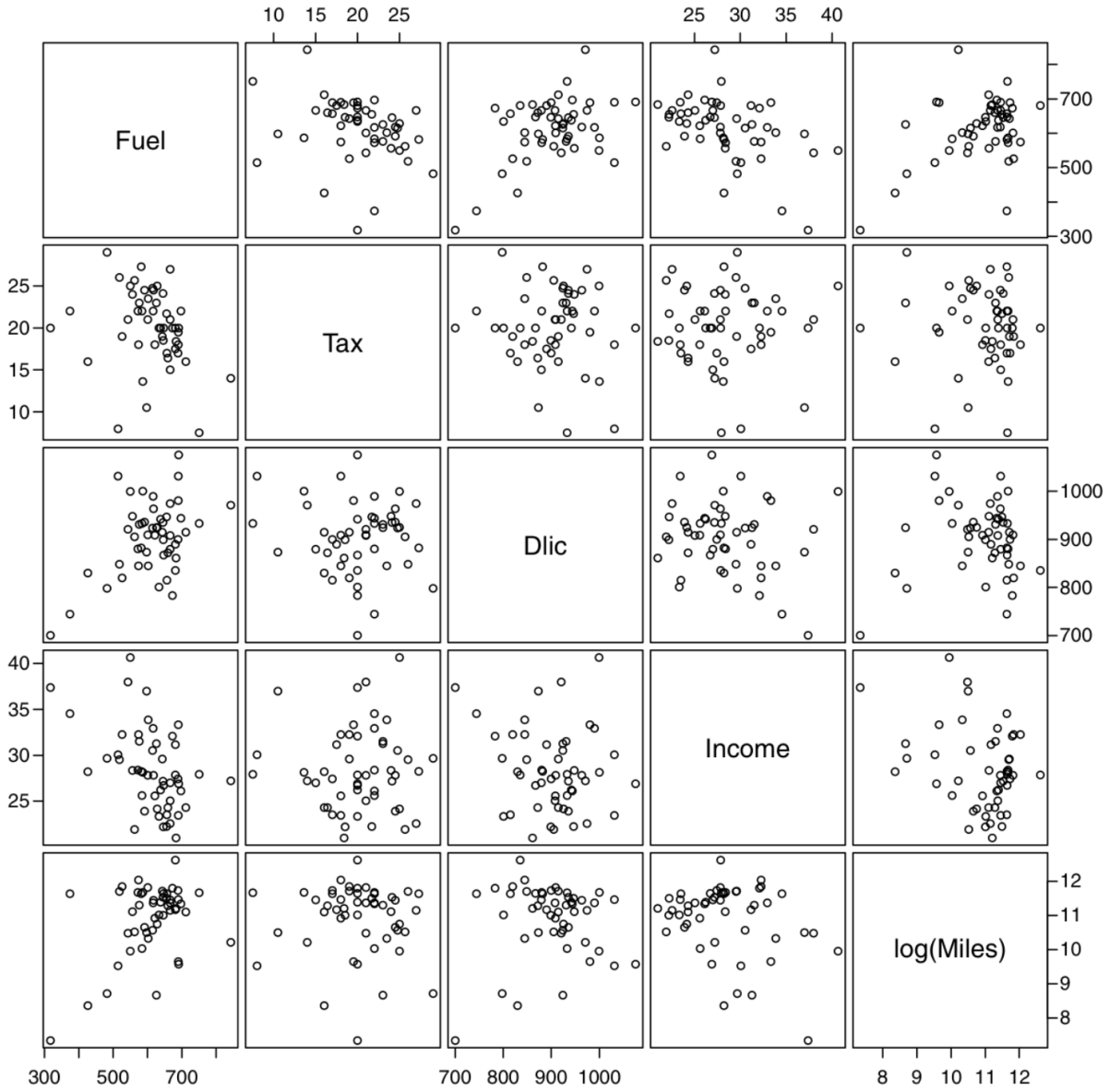


Figure 1.11 Scatterplot matrix for the fuel data.

Table 3.1 Summary Statistics for the Fuel Data

	N	Average	Std Dev	Min	Max
Tax	51	20.15	4.54	7.50	29.00
Dlic	51	903.68	72.86	700.20	1075.29
Income	51	28.40	4.45	20.99	40.64
log(Miles)	51	10.91	1.03	7.34	12.61
Fuel	51	613.13	88.96	317.49	842.79

Table 3.2 Sample Correlations for the Fuel Data

	Tax	Dlic	Income	log(Miles)	Fuel
Tax	1.0000	-0.0858	-0.0107	-0.0437	-0.2594
Dlic	-0.0858	1.0000	-0.1760	0.0306	0.4685
Income	-0.0107	-0.1760	1.0000	-0.2959	-0.4644
log(Miles)	-0.0437	0.0306	-0.2959	1.0000	0.4220
Fuel	-0.2594	0.4685	-0.4644	0.4220	1.0000

Data and Matrix Notation

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}$$

For fuel consumption dataset,

$$\mathbf{X} = \begin{pmatrix} 1 & 18.00 & 1031.38 & 23.471 & 16.5271 \\ 1 & 8.00 & 1031.64 & 30.064 & 13.7343 \\ 1 & 18.00 & 908.597 & 25.578 & 15.7536 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 25.65 & 904.894 & 21.915 & 15.1751 \\ 1 & 27.30 & 882.329 & 28.232 & 16.7817 \\ 1 & 14.00 & 970.753 & 27.230 & 14.7362 \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} 690.264 \\ 514.279 \\ 621.475 \\ \vdots \\ 562.411 \\ 581.794 \\ 842.792 \end{pmatrix}$$

Now define MLR model :

$$E(Y|X = x_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = x_i'\beta$$

This is the mean function evaluated at x_i.

The mean function across all n observations is

$$E(Y_{n\times 1}|X_{n\times(p+1)}) = X_{n\times(p+1)}\beta_{(p+1)\times 1} \quad (p+1 \text{ allows for intercept})$$

What about the errors in the model?

Define elementwise

$$e_i = y_i - E(Y|X = x_i) = y_i - x_i'\beta$$

and note

$$e = \Big(e_1, \ldots, e_n\Big)'$$

We assume

$$\left.\begin{array}{c} E(e|X) = 0_{n\times 1} \\ \text{var}(e|X) = \sigma^2 I_{n\times n} \end{array}\right] \text{ Key Assumptions}$$

I: independence of errors and constant variance

The adjective linear means the model is linear in its parameters $\beta_0, \ldots, \beta_p$.
Examples:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$

$$Y = \beta_0 + \beta_1 \log(X) + e$$

Models are written out as a function of the r.v.'s X and Y.

Both models are linear even though the relationship between Y and X is not linear.

Some non-linear models can be turned linear.

Example:

Non-linear model:

$$E(Y|X) = \frac{X}{\alpha X + \beta}$$

Take reciprocal:

$$E\left(\frac{1}{Y}\Big|X\right) = \alpha + \beta\left(\frac{1}{X}\right)$$

$$Y' = \frac{1}{Y}, \quad X' = \frac{1}{X}$$

$$\Rightarrow E(Y'|X') = \alpha + \beta X$$

Another Example

Volume of trees $\approx cr^2h$

or more generally

$$\alpha r^{\beta_1} h^{\beta_2}$$

$\Rightarrow$ Non-linear model:

$$E(\text{vol}|r,h) = \alpha r^{\beta_1} h^{\beta_2}$$

Take logarithm:

$$\underbrace{E(\log(\text{vol})|r,h)}_{Y} = \underbrace{\log(\alpha)}_{\beta_0} + \underbrace{\beta_1 \log(r)}_{x_1} + \underbrace{\beta_2 \log(h)}_{x_2}$$

Which of the following models are linear?

(a) $Y = \beta_0 + \beta_1^{X_1} + \varepsilon$

(b) $Y = \beta_0 \beta_1^{X_1} \varepsilon$

(c) $Y = \beta_0 + \beta_1 e^X + \varepsilon$ ..... Linear

(d) $Y = \beta_0 + \beta_1 X^2 + \beta_2 \log(X) + \varepsilon$ ..... Linear

Which below can be turned linear after transformation?

(a) $Y = \beta_0 + \beta_1^{X_1} + \varepsilon$

(b) $Y = \beta_0 \beta_1^{X_1} \varepsilon$

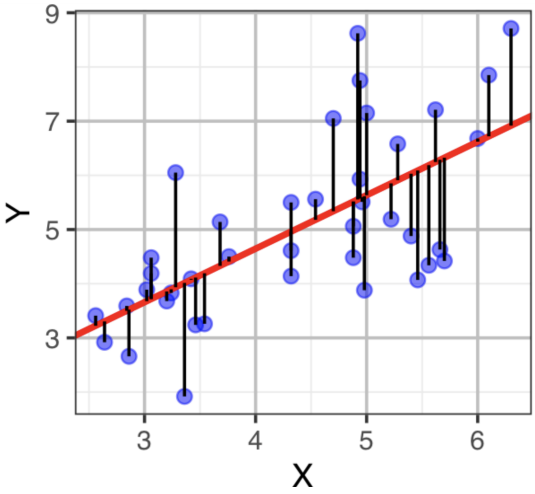
Ans: (b)

Ordinary Least Square Estimators

Legendre

Recall for SLR, the least squares estimate $(\hat{\beta}_0, \hat{\beta}_1)$ for (β_0, β_1) is the intercept and slope of the straight line with the minimum sum of squared vertical distances to the data points

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$



MLR is just like SLR. The least squares estimate $(\hat{\beta}_0, \ldots, \hat{\beta}_p)$ for $(\beta_0, \ldots, \beta_p)$ is the intercept and slopes of the (hyper)plane with the minimum sum of squared vertical distance to the data points

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2.$$

To find $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ minimize

$$\text{RSS}(\beta_0, \beta_1, \ldots, \beta_p) = L(\beta_0, \beta_1, \ldots, \beta_p) = \sum_{i=1}^{n} e_i^2$$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$

Take partial derivatives and equate to 0

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})$$

$$\frac{\partial L}{\partial \beta_k} = -2 \sum_{i=1}^{n} x_{ik} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) \text{ for } k = 1, 2, \ldots, p$$

OLS estimates are the solution to the following system of equations, called normal equations:

$$\begin{aligned}
\beta_0 n + \beta_1 \sum_{i=1}^{n} x_{i1} + \cdots + \beta_p \sum_{i=1}^{n} x_{ip} &= \sum_{i=1}^{n} y_i \\
\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} x_{i1}^2 + \cdots + \beta_p \sum_{i=1}^{n} x_{i1} x_{ip} &= \sum_{i=1}^{n} x_{i1} y_i \\
&\vdots \\
\beta_0 \sum_{i=1}^{n} x_{ik} + \beta_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \cdots + \beta_p \sum_{i=1}^{n} x_{ik} x_{ip} &= \sum_{i=1}^{n} x_{ik} y_i \\
&\vdots \\
\beta_0 \underbrace{\sum_{i=1}^{n} x_{ip}}_{known} + \beta_1 \underbrace{\sum_{i=1}^{n} x_{ip} x_{i1}}_{known} + \cdots + \beta_p \underbrace{\sum_{i=1}^{n} x_{ip}^2}_{known} &= \underbrace{\sum_{i=1}^{n} x_{ip} y_i}_{known}
\end{aligned}$$

Usually, we write this in matrix form:

$$\text{RSS}(\beta) = \sum_{i=1}^{n} e_i^2 = (Y_{n\times 1} - X\beta_{n\times 1})'(Y - X\beta) \text{(matrix notation of observables)}$$

Using the distributive property for matrices:

$$\text{RSS}(\beta) = Y'Y + \beta'(X'X)\beta - 2Y'X\beta$$

Differentiate with respect to β and set to 0 to solve the normal equations:

$$\Rightarrow X'X\beta = X'Y$$

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

Let's look more closely at how normal equations arise:

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2X'(\underbrace{Y - X\beta}_{e}) = 0$$

$$\Rightarrow X'X\beta = X'Y$$

What is this saying?
That the solution to β will produce $\hat{e} = (Y - X\hat{\beta})$ that are orthogonal to the column space of X.

Properties of OLS estimators

Under Key Assumptions:

$$E(\hat{\beta}|X) = E[(X'X)^{-1}X'Y|X]$$

$$= [(X'X)^{-1}X']E(Y|X)$$

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

$$= \beta$$

$\therefore \hat{\beta}$ is unbiased for β if the model is true.

$$\text{Var}(\hat{\beta}|X) = \text{Var}[(X'X)^{-1}X'Y|X]$$

$$= (X'X)^{-1}X'[\text{Var}(Y|X)]X(X'X)^{-1}$$

$$= (X'X)^{-1}X'[\sigma^2 I]X(X'X)^{-1}$$

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

Gauss-Markov Theorem

Under key assumptions, OLS estimator is efficient in the class of linear unbiased estimators.
i.e., for any unbiased estimator b that is linear in Y,

$$\text{Var}(b|X) \geq \text{Var}(\hat{\beta}|X)$$

Sketch of Proof
Since b is assumed linear in Y,
we can write it as

$$b = CY \text{for some matrix } C.$$

Let

$$D = C - A \quad \text{or} \quad C = D + A \quad \text{where} \quad A = (X'X)^{-1}X'$$

Then,

$$b = (D + A)Y$$

$$= DY + AY$$

$$= D(Xb + e) + \hat{\beta}$$

(since $Y = X\beta + e$ and $AY = (X'X)^{-1}X'Y = \hat{\beta}$)

$$= DXb + De + \hat{\beta}$$

$$E(b|X) = DXb + E(De|X) + E(\hat{\beta}|X)$$

Since both b and $\hat{\beta}$ are unbiased and since

$$E(De|X) = DE(e|X) = 0$$

$$\Rightarrow DXb = 0$$

$$\therefore b = De + \hat{\beta} \text{ and } b - \beta = De + (b - \beta) = (D + A)e$$

$$\therefore \text{Var}(b|X) = \text{Var}(b - \beta|X) = \text{Var}((D + A)e|X)$$

$$= (D + A)\text{Var}(e|X)(D' + A')$$

$$= \sigma^2(D + A)(D' + A') \quad \text{since Var}(e|X) = \sigma^2 I$$

$$= \sigma^2(DD' + AD' + DA' + AA')$$

$$\text{But } DA' = DX(X'X)^{-1} = 0 \quad \text{since } DX = 0$$

$$\text{Also } AA' = (X'X)^{-1}$$

$$\therefore \text{Var}(b|X) = \sigma^2(DD' + (X'X)^{-1}) \geq \sigma^2(X'X)^{-1}$$

Since $D'D$ is positive semi-definite.

With fixed x	**With random** x
SR1: $y = \beta_1 + \beta_2 x + e$ with x fixed	A10.1: $y = \beta_1 + \beta_2 x + e$ with x, y, e random
SR2: $E(e) = 0$	A10.2: (x, y) obtained from IID sampling
SR3: $\text{Var}(e) = \sigma^2$	A10.3: $E(e\|X) = 0$
SR4: $\text{Cov}(e_i, e_j) = 0$	A10.4: x takes on at least two values
SR5: x takes on at least two values	A10.5: $\text{Var}(e\|X) = \sigma^2$
SR6: e is normal	A10.6: e is normal

- Note that A10.2 implies SR4 (and A10.5?)
- Note that A10.3 implies both $\text{Cov}(x, e) = 0$ and $E(e) = 0$
  - This assumption is a critical one.
  - Instead of assuming that x is a fixed value and e is random, we make the properties of e conditional on the particular outcome of x.
  - This allows us to operate in very much the same way as if x is fixed, as long as A10.3 holds.

SLR in Matrix Notation

For simple regression, $\mathbf{X}$ and $\mathbf{Y}$ are given by

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X}) = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

By direct multiplication, $(\mathbf{X}'\mathbf{X})^{-1}$ can be shown to be

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{SXX} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

so that

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{SXX} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{SXY}{SXX} \end{pmatrix} \begin{matrix} \leftarrow \hat{\beta}_0 \\ \leftarrow \hat{\beta}_1 \end{matrix}$$

Fuel Consumption Data Example

$$E(\text{Fuel}|X) = \beta_0 + \beta_1 \text{Tax} + \beta_2 \text{Dlic} + \beta_3 \text{Income} + \beta_4 \log(\text{Miles})$$

$(X'X)^{-1}$ given by:

	Intercept	Tax	Dlic	Income	log(Miles)
Intercept	$9.02e+00$	$-2.85e-02$	$-4.08e-03$	$-5.98e-02$	$-2.79e-01$
Tax	$-2.85e-02$	$9.79e-04$	$5.60e-06$	$4.26e-05$	$2.31e-04$
Dlic	$-4.08e-03$	$5.60e-06$	$3.92e-06$	$1.19e-05$	$7.79e-06$
Income	$-5.98e-02$	$4.26e-05$	$1.19e-05$	$1.14e-03$	$1.44e-03$
log(Miles)	$-2.79e-01$	$2.31e-04$	$7.79e-06$	$1.44e-03$	$2.07e-02$

Notice: elements of $(\mathbf{X'X})^{-1}$ differ by several orders of magnitude
$\Rightarrow$ numerical instabilities.

Table 3.3 Multiple Linear Regression Summary in the Fuel Data

	Estimate ($\hat{\beta}_j$)	Std. Error ($\text{Var}(\hat{\beta}_j)$)	t-Value	$\Pr(>\|t\|)$
(Intercept)	154.1928	194.9062	0.79	0.4329
Tax	-4.2280	2.0301	-2.08	0.0429
Dlic	0.4719	0.1285	3.67	0.0006
Income	-6.1353	2.1936	-2.80	0.0075
log(Miles)	26.7552	9.3374	2.87	0.0063

$\hat{\sigma} = 64.8912$ with 46 df, $R^2 = 0.5105$.

Note: $SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)} = \sqrt{\sigma^2 \left((\mathbf{X'X})^{-1}\right)_{jj}}$

where jj : jth diagonal element of the variance-covariance matrix $\text{Var}(\hat{\beta})$.

Estimating σ^2

Let's look at model fitted values:

$$\hat{Y} = X\hat{\beta} \quad \text{(plugging in } \hat{\beta} \text{ for } \beta\text{)}$$

$$= X \underbrace{(X'X)^{-1}X'}_{\hat{\beta}} Y = HY$$

Here, $X(X'X)^{-1}X'$ is $H \rightarrow$ hat matrix.
It projects Y onto the column space of X.
We may split Y into: $\hat{Y} = HY$ and $\hat{e} = (I - H)Y$

$$\Rightarrow E(\hat{Y}) = X\beta, \quad \text{Var}(\hat{Y}) = \sigma^2 H \quad \text{(because } H^2 = H\text{)}$$

$$\textit{and} \quad \Rightarrow E(\hat{e}) = 0, \quad \text{Var}(\hat{e}) = \sigma^2(I - H)$$

$$\therefore \hat{e} \text{ can be used to learn about } \sigma^2.$$

Specifically,

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \|\hat{e}\|^2$$

has

$$E(RSS) = \sigma^2(n - p)$$

(where p or $(p + 1)$ depends on whether the model includes an intercept)

$$\therefore \frac{\|\hat{e}\|^2}{n - p} \text{ is an unbiased estimator of } \sigma^2.$$

Coefficient of Determination

Let's define $\bar{x}_1$: column mean for x_1

$$\mathcal{X} = \begin{bmatrix} (x_{11} - \bar{x}_1) & \dots & (x_{1p} - \bar{x}_p) \\ (x_{21} - \bar{x}_1) & \dots & (x_{2p} - \bar{x}_p) \\ \vdots & \ddots & \vdots \\ (x_{n1} - \bar{x}_1) & \dots & (x_{np} - \bar{x}_p) \end{bmatrix}$$

(based on observed data from sample)

We can similarly define $\mathcal{Y}$ and

$$\hat{\beta}^* = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'\mathcal{Y}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}^*\bar{x}$$

Then, The proof of this is not trivial

$$RSS = \hat{e}'\hat{e}$$

$$= (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= (\mathcal{Y} - \mathcal{X}\hat{\beta}^*)(\mathcal{Y} - \mathcal{X}\hat{\beta}^*)$$

$$= SYY - \hat{\beta}^*(\mathcal{X}'\mathcal{X})^{-1}\hat{\beta}^*$$

$$= SYY - SS_{reg}$$

$$\Rightarrow R^2 = \frac{SS_{reg}}{SYY} = 1 - \frac{RSS}{SYY}$$

proportion of variability in Y explained by the regression model.

For fuel consumption data:

$$R^2 = 1 - \frac{193700}{395694} = 0.510$$

Hypothesis Testing for One Coefficient

Let's assume we want to test:

$$H_0 : \beta_j = 0 \quad \text{(all others arbitrary)}$$

$$H_1 : \beta_j \neq 0 \quad \text{(all others arbitrary)}$$

This is inference. We have some choices:

1. Add structure to model (e.g., Normality assumption)
2. Central Limit Theorem (CLT)
3. Bootstrap

Let's focus on (1):

$$Y = X\beta + e \qquad (e \sim N(0, \sigma^2 I))$$

When we do this, we can show,

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 (X'X)^{-1}_{jj})$$

$\therefore$ we can create a test statistic for testing H_0:

$$Z = \frac{\hat{\beta}_j - 0}{\underbrace{\sigma\sqrt{(X'X)^{-1}_{jj}}}_{SE(\hat{\beta}_j)}} \overset{H_0}{\sim} N(0,1)$$

This assumes knowledge of σ^2 . If we use $\hat{\sigma}^2$ to estimate σ^2,

$$\Rightarrow T = \frac{\hat{\beta}_j - 0}{\hat{\sigma}\sqrt{(X'X)^{-1}_{jj}}} \overset{H_0}{\sim} t_{n-p} \quad n-p\text{: df associated with } \hat{\sigma}^2$$

Table 3.3 Multiple Linear Regression Summary in the Fuel Data

	Estimate	Std. Error	t-Value	Pr(> |t|)(P-values)
(Intercept)	154.1928	194.9062	0.79	0.4329
Tax	-4.2280	2.0301	-2.08	0.0429
Dlic	0.4719	0.1285	3.67	0.0006
Income	-6.1353	2.1936	-2.80	0.0075
log(Miles)	26.7552	9.3374	2.87	0.0063

$\hat{\sigma} = 64.8912$ with 46 df, $R^2 = 0.5105$. Red values: Significant at $\alpha = 0.05$

Note: For a one-sided alternative H_1 (for Tax), the p-value is:

$$p = \frac{0.043}{2} = 0.022$$

Predictions and Fitted Values

Let x^* be a vector of new regressors/covariates.

Want to predict Y at $X = x^*$.

Point prediction:

$$\tilde{y}^* = x^{*\prime}\hat{\beta}$$

Associated Standard Error

$$sepred(\tilde{y}^*|X = x^*) = \underbrace{\hat{\sigma}\sqrt{1 + x^{*\prime}(X'X)^{-1}x^*}}_{\text{generalization of result from SLR.}}$$

Fitted values

Estimated average for all units where $X = x$

$$\hat{E}(Y|X = x) = x'\hat{\beta}$$

$$sefit = \hat{\sigma}\sqrt{x'(X'X)^{-1}x}$$

Relationship between the two:

$$sepred(\tilde{y}^*|X = x^*) = \sqrt{\hat{\sigma}^2 + sefit(\tilde{y}^*|X = x^*)^2} \quad \text{[a curiosity]}$$