PUBH 7405 Block 8 Interpretation of Main Effects

Interpretation of Main Effects

Understanding Parameter Estimates

Recall:

$$\hat{E}(\text{Fuel} \mid X) = 154.19 - 4.23\,\text{Tax} + 0.47\,\text{Dlic} - 6.14\,\text{Income} + 26.76\,(\log \text{Miles})$$

slopes/partial slopes: $\hat{\beta}_j$ coefficients ($j = 1, \ldots, 4$) [not considering intercept]
They have units.
Consider LHS: Fuel (gallons)
$\therefore$ RHS must also be in gallons. $\hat{\beta}_0 = 154.19$ gal
[expected Fuel consumption in a state with no taxes, no income, no roads]
Take Income for example: $\rightarrow$ (thousands of \$)
$\therefore$ $\hat{\beta}_3$ units must be gallons per person per thousand \$ of income.
Similarly: $\hat{\beta}_1$ for Tax is gallons per person per cent of tax.

Rate of change

Slopes usually interpreted as rates of change:

$$\hat{\beta}_1 : \text{Increasing Tax rate by 1 cent} \text{with all other covariates held fixed.}$$

Can visualize this by fixing other covariates at their mean values.

$$\overline{\text{Dlic}} = 903.68 \quad ; \quad \overline{\text{Income}} = 28.4 \quad ; \quad \log(\overline{\text{Miles}}) = 10.91$$

$$\Rightarrow \hat{E}(\text{Fuel} \mid \text{Tax} = \text{tax}, \text{others at sample means})$$

$$= \hat{\beta}_0 + \hat{\beta}_1 \text{tax} + \hat{\beta}_2 \overline{\text{Dlic}} + \hat{\beta}_3 \overline{\text{Income}} + \hat{\beta}_4 \log(\overline{\text{Miles}})$$

$$= 154.19 - 4.23\,\text{tax} + 0.47\,(903.68) - 6.14\,(28.4) + 26.76\,(10.91)$$

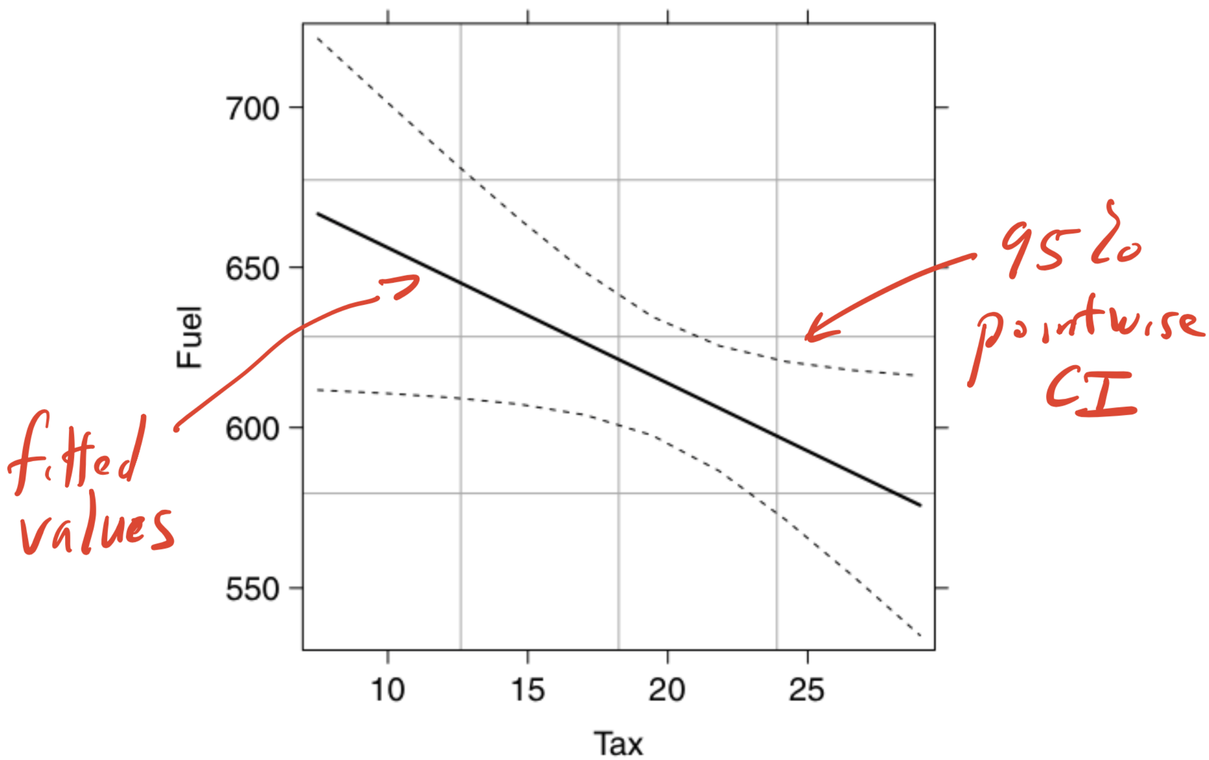

Figure 4.1 Effects plot for `Tax` in the fuel consumption data.

$\Rightarrow$ Effect of higher Tax is lower average Fuel consumption

[not saying anything causal]

Signs of Estimates

Indicates direction of relationship between covariate and response
after adjusting for all other covariates in model.
Caveat: high correlation between covariates can alter both magnitude and sign of an estimated coefficient depending on the other covariates in model.

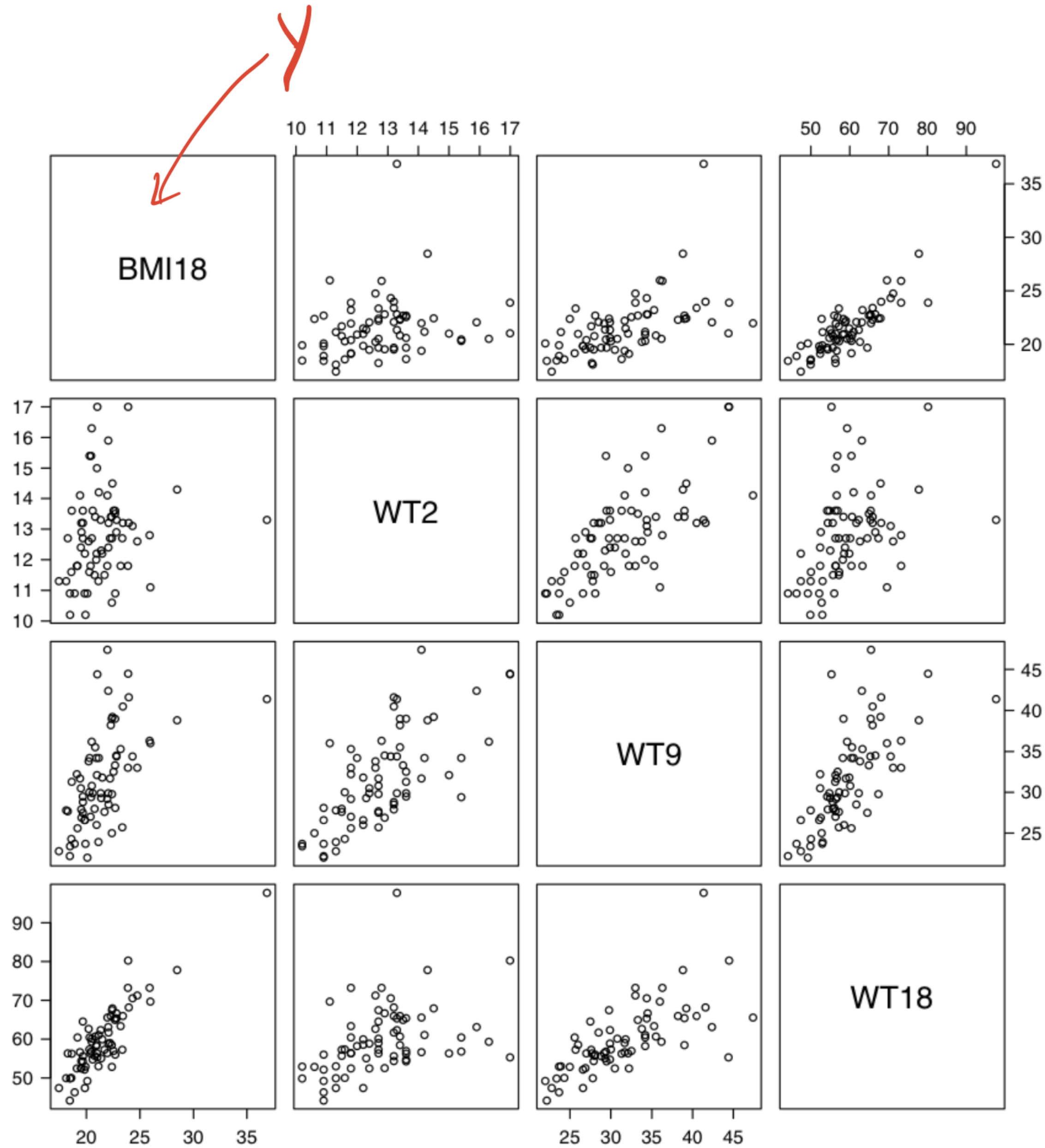

Figure 4.2 Scatterplot matrix for the girls in the Berkeley Guidance Study.

WT2, WT9, WT18 are covariates

Table 4.1 Regression of BMI18 on Different Combinations of Three Weight Variables for the n = 70 Girls in the Berkeley Guidance Study

Regressor	Model 1	Model 2	Model 3
(Intercept)	8.298*	8.298*	8.298*
WT2	–0.383* suspect	–0.065 NS	–0.383*
WT9	0.032		0.032
WT18	0.287*		0.287*
DW9		0.318*	Aliased
DW18		0.287*	Aliased

*Indicates p-value < 0.05.

WT2, WT9, WT18: correlated?

not discussing yet

WT2 = Weight at age 2

DW9 = WT9 − WT2 = Weight gain from age 2 to 9

DW18 = WT18 − WT9 = Weight gain from age 9 to 18

Redefine covariates

Collinearity

Let $X_{n\times p}$ be data matrix of covariates from sample.
If we can find a vector of constants "a" such that $X_a \approx 0$

$$\Rightarrow \quad \text{Covariates are collinear}$$

if $X_a = 0$

$$\Rightarrow \quad \text{over-parameterized model}$$

1988-2011

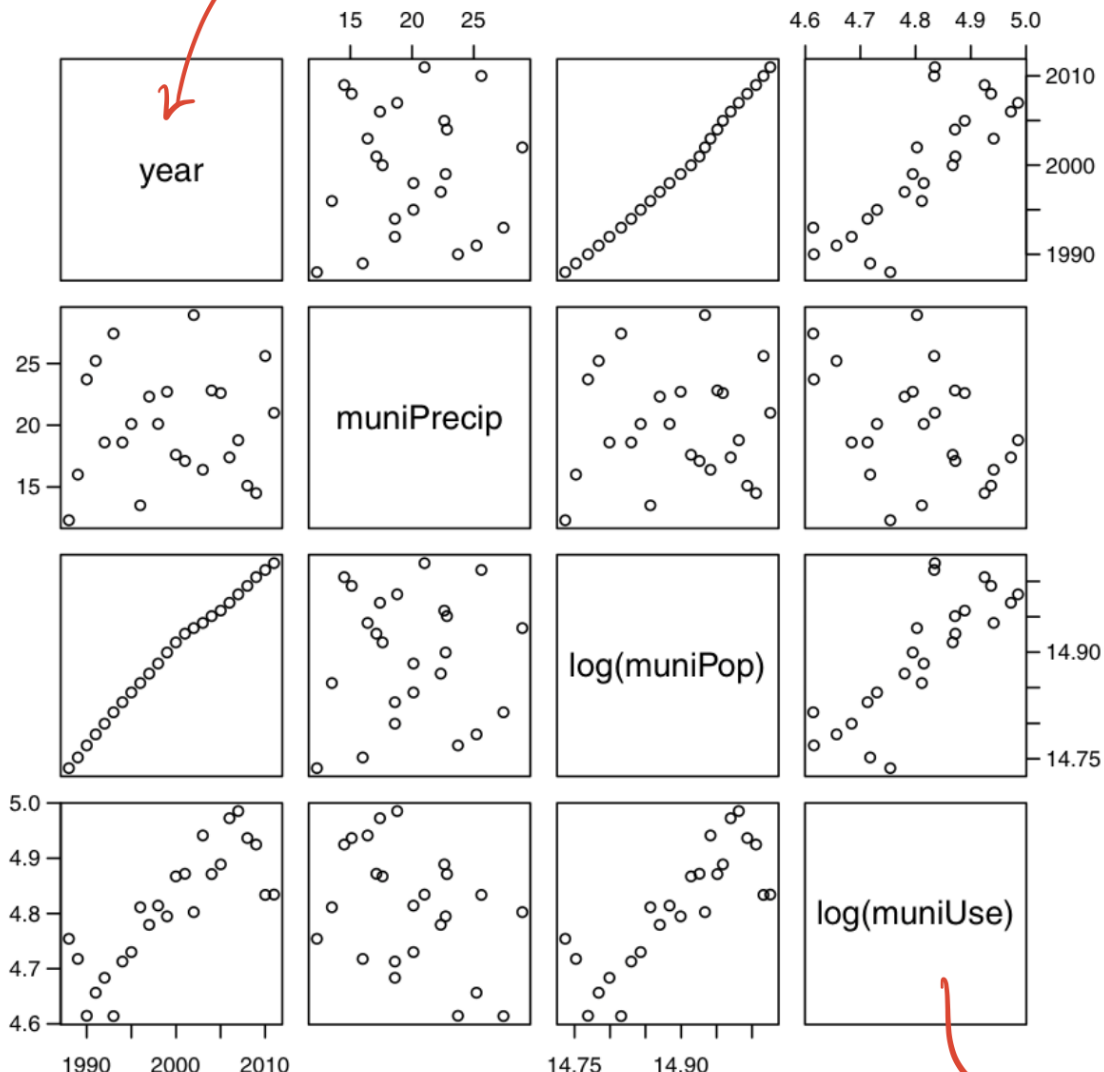

Figure 4.3 Scatterplot matrix for the Minnesota water use data.

y

Suffers from collinearity

Table 4.2 Regression of log(`muniUse`) on Different Combinations of Regressors for the Minnesota Water Use Data

Regressor	Model 1	Model 2	Model 3
(Intercept)	-20.0480^*	-20.1584^*	-1.2784
`year`	0.0124^*	0.0126^*	-0.0111
`muniPrecip`		-0.0099^*	-0.0106^*
log(`muniPop`)			1.9174

*Indicates p-value < 0.01.

Very similar
corr(year, muniPrecip) ≈ 0

Regressors On Logarithmic Scale

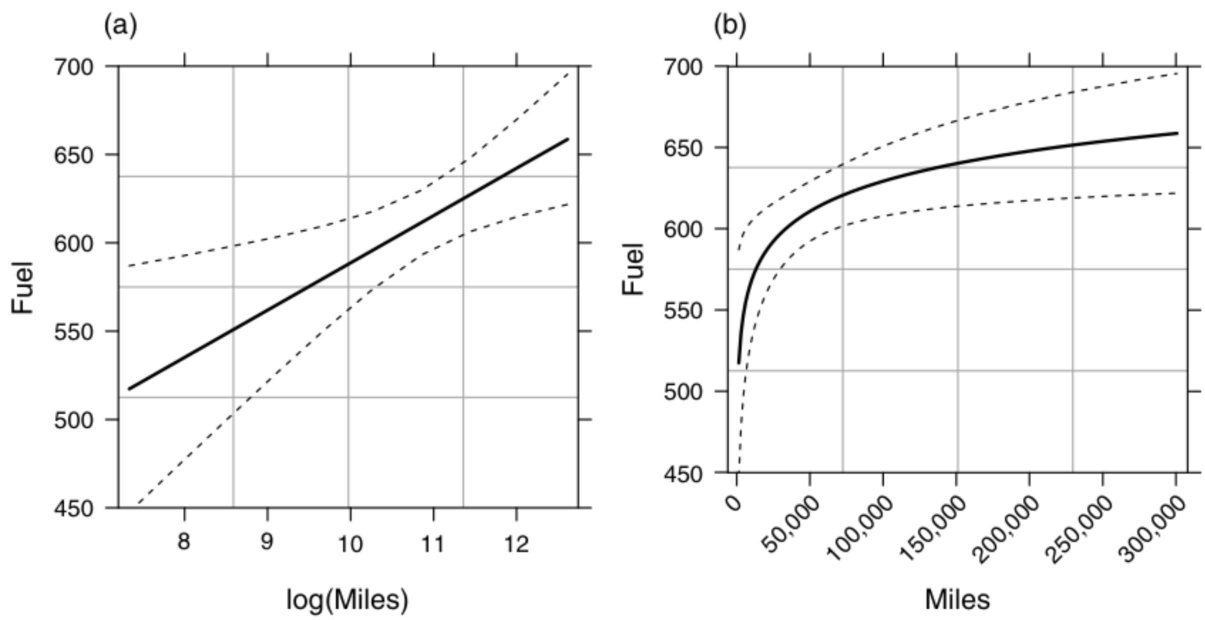

Figure 4.4 (a) Effects plot for log(`Miles`) in the fuel consumption data. (b) The horizonal axis is given in the more useful scale of `Miles`, and thus the fitted effect is a curve rather than a straight line.

Usual effect of logarithms:

fitted effects that change most rapidly when predictor is small

Response on Logarithmic Scale

Consider model

$$E(\log(Y) \mid X_j = x_j, \underbrace{X_{(j)} = x_{(j)}}_{\text{predictors excluding}}) = \beta_0 + \beta_j x_j + \beta'_{(j)} \underbrace{x_{(j)}}_{\textit{vector}}$$

Now approximate LHS by log of expected value:

$$\log E(Y \mid X_j = x_j, X_{(j)} = x_{(j)}) \approx E\left[\log(Y) \mid X_j = x_j, X_{(j)} = x_{(j)}\right]$$

Exponentiate both sides:

$$\begin{aligned}
E\left[Y \mid X_j = x_j, X_{(j)} = x_{(j)}\right] &\approx \exp\left(E\left[\log(Y) \mid X_j = x_j, X_{(j)} = x_{(j)}\right]\right) \\
&= \exp\left(\beta_0 + \beta_j x_j + \beta'_{(j)} x_{(j)}\right) \\
&= \exp(\beta_j x_j)\exp\left(\beta_0 + \beta'_{(j)} x_{(j)}\right) \\
\Rightarrow E\left[Y \mid X_j = x_j + 1, X_{(j)} = x_{(j)}\right] &\approx \exp(\beta_j(x_j+1))\exp\left(\beta_0 + \beta'_{(j)} x_{(j)}\right) \\
&= \exp(\beta_j)\exp(\beta_0 + \beta_j x_j + \beta'_{(j)} x_{(j)}) \\
&= \exp(\beta_j)\left[E\left(Y \mid X_j = x_j, X_{(j)} = x_{(j)}\right)\right]
\end{aligned}$$

$\therefore$ increasing any x_j by 1 will multiply the mean of Y by approximately $\exp(\beta_j)$.
Can express as a percent change:

$$\begin{aligned}
&100 \times \frac{E\left[Y \mid X_j = x_j + 1, X_{(j)} = \chi_{(j)}\right] - E\left[Y \mid X_j = x_j, X_{(j)} = \chi_{(j)}\right]}{E\left[Y \mid X_j = x_j, X_{(j)} = \chi_{(j)}\right]} \\
&= 100\left(\exp(\beta_j) - 1\right)
\end{aligned}$$

Example

If $\beta_j = 0.30 \qquad \Rightarrow 100\,(\exp(\beta_j) - 1) = 34\%$
If $\beta_j = -0.20 \quad \Rightarrow 100\,(\exp(\beta_j) - 1) = -18\%$
Note:
If both Y and x_j are on the log scale, then $x_j = x_{j+1}$

$$\Rightarrow \text{multiply } x_j \text{ by } e = 2.718\ldots$$

(rarely makes sense)

Dropping Regressors

If regressors are changed, then so are parameters and their interpretations (usually).
If $E\,(Y \mid X_1 = x_1, X_2 = x_2)$ is correct $= \beta_0 + \beta_1 x_1 + \beta_2 x_2$
What can we say about $E(Y \mid X_1 = x_1)$?
Can write:

$$E(Y \mid X_1 = x_1) = E\,[E(Y \mid X_1 = x_1, X_2) \mid X_1 = x_1]$$

$$= \beta_0 + \beta_1' x_1 + \underbrace{\beta_2' E(X_2 \mid X_1 = x_1)}_{\therefore \text{ cannot simply drop } X_2 \text{ regressors from correct model}}$$

Variances when Regressors Dropped

$$\mathrm{var}(Y \mid X_1 = x_1)$$

$$= E\,[\mathrm{var}(Y \mid X_1 = x_1, X_2) \mid X_1 = x_1]$$

$$+\mathrm{var}\,[E(Y \mid X_1 = x_1, X_2) \mid X_1 = x_1]$$

$$= \sigma^2 + \beta_2^2\,\mathrm{var}(X_2 \mid X_1 = x_1)\beta_2$$

Sampling From a Normal Population

Suppose

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \text{cov}(X,Y) \\ \text{cov}(X,Y) & \sigma_Y^2 \end{pmatrix} \right)$$

$\therefore$ assume data pairs $\{(x_i, y_i); i = 1, \ldots, n\}$ are realizations of bivariate normal random variables X, Y.

Can show

$$Y_i \mid X_i \sim \mathcal{N}\left(\mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(X_i - \mu_X), \sigma_Y^2(1 - \rho_{XY}^2) \right)$$

Where

$$\text{cov}(X,Y) = \rho_{XY}\sigma_X\sigma_Y$$

Now define

$$\beta_0 = \mu_Y - \beta_1\mu_X$$

$$\beta_1 = \rho_{XY}\frac{\sigma_Y}{\sigma_X}$$

$$\sigma^2 = \sigma_Y^2(1 - \rho_{XY}^2)$$

$$\Rightarrow Y_i \mid X_i \sim \mathcal{N}(\beta_0 + \beta_1 X_i, \sigma^2)$$

$\equiv$ SLR with normality assumption added.

$\sigma^2 = \sigma_Y^2(1 - \rho_{XY}^2)$ called residual variance

(part of Y not explained by X)

Usual sample estimates under random sampling

$$\hat{\mu}_X = \bar{x} \quad \hat{\mu}_Y = \bar{y}$$

$$\hat{\sigma}_X^2 = S_X^2 \quad \hat{\sigma}_Y^2 = S_Y^2$$

$$\hat{\rho}_{XY} = r_{XY}$$

Maximum Likelihood Estimators (MLE)

$$Y_i \mid X_i = \mathcal{N}(\beta' X_i, \sigma^2) \quad \text{[more general MLR case]}$$

$$f(Y_i \mid X_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(Y_i - \beta' X_i)^2}{2\sigma^2}\right]$$

from sample of (independent) data $\{(x_i, y_i); i = 1, \ldots, n\}$

$$L(\beta, \sigma^2 \mid y_1, \ldots, y_n) = \prod_{i=1}^{n} f(y_i \mid x_i; \beta, \sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta' x_i)^2\right]$$

$$\log L(\beta, \sigma^2 \mid y_1, \ldots, y_n) = -\frac{n}{2}(\log 2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta' x_i)^2$$

What this shows is that

$$\hat{\beta}^{\text{MLE}} = \hat{\beta}^{\text{OLS}}$$

Then

$$\hat{\sigma}^2_{\text{MLE}} = \text{RSS}/n \quad \left[\text{not } \frac{\text{RSS}}{(n - (p+1))}\right]$$

MLE theory has some attractive theoretical properties

e.g., UMVUE, normality or asymptotic normality

MLR and R^2

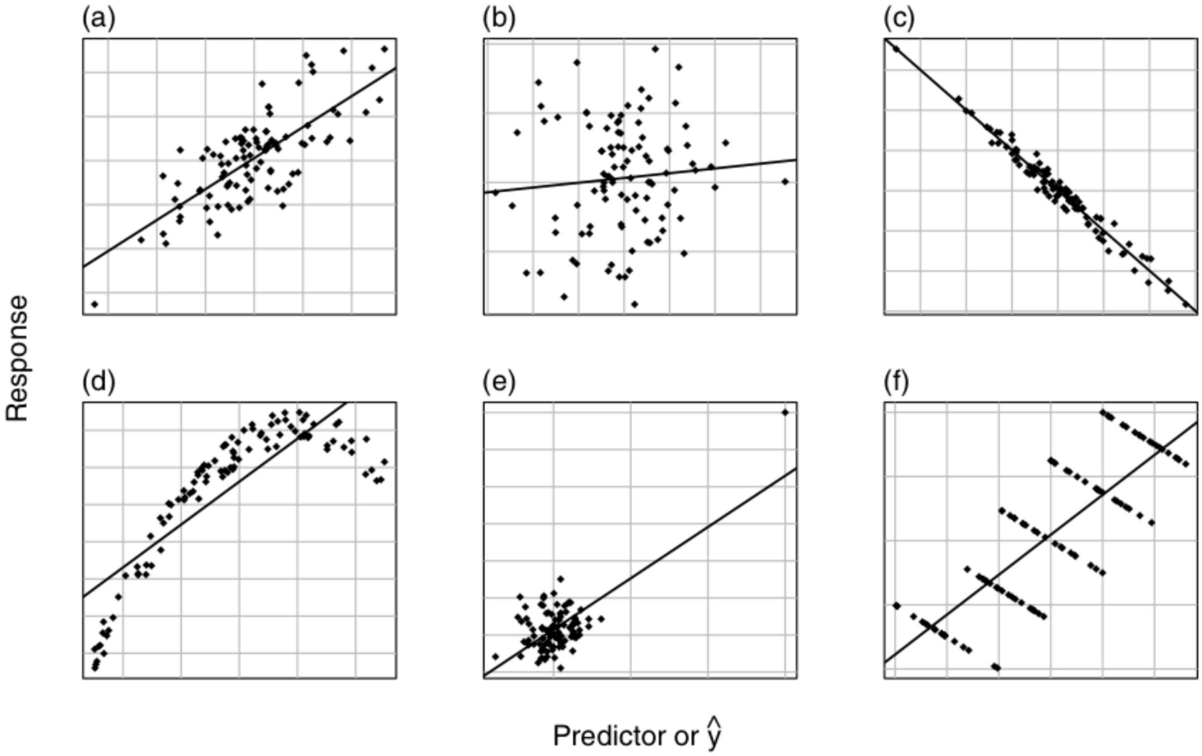

Figure 4.6 Six summary graphs. R^2 is an appropriate measure for a–c, but inappropriate for d–f.

$$R^2 \equiv \text{corr}(Y, \hat{Y}) \quad \text{[one can show this]}$$

Regression through Origin and R^2

$$\text{prop of variability explained} = 1 - \frac{\text{RSS}}{\sum_{i=1}^{n} y_i^2}$$

Not invariant under location change.

(e.g., going from °F to °C, R^2 changes)

$\therefore$ R^2 use here not recommended.