4.1.3 Berkeley Guidance study

- Remark:
  - Aliased:
    - Some predictors are linear combinations of other predictors
  - Multicollinearity:
    - Some predictors are highly correlated.
    - Perfect multicollinearity (correlation=1) is equivalent to being aliased.

$$X = \begin{array}{c} \begin{matrix} x_1 & x_2 & z \end{matrix} \\ \begin{pmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix} \end{array}$$

$z = x_1 + x_2$

- Mathematically,
  - When two columns of Matrix X are similar (correlated predictors), $(X'X)^{-1}$ is unstable.
  - Reason: $(X'X)$'s determinant is close to 0.
  - Therefore
    - Estimator of β (i.e., $(X'X)^{-1}X'Y$) is unstable
    - Variance $\sigma^2(X'X)^{-1}$ can be very large

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

- Intuitively,
  1. Soma=1.59-0.116 WT2+0.056 WT9+0.048 WT18
  2. Soma=1.59-(0.116+b) WT2+(0.056+b) WT9+0.048 WT18 - b DW9
  - Any b in model 2 is essentially model 1!
  - Therefore
    - Estimator of β, $(X'X)^{-1}X'Y$ is unstable
    - Variance $\sigma^2(X'X)^{-1}$ can be very large

Conclusion: Check for multicollinearity by studying predictors' correlation!

4.1.6 Dropping terms

- What happens when a bigger model is fit to the data from a smaller model?
  - Data:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+e$ $+0X_3+0X_4$
  - Model:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+e$

$$\boxed{\textit{Answer}:\quad \hat{\beta}_3\approx 0,\hat{\beta}_4\approx 0}$$

- What happens when a smaller model is fit to the data from a bigger model?

Conditional Expectation

- Tower property

$$\boxed{E\left(\underline{E(Y \mid X_1, X_2)} \mid X_2\right) = E(Y \mid X_2)}$$

Proof: (optional)

$$E\left(E(Y \mid X_1, X_2) \mid X_2\right)$$

$$= E\left(\int \frac{y f_{Y,X_1,X_2}(y, X_1, X_2)}{f_{X_1,X_2}(X_1, X_2)} dy \,\middle|\, X_2 \right)$$

$$= \int \frac{f_{X_1,X_2}(X_1, X_2)}{f_{X_2}(X_2)} \int \frac{y f_{Y,X_1,X_2}(y, X_1, X_2)}{f_{X_1,X_2}(X_1, X_2)} dy\, dx_1$$

$$= \int \frac{y f_{Y,X_2}(y, X_2)}{f_{X_2}(X_2)} dy$$

$$= E(Y \mid X_2)$$

4.1.6 Dropping terms

- Observational analysis
  - Variables are observed via sampling.
  - Beyond the control of experimenter.
  - Cannot avoid lurking variable
    - Variables that are useful but are ignored in the regression model
  - True relationship:

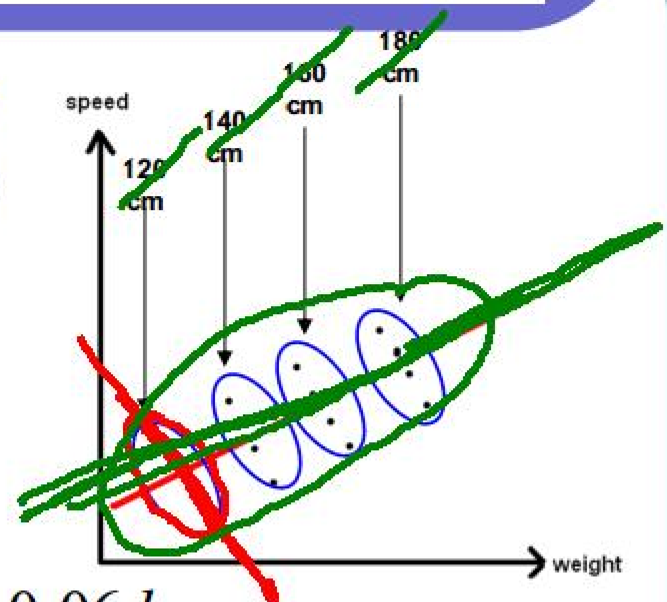
    $E(Y \mid X = x, L = l) = \beta_0 + \beta_1 x + \delta l$
  - Wrong model used (L is lurking variable)

$$E(Y \mid X = x) = \beta_0 + \beta_1 x + \delta \underline{E(L \mid X = x)}$$

$$= (\beta_0 + \delta\gamma_0) + (\beta_1 + \delta\gamma_0)x \qquad \text{if } E(L \mid X = x) = \gamma_0 + \gamma_1 x$$

$$E(\underline{\beta_0 + \beta_1 x} + \delta l \mid X = x)$$

- Example of lurking variable
  - Y: Maximum running speed (m/s)
  - X: Weight (lbs)
  - L: Height (cm)



true

- True model

$$E(Y \mid X = w, L = h) = 2 - 0.05w + 0.06h$$

- Wrong model used

$$E(Y \mid X = w) = 2 - 0.05w + 0.06\overbrace{E(L \mid X = w)}^{1.5w}$$

$$= 2 + 0.04w \qquad [\text{if } E(L \mid X = w) = 1.5w]$$

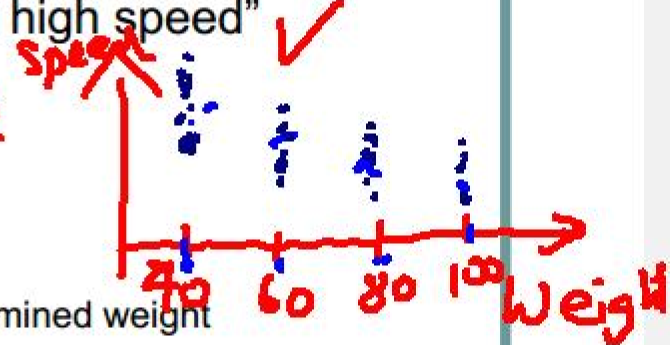
estimated

4.2 Experimentation v.s. Observation

- In observational study
  - May have unknown effect of lurking variables.
  - Can't draw causal conclusion
    - e.g. Cannot say "Higher weight cause people to run faster!"
  - Only can say about association
    - e.g. Can say "High weight is associated with high speed"

- In experiments

Speed v.s. weight



  - Have control on every aspect.
    - e.g.
      - Randomly assign people to achieve some pre-determined weight
      - Measure their running speed.
  - Lurking variables' effect are averaged out by random assignment.
  - Can draw causal conclusion
    - e.g. "higher weight causes lower speed"

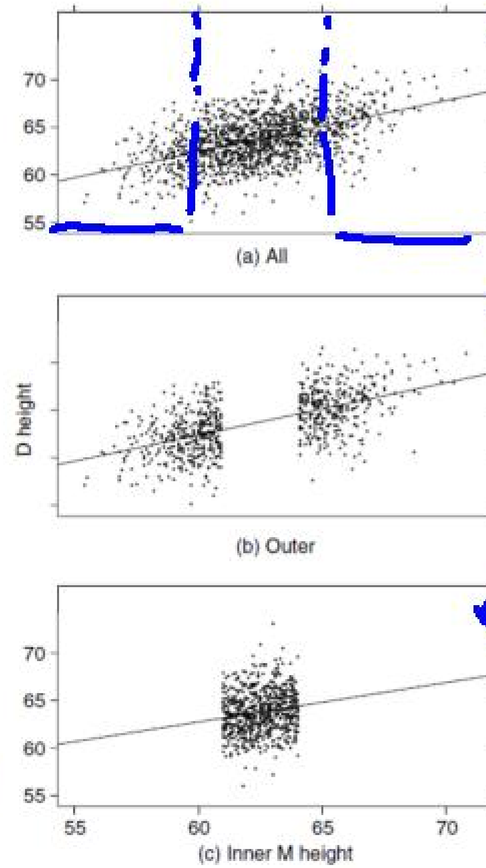
4.4 More on R^2

- R^2 tends to be large if the X are dispersed
- R^2 tends to be small if the X are concentrated
- Therefore, need to be careful about sampling!

$$SYY = SSreg + RSS$$

$$\Sigma(Y_i-\bar{Y})^2 \quad \Sigma(\hat{Y}_i-\bar{Y})^2 \quad \Sigma(Y_i-\hat{Y}_i)^2$$

$$R^2 = \frac{SSreg}{SYY} \in [0,1]$$



← whole data

$R^2=0.24$

← ignore middle

$R^2=0.37$

← only middle

$R^2=0.027$

- R^2 is useful to measure goodness of regression fit if and only if the scatterplot looks like a sample from a bivariate normal distribution (elliptical shaped)

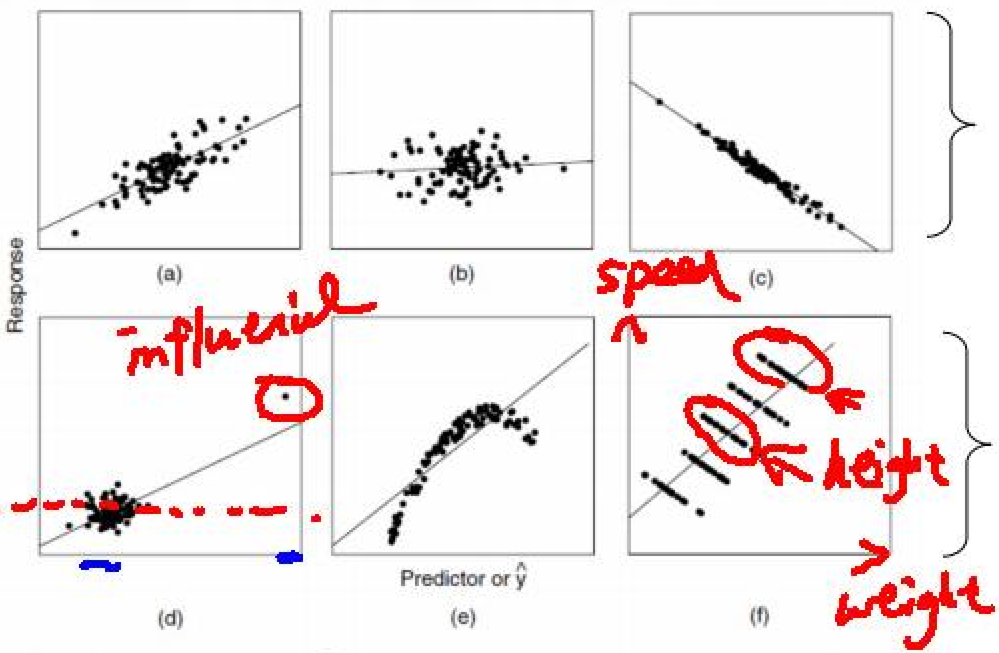


FIG. 4.3 Six summary graphs. R^2 is an appropriate measure for a–c, but inappropriate for d–f.

useful

Not useful

- **Leverage point**
- **Curve**
- **Lurking variable**