Chapter 4

Drawing Conclusion

4.1 Understanding parameter estimates parameters=($\beta_0, \beta_1, \ldots \beta_p, \sigma^2$)

$$E(Y|X) = \beta_0 X_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$E(Fuel|X) = 154.19 - 4.23\,Tax + 0.47\,Dlic - 6.14\,Income + 18.54\log(Miles)$$

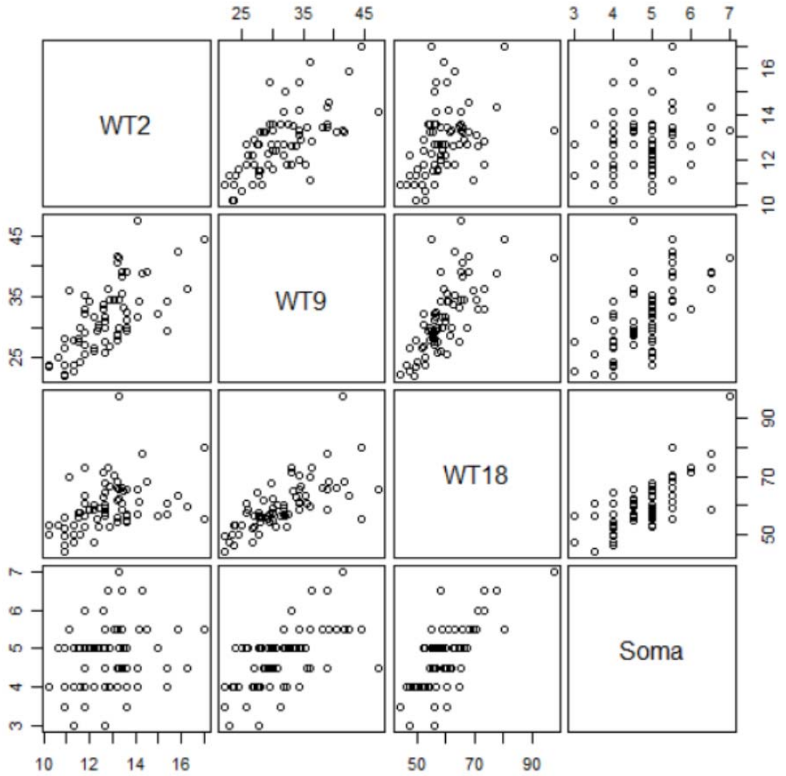
- Unit of βs:
  - unit of y/unit of x. (e.g. gallon/$1000 for -6.14)
- Unit of σ^2:
  - (unit of y)^2
- Meaning of β_i:
  - Rate of change of y on x_i , after adjusting for other variables
  - e.g.
    - Fuel decreased by 6.14 gallon when Income increased by $1000
    - Fuel increased by 18.54 gallon when Miles is doubled ($\log_2 2x = 1 + \log(x)$)
- Meaning of $\sigma^{2:}$
  - Variability that can't be explained by the regression line.

4.1.3 Berkeley Guidance study

- To Illustrate: interpretation of regression depends on other terms

- Background
  - Study the growth of boys and girls
  - Y=Soma=Body type -- 1 to 7 (thin to fat)
  - Predictors:
    - WT2 = weight at age 2
    - WT9 = weight at age 9
    - WT18 = weight at age 18

- First thing to do: Scatterplot
  - Roughly linear relationship



4.1.3 Berkeley Guidance study

- Second thing to do: Fit a regression
  - Soma=1.59-0.116 WT2+0.056 WT9+0.048 WT18
- What?
  - Heavier at age 2 → thinner now? (-0.116)
  - Explanations:
    - p-value = 0.06, not very significant
    - May be due to correlations between the terms (WT2, WT9 and WT18 are correlated)

4.1.3 Berkeley Guidance study

- Second thing to do: Fit a regression
  - Soma=1.59-0.116 WT2+0.056 WT9+0.048 WT18
- Modification: Define new terms
  - DW9 = WT9 - WT2 = Weight gain from age 2 to 9
  - DW18 = WT18 - WT9 = Weight gain from age 9 to 18
  - Hope the correlation between WT2, DW9 and DW18 are smaller
  - Soma=1.59 - 0.011 WT2+0.105 DW9+0.048 DW18
    - Now the coefficient -0.0111 is not significant
    - Soma depends on DW9 and DW18 but not WT2
- Conclusion:
  - Interpretation of the effect of a variable depends on other variables.

4.1.3 Berkeley Guidance study

- Remark: If all five variables are included. We get
  - Soma=1.59-0.116 WT2+0.056 WT9+0.048 WT18+ NA DW9 +NA DW18
  - Since the model is unidentifiable!
    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!
    - Reason: DW9 and DW18 are linear combination of WT2, WT9 and WT18, they do not contribute to any extra information
    - In this case the terms are said to be aliased. Some irrelevant terms need to be dropped out.
  - Try the R-code
    - x1=rnorm(100); x2=rnorm(100); e=rnorm(100,0,0.1);y=1+2*x1+3*x2+e; z=x1+2*x2
    - lm(y~x1+x2)
    - lm(y~x1+x2+z)
    - lm(y~z+x1+x2)

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.
  - 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
  - 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:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+e$
  - Model:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+e$

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

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

4.1.6 Dropping terms

- What happens when a smaller model is fit to the data from a bigger model?
  - Data (truth):
    - A=Area, L=Length, W=Width
    - Area=Length*Width
    - $E(\log(A)|L,W)=0+(1)\log(L)+(1)\log(W)$
  - Model:
    - $E(\log(A)|L)=\beta_0+\beta_1\log(L)$

4.1.6 Dropping terms (Mean function)

- Data: E(log(A)|L,W)=0+(1)log(L)+(1)log(W)
- Model: E(log(A)|L)= β_0+ $\beta_1$log(L)

- From the true model, we get
  - E(log(A)|L)=E(E(log(A)|L,W)|L) (Tower property)

    =E(log(L)+log(W)|L) (True relationship)

    =log(L)+E(log(W)|L) (E(log(L)|L)=L)
  - If L and W are independent, E(log(W)|L)= E(log(W))= c,
    - Data: E(log(A)|L)=c+log(L), $\left(\hat{\beta}_0 \to c,\ \ \hat{\beta}_1 \to 1\right)$
  - If E(log(W)|L)=d_0+$d_1$log(L),
    - Data: E(log(A)|L)=d_0+(1+d_1)log(L) $\left(\hat{\beta}_0 \to d_0,\ \ \hat{\beta}_1 \to (1+d_1)\right)$

Conditional Expectation

- Tower property

$$\boxed{E\big(E(Y \mid X_1, X_2) \mid X_2\big) = E(Y \mid X_2)}$$

Proof:

$$
\begin{aligned}
& E\big(E(Y \mid X_1, X_2) \mid X_2\big) \\
&= E\left(\int \frac{y f_{Y,X_1,X_2}(y, X_1, X_2)}{f_{X_1,X_2}(X_1, X_2)} dy \,\middle|\, X_2 \right) \\
&= \int \frac{f_{X_1,X_2}(X_1, X_2)}{f_{X_2}(X_2)} \int \frac{y f_{Y,X_1,X_2}(y, X_1, X_2)}{f_{X_1,X_2}(X_1, X_2)} dy\, dx_1 \\
&= \int \frac{y f_{Y,X_2}(y, X_2)}{f_{X_2}(X_2)} dy \\
&= E(Y \mid X_2)
\end{aligned}
$$

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:

$$E(Y \mid X = x, L = l) = \beta_0 + \beta_1 x + \delta\, l$$

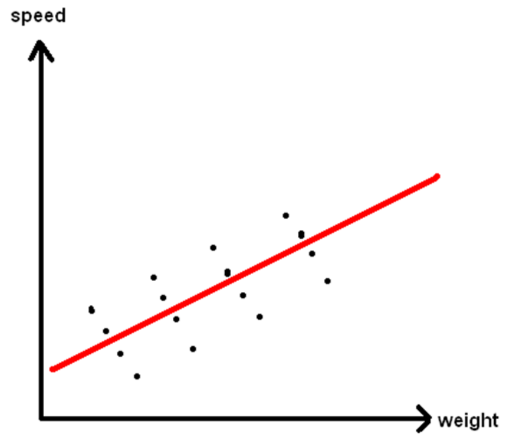
  - Wrong model used (L is lurking variable)

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

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

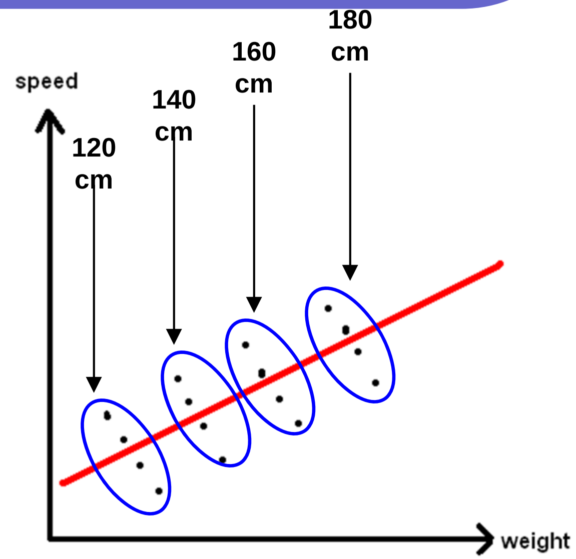
4.1.6 Dropping terms

- Example of lurking variable
  - Study:
    - Y: Maximum running speed
    - X: Weight
  - What you might expect
    - $Y=\beta_0 + \beta_1 X + e, \quad \beta_1 < 0$
  - What you found:



4.1.6 Dropping terms

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

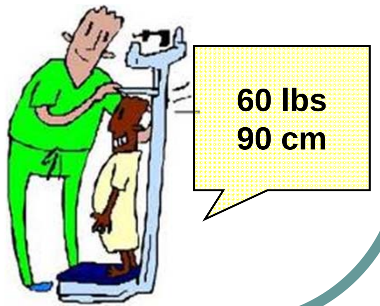


- True model

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

- Wrong model used

$$E(Y \mid X = w) = 2 - 0.05w + 0.06E(L \mid X = w)$$
$$= 2 + 0.04w \qquad [\text{if } E(L \mid X = w) = 1.5w]$$



4.1.6 Dropping terms

- Conclusion
  - If a simple model is fit to data with complicated structure,
    - Estimated parameter may not tell the true effect of a variable.
  - Therefore, when we get non-null residual plots, beware of
    - The non-linear relationship between response and terms
    - Useful predictor/terms that are not included in the model but are correlated with other terms in the model.
      - These variables are called lurking variables
      - E.g.
        - Data: $E(\log(A)|L,W)=0+(1)\log(L)+(1)\log(W)$
        - Model: $E(\log(A)|L)=\beta_0+\beta_1\log(L)$
        - W is the lurking variable

4.2 Experimentation v.s. Observation

- Experimental analysis
  - Predictors (X) under the control of the experimenter.
  - Assignment based on randomization scheme.
  - Examples
    - Agricultural study: amount of fertilizers, water, space…

- Observational analysis
  - Variables are observed via sampling.
  - Beyond the control of experimenter.
  - Examples
    - Agricultural study: Soil fertility, temperature

4.2 Experimentation v.s. Observation

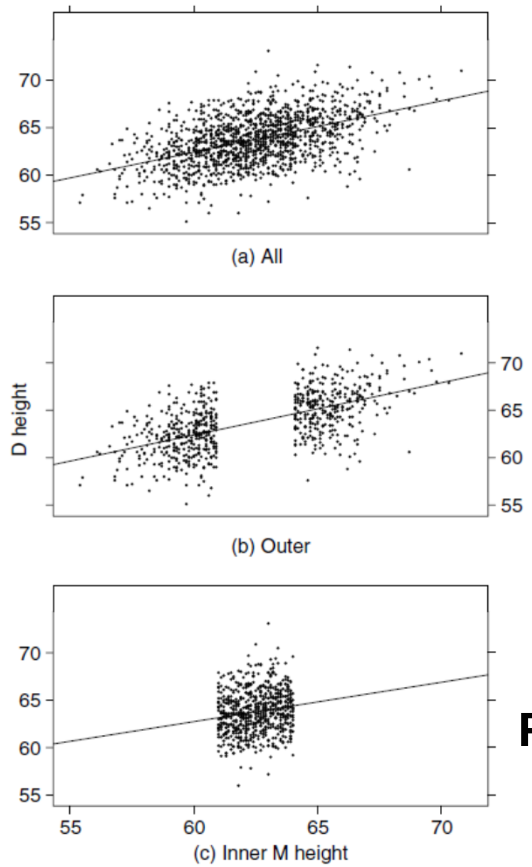
- In observational study
  - May have unknown effect of lurking variables.
  - Can’t draw causal conclusion
    - e.g. Cannot say “Higher weight cause people to run faster!”
  - Only can say about association
    - e.g. Can say “High weight is associated with high speed”
- In experiments
  - Have control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

4.2 Experimentation v.s. Observation

- More example: “Mobile phone decrease brain activity?”
  - Observational study
    - Find a group of people
      - Measure their brain activities (Y)
      - Measure their habit of using mobile phone – hours/week (X)
      - Regress Brain activities (Y) on time (X)
    - Only: “More phone usage is associated with lower brain activities ”
  - Experiments
    - Find a group of people
      - Randomly assign them into groups
        - Randomization sometimes helps average out unknown lurking variables.
      - For each group, force them to use mobile phone for different amount of time (X)
        - Sometime not ethical to use…
      - Measure brain activities (Y)
      - Regress Brain activities (Y) on time (X)
    - Ok: “More phone usage cause lower brain activities”

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!



R^2=0.24

R^2=0.37

R^2=0.027

4.4 More on R^2

- 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)

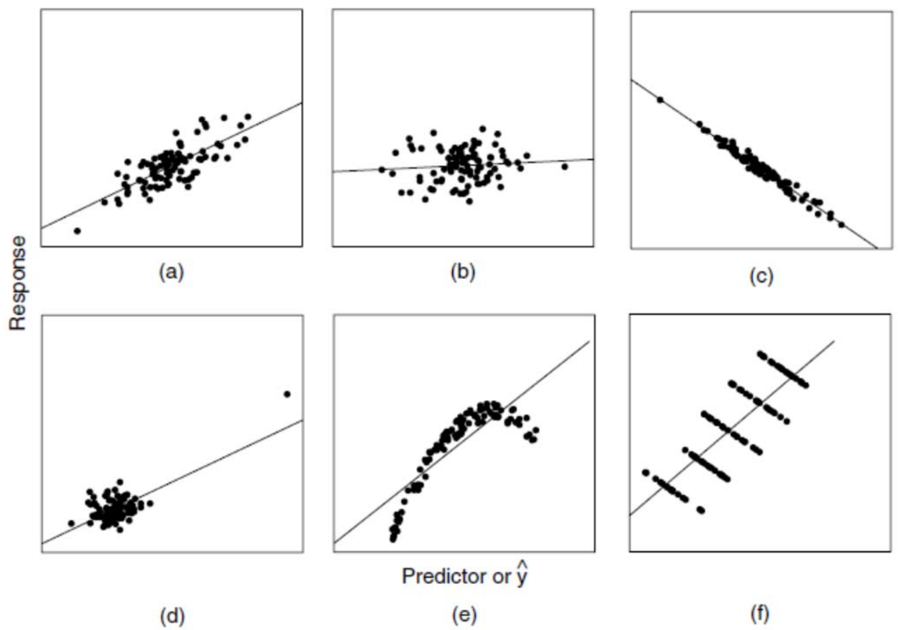


FIG. 4.3 Six summary graphs. R^2 is an appropriate measure for a–c, but inappropriate for d–f.

useful

Not useful

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

Summary of Chapter 4

- Result of regression depends on
  - which predictors are included in the model
  - the relationship between the terms
- Drawing conclusions
  - Observational studies – association
  - Experiments – causal relationship
- New vocabularies
  - Aliased
  - Multicollinearity
  - Lurking variables