Preamble / Motivation / ...

- Easy to carry out (just click!)
- Easy to be "glib" about what it accomplishes
- BUT ... WHY use it ??? HOW to explain to father-in-law?

• ***If interested in separate contributions of each of several variables...***

are there any situations where one can assess them one at a time? *i.e.*

assess a particular X while ignoring the others ...
assess a different X while ignoring the others ... ?

or does one have to assess them simultaneously ?

• ***If interested in ("net") contribution of ONE particular variable...***

are there situations where one can assess it while ignoring the others ... ?

or does one always have consider the other X's as well ?

Answers ... Illustrated by examples

- birthweight as function of gestational age and gender
- weight in relation to age and height
- breast milk and subsequent IQ in children born preterm
- increase in heating costs after adding a room to a house
- decrease in longevity if greater amount of sexual activity

Multiple Regression Equation

$$Y_{X1\ \ X2\ ...} = \mu_{Y\ |\ X1\ \ X2\ ...} + \varepsilon$$

$$\mu_{Y\ |\ X1\ \ X2\ ...} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...$$

How to describe it ...

in words / symbols

$\mu_{Y\ |\ X1\ X2}$ .. as a function of X1 X2 ..
(don't forget the ε's with SD σ about μ's)

geometrically

"plane" or "surface" of means

(in case of 2 X's) without leaving "2-D"

as contour map (cf web page)

using links to simpler procedure..

as a sequence of simple linear regressions
(but be careful: see my notes on Ch 2/9 of M&M)

Meaning of β_i

$$\beta_i = \frac{\delta \mu_{Y \mid X1 \; X2 \ldots}}{\delta X_i}$$

difference in μ_Y for a 1 unit difference in X_i but no difference in other X 's , i.e. all other X 's held "constant"

Main Purposes

- Summarization / Description
- Adjustment (Bias Reduction)
- Increased Precision of estimates of specific β_i 's
  (by removing extraneous variation)
- Prediction
- Interpolation / Smoothing
  "borrowing strength" (e.g. estimates of outcome of prostate cancer if sparse data in some age-histologic grade "cells")
- Polynomial Regression
  (several powers of 1 X -- each power is a *term* in regression; can also have other X's in equation)

Assumptions

see G&S page 54 [page 58 in 2nd ed] ;
see also comments in my notes on ch. 3

Parameters / Estimates of these *(by computer!)*

1. β_0 — $b_0 \pm t \, SE[b_0]$

   [β_0 seldom of interest]

2. β_i — $b_i \pm t \, SE[b_i]$

3. $\sigma_{Y \mid X1 \, X2 \ldots}$ — $\sqrt{\dfrac{\Sigma (y_i - [b_0 + b_1 x_i + b_2 x_2 \ldots])^2}{n - \# \text{ of b's fitted}}}$

   ("Root Mean Squared Error")

4. $\mu_{Y \mid X1 \, X2 \ldots}$ — $b_0 + b_1 X_1 + b_2 X_2 + \ldots$

   $\pm$ t SE[thereof]

5. $Y_{\mid X1 \, X2 \ldots}$ — $b_0 + b_1 X_1 + b_2 X_2 + \ldots$

   $\pm t \, SE[b_0 + b_1 X_1 + b_2 X_2 + \ldots + \varepsilon]$

   (Interval for **Y|X wider** than for $\mu_{Y|X}$)

Multiple Correlation Coefficient

- a helpful way to look at least squares estimate (scalar)

$R_{\text{Y and best linear combination of X's}}$