G&S Chapter 2: Simple Linear Regression...

- see my notes on G&S Chapter 2

- see my notes on M&M Chapters 2/9

Parameters of interest	Estimates of these
1. $_0$	b_0 $\pm$ t SE[b_0]
2. $_1$	b_1 $\pm$ t SE[b_1]
3. $_{Y|X}$	$\sqrt{\dfrac{(y_i - [b_0 + b_1 x_i])^2}{n - 2}}$

("Root Mean Squared Error")

4*. $\mu_{Y|X} = {}_0 + {}_1 X$ $b_0 + b_1 X$ $\pm$ t SE[$b_0 + b_1 X$]

5*. $Y_{|X} = {}_0 + {}_1 X + $ $b_0 + b_1 X$ $\pm$ t SE[$b_0 + b_1 X + $]

Notes / Practicalities:

• Finding these on printouts

• Assumptions - see notes

• Notation for Residual variation

  - MSE = average squared "error"

  - Root Mean Squared Error = MSE

      = square root of average squared error

      = "average" error (roughly!)

      = "average" residual

  - why n-2 df? see notes

• Precision of parameter estimates

  - what factors affect it

4* and 5*

Not covered in G&S1 (see next page)

Covered in G&S2

• Confidence interval for the **Line** of Means (p37)

• Confidence interval for an **Observation** (p38-9)

"Estimate of **mean** Y at specific X_0 value"

"Estimate of **individual** Y at specific X_0 value"
    (X_0 *value not necessarily in dataset)*

- Estimate of $\mu(Y \mid$ specific X_0 value)

  - SE of estimate (and thus width of CI) is larger the further X_0 is from $\bar{x}$, the mean of x values in dataset.

  - CI's are for $\mu(Y)$ at a specific X_0 value not for entire line (see Neter)

- "Prediction" of "new" Y value at $X = X_0$

  - best (point) estimate is estimate of $\mu(Y \mid$ specific value of X)

  BUT…

  Y is an INDIVIDUAL value (contains an) !
  - not influenced by size (n) of dataset used to estimate $\mu(Y \mid X_0)$

  - will vary around true $\mu(Y \mid X_0)$ according to the SD of possible 's, i.e., the SD of INDIVIDUAL Y's at $X=X_0$

SO…

  Uncertainty re "new" Y is amalgam of

  - SD of INDIVIDUALS

  - uncertainty in estimate of $\mu(Y \mid X_0)$

  and is (almost) independent of n used to estimate $\mu(Y \mid X_0)$

- Obtaining these CI's from packages

- Examples:

  Alcohol and Eye Movement

  Noninvasive prediction
  (see "Appropriate use of prediction bands" letter to NEJM … on web page under Resources)

"CENTERED" version of regression equation

ANALYSIS of VARIANCE TABLE;

   R^2 …   F test …

Method Comparisons

- DON'T use correlation or regression
  (see article on Method Agreement)