11.7 Prediction in SLR

- Consider a slightly different problem. We wish to predict the 20-km race time for a single skier who has a 10½-min. time to exhaustion.
- Here we are not estimating a fixed quantity (like $E(Y)$) but rather predicting a r.v. (call it Y^*) at a specific value x^* of x.
- A natural prediction formula would be

- Consider then the prediction error

- This is the difference between two ____________ ________ r.v.'s, so it has a ____________ distribution.

$E[y^* - \hat{y}^*] =$

And $V[y^* - \hat{y}^*] =$

- Similarly to past arguments,

and replacing σ^2 by its estimator MSE, Hence

Isolating y^* in the middle with algebra, we get a $100(1-\alpha)\%$ <u>prediction interval</u> for y^* corresponding to $x = x^*$:

<u>Note:</u> This is almost equivalent to the CI formula for $E(Y)$ at x^*, except for the extra "$1+$" in the standard error part.

- Thus a $100(1-\alpha)\%$ CI for $E(Y)$ and a $100(1-\alpha)\%$ PI for y^*, for the same x^*, will have the same __________ but the PI will be

__________.

Example 1 again: Find a 90% prediction interval for the 20-km race time of a skier with time to exhaustion of $10\frac{1}{2}$ minutes.

11.10 Matrix Approach to Linear Regression

- We have focused on simple linear regression (one x variable).
- When there is more than one x variable, this is multiple linear regression (MLR).

- The estimation results must be expressed using matrix notation.

General MLR Model

- Here, x_{ij} is the value of the j-th independent variable for the i-th observation.
- This model may be expressed by the matrix equation:

where

- The $(k+1)\times 1$ vector of least squares estimates is thus:

Result: In MLR, the least-squares estimator of $\underline{\beta}$ is given by the formula:

Proof (only for $k=1$ case):

Note that $\hat{\underset{\sim}{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ and if we set $(\underset{\sim}{X}'\underset{\sim}{X})\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{Y}$ to be the least-squares equations, we get

which are exactly the least-squares equations we derived via differentiation in Sec. 11.3.

- Solving this matrix equation for $\hat{\underset{\sim}{\beta}}$ gives

- We will not prove it, but this formula holds true for any value of k.

<u>Example 1</u>: Let Y = graduation rate (in %) of a sample of 22 small colleges.

X_1 = median SAT score of accepted applicants

X_2 = per capita student expenses (in \$1000s)

- See course web page for data and calculations:

Example 2: Fit a quadratic regression model to the ski-racer data (Sec. 11.3 notes).

- Note

- Is the quadratic model needed, or is the linear regression sufficient? We could do inference about ______.

- We have seen that in the SLR setting,

$$\underset{\sim}{X}'\underset{\sim}{X} =$$

- It can be shown fairly easily that

- These are the same expressions as we derived for:

- In general in MLR, variances and covariances of the $\hat{\beta}_i$'s will involve elements of the $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ matrix.