Statistics 135 Spring 2004. Least Squares Regression Theorems

Least Squares Regression.

Simple Linear Regression:

- Data Generating Model: Let $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

  Observations: $\{(y_1, x_1,), \ldots, (y_n, x_n)\}$.

Estimates:

- With $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, let $s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ and $s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$. (Note that these definitions are different than s_{xx} and s_{xy} in Rice, p526)

  The maximum likelihood estimates of β_0, β_1 are given by

  $$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

  $$\hat{\beta}_1 = s_{xy}/s_{xx}.$$

  Theorem 1.

  (a) $E(y_i|x_i) = \beta_0 + \beta_1 x_i$
  (b) $var(y_i|x_i) = \sigma^2$
  (c) $y_i|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

  Theorem 2. With $\sigma^2_{\hat{\beta}_0} = \sigma^2\left(1/n + \bar{x}^2/s_{xx}\right)$ and $\sigma^2_{\hat{\beta}_1} = \sigma^2/s_{xx}$,

  (a) $\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})$
  (b) $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$

  Theorem 3. Let $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Then,

  (a) $RSS/\sigma^2 \sim \chi^2_{n-2}$,
  (b) RSS independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

  Theorem 4. With $s^2 = RSS/(n-2)$, let $s^2_{\hat{\beta}_0} = s^2\left(1/n + \bar{x}^2/s_{xx}\right)$, and $s^2_{\hat{\beta}_1} = s^2/s_{xx}$,

  (a) $(\hat{\beta}_0 - \beta_0)/s_{\hat{\beta}_0} \sim t_{n-2}$
  (b) $(\hat{\beta}_1 - \beta_1)/s_{\hat{\beta}_1} \sim t_{n-2}$

Theorem 5. If $\epsilon \overset{iid}{\sim} N(0, \sigma^2)$,

(a) $e_i = (y_i - \hat{y}_i) = [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]$ follows a normal distribution with mean zero.

(b) For large n and $i \neq j$, $cov(e_i, e_j) \approx 0$.

Multiple Linear Regression:

- Data Generating Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

  Observations: $(y_i, x_{1i}, x_{2i}, \ldots, x_{ki}), i = 1, \ldots, n$.

  The MLR model can be written in matrix form as $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Estimates:

- The least squares estimate of β is given by $\hat{\beta} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'Y}$.

Theorem 7. Let $\mathbf{v}$ be a random vector with $\mathrm{E}(\mathbf{v}) = \mu$ and $\mathrm{cov}(\mathbf{v}) = \mathbf{\Sigma}$.
Suppose $\mathbf{u} = \mathbf{c} + \mathbf{A}\mathbf{v}$, where $\mathbf{c}$ an $\mathbf{A}$ are a constant vector and matrix, respectively.

(a) $\mathrm{E}(\mathbf{u}) = \mathbf{c} + \mathbf{A}\mu$

(b) $cov(\mathbf{u}) = \mathbf{A\Sigma A'}$.

Theorem 8.

(a) $\mathrm{E}(\hat{\beta}) = \beta$

(b) $cov(\hat{\beta}) = \sigma^2(\mathbf{X'X})^{-1}$

(c) $\hat{\beta} \sim N(\beta, cov(\hat{\beta}))$.

(d) Let c_{ii} be the i:th diagonal element of $(\mathbf{X'X})^{-1}$. The standard error of $\hat{\beta}_i$ is estimated by $s_{\hat{\beta}_i} = s\sqrt{c_{ii}}$, and a $100(1-\alpha)\%$ CI for β_i is given by $\hat{\beta}_i \pm t_{\alpha/2}(df)s_{\hat{\beta}_i}$. Here, $df = n - (k+1)$ and s is as defined in Theorem 9.

Theorem 9. Let $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'}$.

(a) $\hat{\mathbf{Y}} = \mathbf{PY}$

(b) $\mathbf{e} = (\mathbf{Y} - \hat{\mathbf{Y}}) = (\mathbf{I} - \mathbf{P})\mathbf{Y}$

(c) $E(\mathbf{e}) = \mathbf{0}$

(d) $cov(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{P})$

(e) $s^2 = ||\mathbf{Y} - \hat{\mathbf{Y}}||^2/(n - (k+1))$.