Simple linear regression:

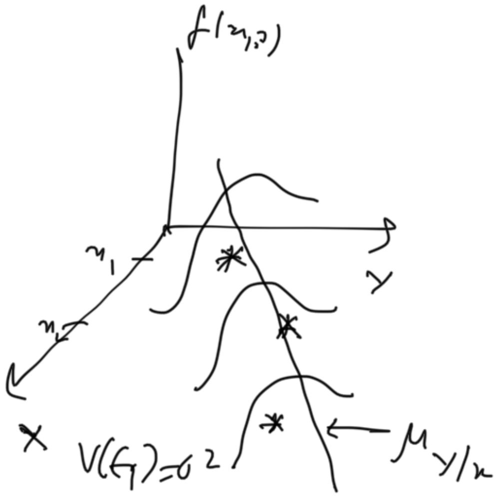
$Y \rightarrow$ Water temp.

$X \rightarrow$ depth

r.v. $Y|x \equiv Y \mid X = x$

$x_1, \dots, x_n$

$Y|x_1, \dots, Y|x_n$
$\quad\; = \qquad\quad\; =$
$\quad\; y_1 \qquad\quad\; y_n$



$$\mu_{y|x} = \beta_0 + \beta_1 x$$

$E(\epsilon_i) = 0$

$E(\epsilon_i \epsilon_j) = 0 \quad i \neq j$, $V(\epsilon_i) = \sigma^2$

$$y_i = Y|x_i = \mu_{y|x_i} + \epsilon_i \quad , i = 1, \dots, n$$

$$L = \sum_{i=1}^{n} \epsilon_i^2$$

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat\beta_0, \hat\beta_1} = 0, \quad \left.\frac{\partial L}{\partial \beta_1}\right|_{\hat\beta_0, \hat\beta_1} = 0$$

$$\Rightarrow \hat\beta_0 = \bar y - \hat\beta_1 \bar x$$

$\hat y_i = \hat\beta_0 + \hat\beta_1 x_i$

$$\hat\beta_1 = \frac{S_{xy}}{S_{xx}}, \text{ where } S_{xy} = \sum_{i=1}^{n} (y_i - \bar y)(x_i - \bar x) = \sum_{i=1}^{n} y_i (x_i - \bar x)$$

residuals $e_i = y_i - \hat y_i$

$$E(\hat\beta_1) = \beta_1 \; ; \; E(\hat\beta_0) = \beta_0, \; V(\hat\beta_1) = \frac{\sigma^2}{S_{xx}},$$

$$V(\hat\beta_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar x^2}{S_{xx}} \right), \; Cov(\hat\beta_0, \hat\beta_1) = -\frac{\sigma^2 \bar x}{S_{xx}}$$

Hypo testing on slope and intercept

$\epsilon_i \sim NID(0, \sigma^2)$

$H : \beta = \beta$ vs $H : \beta \neq \beta$

$H_0: \beta_1 = \beta_{10}$ vs $H_1: \beta_1 \neq \beta_{10}$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

σ^2 known $\quad Z_0 = \dfrac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}$ under H_0

$100(1-\alpha)\%$ CI for β_1 is $\left(\hat{\beta}_1 \pm z_{\frac{\alpha}{2}}\, se(\hat{\beta}_1)\right)$

σ^2 unknown $\qquad U = \dfrac{(n-2)MS_E}{\sigma^2} \sim \chi^2_{n-2}$

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{MS_E}{S_{xx}}}} \sim t_{n-2}$$

$H_0: \beta_0 = \beta_{00}$ vs $H_1: \beta_0 \neq \beta_{00}$

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{MS_E\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right)}}$$

CI for mean response at specific x (say x_0)

$$\hat{y}_0 = \hat{E}(y|x_0) \sim N\left(y_0, \sigma^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)\right)$$

$100(1-\alpha)\%$ CI about the true regression line at $x = x_0$ is

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2}\sqrt{MS_E\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$100(1-\alpha)\%$ prediction interval on future obs. at $x = x_0$ is

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2}\sqrt{MS_E\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$\ldots - t_{\frac{\alpha}{2}, n-2} \sqrt{MSE\left(1 + \frac{1}{n} + \frac{\ldots}{S_{xx}}\right)}$$

Lack of Fit test:

H_0 : the model adequately fit the data

H_1 : " " does not fit the data

$$SS_E = SS_{PE} + SS_{LOF}$$

SS_{PE}

$$\begin{array}{ll} y_{11}, \ldots, y_{1n_1} & x_1 \\ \vdots & \vdots \\ y_{m1}, \ldots, y_{mn_m} & x_m \end{array} \Bigg] \text{ m distinct level of } x$$

$$SS_{PE} = \sum_{i=1}^{m} \sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2 \quad \text{df } n_e = \sum_{i=1}^{m} 1.(n_i - 1) = n - m$$

$$SS_{LOF} = SS_E - SS_{PE} \quad \text{df } n - 2 - n_e = m - 2$$

test stat

$$F_0 = \frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)} = \frac{MS_{LOF}}{MS_{PE}}$$

reject H_0 if $F_0 > F_{\alpha, m-2, n-m}$

$$y = \beta_0 e^{\beta_1 x} \epsilon$$

$$\ln y = \ln \beta_0 + \beta_1 x + \ln \epsilon$$

$$y = \frac{1}{\beta_0 + \beta_1 x + \epsilon} \qquad y^* = \frac{1}{y}$$

e " -

$$\ln y^* = \beta_0 + \beta_1 x + \epsilon$$

Correlation analysis:

X, Y both r.v.

$$(y_i, x_i) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$

$$Y|x \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x - \mu_2), \sigma_1^2(1-\rho^2)\right)$$

$$E(Y|x) = \beta_0 + \beta_1 x,$$

where $\beta_0 = \mu_1 - \mu_2 \rho \frac{\sigma_1}{\sigma_2}$

$\beta_1 = \rho \frac{\sigma_1}{\sigma_2}$

mle of β_0 and β_1

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\rho} = r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

$\equiv H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$

test stat $t_0 = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$ if H_0 is true

reject H_0 if $|t_0| > t_{\frac{\alpha}{2}, n-2}$