residual: $\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{E}(Y|X=x_i) = y_i - \hat\beta_0 - \hat\beta_1 x_i$ OLS $RSS(\beta_0,\beta_1) = \sum[y_i - (\beta_0+\beta_1 x_i)]^2$. $\hat\beta_0 = \bar{y} - \hat\beta_1\bar{x}$, $\hat\beta_1 = \frac{SXY}{SXX}$

$SXX = \sum(x_i-\bar{x})^2 = \sum(x_i-\bar{x})x_i$, $SD_x^2 = SXX/(n-1)$, $SXY = \sum(x_i-\bar{x})(y_i-\bar{y}) = \sum(x_i-\bar{x})y_i$, $s_{xy} = SXY/(n-1)$, $r_{xy} = \frac{s_{xy}}{SD_x SD_y}$

$\hat\sigma^2 = RSS/(n-2)$, $df = n-2$, $\hat\sigma$ is standard error of regression. $RSS = SYY - \hat\beta_1^2 SXX$ sample covariance sample correlation

properties of LSE let $c_i = (x_i-\bar{x})/SXX$, $\sum(x_i-\bar{x})\bar{y}=0$. then $\hat\beta_1 = \sum\frac{x_i-\bar{x}}{SXX}y_i = \sum c_i y_i$ & fitted line passes $(\bar{x},\bar{y})$. CH2

① $E(\hat\beta_1|X) = E(\sum c_i y_i|X=x_i) = \sum c_i E(y_i|X=x_i) = \sum c_i(\beta_0+\beta_1 x_i) = \beta_0\sum c_i + \beta_1\sum c_i x_i = \beta_0\cdot 0 + \beta_1\cdot 1 = \beta_1$

② $E(\hat\beta_0|X) = E(\bar{y}|X) - \beta_1\bar{x} = \beta_0$ ③ $E(\hat\sigma^2|X) = \sigma^2$ Var of LSE ① $Var(\hat\beta_1|X) = \sigma^2/SXX$ ② $Var(\hat\beta_0|X) = \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{SXX})$

③ $Cov(\hat\beta_0,\hat\beta_1|X) = -\sigma^2\frac{\bar{x}}{SXX}$, ④ $\rho(\hat\beta_0,\hat\beta_1|X) = -\bar{x}\cdot(SXX/n+\bar{x}^2)^{-\frac{1}{2}}$ Replace σ^2 by $\hat\sigma^2$ get $\hat{Var}(\hat\beta_1|X)$ & $\hat{Var}(\hat\beta_0|X)$

& $se(\hat\beta_1|X) = \sqrt{\hat{Var}(\hat\beta_1)}$, $se(\hat\beta_0|X) = \sqrt{\hat{Var}(\hat\beta_0)}$ NOTE: $\rho(\hat\beta_0,\hat\beta_1|X) = Cov(\hat\beta_0,\hat\beta_1|X)/\sqrt{Var(\hat\beta_0)Var(\hat\beta_1)}$

$RSS = SYY - SXY^2/SXX$, $SS_{reg} = RSS_1 - RSS_2 = SYY - (SYY - SXY^2/SXX) = SXY^2/SXX$, $df = (n-1)-(n-2) = 1$.

	df	SS	MS	F	p-value
Reg	1 p	SSreg	$\frac{SSreg}{1 \text{ or } p}$	$MSreg/\hat\sigma^2$	
Res	$n-2$, $n-(p+1)$	RSS	$\hat\sigma^2 = \frac{RSS}{n-(p+1)}$ / $\frac{RSS}{n-2}$		
Tot	$n-1$	SYY			

if $F_{obs} > F_{(\alpha,1,n-2)}$, reject NH; $F_{obs} \le F_{(\alpha,1,n-2)}$, don't reject

coeff of Determination, $R^2 = \frac{SSreg}{SYY} = \frac{SXY^2}{SXX\,SYY} = r_{xy}^2$

CI: $100(1-\alpha)\%$ for θ is $(\hat\theta - t(\frac{\alpha}{2},d)se(\hat\theta), \hat\theta + t(\frac{\alpha}{2},d)se(\hat\theta))$

Prediction ≠ Fitted value

① $\tilde{y}_* = \hat\beta_0 + \hat\beta_1 x_*$. $Var(\tilde{y}_* - y_*|X,x_*) = Var(y_*|x_*) + Var(\tilde{y}_*|X,x_*) = \sigma^2 + \sigma^2(\frac{1}{n} + \frac{(x_*-\bar{x})^2}{SXX})$

sepred$(\tilde{y}_* - y_*|X,x_*) = \hat\sigma(1 + \frac{1}{n} + \frac{(x_*-\bar{x})^2}{SXX})^{\frac{1}{2}}$ prediction Interval: $\tilde{y}_* \pm t(\frac{\alpha}{2},n-2)$ sepred

② sefit$(\hat{y}|X,x) = \hat\sigma(\frac{1}{n} + \frac{(x-\bar{x})^2}{SXX})^{\frac{1}{2}}$, CI: $(\hat\beta_0 + \hat\beta_1 x) \pm$ sefit$(\hat{y}|X,x)[2F(\alpha;2,n-2)]$

MLR: $Y = (y_1,\cdots,y_n)'$

$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{np}\end{pmatrix}$ x_i — ith row of X, $e = (e_1,\cdots,e_n)'$, $\beta = (\beta_0,\cdots,\beta_p)'$. $Y = X\beta + e$ with $E(e)=0$, $Var(e) = \sigma^2 I_n$, $e \sim N(0,\sigma^2 I_n)$

$\hat{e} = Y - \hat{Y}$. $\hat\sigma^2 = RSS/(n-(p+1))$.

$RSS(\beta) = (Y-X\beta)'(Y-X\beta) = Y'Y + \beta'(X'X)\beta - 2Y'X\beta \Rightarrow \hat\beta = (X'X)^{-1}X'Y$ $E(\hat\beta|X) = \beta$, $Var(\hat\beta|X) = \sigma^2(X'X)^{-1}$ CH3

$RSS = RSS(\hat\beta) = Y'Y - \hat\beta'X'X\hat\beta = Y'Y - \hat{Y}'\hat{Y}$ with $\hat{Y} = X\hat\beta$

Similar ANOVA Pred & FittedV $\hat{y}_* = x_*'\hat\beta$, sepred $= \hat\sigma\sqrt{1 + x_*'(X'X)^{-1}x_*}$ $\hat{E}(Y|X=x) = \hat{y} = x'\hat\beta$, sefit$(\hat{y}|x) = \hat\sigma\sqrt{x'(X'X)^{-1}x}$

Sampling from Normal Population: $\binom{x_i}{y_i} \sim N\left(\binom{\mu_x}{\mu_y}, \begin{pmatrix}\sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$ $y_i|x_i \sim N(\mu_y + \rho_{xy}\frac{\sigma_y}{\sigma_x}(x_i - \mu_x), \sigma_y^2(1-\rho_{xy}^2))$

define $\beta_0 = \mu_y - \beta_1\mu_x$, $\beta_1 = \rho_{xy}\frac{\sigma_y}{\sigma_x}$, $\sigma^2 = \sigma_y^2(1-\rho_{xy}^2)$ so $y_i|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \Rightarrow \hat\mu_x = \bar{x}$, $\hat\mu_y = \bar{y}$, $\hat\sigma_x = \frac{SXX}{n-1}$, $\hat\sigma_y = \frac{SYY}{n-1}$, $\hat\rho_{xy} = \frac{SXY}{\sqrt{SXX\,SYY}}$

plug in to get $\hat\beta_0, \hat\beta_1 \Rightarrow$ OLS est. CH4

WLS $Var(Y|X=x_i) = Var(e_i) = \frac{\sigma^2}{w_i}$ matrix form: $Y = X\beta + e$, $Var(e) = \sigma^2 W^{-1}$, W is diagonal with $w_1,\cdots,w_n$

$\hat\beta$ is to minimize $RSS(\beta) = \sum w_i(y_i - x_i'\beta)^2 = (Y-X\beta)'W(Y-X\beta)$. How? ① $Var(W^{\frac{1}{2}}e) = W^{\frac{1}{2}}Var(e)W^{\frac{1}{2}} = W^{\frac{1}{2}}(\sigma^2 W^{-1})W^{\frac{1}{2}} = \sigma^2 I$ ② $W^{\frac{1}{2}}Y = W^{\frac{1}{2}}X\beta + W^{\frac{1}{2}}e$, let $Z = W^{\frac{1}{2}}Y$, $M = W^{\frac{1}{2}}X$, $d = W^{\frac{1}{2}}e$ then $Z = M\beta + d$ ③ $\hat\beta = (M'M)^{-1}M'Z = (X'WX)^{-1}(X'WY)$

LOF idea: ① Good $\Rightarrow E(\hat\sigma^2) \approx \sigma^2$ ② not, $\hat\sigma^2 > \sigma^2$. Two cases: ① σ^2 known, if no LOF (NH). $X^2 = RSS/\sigma^2 = (n-(p+1))\hat\sigma^2/\sigma^2$

p-value $= P(\chi^2 \ge X_{obs}|NH)$ ② unknown (replicates to do this). SS_{pe}, $df = 2+0+1+3 = 6$ (e.g. 3↑ $x_i = 1$, 1↑2, 2↑3, 4↑4)

$\hat\sigma_{pe}^2 = \frac{SS_{pe}}{df_{pe}}$. $RSS = \sum_i\sum_j(y_{ij} - \hat\beta_0 - \hat\beta_1 x_i)^2 = \sum_i\sum_j(y_{ij} - \bar{y}_i + \bar{y}_i - \hat\beta_0 - \hat\beta_1 x_i)^2 = \sum\sum(y_{ij}-\bar{y}_i)^2 + \sum_i n_i(\bar{y}_i - \hat\beta_0 - \hat\beta_1 x_i)^2 + 2\sum_i(\sum_j(y_{ij}-\bar{y}_i))(\bar{y}_i - \hat\beta_0 - \hat\beta_1 x_i)$

$= \cdots$ 前两项 $= SS_{pe} + SS_{lof} = SS_{pe} + RSS_{wls}$

General F-test: $F = \frac{(RSS_{NH} - RSS_{AH})/(df_{NH} - df_{AH})}{RSS_{AH}/df_{AH}}$ v.s. $F(\alpha, df_{NH} - df_{AH}, df_{AH})$ CH5

$F = (SS_{lof}/df_{lof})/(SS_{pe}/df_{pe})$ v.s. $F(\alpha, df_{lof}, df_{pe})$ $\frac{dE(Y|X)}{dx} = \beta_1 + 2\beta_2 X = 0 \Rightarrow X_m = -\beta_1/2\beta_2$

Polynomial: 2 predictor $E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2$, solve

THE Δ METHOD. suppose $\hat\theta \overset{a}{\sim} N(\theta,\Sigma)$, $g(\theta)$ is cont. func. of θ, then when $n\to\infty$, have $E[g(\hat\theta)] \approx g(\theta)$, $Var[g(\hat\theta)] \approx \dot{g}(\theta)'\Sigma\dot{g}(\theta)$

where $\dot{g}(\theta) = \frac{\partial g}{\partial\theta} = (\frac{\partial g}{\partial\theta_1},\cdots,\frac{\partial g}{\partial\theta_k})'$ E.g. $\beta = (\beta_0,\beta_1,\beta_2)'$, $\hat\beta = (\hat\beta_0,\hat\beta_1,\hat\beta_2)'$, for large n, $\hat\beta \overset{a}{\sim} N(\beta, \sigma^2(X'X)^{-1})$, $\widehat{cov}(\hat\beta) = \hat\sigma^2(X'X)^{-1}$

$g(\beta) = \frac{-\beta_1}{2\beta_2} \Rightarrow \dot{g}(\beta) = (0, \frac{-1}{2\beta_2}, \beta_1/2\beta_2^2)$. $Var(g(\hat\beta)) = \dot{g}(\hat\beta)'\widehat{cov}(\hat\beta)\dot{g}(\hat\beta) = \frac{1}{4\hat\beta_2^2}(Var(\hat\beta_1) + \frac{\hat\beta_1^2}{\hat\beta_2^2}Var(\hat\beta_2) - \frac{2\hat\beta_1}{\hat\beta_2}Cov(\hat\beta_1,\hat\beta_2))$

$E(Y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$

$P(\frac{|g(\hat\beta) - g(\beta)|}{\sqrt{\dot{g}(\hat\beta)'\hat\Sigma\dot{g}(\hat\beta)}} < z_{\frac{\alpha}{2}}) = 1-\alpha$ $N(0,1)$ $1-\alpha$ $\frac{\alpha}{2}$ Another eg. solve for x_1: $x_m = g(\beta; x_2) = -\frac{\beta_1 + \beta_5 x_2}{2\beta_3}$. $\frac{\partial x_m}{\partial\beta} = \dot{g}(\beta, x_2) = (\cdots)'$

$Var(\hat{x}_m) = \dot{g}(\cdots)'\widehat{cov}(\hat\beta)\dot{g}(\cdots) \Rightarrow 100(1-\alpha)\%$ CI for $x_m = \hat{x}_m \pm z_{\frac{\alpha}{2}}\sqrt{Var(\hat{x}_m)}$

Factors Rule: d levels, d dummy variables; if the intercept is the mean function, at most d−1 of the d.v. can be used in mean function. def: jth d.v. U_j, $j = 1\cdots 5$. $u_{ij} = \begin{cases}1 & \text{if } D_i = j\text{th category of } D \\ 0 & \text{o.w.}\end{cases}$ CH6

Model: $E(TS|D) = \beta_1 U_1 + \cdots + \beta_5 U_5$. ①

β_j: population mean for all species with D index j. TS: total time of sleep

One way ANOVA model: $E(TS|D) = \eta_0 + \eta_2 U_2 + \cdots + \eta_5 U_5$. where $\eta_0 = \beta_1$, $\eta_0 + \eta_2 = \beta_2$, $\cdots$, $\eta_0 + \eta_5 = \beta_5$ ②

① NH: $\beta_j = 0\ \forall j$ ② NH: $E(TS|D) = \eta_0$ Adding a term log(BodyWt) $E(TS|\log(BdWt) = x, D = j) = \beta_{0j} + \beta_{1j}x$

4 situations: ① Most general: every level of D has a slope & intercept. So $E(TS\cdots) = \sum_{j=1}^d(\beta_{0j}U_j + \beta_{1j}U_j x)$

Similarly $E(\cdots) = \eta_0 + \eta_1 x + \sum_{j=2}^d(\eta_{0j}U_j + \eta_{1j}U_j x)$ ② Parallel regressions: $\beta_{11} = \cdots \beta_{1d}$ or $\eta_{12} = \cdots = \eta_{1d} = 0$. Each lvl of D has its own intercept. ③ Common intercept. $\beta_{01} = \cdots = \beta_{0d}$ or $\eta_{02} = \cdots = \eta_{0d} = 0$ ④ Coincident reg. line. all lines are the same.

A. Power transformation (on both, on either) $\psi(U,\lambda) = U^\lambda$, λ: −2 to 2 $\lambda \in \{-2,-1,-\frac{1}{2},0,\frac{1}{2},\frac{3}{2},1,2\}$

~~Trans~~ → to linear

B. scaled PTrans (on predictor) $\psi_S(X,\lambda) = \begin{cases}(X^\lambda - 1)/\lambda & \text{if } \lambda \ne 0 \\ \log_e X & \text{if } \lambda = 0\end{cases}$ anyway, λ has to minimize RSS (For linearity)

C. modified PTrans (on response) $gm(Y) = \exp\{\frac{1}{n}\sum_{i=1}^n \log_e y_i\}$

for $Y>0$, $\psi_M(Y,\lambda_y) = \psi_S(Y,\lambda_y) \times gm(Y)^{1-\lambda_y} = \begin{cases} gm(Y)^{1-\lambda_y} \times (Y^{\lambda_y} - 1)/\lambda_y & \text{if } \lambda \ne 0 \\ gm(Y) \times \log(Y) & \text{if } \lambda = 0\end{cases}$

B-C method assumes $E(\psi_M(Y,\lambda_y)|X=x) = \beta'x$ Might not work all the time; may not possible for "instant corrections"

goal: not for linearity, but for normality.

~~e.g. A~~ CH7

$p' = p+1$ ~~number of terms~~

p: $x_1, x_2, \cdots, x_p$

p': $\beta_0, \beta_1, \cdots, \beta_p$

Reg. diagnostics: residuals: $\hat\beta = (X'X)^{-1}X'Y$, $\hat{Y} = X\hat\beta = X(X'X)^{-1}X'Y$, define hat matrix $H = X(X'X)^{-1}X'$

s.t. $HY = \hat{Y}$. residuals: $\hat{e} = Y - \hat{Y} = Y - HY = (I-H)Y$. H is idempotent project matrix: $H' = H$, $HH = H$

assumptions for e (stat. errors) $E(e) = 0$, $Cov(e) = \sigma^2 I$, h_{ii} is ith diagonal element of H. $HX = X$. (leverage value)

$\boxed{Var(\hat{e}_i) = \sigma^2(1-h_{ii})}$ ⇒ this is not the same. if intercept included $\sum_{i=1}^n \hat{e}_i = 0$.

HAT matrix: $HH = X(X'X)^{-1}X' \cdot X(X'X)^{-1}X' = H$, $(I-H)X = 0 \Rightarrow E(\hat{e}) = 0$, $Cov(\hat{e},\hat{Y}) = Cov((I-H)Y, HY) = \sigma^2 H(I-H) = 0$

$Cov(Y) = \sigma^2 I$, $Cov(\hat{Y}) = \sigma^2 HH' = \sigma^2 H$, $Cov(\hat{e}) = \sigma^2(I-H)(I-H)' = \sigma^2(I-H)$ note: $Cov(\hat{e}) = Cov(Y - \hat{Y}) = Cov(Y) - Cov(\hat{Y})$

$h_{ij} = x_i'(X'X)^{-1}x_j = x_j'(X'X)^{-1}x_i = h_{ji}$. $\sum h_{ij} = p'$ if mean function has an intercept. $\sum h_{ij} = \sum h_{ji} = 1$

$\frac{1}{n} \le h_{ij} \le \frac{1}{r_i}$, r_i is # of rows of X that are identical to x_i.

$h_{ii}\uparrow$, $Var(\hat{e}_i)\downarrow$. $Var(\hat{e}_i) = 0 \Rightarrow$ ith (only) observation is used to get $\hat{y}_i$. High h_{ii} → an unusual x_i (but ≠ outlier)

$h_{ii} = h_{ij}^2$, i.e. $h_{ii}(1-h_{ii}) = \sum_{j\ne i}h_{ij}^2$. $\hat{y}_i = \sum_{j=1} h_{ij}y_j = h_{ii}y_i + \sum_{j\ne i}h_{ij}y_j$. CH 8

as $h_{ii} \to 1$, $\hat{y}_i \to y_i$, $\hat{y}_i$ is mostly determined by y_i only. fitted value $\hat{Y} = X\hat\beta = HY$

When doing WLS, $Var(e) = \sigma^2 W^{-1}$, then $H = W^{\frac{1}{2}}X(X'WX)^{-1}X'W^{\frac{1}{2}}$, variance function 一样"粗细", not constant but looks like constant.

def $\hat{e}_i = \sqrt{w_i}(y_i - \hat{y}_i)$ When model correct: null plot. When ~~E~~ Incorrect, nonlinearity, ~~not~~ inconstant variance ··· (or WLS, but hard to determine weights)

Variance stablizing transformation.

WHAT TO DO with nonconstant variance? $Var(Y|X=x) = \sigma^2 g(E(Y|X=x))$. g is usually ① $\log(Y)$ common, when response is counts/prices. ② $\sqrt{Y}$, mild. when log is too much ③ Y^{-1}, "time to an event".

Outlier Test sps ith case: suspected. $U = \begin{cases}u_j = 0 & \text{for } j \ne i \\ u_i = 1\end{cases}$, $E(Y|X) = X\beta + \delta U$. $\hat\delta$ is estimated mean shift

do NH: $\delta = 0$, AH: $\delta \ne 0$. Two side t-test. $df = n - p' - 1$ where $p' = p+1$ ~~p number of variables~~

Alt. approach: ① delete ith case ② est. β, σ^2. denote as $\hat\beta_{(i)}$, $\hat\sigma_{(i)}^2$, $df = n-p'-1$ ③ Compute fitted value for the deleted case $\hat{y}_i(i) = x_i'\hat\beta_{(i)}$, since y_i & $\hat{y}_i(i)$ indpt. $Var(y_i - \hat{y}_{i(i)}) = Var(y_i) + Var(\hat{y}_i(i))$

$= \sigma^2 + \sigma^2 x_i'(X_{(i)}'X_{(i)})^{-1}x_i$, $X_{(i)}$ is matrix X with i-th row deleted. ④ we have $E(y_i) = x_i'\beta + \delta$,

$E(\hat{y}_i(i)) = E(x_i'\hat\beta_{(i)}) = x_i'\beta \Rightarrow E(y_i - \hat{y}_i(i)) = \delta$. t for $\delta = 0$ is $t = \frac{y_i - \hat{y}_i(i)}{\hat\sigma_{(i)}\sqrt{1 + x_i'(X_{(i)}'X_{(i)})^{-1}x_i}}$ (use $\hat\sigma_{(i)}$ to replace σ)

求 p-value, p↓ → reject NH. 是 outlier.

WHY prefer Alt. approach? ① nice formula for t_i. ② define stddzd residual $r_i = \frac{\hat{e}_i}{\hat\sigma(1-h_{ii})^{\frac{1}{2}}}$

try to make all r_i's to have same variance.

Then $t_i = \frac{\hat{e}_i}{\hat\sigma_{(i)}\sqrt{1-h_{ii}}} = r_i\left(\frac{n-p'-1}{n-p'-r_i^2}\right)^{\frac{1}{2}}$ CH9

Bonferroni Adjustment. perform n tests, change sig. level α to $\frac{\alpha}{n}$. then overall sig level won't > α. we could also multiple p-value by n. E.g. Forbe's data, case 12 suspected.

$\hat{e}_{12} = 1.36$, $\hat\sigma = 0.379$, $h_{12,12} = 0.0639$. $r_{12} = \frac{1.36}{0.379\sqrt{1-0.0639}} = 3.7078 \Rightarrow t_{12} = 3.7078\left(\frac{17-2-1}{17-2-3.7078^2}\right)^{\frac{1}{2}} = 12$

p-value is 6.13×10^{-9} (df = 14), 6.13×10^{-9} times(n=17) ⇒ $1.04\times10^{-7} \ll 0.05$ (NH reject) 是 outlier

Cook's Distance. $D_i = \frac{(\hat\beta_{(i)} - \hat\beta)'(X'X)(\hat\beta_{(i)} - \hat\beta)}{p'\hat\sigma^2} = \frac{(\hat{Y}_{(i)} - \hat{Y})'(\hat{Y}_{(i)} - \hat{Y})}{p'\hat\sigma^2} = \frac{1}{p'}r_i^2\frac{h_{ii}}{1-h_{ii}}$

large D_i, potential problem

Normal prob. plot. (Q-Q plot) (check for normality of e_i)

Collinearity: $c_1X_1 + c_2X_2 = c_0 \Rightarrow r_{ij}^2 = 1 \Rightarrow$ collinearity

$c_1X_1 + \cdots + c_pX_p \approx c_0 \Rightarrow$ max of r_{ij}^2 close to 1 ⇒ approximate collinearity.

But Harm: ① if Cd exists, $(X'X)^{-1}$ not exist, no fitting done, need to drop terms.

② If AC exists, $var(\hat\beta)$ too large.

To see this, 2-term model. $E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ $Var(\hat\beta_j) = \frac{\sigma^2}{1-r_{12}^2}\frac{1}{SX_jX_j}$

where $SX_jX_j = \sum(x_{ji} - \bar{x}_j)^2$

Criteria: $f_1(RSS) + f_2(P)$ minimize it.

AIC: $n\log(\frac{RSS}{n}) + 2p$

BIC: $n\log(\frac{RSS}{n}) + p\log(n)$

C_p: $\frac{RSS}{\hat\sigma^2} + 2p - n$

CH10

CV: $\sum_{i=1}^n(y_i - \hat{y}_{i(i)})^2 = \sum_{i=1}^n\frac{\hat{e}_i^2}{(1-h_{ii})^2}$

4 Decimal digits!

Practice Exam: 1(d). MLR to predict heart attack: responses contrained to [0,1], difficult to observe.

4. $Y_t = a_0 + \sum_{k=1}^{K} a_k \sin(\omega_k t) + \cdots b_k \cos - + e_t$ (a). K cannot be OLS estimated B/c $K \to \infty$, model will be perfectly fitted. suggest method: choose K's to do AIC/BIC model comparison.

(b). simpler version $Y_t = 3 + b\cos(4t) + e_t$. derive OLS for b. $\frac{\partial}{\partial b} \sum_{t=1}^{n} [y_t - (3 + b\cos(4t))]^2 = -2\sum_{t=1}^{n} [y_t - 3 - b\cos(4t)]\cos(4t) = 0$

$\therefore \hat{b} = \frac{\sum_{t=1}^{n} (y_t - 3)\cos(4t)}{\sum_{t=1}^{n} \cos^2(4t)}$ (c). $Y_t = 3 + 2\cos(\omega t) + e_t$. $\frac{\partial}{\partial \omega} \sum_{t=1}^{n} [y_t - 3 - 2\cos(\omega t)]^2 = 4\sum_{t=1}^{n} (y_t - 3 - 2\cos(\omega t))\, t \sin(\omega t)$

not a closed form. use some numerical methods.

5. suspected case y_4, y_9 with δ & 2δ. Define $U = (u_1, \cdots, u_n)$ where $u_4 = 1, u_9 = 2$, others are zero.

model $E(Y|X=x) = \beta_0 + \beta_1 X + \delta U$. use OLS to estimate δ. denoted as $\hat{\delta}$, calculate se of $\hat{\delta}$. $se(\hat{\delta})$

$H_0: \delta = 0$ v.s. $H_1: \delta \neq 0$. $|t| = \left|\frac{\hat{\delta}}{se(\hat{\delta})}\right|$ and we compare $|t|$ with $t_{0.025}(n-3)$ to determine reject or accept H_0.

6. (a). $Var(\hat{\theta}) = Var(2\hat{\beta}_1 - \hat{\beta}_0) = Var(2\hat{\beta}_1) + Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_1, \hat{\beta}_0)$ Use formula $\Rightarrow \hat{\sigma}^2 (X'X)^{-1}$ plug in each number in the Mtrx $(X'X)^{-1}$

$H_0: \theta = 0$ v.s. $H_1: \theta \neq 0$. $t = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{se(\hat{\theta})} = -0.0146963$ $\because |t| < 1.96$ don't reject.

(b). $g = \hat{\beta}_1 / \hat{\beta}_0$. $(\frac{\partial g}{\partial \beta})' = (0, 1/\hat{\beta}_2, -\hat{\beta}_1/\hat{\beta}_2^2)$ $Var(g(\hat{\beta})) = (0\ \frac{1}{\hat{\beta}_2}\ \frac{-\hat{\beta}_1}{\hat{\beta}_2^2})(\hat{\sigma}^2 (X'X)^{-1})\begin{pmatrix} 0 \\ \vdots \\ \vdots \end{pmatrix} \approx 0.06021$

(d). LOF test can tell us whether model fits data. not the curvature exists.

(e). x_i^2 should be included since its p-value $0.02564 < 0.05$. $H_0: \beta_3 = 0$ is rejected.

Quiz 2 Recap:

(a). 32 students, for ith students. $u_{ij} = \begin{cases} 1 & \text{if } D_i = j \\ 0 & \text{o.w.} \end{cases}$ The linear model with intercept that contains interaction between library hours (X) & departments (D) is $y_i = \eta_{01} + \eta_{11} x_i + \sum_{j=2}^{4} (\eta_{0j} u_{ij} + \eta_{1j} u_{ij} x_i)$, $i = 1, \cdots, 32$

where η_{01} & η_{11} are intercept and slope effects of library hours for STAT. For $j = 2, 3, 4$. η_{0j} is the difference b/w intercepts of department of level j & STAT, η_{1j} is the difference between library hour slope effects of department at level j and STAT.

(b). $H_0: \eta_{12} = \eta_{13} = \eta_{14} = 0$ i.e. interaction effects between D & X is not significant.

$y_i = \eta_{01} + \eta_{11} x_i + \sum_{j=2}^{4} \eta_{0j} u_{ij}$, $i = 1, \cdots, 32$.

(c). Model III is "same intercept & slope for departments, i.e. only 2 reg. parameters".

$df_{M_1} = 32 - 8 = 24$, $df_{M_2} = 32 - 5 = 27$, $df_{M_3} = 32 - 2 = 30$.

$F = \frac{(RSS_{II} - RSS_I)/(27-24)}{RSS_I / 24} = \cdots$ compare.

(d).

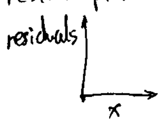
2. (a): $X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \\ 1 & x_n - \bar{x} \end{pmatrix}$ $X^T X = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^{n} (x_i - \bar{x})^2 \end{pmatrix}$ $h_{ij} = (1, x_i - \bar{x}) \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum (\ )^2} \end{pmatrix} \begin{pmatrix} 1 \\ x_j - \bar{x} \end{pmatrix}$

$= \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum (\ )^2} \geq \frac{1}{n}$

(b). replicates $h_{ii} = h_{ik}$, note $H^2 = H$ i.e. $h_{ii} = \sum_{j=1}^{n} h_{ij}^2 \geq r_i h_{ii}^2 \Rightarrow h_{ii} \leq \frac{1}{r_i}$

$x_i = x_k$

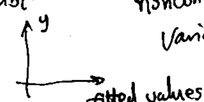
residual plot



nonlinearity trend / suspected outlier

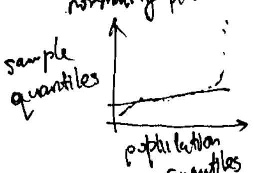
influential point

fitted value plot



nonconstant variance

normality plot



normality is in doubt & suspected outlier