Lecture 07 Adv. Data Adj. 7-1
21-june-2016 Tuesday

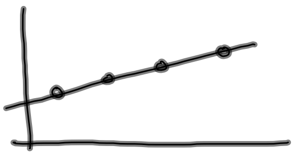
We meet Thurs, 23rd

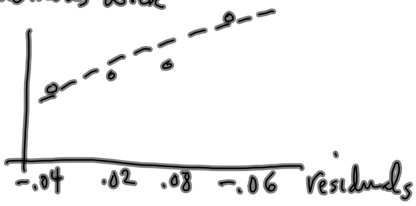
errors: random ☺
systematic 😐
blunder ☹

blunder detection
gross error detection
outlier detection
robust estimation
} techniques: Data Snooping, IRLS, L1 norm minimization, RANSAC

Big Idea: large redundancy + well distributed
then all methods work well

small redundancy OR not well distributed
then NO methods work





Jun 21-7:11 PM

instead of looking @ resid. magnitudes 7-2
look @ std. residuals: $\frac{v_i}{\sigma_{v_i}}$ $\Sigma_{vv} = \sigma_0^2 Q_{vv}$

$\frac{v_i}{\sigma_{v_i}}$ -73.0 23.9 95.6 <u>-109.5</u>

Ind. Obs. model $v = f - B\delta$

$= f - BN^{-1}t$

$= f - BN^{-1}B^TWf$

$= [I - BN^{-1}B^TW]f$

$= [QW - BN^{-1}B^TW]f$

$= \underbrace{[Q - BN^{-1}B^T]}_{Q_{vv}}Wf$

$v = Q_{vv}Wf$

$$\boxed{v = \underbrace{Q_{vv}W}_{\overline{W}}(-\ell)}$$

Jun 21-7:12 PM

look at $(Q_{vv}W)^2$ 7-3

$(Q - BN^{-1}B^T)W(Q - BN^{-1}B^T)W$

$(I - BN^{-1}B^TW)(I - BN^{-1}B^TW)$

$I + BN^{-1}\underbrace{B^TWB}\underbrace{N^{-1}}B^TW - BN^{-1}B^TW - BN^{-1}B^TW$

$I - BN^{-1}B^TW$

$(Q - BN^{-1}B^T)W = Q_{vv}W$

$(Q_{vv}W)^2 = Q_{vv}W$ <u>Idempotent</u>

Trace matrix = sum of diag. elements

Trace of Idempotent matrix = rank

$Tr(A+B) = Tr(A) + Tr(B)$

$Tr(EF) = Tr(FE)$

$Tr(\bar{W}) = Tr(Q_{vv}W) = Tr[(Q - BN^{-1}B^T)W] = Tr[I - BN^{-1}B^TW]$

$= Tr(I_n) - Tr(BN^{-1}\underline{B^TW})$

$= Tr(I_n) - Tr(B^TWBN^{-1}) = n - u = r$

$Tr(I_u)$

Rank of $\bar{W} = r$

Jun 21-7:12 PM

diagonal elements of $\bar{W}$: r_i redundancy numbers 7-4

$Tr(\bar{W}) = r_1 + r_2 + \dots + r_n = r$

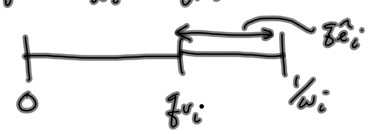
assume W diagonal

$r_i = q_{v_i} \cdot w_i$, $q_{v_i} = r_i / w_i$ ←

$Q_v = Q - \underline{BN^{-1}B^T} \rightarrow Q_{\hat{\ell}\hat{\ell}}$

$q_{v_i} = q_i - q_{\hat{\ell}_i}$

$q_{v_i} = \frac{1}{w_i} - q_{\hat{\ell}_i}$ all are + values



$\Rightarrow$

$0 \le q_{v_i} \le 1/w_i$

$0 \le q_{v_i} w_i \le 1$

$\boxed{0 \le r_i \le 1}$ redundancy number between 0 & 1

$w_i = \sigma_0^2/\sigma_i^2$, $\sigma_{v_i} = \sigma_0\sqrt{q_{v_i}}$, $\sigma_{v_i} = \sigma_0\sqrt{\frac{r_i}{w_i}} = \sigma_0\sqrt{\frac{\sigma_i^2}{\sigma_0^2} \cdot r_i} = \frac{\sigma_0}{\sigma_0}\sigma_i\sqrt{r_i}$

$\boxed{\sigma_{v_i} = \sigma_i\sqrt{r_i}}$

Jun 21-7:12 PM

7-5

r_i near ZERO, σ_{v_i} small ⇒ residual small

r_i near ONE, $\sigma_{v_i} \approx \sigma_i$ ⇒ residual will approximate the actual obs. error

r_i near ZERO, obs. errors are "absorbed in adjustment"

r_i near ONE, obs. errors are revealed in the residual

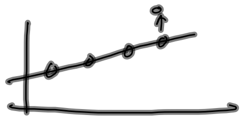
---

already, we have powerful design tool

make sure that all r_i > some threshold, for example

$r_i > 0.8$

back to line fit



r_i 0.3, 0.7, 0.7, 0.3

Jun 21-7:12 PM

7-6

n	3	4	5	10	1000
	.17	.3	.4	.65	.99
	.67	.7	.7	.75	
	.17	.7	.8	.82	
		.3	.7	.87	
			.4	.89	
				.89	
				.87	
				.82	
				.75	
				.65	.99

red. num. fitting n pts. to line

all of prior disc. (a) std. residuals useful for B.D.

(b) Keep r_i > Threshold

---

famous guy BAARDA invented data snooping

called B-method of testing

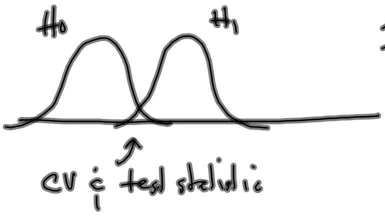
Reliability: ability of test or algo. to detect & elim. blunders of given magnitude

key element of B-method is to connect global test to individual tests on std. residuals: reliability consistent

Jun 21-7:12 PM

Hypothesis Tests

7-7



2 types of error

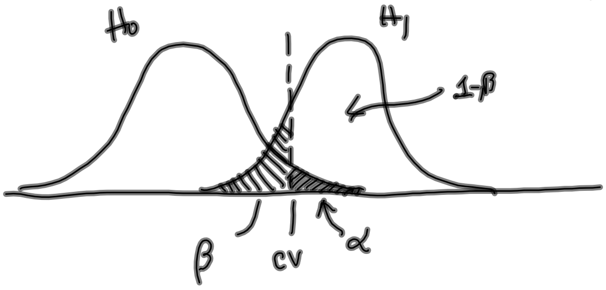
Type I : REJECT Ho when TRUE

Type II : ACCEPT Ho when FALSE

prob(Type I error) = α

prob(Type II error) = β $1-\beta$: power of test

warning some researchers switch meaning of β & $1-\beta$



Jun 21-7:12 PM

naive approach :

7-8

if you don't want to reject many obs : α : small

β will be large, let many blunders through

if you don't want to accept blunders, β = small

but α will be large, you reject many good obs.

long experience yields

$$\alpha_0 = .001$$
$$\beta_0 = 0.2$$

Jun 21-7:12 PM

Jun 21-7:12 PM

Jun 21-7:12 PM

Jun 21-7:12 PM

Jun 21-7:12 PM