Lec 2 Simple Linear Regression

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ $\varepsilon_i \sim N(0, \sigma^2)$ 随机误差项.

LINE: Linear Independent Normally Equal var.

β_1 的释义: change in mean of prob distribution of Y per unit increase in X

$b_1 = \hat{\beta}_1 = \frac{\sum_i (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_i (X_i-\bar{X})^2}$ $b_0 = \hat{\beta}_0 = \bar{Y} - b_1\bar{X}$

求得方法: $Q = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$ 对 β_0, β_1 求偏导数.

$e_i = Y_i - \hat{Y}_i \Rightarrow \sum e_i = 0, \sum X_i e_i = 0. \sum \hat{Y}_i e_i = 0$

$df(e_i) = n-2$ $s^2 = \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{SSE}{n-2} = MSE.$

MLE: $\hat{\beta}_0^{mle} = b_0$ $\hat{\beta}_1^{mle} = b_1$ $\hat{\sigma}^{2\,mle} = \frac{\sum e_i^2}{n} \neq s^2$

Lec 3 Statistical Inference & Power.

β_1 统计推断比 β_0 更重要.

$b_1 \sim N(\beta_1, \sigma^2(b_1))$ $\sigma^2(b_1) = \frac{\sigma^2}{S_{XX}}$ $S_{XX} = \sum_i (X_i - \bar{X})^2$

$s(b_1) = \frac{s}{\sqrt{S_{XX}}}$ $s = \left(\frac{\sum e_i^2}{n-2}\right)^{\frac{1}{2}}$

对 β_1 的检验 $T = \frac{b_1 - 0}{s(b_1)} \overset{H_0}{\sim} t(n-2)$ $H_0: \beta_1 = 0$

decision rule: $|t_0| > t_{n-2, 1-\frac{\alpha}{2}}$

accept H_0	reject H_0
① Commit Type II Error	① Type I Error
② X与Y无 Linear relationship	② X与Y有 Linear
③ X与Y有非线性关系	③ X与Y还有非线性但未发现.

CI: $b_1 \pm t_{n-2, 1-\frac{\alpha}{2}} s(b_1)$

β_0 检验与推断:

$b_0 \sim N(\beta_0, \sigma^2(b_0))$ $\sigma^2(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)$ ($\bar{Y}$, $b_1\bar{X}$) $cov(\bar{Y}, b_1) = 0$ ↗ 上下平移 不会影响slope.

$(b_0 - \beta_0)/s(b_0) \sim t(n-2)$ $s(b_0) = s\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}}$

CI: $b_0 \pm t_{n-2, 1-\frac{\alpha}{2}} s(b_0)$

BLUE: Best Linear Unbiased Estimator.

证明 b_1 是 BLUE, $\forall \hat{\beta}_1 = \sum d_i Y_i$ 由 $E(\hat{\beta}_1) = \beta_1$ 得 $\sum d_i = 0$ $\sum d_i X_i = 1$

然后再证 $Var(\hat{\beta}_1) = \sum d_i^2 \sigma^2 \geq \frac{\sigma^2}{S_{XX}}$.

Bonferroni Correction. 对 β_0, β_1 联合推断时, α各自减半.

Power: $1 - P(\text{Type II Error})$

用偏T分布来像计算power $power(\beta_1) = P(T > t_{n-2, 1-\frac{\alpha}{2}}) + P(T < -t_{n-2, 1-\frac{\alpha}{2}})$

其中 $T \sim t(n-2, \beta_1/\sigma(b_1))$ $\sigma(b_1)$ 用 $s(b_1)$ 估.

Lec 4 Prediction & Anova.

Estimation and prediction on Y.

	Y_h	$Y_{h(new)}$
公式	$Y_h = \beta_0 + \beta_1 X_h$	$Y_{h(new)} = \beta_0 + \beta_1 X_h + \varepsilon$
$\hat{u}$	$\hat{u}_h = b_0 + b_1 X_h$	$\hat{u}_h = b_0 + b_1 X_h$
s^2	$s^2(\hat{u}_h) = s^2\left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]$	$s^2(Y_{h(new)} - \hat{u}_h) = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]$
CI	$\frac{\hat{u}_h - u_h}{s(\hat{u}_h)} \sim t(n-2)$	$\frac{Y_{h(new)} - \hat{u}_h}{s(pred)} \sim t(n-2)$
	$\hat{u}_h \pm t_{n-2, 1-\frac{\alpha}{2}} s(\hat{u}_h)$	$\hat{u}_h \pm t_{n-2, 1-\frac{\alpha}{2}} s(pred)$
正态性依赖	无	$\sigma^2(d_h) = Var(Y_{h(new)}) + Var(\hat{u}_h)$

mean of m obs: $s^2(predmean) = MSE\left(\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_i (X_i - \bar{X})^2}\right)$

Confidence Band: $P[L(x) < \beta_0 + \beta_1 X < U(x)$ for all $x] = 1 - \alpha$

$(L(x)\ U(x)) = \hat{u}_x \pm W s(\hat{u}_x)$ $W = \sqrt{2F_{1-\alpha, 2, n-2}}$ W更宽

Confidence Interval < Confidence Band < Pred Interval.

Anova Table. (一元线性回归)

Source	df	SS	MS	F value	Pr(>F)
Reg	1	$\sum(\hat{Y}_i - \bar{Y})^2$	SSR/dfR	MSR/MSE	
Error	n-2	$\sum(Y_i - \hat{Y}_i)^2$	SSE/dfE		
Total	n-1	$\sum(Y_i - \bar{Y})^2$	SST/dfT		

F-test: $F = \frac{MSR}{MSE} \sim F_{dfR, dfE} = F_{1, n-2}$ $H_0: \beta_1 = 0$

reject H_0 if $P(F > F_0 | F_{1,n-2}) < \alpha$

与 t-test 关系: $F = T^2$, under SLR, p-value are same.

相关系数(r)与决定系数(R):

$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(\sum(X_i-\bar{X})^2 \sum(Y_i-\bar{Y})^2)^{\frac{1}{2}}}$ $R^2 = \frac{SSR}{SST}$

under SLR: $r^2 = R^2$ ($r^2 = b_1^2 \frac{\sigma_X^2}{\sigma_Y^2} = \frac{SSR}{SST}$)

$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \sim t_{n-2}$

4种对 $H_0: \beta_1 = 0$ 的检验:

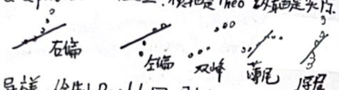
① Individual parameter t-test ② t-test for correlation coff ρ

③ Anova F test ④ GLT.

Lec 5 Model Assumption and Diagnostics

		Important Stat	
L	$E(Y_i) = \beta_0 + \beta_1 X_i$	Mean:	Location
I	$\varepsilon_1 \cdots \varepsilon_n$ are ind	Std dev:	Variability
N	$\varepsilon_1 \cdots \varepsilon_n \sim N(0, \sigma^2)$	Skewness:	Lack of sym (偏度)
E	$Var(Y_i) = \sigma^2$	Kurtosis(峰度):	Heavy tail { 尖峰厚尾 / 低峰薄尾
t对异方差敏感.		Range:	范围
F对异方差不敏感.		左偏 or 右偏	{ mean在峰值哪一侧 / 拖尾方向

Q-Q plot: 正态性检验. 横轴是 Theo 纵轴是实际.



异方差: 绘制 Residuals 图. 异方差不影响无偏但不是BLUE

Equal var检验: Bartlett's test (Likelihood ratio test)

Levene's test (2 sample t-test / 多组 Anova) Brown-Forsythe

Brensch-Pagan Test (test $\sigma_i^2 = h(\vec{Z}_i^T \vec{\alpha})$) $H_0: \alpha_1 = \cdots = \alpha_p = 0$

Normality检验: Shapiro-Walk, Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling

Dependence检验: Durbin-Watson test, Ljung-Box test.

Lec 6 Remedies & Miscellaneous Topics

概念: outlier(离群点): 偏离回归线大的点 (high residual)

leverage(高杠杆点): $h_{ii} = \frac{\partial \hat{Y}_i}{\partial Y_i}$ 大的点 (远离 $\bar{X}$)

high influential = outlier + leverage.

对 h_{ii}: $h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{j=1}^n (X_j - \bar{X})^2}$ outlier

Cook's D: $D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p s^2} = \boxed{\frac{e_i^2}{p s^2}} \boxed{\frac{h_{ii}}{(1-h_{ii})^2}}$ high leverage.

失拟检验: 线性模型是否很好地拟合了模型

$H_0: U_i = \beta_0 + \beta_1 X_i$ $H_1: U_i \neq \beta_0 + \beta_1 X_i \Rightarrow$ $Y_{ij} = \beta_0 + \beta_1 X_i + \varepsilon_{ij}$ reduced; $Y_{ij} = U_{(j)} + \varepsilon_{ij}$ Full

在失拟检验下可以将残差分析细化.

SSE { $SSLF = \sum\sum(\bar{Y}_i - \hat{Y}_{ij})^2$ df = c-2 $MSLF = \frac{SSLF}{c-2}$; $SSPE = \sum\sum(Y_{ij} - \bar{Y}_i)^2$ df = n-c $MSPE = \frac{SSPE}{n-c}$

$F = \frac{MSLF}{MSPE} \sim F_{c-2, n-c}$

非线性补救: 使用非线性拟合. 异方差补救: 加权.

非正态: 做变换

常用变换: ① 方差稳定. $E(Y) = U_x$, $Var(Y) = h(U_x)$

$f(Y) \approx f(U_x) + f'(U_x)(Y - U_x)$ $Var f(Y) \approx (f'(U_x))^2 h(U_x)$

Aim to have $(f'(U_x))^2 h(U_x) = constant \Rightarrow f'(u) = \frac{c}{\sqrt{h(u)}}$

$f(u) = \int \frac{c\,du}{\sqrt{h(u)}}$ when $h(u) = u^2$, $f(u) = \log u$; when $h(u) = u^{2r}$, $f(u) = cu^{1-r}$

② Box-cox

$Y^* = \frac{Y^\lambda - 1}{\lambda}$ 找 min SSE(λ) 先进行探索 决定是否引入.

变换带来的问题是: 可解释性↓

log变换后 解释平均变化的百分比

$f(E(Y)) \neq E(f(Y))$ doesn't answer of that interest.

[illegible] (不是必然的) [illegible]: [illegible] γ越大 [illegible] 越小..

Lec 7 Matrix Approach to Linear Regression & MLR

$Y = X\beta + \varepsilon$ $Y_{n\times1}$ $X_{n\times2}$ $\beta_{2\times1}$ $\varepsilon_{n\times1}$

$\sigma^2(\varepsilon) = \sigma^2 I_n$ $Y \sim N(X\beta, \sigma^2 I_n)$

$\vec{b} = \hat{\beta} = (X^TX)^{-1}X^TY$ Normal Eq. 正定

求 $\vec{b}$: $Q(\beta) = \|Y - X\beta\|^2$ $\frac{\partial Q}{\partial \beta} = -2\boxed{X^T(Y - X\beta) = 0}$ $\frac{\partial^2 Q}{\partial \beta^2} = 2X^TX$

$\hat{Y} = HY$ $H = X(X^TX)^{-1}X^T$ $\vec{e} = Y - \hat{Y} = (I - H)Y$

$\vec{b} \sim N(\beta, \sigma^2(X^TX)^{-1})$ $s^2(b) = s^2(X^TX)^{-1}$ $s^2 = \frac{e^Te}{n-2} = \frac{Y^T(I-H)Y}{n-2}$.

MLR: $Y = X\beta + \varepsilon$ $Y_{n\times1}$ $X_{n\times p}$ $\beta_{p\times1}$ $\varepsilon_{n\times1}$

β_i: partial coff 实际含义是其它值固定时单位自变量的响应.

MLR中需要注意 $rank(X) = p$ P是解释变量个数+1

$s^2 = \frac{Y^T(I-H)Y}{n-p}$ $s^2 = MSE$.

Gauss-Markov Theorem: Linearity. Randomness, Non-collinearity, Exogeneity (外生性), Homoscedasticity

under Gauss-Markov Thm ⇒ BLUE. 不要求正态性. ⇔ $rank(X) = p$

H的含义: 建立 $\hat{Y}$ 与 Y 的映射. $h_{ii} \in (0,1) = \frac{\partial \hat{Y}_i}{\partial Y_i}$

$Var(e_i) = \sigma^2(1 - h_{ii})$ $cov(e_i, e_j) = -\sigma^2 h_{ij}$

残差中间大, 两边小, 拟合能力中间弱, 预测能力两端弱

Anova Table: df(Reg) = p-1 df(Error) = n-p df(T) = n-1

Anova F-test: $H_0: \beta_1 = \cdots = \beta_{p-1} = 0$ $F^* = \frac{MSM}{MSE}$

$F^* \overset{H_0}{\sim} F_{p-1, n-p}$ reject H_0 means there is one or more not zero.

$R^2 = r^2 = \frac{SSM}{SST} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$

adjusted R^2: $R_a^2 = 1 - \frac{MSE}{MST} = \frac{R^2}{1-R^2} \cdot \frac{n-p}{p-1}$

R^2 会随着 p 增加而增加, R_a^2 则对解释变量做了 trade-off

R_a^2 与 $\hat{\sigma}^2$ 是此消彼长的关系.

Lec 8.

模型选择: { 预测新值: 共线性允许, 可以引入更多变量; 评估参数: 尽量避免共线性 (collinearity)

为什么模型参数显著但 R^2 小: n 很大时即使 R^2 小 也能显著.

MLR 中的检验与推断:

对 $\vec{b}$: $\vec{b} = (b_0 \cdots b_{p-1})^T \sim N(\beta, \sigma^2(X^TX)^{-1})$

$b_k \sim N(\beta_k, \sigma^2((X^TX)^{-1}_{k,k}))$ 最后加进模型

$t^* = \frac{b_k}{s(b_k)}$ $t^* \overset{H_0}{\sim} t_{n-p}$ $H_0: \beta_k = 0$ "variable added last"

$s^2(b_k) = MSE((X^TX)^{-1}_{k,k})$ $s^2(b) = MSE(X^TX)^{-1}$

CI: $b_k \pm t_{1-\frac{\alpha}{2}, n-p} s(b_k)$

对 $\hat{Y}$:

$\hat{u}_h = X_h\vec{b} = X_h(X^TX)^{-1}X^TY \sim N(u_h, \sigma^2 X_h(X^TX)^{-1}X_h^T)$

$s^2(\hat{u}_h) = s^2 X_h(X^TX)^{-1}X_h^T$ CI: $t_{1-\alpha/2, n-p}\, s(\hat{u}_h)$

predict: $\hat{Y}_h = X_{(h)}(X^TX)^{-1}X^TY \quad Y_h = X_{(h)}\beta + \Sigma$.

$\hat{Y}_h - Y_h \sim N(0, \sigma^2 + \sigma^2 X_{(h)}(X^TX)^{-1}X_{(h)}^T)$

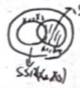
$S^2(\hat{Y}_h - Y_h) = S^2[1 + X_{(h)}(X^TX)^{-1}X_{(h)}^T]$

CI: $\hat{Y}_h \pm t_{\frac{\alpha}{2}, n-p}\, S(\hat{Y}_h - Y_h)$

MLR t-test 与 F-test 区别:

Anova table 顺序进入 t-test 是最后一个进入.

Lec9 Extra Sum of Squares & General Linear Test.



- Type I: 顺序进入
- Type II: hierachical (低阶先进一阶)
- Type III: unique.

Term	Type I	Type II	III
A	$SSR(A) = SSE(1) - SSE(A)$	$SSR(A\|B) = SSE(B) - SSE(A,B)$	总认为最后进
B	$SSR(B\|A) = SSE(A) - SSE(A,B)$	$SSR(B\|A) = SSE(A) - SSE(A,B)$	
AB	$SSR(AB\|A,B) = SSE(A,B) - SSE(A,B,AB)$	no interaction	

选择方式: 有顺序选 I, 无序且高阶项选 III, 否则 II.
III 不满足相加和 = SST 的要求.

GLT: (Partial) F-test

$H_0: \beta_{a_1} = \cdots = \beta_{a_r} = 0 \quad a_1 \cdots a_r$ 为 $\{1, \cdots p-1\}$ 中的 r 个值.

$$F^* = \frac{(SSE(R) - SSE(F)) / (df_E(R) - df_E(F))}{SSE(F)/df_E(F)} \overset{H_0}{\sim} F_{df_E(R)-df_E(F),\, df_E(F)}$$

$df_E(F) = n - p$ 注意 p 是解释变量数 +1

t-test 与 F-test 联系: R 与 F 只差一个参数时, General Linear Test 基于 $SSR(X_i | X_1 \cdots X_{i-1} X_{i+1} \cdots X_{p-1})$ $F_{1,n-p} = t^2_{n-p}$

当 F-R 包含所有参数时, 退化为 Anova F-test.

当 Type I 在改变顺序后 SS 不变 ⇒ 不存在共线.

GLT 一般型式 $C_{q\times p} \quad C\beta = t$ q 为约束秩.

$$F = \frac{(C\hat{\beta} - t)^T [C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta} - t)}{q s^2} \sim F_{q, n-p}.$$

偏 R^2: $R^2_{Y k | 1, \cdots k-1, k+1 \cdots p-1} = \frac{SSR(X_k | X_1 \cdots X_{k-1} X_{k+1} \cdots X_{p-1})}{SSE(X_1 \cdots X_{p-1})}$

$\frac{①}{① + ②} = R^2_{Y1|2}$

与 R^2 关系: $R^2_{Yk|1 \cdots k-1\, k+1 \cdots p-1} = R^2(Y|1, \cdots k-1, k+1 \cdots p-1,\ k|1, \cdots k-1\, k+1 \cdots p-1)$

标准化: Centraling + Scaling $X^{*T}X^* = r_{XX} \quad X^{*T}Y^* = r_{YX}$

优势: 回归系数可比, 计算稳定, 回归系数更直观.

抑制变量 Suppressor Variable:

$SSR(X_2|X_1) > SSR(X_1)$ The general idea is that suppressor variable will suppress irrelevant variance of another independent variables.

Lec10 Multicollinearity and Polynomial Reg. Interaction Model

三种情况 ↑ 出现标志: 系数估计乘常, 整体显著但个体不显著.

① Zero Collinearity

e.g. X 是正交阵

$b_j = \frac{X_j^TY}{\|X_j\|^2} \quad t = \frac{b_j}{sd(b_j)} = \frac{X_j^TY / \|X_j\|^2}{\sqrt{MSE / \|X_j\|^2}}$ b_j 不受其它 X 影响 t 会影响.

Type I 与 Type III SS of X_j 相同 element.

p-value for testing β_j will change depend on MSE. 但 $Var(b_X)$ 本身不变

② 线性相关.

$CX_j = \text{const}$ for $\forall j$ ⇒ $rank(X^TX) < p$ 不可逆 b 不可求. 需要删除变量以进一步计算.

③ 一般情况.

$(X^TX)^{-1}$ unstable

making it difficult to interpret coff

Type I SS 与 Type III SS 不同, Anova F-test 与 t-test 不同.

$Var(b) = \sigma^2(X^TX)^{-1}$ inflates. R^2↑ ⇒ 过拟合风险.

目的:
- prediction: increase sample size 可以解决
- explanation: 存在问题

从解释性出发选择模型: 先看系数估计是否合理再看 R_a^2

补救措施: 收集更多 data. Pca. 选择变量. transform.

Polynomial Reg: 引入高次解释变量

问题. X 和 X^α 相关 ⇒ Centring 来补救.

交互模型:

Case 1: binary + Continues.

$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \Sigma \quad X_1 = 0$ or 1 分成两组,

$H_0: \beta_1 = \beta_3 = 0$ test 2 groups are same

$H_0: \beta_1 = 0$ test intercepts are same 做 GLT-test.

$H_0: \beta_3 = 0$ test slopes are the same

Case 2: 双连续

$Y = \beta_0 + (\beta_1 + \beta_3X_2)X_1 + \beta_2X_2 + \Sigma$

$Y = \beta_0 + (\beta_2 + \beta_3X_1)X_2 + \beta_1X_1 + \Sigma$

Lec11 Mode Selection and Diagnostics.

常规: K-fold 交叉验证, Subset Selection methods, Regularization and shrinkage methods, 包括岭回归和 Lasso.

Subset Selection: e.g. max R_a^2 or min MSE

$P \leq 30, 40$ 穷举搜索. P 较大: 贪心搜索.

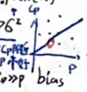
Mallow's C_p: mean square error 之和.

$$\sum_{i=1}^n E(\hat{Y}_i^P - \mu_i)^2 = \sum_{i=1}^n \underbrace{(E(\hat{Y}_i^P) - \mu_i)^2}_{bias} + \sum_{i=1}^n \underbrace{Var(\hat{Y}_i^P)}_{var}$$

$$= E(SSE(P)) - (n-p)\sigma^2 + p\sigma^2$$

$$\Gamma_p = \frac{\sum_{i=1}^n E(\hat{Y}_i^P - \mu_i)^2}{\sigma^2} = \frac{E(SSE(P))}{\sigma^2} - (n - 2p)$$

$C_p = \hat{\Gamma}_p = \frac{SSE(P)}{\hat{\sigma}^2} - (n-2p) \quad \hat{\sigma}^2 = MSE(P)$



$C_p \gg p$ bias

$C_p \ll p$ overfitting

$C_p = p$ 下直选较小的.

AIC: $AIC(p) = n\log(\frac{SSE(P)}{n}) + 2P$ minimize $AIC(P)$ 似然

BIC: $BIC(P) = n\log(\frac{SSE(P)}{n}) + \log(n)P$ minimize $BIC(P)$ 惩罚项更大 比 AIC 选的 P 更小

Press: $PRESS(P) = \sum_{i=1}^n (Y_i - \hat{Y}_{i(-i)})^2 \quad Y_i - \hat{Y}_{i(-i)} = \frac{e_i}{1 - h_{ii}}$ cross-validation

$\hat{Y}_{i(-i)} = (1, X_{i1}, \cdots X_{i(p-1)})\hat{\beta}_{(-i)} \quad \hat{\beta}_{(-i)} = (X_{(-i)}^TX_{(-i)})^{-1}X_{(-i)}^TY_{(-i)}$

每次去掉 i, 得到大模型, 计算与 Y_i 的差.

三种 R^2: $R^2 = 1 - \frac{SSE}{SST} \quad R^2_a = 1 - \frac{MSE}{MST} \quad R^2_P = 1 - \frac{PRESS}{SST}$

noise 太多. $R^2_P < 0$ ⇒ 模型过拟合. $R^2_a < 0$ ⇒ SSE > SST 模型欠拟 则加 X_i^α.

$R_P^2 \ll R^2$ ⇒ 出现过拟合.

Avplots: res $Y|X_{-i}$ vs. res $X_i|X_{-i}$ 图能看到明显 pattern residual plots 也可以近似看出来.

离群值诊断: (student) residuals $e_i^* = \frac{e_i}{\sqrt{MSE(1-h_{ii})}}$ 标准化.

强影响力点: $DFFIT = \hat{Y}_i - \hat{Y}_{i(-i)} = \frac{h_{ii}}{1-h_{ii}}e_i$ difference caused to fitted value.

Lec12 Remedies & Single Factor Analysis of Variance.

解决异方差.

Weighted Reg: minimize $\sum w_ie_i^2$ optimal: $w_i = \frac{k}{\sigma_i^2}$

Case①: Var known $w_i = \frac{1}{\sigma_i^2}$ $W = \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{bmatrix}$ $(W^{\frac{1}{2}}Y) = (W^{\frac{1}{2}}X)\beta + W^{\frac{1}{2}}\Sigma$

$Var(W^{\frac{1}{2}}\Sigma) = I \quad b_w = (X^TWX)^{-1}(X^TWY) \quad Var(b_w) = (X^TWX)^{-1}$

Case②: 比例已知 $W = k\begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{bmatrix}$ $Var(Y^*) = kI \quad Var\, b_w = k(X^TWX)^{-1}$

$\hat{k} = MSE_{wls} = \frac{\sum w_ie_i^2}{n-p}$.

Case③: 先 OLS, 再用 residual 估 $w_i = \frac{1}{\hat{\sigma}_i^2}$

wls 后的 $\hat{\sigma}$ 与 ols 不可直接比较.

Bias-Var trade off.

岭回归: 把 $b = \hat{\beta}$ 中的 X^TX 变为 $(X^TX + \lambda I)$ b 和 Var(b) 都变小.

Ridge is to minimize $(Y - Z\beta)^T(Y - Z\beta) + \lambda\sum\beta_j^2$ Z 为 std X

LASSO is to minimize $(Y - Z\beta)^T(Y - Z\beta) + \lambda\sum|\beta_j|$

Factor Anova 诊断 & 补救

Normality: Wilk-Shapiro, Anderson-Darling, kolmogorove-Smirnov

Homogeneity of var: Bartlett's, Modified Levene's test, Hartley's test

补救措施:

Delete Outliers, Transformation, Use weight, Nonparameter.

Lec13 Single Factor Anova & Inference.

Goals: focus on mean responses for diff level { 是否有差异 / 差异大小.

Cell means model:

$Y_{ij} = \mu_i + \varepsilon_{ij} \quad \hat{\mu}_i = \bar{Y}_i \quad s_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2/(n_i - 1)$

pooled estimate $s^2 = \sum(n_i - 1)s_i^2 / \sum(n_i - 1)$ r+1 参数 (含 σ^2)

Anova Table r 组.

Source	df	SS	MS
Model	$r-1$	$\sum_i n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	SSR/df_R
Error	$n_T - r$	$\sum_{i,j}(Y_{ij} - \bar{Y}_{i\cdot})^2$	SSE/df_E
Total	$n_T - 1$	$\sum_{i,j}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	SST/df_T

$F^* = \frac{MSR}{MSE} \overset{H_0}{\sim} F(r-1, n_T - r) \quad H_0: \mu_1 = \cdots = \mu_r = \mu$.

Coding $X = \begin{bmatrix} \vec{1}_{n_1} & & \\ & \vec{1}_{n_2} & \\ & & \ddots \\ & & & \vec{1}_{n_r} \end{bmatrix}$ no interpret. $\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix}$

Factor Effect Model:

$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ (r+2 个参数)

$X = \begin{bmatrix} 1 & 1 & & \\ 1 & & \ddots & \\ \vdots & & & \\ 1 & & & 1 \end{bmatrix} \quad \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_r \end{bmatrix}$

X 变为 $\begin{bmatrix} 1_{n_1} & 1_{n_1} & & \\ 1_{n_2} & & 1_{n_2} & \\ \vdots & & & 1_{n_{r-1}} \\ 1_{n_r} & -1_{n_r} & \cdots & -1_{n_r} \end{bmatrix}$

X^TX 无 inverse ⇒ 引入对 τ 限制 e.g. $\sum\tau_i = 0$

其它限制如 $\tau_r = 0 \quad \tau_1 = 0$ 等. R 中用的.

推断: $\bar{Y}_{i\cdot} \sim N(\mu_i, \sigma^2/n_i)$ CI: $\bar{Y}_i \pm t_c S/\sqrt{n_i} \quad t_c = (1 - \frac{\alpha}{2}, n_T - r)$

多重比较谬误: 多次实验中 α inflation ⇒ 未控制总体 α.

常用方法: Bonferroni, stepwise procedure

Test diff in Mean:

- 控制 family α? Fisher Least Significant Difference (LSD)
- r 是否很大? Bonferroni
- 成对? Tukey's HSD $q = \frac{\bar{y}_{max} - \bar{y}_{min}}{s/\sqrt{n}} \quad t_c \to q_c/\sqrt{2}$
- or linear comb Scheffe's method $\sum c_i\mu_i = 0,\ \sum c_i = 0 \quad t_c \to \sqrt{(a-1)F(1-\alpha; a-1, n_T - r)}$

Lec14 Two-way Anova

假设每组都有 n 个

Cell means Model: $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \quad \bar{Y}_{ij} = \sum_k Y_{ijk}/n \quad s_{ij}^2 = \sum_k \frac{(Y_{ijk} - \bar{Y}_{ij})^2}{n-1}$

评价因素效应: 对 $\bar{\mu}_{i\cdot}$ 和 $\bar{\mu}_{\cdot j}$ 判断 AB 是否相互影响 ⇒ 平行线检测.

因素效应模型: $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ 待估参数 $a-1 + b-1 + (a-1)(b-1) + 1(\sigma^2)$

Two way anova

SS	df	MS	E(MS)
$SSA = \sum_{ijk}\hat{\alpha}_i^2 = \sum_{ijk}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$a-1$	$MSA = SSA/df_A$	$\sigma^2 + \frac{bn}{a-1}\sum\alpha_i^2$
$SSB = \sum_{ijk}\hat{\beta}_j^2$	$b-1$	$MSB = SSB/df_B$	$\sigma^2 + \frac{an}{b-1}\sum\beta_j^2$
$SSAB = \sum_{ijk}(\hat{\alpha\beta})_{ij}^2$	$(a-1)(b-1)$	$MSAB = SSAB/df_{AB}$	$\sigma^2 + \frac{n}{(a-1)(b-1)}\sum(\alpha\beta)^2_{ij}$
$SSE = \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$ab(n-1)$	$MSE = SSE/df_E$	
$SST = \sum_{ijk}(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$	$abn - 1 = n_T - 1$	$MST = SST/df_T$	

$df_{SST} - df_{SSR}$ 算出来的.