1. 概率论

§1. 特殊分布性质

1. χ^2: $X_i \overset{iid}{\sim} N(0,1)$, $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$, $\mu=n$, $\sigma^2=2n$, $\frac{1}{\Gamma(\frac{n}{2})2^{n/2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$, $\Gamma(t)=\int_0^\infty y^{t-1}e^{-y}dy$
2. t: $X\sim N(0,1)$, $Y\sim\chi^2(n)$ 独, $\frac{X}{\sqrt{Y/n}} \sim t(n)$, $\mu=0$, $\sigma^2=\frac{n}{n-2}$
3. F: ① $X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$ 独, $\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$
   ② $F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$ $N_{0.995}=2.576$
4. N: $N_{0.95}=1.645$ $N_{0.975}=1.96$ $N_{0.99}=2.326$

§2. 重要定理

1. 若X连续, F(x)严格增, 则 $F_X(X)\sim Unif[0,1]$
2. C-S inequality: $[E(XY)]^2 \le E[X^2]E[Y^2]$
3. Law of LN: $\bar{Z}=\frac1n\sum_{i=1}^n X_i \to_p \mu$
4. CoLT: $\frac{\bar X-\mu}{\sigma/\sqrt n} \to_d N(0,1)$ $(n>30)$
5. 重要样本分布 $S_p^2=\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$

$S=\frac{1}{n-1}\sum(X_i-\bar X)^2$ one population two population

$\bar X$ σ^2已知 $\frac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$ $\frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$

独 σ^2未知 $\frac{\bar X-\mu}{S/\sqrt n}\sim t(n-1)$ $\frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$

S^2 $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

2. SLR

§1 模型建立

1. 假设: $E(e_i)=0$, $var(e_i)=const$, e_i间 uncorre…
2. vertical line?: ①平等 treat x, y ②试图 pred y from x
3. 选 h(·): ①=1: pointwise median ②=2: … mean
4. 选 q=2?: ① conditional mean ②MSE常见 ③ var↓

§2 模型求解 $s.d.(x)=\sqrt{\frac{1}{n-1}S_{xx}}$ $s.d.(y)=\sqrt{\frac{1}{n-1}S_{yy}}$

1. 目标: min $RSS := \sum_i (y_i-\beta_0-\beta_1x_i)^2$
2. Notation: ① $S_{xx}=\sum_{i=1}^n(X_i-\bar X)^2=\sum_{i=1}^n X_i^2-n\bar x^2$
   ② $S_{xy}=\sum_{i=1}^n(X_i-\bar X)(Y_i-\bar Y)=\sum_{i=1}^n x_iy_i-n\bar x\bar y=\sum_{i=1}^n(x_i-\bar x)y_i$
3. 求解 b_0, b_1 ($\hat\beta_0,\hat\beta_1$)

$$\begin{cases}\frac{\partial RSS}{\partial\beta_0}=-2\sum_i(y_i-\beta_0-\beta_1x_i)=0\\ \frac{\partial RSS}{\partial\beta_1}=-2\sum_i(y_i-\beta_0-\beta_1x_i)x_i=0\end{cases}\Rightarrow\begin{cases}\sum y_i = nb_0+b_1\sum x_i & ①\\ \sum x_iy_i=b_0\sum x_i+b_1\sum x_i^2 & ②\end{cases}$$

由①: $\bar y=b_0+b_1\bar x \Rightarrow b_0=\bar y-b_1\bar x$, 由②: $\sum x_iy_i=(\bar y-b_1\bar x)n\bar x+b_1\sum x_i^2$
$=n\bar x\bar y-b_1n\bar x^2+b_1\sum x_i^2 \Rightarrow b_1=\frac{\sum x_iy_i-n\bar x\bar y}{\sum x_i^2-n\bar x^2}=\frac{\sum(x_i-\bar x)(y_i-\bar y)}{\sum(x_i-\bar x)^2}=\frac{S_{xy}}{S_{xx}}$

§3 模型性质 (仅在有 intercept 时成立)

① $\sum\hat e_i=\sum_{i=1}^n(y_i-(\bar y-b_1\bar x)-b_1x_i)=\cdots=0$
② $\sum_{i=1}^n \hat e_i x_i=\sum_{i=1}^n(y_i-b_0-b_1x_i)x_i=\frac{\partial RSS}{\partial\beta_1}=0$
③ $\sum_{i=1}^n\hat e_i\hat y_i=\sum_{i=1}^n(y_i-b_0-b_1x_i)(b_0+b_1x_i)=b_0\frac{\partial RSS}{\partial b_0}+b_1\frac{\partial RSS}{\partial b_1}=0$
④ $\sum_{i=1}^n\hat y_i=\sum_{i=1}^n(\bar y-b_1\bar x+b_1x_i)=n\bar y-b_1n\bar x+b_1\sum x_i=\sum_{i=1}^n y_i$
⑤ $S^2=\hat\sigma^2=\frac{RSS}{n-2}=\frac{1}{n-2}\sum_{i=1}^n\hat e_i^2$, $\hat e_i=y_i-(\hat\beta_0+\hat\beta_1x_i)$

§4 参数分布 $\frac{(n-2)S^2}{\sigma^2}\sim\chi^2(n-2)$

① $E(\hat\beta_1)=\frac{1}{S_{xx}}\sum(x_i-\bar x)\underbrace{(\beta_0+\beta_1x_i)}_{E(y_i)} \overset{全部展开}{=}\cdots=\beta_1$
② $E(\hat\beta_0)=E(\bar y)-E(\hat\beta_1)\bar x=\frac1n\sum E(y_i)-\beta_1\bar x=\cdots=\beta_0$
③ $Var(\hat\beta_1)=Var(\frac{\sum(x_i-\bar x)y_i}{S_{xx}})=\frac{1}{S_{xx}^2}\sum_{i=1}^n[(x_i-\bar x)^2Var(y_i)]=\frac{\sigma^2}{S_{xx}}$
④ $Var(\hat\beta_0)=Var(\bar y-\hat\beta_1\bar x)=Var(\bar y)-2Cov(\bar y,\hat\beta_1\bar x)+Var(\hat\beta_1\bar x)$
(1) $=\frac{1}{n^2}\sum Var(y_i)=\frac{\sigma^2}{n}$ (3) $=\bar x^2Var(b_1)=\bar x^2\frac{\sigma^2}{S_{xx}}$
(2) $=\sum_{i=1}^n\sum_{j=1}^n\frac1n\frac{x_j-\bar x}{S_{xx}}\cdot\bar x\cdot Cov(y_i,y_j)=0$
$Var(\hat\beta_0)=\sigma^2(\frac1n+\frac{\bar x^2}{S_{xx}})$
⑤ $Cov(\hat\beta_0,\hat\beta_1)=Cov(\bar y,\hat\beta_1)-\bar x Cov(\hat\beta_1,\hat\beta_1)=-\frac{\bar x\sigma^2}{S_{xx}}$

3. CI 与 T-test

§1 关于 β_1 (σ^2未知) ($H_0:\beta_1=\beta_1^0$)

1. $\hat\beta_1$分布: $\hat\beta_1\sim N(\beta_1,\frac{\sigma^2}{\sum(x_i-\bar x)^2})$
2. T统计量: $T=\frac{\hat\beta_1-\beta_1^0}{S/\sqrt{\sum(X_i-\bar X)^2}}\sim t(n-2)$ (H_0下)
3. SE: $s.e.(\hat\beta_1)=\sqrt{S^2/\sum(X_i-\bar X)^2}$
4. CI: $\hat\beta_1\pm S/\sqrt{S_{xx}}\cdot t(1-\frac\alpha2,n-2)$
5. P: $\hat p=\inf\{\alpha\in[0,1]:|T|>t_{1-\frac\alpha2}(n-2)\}=1-\Phi_{t,n-2}(|T|)+\Phi_{t,n-2}(-|T|)$

§2 关于 β_0 (σ^2未知) ($H_0:\beta_0=\beta_0^0$)

1. $\hat\beta_0$分布: $\hat\beta_0\sim N(\beta_0,\sigma^2(\frac1n+\frac{\bar x^2}{S_{xx}}))$
2. T统计量: $T=\frac{\hat\beta_0-\beta_0^0}{S\sqrt{\frac1n+\frac{\bar x^2}{S_{xx}}}}\sim t(n-2)$ (H_0下)
3. SE: $s.e.(\hat\beta_0)=S\sqrt{\frac1n+\frac{\bar x^2}{S_{xx}}}$
4. CI: $\hat\beta_0\pm S\sqrt{\frac1n+\frac{\bar x^2}{S_{xx}}}\cdot t(1-\frac\alpha2,n-2)$
5. P: $\hat p=1-\Phi_{t,n-2}(|T|)+\Phi_{t,n-2}(-|T|)$

§3 关于 Fitted (predicted) value μ_0

1. $\hat\mu_0$分布: $\hat\mu_0\sim N(\beta_0+\beta_1x_0,\sigma^2[\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}])$
   $Var(\hat\mu_0)=Var(\hat\beta_0)+x_0^2Var(\hat\beta_1)+2x_0Cov(\hat\beta_0,\hat\beta_1)\cdots$
2. T统计量: $T=\frac{\hat\mu_0-\mu_0}{S\sqrt{\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}}\sim t(n-2)$
3. SE: $s.e.(\hat\mu_0)=S\sqrt{\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}$
4. CI: $\hat\beta_0+\hat\beta_1x_0\pm S\sqrt{\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}\cdot t(1-\frac\alpha2,n-2)$

§4 关于 True observation / prediction error

1. $y_p-\hat y_p$分布: $y_p-\hat y_p\sim N(0,\sigma^2[1+\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}])$
   $Var(y_p-\hat y_p)=Var(\underbrace{\mu_p}_{const}-\hat\mu_p)+Var(\varepsilon_p)+2\underbrace{Cov(\cdots)}_{=0}$
2. T统计量: $T=\frac{y_p-\hat y_p}{S\sqrt{1+\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}}\sim t(n-2)$
3. SE: $s.e.(y_p-\hat y_p)=S\sqrt{1+\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}$
4. CI: $\hat\beta_0+\hat\beta_1x_0\pm S\sqrt{1+\frac1n+\frac{(x_0-\bar x)^2}{S_{xx}}}\cdot t(1-\frac\alpha2,n-2)$

§5 双样本 difference in mean (var未知但相同)

$\bar X=\frac{n_2}{n_1+n_2}$ $S_{xx}=\frac{n_1n_2}{n_1+n_2}$ $\bar y=\frac{\sum^{n_1}y_{1i}+\sum^{n_2}y_{2i}}{n_1+n_2}$

$\hat\beta_1=\bar y_2-\bar y_1$, $\hat\beta_0=\bar y_1$ $S^2=S_p^2$ $T=\frac{\bar y_2-\bar y_1}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

4. R² 与 ANOVA F-test

§1 相关统计量

1. TSS/SST: $TSS=\sum_{i=1}^n(y_i-\bar y)^2$ ($=S_{yy}$) d.f.=n-1
2. SSE/RSS: $SSE=\sum_{i=1}^n(y_i-\hat y_i)^2$ d.f.=n-2
3. SSR: $SSR=\sum_{i=1}^n(\hat y_i-\bar y)^2(=\hat\beta_1^2S_{xx})$ d.f.=1
4. TSS=SSE+SSR $\sum(y_i-\bar y)^2=\sum(y_i-\hat y_i+\hat y_i-\bar y)^2$
   $=SSE+SSR+2\underline{\sum(y_i-\hat y_i)(\hat y_i-\bar y)}$
   $=\sum(y_i-\hat y_i)\hat y_i-\sum(y_i-\hat y_i)\bar y=\sum\hat e_i\hat y_i-\bar y\sum\hat e_i=0$

§2 Coefficient of determination R^2 $=corr(x,y)$

$R^2=\frac{SSR}{TSS}=1-\frac{SSE}{TSS}=(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}})^2=\rho^2=r^2$

not resistent to outliers; no intercept → x meaningful

$r_{\hat e,\hat y}=\sum\hat e_i\hat y_i-n\bar{\hat e}\bar{\hat y}=0$ $r_{\hat e,x}=0$

§3 ANOVA-F-test

1. MSR: $MSR=\frac{SSR}{1}=\hat\beta_1^2S_{xx}$, $E(MSR)=\sigma^2+\beta_1^2S_{xx}$
2. MSE: $MSE=\frac{SSE}{n-2}=\frac{\sum\hat e_i^2}{n-2}$, $E(MSE)=E(S^2)=\sigma^2$
3. F-test: $\frac{SSE}{\sigma^2}\sim\chi^2(n-2)$, $\frac{TSS}{\sigma^2}\sim\chi^2(n-1)$, $\frac{SSR}{\sigma^2}\sim\chi^2(1)$
   H_0下, $\frac{y_i-\beta_0}{\sigma}\sim N(0,1)\Rightarrow(\cdot)^2\sim\chi^2_{(1)}\Rightarrow\sum(\cdot)^2\sim\chi^2_{(n)}$
   $\Rightarrow\frac{TSS}{\sigma^2}\sim\chi^2_{(n-1)}$ $(\hat\beta_0=\bar y)$; $\frac{\hat\beta_1}{\sqrt{\sigma^2/S_{xx}}}\sim N(0,1)\Rightarrow(\cdot)^2\sim\chi^2_{(1)}$
   $F=\frac{SSR/1}{SSE/(n-2)}=\frac{MSR}{MSE}\sim F_{1,n-2}$
4. $t_{obs}^2=F_{obs}$ $t_{obs}=\frac{\hat\beta_1}{S/\sqrt{S_{xx}}}=\frac{\hat\beta_1}{\sqrt{MSE}/\sqrt{S_{xx}}}$; $(\cdot)^2=\frac{MSR}{MSE}$
5. ANOVA table

	SS	d.f	MS	F
Regression line	$b_1^2S_{xx}$	1	$b_1^2S_{xx}$	$b_1^2S_{xx}/S^2$
Error	$\sum\hat e_i^2$	n-2	S^2	
Total	$\sum(y_i-\bar y)^2$	n-1		

5. Regression diagnostics

(MLR)

§1 Check 2: leverage points

1. $\hat y_i=\sum_{j=1}^n(\frac1n+\frac{(x_i-\bar x)(x_j-\bar x)}{S_{xx}})y_j=\sum_{j=1}^n h_{ij}y_j$
2. $h_{ii}=\frac1n+\frac{(x_i-\bar x)^2}{S_{xx}}>\frac4n$ $(\frac{2(p+1)}{n})$ ⇒ leverage point
3. ① $\frac1n\sum_i h_{ii}=\frac1n[1+\frac{\sum(x_i-\bar x)^2}{S_{xx}}]=\frac2n$
   ② $\sum_j h_{ij}=1+\frac{(x_i-\bar x)}{S_{xx}}\sum_j(x_j-\bar x)=1$
   ③ $\sum_j h_{ij}^2=\sum\frac{1}{n^2}+2\cdot\frac1n\cdot0+\frac{(x_i-\bar x)^2}{S_{xx}^2}\sum(x_j-\bar x)^2=h_{ii}$

(MLR 也有变化)

§2 Check 1: (Plot) standard residual & 3: outlier

1. residual: $\hat e_i=y_i-\hat y_i=y_i-\sum_j h_{ij}y_j$
2. $Var(\hat e_i)=V[(1-h_{ii})y_i-\sum_{j\ne i}h_{ij}y_j]=\sigma^2[1-h_{ii}]$
   $Var(\hat y_i)=\sum_j V(h_{ij}y_j)=\sigma^2h_{ii}=\sigma^2[\frac1n+\frac{(x_i-\bar x)^2}{S_{xx}}]$
3. Standard residual: $r_i=\frac{\hat e_i}{S\sqrt{1-h_{ii}}}$ >2 / <-2 ⇒ outlier

§3 Check 4: influential points

1. $DFBETA_{ik}=\frac{\hat\beta_k-\hat\beta_{k(i)}}{s.e.(\hat\beta_{k(i)})}$ $|\cdot|\begin{cases}>2/\sqrt n & \text{large}\\ >1 & \text{small data}\end{cases}$
2. $DFFITS_i=\frac{\hat y_i-\hat y_{i(i)}}{s.e.(\hat y_{i(i)})}$ $|\cdot|\begin{cases}>2\sqrt{2/n}\\ >1\end{cases}$ $>2\sqrt{\frac{p+1}{n}}$
3. Cook's dist: $D_i=\frac{\sum_j(\hat y_{j(i)}-\hat y_j)^2}{2S^2}=\frac{r_i^2h_{ii}}{2(1-h_{ii})}>\frac{4}{n-p-1}$

§4 Residual plot 的作用 $cov(\hat e_i,\hat e_j)=-h_{ij}\sigma^2$ $i\ne j$

check for ① linearity ($\hat e_i$ vs x_i) ② missing predictor variables
③ homoscedasticity (check 5) ($\sqrt{|r_i|}$ v.s. $\hat y_i$) (MLR) ④ normality (r_i vs $\hat y_i$)

§5 Check 6: error independent ($\hat e_i$ vs time/dist)

§6 Check 7: normality of error

1. check symmetry: boxplot, histogram, dot plot…
2. check heavy tails: QQ plot $E(X_{(j)})=\frac{j}{n+1}$ (unif(0,1))
   $X_{(k)}$ v.s. $F^{-1}(\frac{k}{n+1})$ light tailed: heavy:
   left skew: right: bimodal:

$\hat\beta_{1w}=\frac{\sum w_i(x_i-\bar x_w)(y_i-\bar y_w)}{\sum w_i(x_i-\bar x_w)^2}$ $\hat\beta_{0w}=\bar y_w-\hat\beta_{1w}\bar x_w$

$\bar x_w=\sum w_ix_i/\sum w_i$ $h_{wij}=[w_j^*+\frac{w_j(x_i-\bar x_w)(x_j-\bar x_w)}{WSXX}]$

$WSXX=\sum w_j(x_j-\bar x_w)^2$ $w_j^*=w_j/\sum w_k$

6. 均值回归 & 变量转换

§1 Regression towards the mean

$\hat{y} = (\bar{y} - \hat{\beta}_1 \bar{x}) + (\overset{=\hat{\beta}_1}{r \frac{S_y}{S_x}}) x \Rightarrow \frac{\hat{y} - \bar{y}}{S_y} = \overset{|r|<1}{\underset{\sim\sim}{r}} \frac{x - \bar{x}}{S_x}$

§2 Transformations （对 x tran.: skew）

1. The Delta method: $z = f(Y) \approx f(\mu) + (Y - \mu) f'(\mu)$

   $E(z) \approx f(\mu)$, $var(z) \approx \sigma_Y^2 [f'(\mu)]^2$

2、假设 $E(Y_i) = \mu_i$, $var(Y_i) \propto V(\mu_i)$ 求 $z = f(Y)$

   $\Rightarrow f(\mu) \propto \int \frac{1}{\sqrt{V(\mu)}} d\mu$

3、Square root transformation ($Y_i \sim Poi(\mu_i)$)

   若 $var(Y) \propto V(\mu) = \mu \Rightarrow f(\mu) \propto \sqrt{\mu}$

   Logarithmic transformation ($Y_i \sim Exp(\lambda)$)

   若 $var(Y) \propto V(\mu) = \mu^2 \Rightarrow f(\mu) \propto \ln \mu$

   Reciprocal transformation

   若 $var(Y) \propto V(\mu) = \mu^4 \Rightarrow f(\mu) \propto \frac{1}{\mu}$

4. 优先级: ① $E(\mathbf{e}) = \mathbf{0}$ ② errors independency ③ constant variance (robust) ④ Normality (对 CI 重要)

5. 对 y tran.: ① nonconst. var. ② nonlin. ③ Nor.

7. SLR 与 MLR 的矩阵表示

§1 Operations

$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$

1. Exp: ① $E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$ ② $E(a\mathbf{x}) = aE(\mathbf{x})$

   ③ $E(A\mathbf{x}) = AE(\mathbf{x})$ ④ $E(\mathbf{a}^T\mathbf{x}) = \mathbf{a}^T E(\mathbf{x})$

2. Var.: ① $Var(\mathbf{x}) = E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T] = E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}^T]$ ② $Var(A\mathbf{x}) = A \cdot var(\mathbf{x}) A^T$

3. gradient: ① $\frac{\partial \mathbf{c}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{c}$ ② $\frac{\partial \mathbf{x}^T A \mathbf{x}}{\partial \mathbf{x}} \overset{(A\ sym)}{=} 2A\mathbf{x}$

   ③ $\frac{d(\mathbf{w}^T X \mathbf{w})}{d\mathbf{w}} = (X + X^T)\mathbf{w}$ ④ $\frac{\partial A^T \mathbf{x}}{\partial \mathbf{x}} = A$

§2 SLR 点估计量的矩阵表示

1. $\mathbf{y}$ 的分布: $\mathbf{y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$ $\Sigma^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu}) \sim N(0, I_n)$

2. $\boldsymbol{\beta}$ 的求解 $\hat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\mathbf{y}$ (若 X^TX 可逆)

   $X^TX = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$, $(X^TX)^{-1} = \frac{1}{S_{xx}} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$

   X^TX 可逆 $\Leftrightarrow det(X^TX) = 0 \Leftrightarrow rank(X^TX) = p+1$

§3 Hat matrix

1. Hat matrix: $H = X(X^TX)^{-1}X^T \Rightarrow \hat{\mathbf{y}} = H\mathbf{y}$

2. 性质: ① idempotent: $H^2 = H$ ② $H^T = H$

   ③ $HX = X \Rightarrow$ ④ $H\mathbf{1} = \mathbf{1} / HJ = J$ (存在 intercept)

3. Idempotent matrix 性质 ① $trace(A) = rank(A)$

   $\exists A = B^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}C^{-1} \overset{A^2=A}{\Rightarrow} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}C^{-1}B^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

   $rank(A) = r = tr(\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}) = tr(\quad) = tr(A)$

   ② idempotent $\Leftrightarrow rank(A) + rank(\underset{idem.}{I - A}) = n$

   $rank(I - A) = tr(I - A) = tr(I) - tr(A) = n - rank(A)$

   ③ H, $I - H$, $\frac{1}{n}J$, $H - \frac{1}{n}J$ 为 idempotent

   ④ idempotent $A = B + C \Rightarrow r(A) = r(B) + r(C)$

4. Leverage h_{ij} 性质 ① $\sum_{i=1}^n h_{ii} = tr(XX^T(X^TX)^{-1}) = p+1$

   ② $H\mathbf{1} = \mathbf{1} \Rightarrow \sum_{j=1}^n h_{ij} = 1$ ③ $H^T = H \Rightarrow h_{ij} = h_{ji}$

   ④ $(\mathbf{e}_i^T H)(\mathbf{e}_i^T H)^T = \sum_{j=1}^n h_{ij}^2 = \mathbf{e}_i^T H H^T \mathbf{e}_i = h_{ii}$

   ⑤ $D_i = \frac{\sum_{j=1}^n (\hat{y}_{j(i)} - \hat{y}_j)^2}{2S^2} = \frac{r_i^2 h_{ii}}{2(1 - h_{ii})}$ $S^2 = \frac{1}{n-p-1}\sum_{i=1}^n (\hat{y}_i - y_i)^2$

   $r_i = \frac{\hat{e}_i}{S\sqrt{1 - h_{ii}}}$, $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} = \mathbf{x}^T(X^TX)^{-1}\mathbf{x}$

   $\hat{\boldsymbol{\beta}}_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \mathbf{y}_{(i)} = (X^TX - \mathbf{x}_i\mathbf{x}_i^T)^{-1}(X^T\mathbf{y} - \mathbf{x}_i y_i)$

   $(A + \mathbf{b}\mathbf{c}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{b}\mathbf{c}^TA^{-1}}{1 + \mathbf{c}^TA^{-1}\mathbf{b}}$ (提出 $\frac{1}{1-h_{ii}}(X^TX)^{-1}\mathbf{x}_i$)

   $\Rightarrow \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{\hat{e}_i}{1 - h_{ii}}(X^TX)^{-1}\mathbf{x}_i$

   $D_i = \frac{[X(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})]^T[X(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})]}{2S^2}$

   $= \frac{\hat{e}_i^2 [X(X^TX)^{-1}\mathbf{x}_i]^T[X(X^TX)^{-1}\mathbf{x}_i]}{(1 - h_{ii})^2 \cdot 2S^2} = \frac{r_i^2 h_{ii}}{(p+1)(1 - h_{ii})}$

5. $\hat{\boldsymbol{\beta}}$ 的性质: ① $E[\hat{\boldsymbol{\beta}}] = (X^TX)^{-1}X^TE[\mathbf{y}] = \boldsymbol{\beta}$

   ② $Var(\hat{\boldsymbol{\beta}}) = (X^TX)^{-1}X^TVar(\mathbf{y})X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$

   $V(\hat{\beta}_i) = \sigma^2 V_{ii}$, $Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 V_{ij}$

   ③ $E(\mathbf{a}^T\hat{\boldsymbol{\beta}}) = \mathbf{a}^T\boldsymbol{\beta}$, $V(\mathbf{a}^T\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{a}^T(X^TX)^{-1}\mathbf{a}$

   ④ MLE $\boldsymbol{\beta}$ 为 the BEST estimator 考虑另一个 $\hat{\mathbf{b}} = M\mathbf{y}$ s.t. $E(\hat{\mathbf{b}}) = \boldsymbol{\beta}$, 定义 $\Delta M = M - (X^TX)^{-1}X^T$ 则 $\hat{\mathbf{b}} = (\Delta M + (X^TX)^{-1}X^T)\mathbf{y}$, $E(\hat{\mathbf{b}}) = \Delta MX\boldsymbol{\beta} + \boldsymbol{\beta}$ 则 $\Delta MX = \mathbf{0}$, 因此 $Var(\hat{\mathbf{b}}) = \sigma^2\Delta M\Delta M^T + \sigma^2(X^TX)^{-1} \geq Var(\hat{\boldsymbol{\beta}})$

6. Fitted value $\hat{\boldsymbol{\mu}}$ / $\hat{\mathbf{y}}$ 的性质 $\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}} = H\mathbf{y}$

   ① $E[\hat{\boldsymbol{\mu}}] = X\boldsymbol{\beta} = \boldsymbol{\mu}$ ② $V(\hat{\boldsymbol{\mu}}) = HV(\mathbf{y})H^T = H\sigma^2$

7. Residual 的性质: $\hat{\mathbf{e}} = \mathbf{y} - X\hat{\boldsymbol{\beta}} = (I - H)\mathbf{y}$

   ① $E(\hat{\mathbf{e}}) = (I - H)E(\mathbf{y}) = (X - X)\boldsymbol{\beta} = \mathbf{0}$

   ② $Var(\hat{\mathbf{e}}) = (I - H)Var(\mathbf{y})(I - H)^T = \sigma^2(I - H)$

   ③ $rank(var(\hat{\mathbf{e}})) = r(I - H) = r(I) - r(H) = n - p - 1$

8. MLR 的 CI 和假设检验

§1 $\hat{\beta}_i$ 的分布

$\hat{\beta}_i \sim N(\beta_i, \sigma^2 V_{ii})$

§2 σ^2 的无偏估计量

$S^2 = \frac{SSE}{n - (p+1)} = \frac{S(\hat{\boldsymbol{\beta}})}{n - (p+1)} = \frac{\sum_i (y_i - \hat{y})^2}{n - (p+1)} = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{n - (p+1)}$

有以下结论: ① $\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{V_{ii}}} \sim N(0, 1)$ $S^2 = f(\hat{\mathbf{e}})$ 证 $V(\begin{bmatrix} \hat{\mathbf{e}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix})$ 中 $Cov(\hat{\mathbf{e}}, \hat{\boldsymbol{\beta}}) = 0$

② $\frac{(n-p-1)s^2}{\sigma^2} \sim \chi_{n-p-1}^2$ ③ $s^2 \perp \hat{\boldsymbol{\beta}}$

§3 β_i 的 confidence interval 和 hypothesis testing

① T 统计量: $T = \frac{\hat{\beta}_i - \beta_i}{s\sqrt{V_{ii}}} \sim t_{n-p-1}$

② β_i 的 CI: $\hat{\beta}_i \pm t_{\alpha/2, n-p-1} \cdot s\sqrt{V_{ii}}$

③ $H_0: \beta_i = 0$, $\hat{T} = \frac{\hat{\beta}_i}{s\sqrt{V_{ii}}}$ reject 若 $|\hat{T}| > t_{\alpha/2, n-p-1}$

§4 Coefficients 的线性组合 θ 的 CI 和 hypothesis testing

假设 $\theta = \mathbf{a}^T\boldsymbol{\beta}$, 则 $\hat{\theta} = \mathbf{a}^T\hat{\boldsymbol{\beta}} \sim N(\theta, \sigma^2\mathbf{a}^T(X^TX)^{-1}\mathbf{a})$

① T 统计量: $T = \frac{\hat{\theta} - \theta}{s\sqrt{\mathbf{a}^T(X^TX)^{-1}\mathbf{a}}} \sim t_{n-p-1}$

② θ 的 CI: $\hat{\theta} \pm t_{\alpha/2, n-p-1} \cdot s\sqrt{\mathbf{a}^T(X^TX)^{-1}\mathbf{a}}$

③ $H_0: \theta = 0$, $\hat{T} = \frac{\hat{\theta}}{s\sqrt{\mathbf{a}^T(X^TX)^{-1}\mathbf{a}}}$ reject 若 $|\hat{T}| > t_{\alpha/2, n-p-1}$

§5 New observation 的 predicting

Prediction error

$y_p - \hat{\mu}_p = \mu_p - \hat{\mu}_p + \varepsilon_p \sim N(0, \sigma^2(1 + \mathbf{a}^T(X^TX)^{-1}\mathbf{a}))$

① T 统计量: $T = \frac{y_p - \hat{\mu}_p}{s\sqrt{1 + \mathbf{a}^T(X^TX)^{-1}\mathbf{a}}} \sim t_{n-p-1}$

② μ_p 的 CI: $\hat{\mu}_p \pm t_{\alpha/2, n-p-1} \cdot s\sqrt{1 + \mathbf{a}^T(X^TX)^{-1}\mathbf{a}}$

9. Partial & ANOVA F-test

§1 MLR ANOVA 中的 statistics

1. TSS / SST $= \mathbf{y}^T\mathbf{y} - \frac{1}{n}\mathbf{y}^TJ\mathbf{y} = \mathbf{y}^T(I - \frac{1}{n}J)\mathbf{y}$

   $rank(I - \frac{1}{n}J) = n - 1 \Rightarrow d.f.(TSS) = n - 1$

2. SSE / RSS $= \hat{\mathbf{e}}^T\hat{\mathbf{e}} = \mathbf{y}^T(I - H)\mathbf{y} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^TX^T\mathbf{y}$

   $rank(I - H) = rank(I) - rank(H) = n - \sum_{i=1}^n h_{ii} = n - p - 1$

   $d.f.(SSE) = n - p - 1$

3. SSR / SSReg $= \mathbf{y}^TH\mathbf{y} - \frac{1}{n}\mathbf{y}^TJ\mathbf{y} = \mathbf{y}^T(H - \frac{1}{n}J)\mathbf{y}$

   $rank(H - \frac{1}{n}J) = rank(H) - rank(\frac{1}{n}J) = p \Rightarrow d.f.(SSR) = p$

4. TSS = SSE + SSR

§2 F-test in an MLR ANOVA table

1. Hypotheses setting: $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

2. F 检验: 若 H_0 为 true, 则 $\frac{TSS}{\sigma^2} \sim \chi^2(n-1)$

   $F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{MSReg}{MSE} \sim F(p, n-p-1)$

3. SSR 与 SSE 独立 $SSE = f((I - H)\mathbf{y})$

   $SSR = [(H - \frac{1}{n}J)\mathbf{y}]^T[(H - \frac{1}{n}J)\mathbf{y}] = f((H - \frac{1}{n}J)\mathbf{y})$

   $Cov((H - \frac{1}{n}J)\mathbf{y}, (I - H)\mathbf{y}) = (H - \frac{1}{n}J)\sigma^2 I_n(I - H) = 0$

4. 拒绝域与 p 值 拒绝 H_0 若 $F > F_{1-\alpha}(p, n-p-1)$

   $\hat{p} = \inf\{\alpha \in [0, 1]: F > F_\alpha(p, n-p-1)\} = 1 - \Phi_{F, p, n-p-1}(F)$

5. ANOVA table

Source	SS	D.F.	MS	F
Reg.	SSR	p	SSR/p	MSR/MSE
Residual	SSE	n-p-1	S^2	
Total	TSS	n-1		

§3 Partial F-test

1. 假设设定: $A\boldsymbol{\beta} = \mathbf{0}$, $rank(A) = l$

2. $\hat{F}_{obs} = \frac{(RSS_{REST} - RSS_{FULL})/l}{RSS_{FULL}/(n-p-1)} \sim F(l, n-p-1)$

3. R 中的 anova(mod) 是基于 sequential SS:

   x_p: $SSR(x_p | x_1 \cdots x_{p-1}) = SSR(x_1, \cdots, x_p) - SSR(x_1, \cdots, x_{p-1})$

4. X 的各列正交 (orthorgonal design) 的情况:

   $\mathbf{x}_j^T\mathbf{x}_k = 0 \Rightarrow X^TX = diag(\mathbf{1}^T\mathbf{1}, \mathbf{x}_1^T\mathbf{x}_1, \cdots, \mathbf{x}_p^T\mathbf{x}_p)$

   $\Rightarrow \hat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} \frac{1}{n} & & & \\ & \frac{1}{\mathbf{x}_1^T\mathbf{x}_1} & & \\ & & \ddots & \\ & & & \frac{1}{\mathbf{x}_p^T\mathbf{x}_p} \end{bmatrix} \begin{bmatrix} \Sigma y_i \\ \Sigma x_{1i} y_i \\ \vdots \\ \Sigma x_{pi} y_i \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}^T\mathbf{y}}{n} \\ \frac{\mathbf{x}_1^T\mathbf{y}}{\mathbf{x}_1^T\mathbf{x}_1} \\ \vdots \\ \frac{\mathbf{x}_p^T\mathbf{y}}{\mathbf{x}_p^T\mathbf{x}_p} \end{bmatrix}$

   $\Rightarrow RSS(\hat{\boldsymbol{\beta}}) = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^TX^T\mathbf{y} = \mathbf{y}^T\mathbf{y} - \sum_{j=0}^p \mathbf{x}_j^T\mathbf{y} \cdot \hat{\beta}_j$

§4 Adjusted R^2

1. $R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS} = \frac{\mathbf{y}^T(H - \frac{1}{n}J)\mathbf{y}}{\mathbf{y}^T(I - \frac{1}{n}J)\mathbf{y}} \neq \rho^2$

2. Adj $R^2 = 1 - (n-1)\frac{MSE}{SST} = 1 - \frac{n-1}{n-p-1}\frac{SSE}{SST} = 1 - \frac{SSE/n-p-1}{SST/n-1}$ 对 small dataset 更有意义

10. Model Selection

§1 Interaction

其中一个 var. 取决于另一个的 value

① 加入 interaction 的 model 的 MSE↓, Adj R^2↑

② n-way interaction 被引入, 则所有 i<n-way 均须被引入

§2 Model selection

$AIC_p = n\ln(SSE_p/n) + 2(p+1)$

1. Akaike's information criterion 越小越好

2. Stepwise regression Step 1: 由 null model 开始

   Step 2: 找出未加入 model 的 p-value 最小 (且小于 α_1) 的加入

   Step 3: 找出加入 model 的 p-value 最大 (且大于 α_1) 的移出

   Step 4: 交替重复 Step 2 & 3 直到不会 improve the model

   α 的常取 [0.05, 0.15], Multicollinearity 可能导致结果差异

§3 K-sample test

① $\bar{y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$

② $\bar{y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^p\sum_{j=1}^{n_i} y_{ij} = \frac{1}{N}\sum_{i=1}^p n_i\bar{y}_{i\cdot}$ ③ $TSS = \sum_{i=1}^p\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2$ (d.f. = n-1)

④ $SS(E) = \sum_{i=1}^p\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$ (n-k) ⑤ $SS(T) = \sum_{i=1}^p n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ (k-1)

§4 General least squares

$var(\boldsymbol{\varepsilon}) = \sigma^2 V$

$V = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T$, 定义 $L = \Lambda^{-\frac{1}{2}}Q^T$, 则

$L\mathbf{y} = LX\boldsymbol{\beta} + L\boldsymbol{\varepsilon}$ 有 $var(L\boldsymbol{\varepsilon}) = \sigma^2 I$

$\hat{\boldsymbol{\beta}}^{GLS} = (X^TV^{-1}X)^{-1}X^TV^{-1}\mathbf{y}$, $Var(\hat{\boldsymbol{\beta}}^{GLS}) = \sigma^2(X^TV^{-1}X)^{-1}$