Syllabus

Review of 250A final.

Q1

(a)(i) $Y \sim N(0, I_p)$

$Y \mid 1^T Y = 0 \sim ?$

Sol. Define $Z = \begin{pmatrix} Y \\ 1^T Y \end{pmatrix} = \begin{pmatrix} I_p \\ 1^T \end{pmatrix} Y$

$\mathrm{Cov}\, Z = \begin{pmatrix} I_p & 1 \\ 1^T & p \end{pmatrix}$, $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

$\Rightarrow Z_1 \mid Z_2 = 0 \sim N(0, I - \frac{11^T}{p})$

$x^T \left(I - \frac{11^T}{p}\right) x = \sum_i x_i^2 - p\bar{x}^2$

$= \sum_i (x_i - \bar{x})^2 \geq 0$

(ii) $Y^T I Y \mid a^T Y = 0 \sim ?$ $(a \neq 0)$

Sol. Let $\underset{p\times 1}{Z} = \underset{p\times p}{\begin{pmatrix} \frac{a^T}{\|a\|_2} \\ Q_{(p-1)\times p} \end{pmatrix}} \underset{p\times 1}{Y}$, $QQ^T = I$

$= AY = b = \begin{pmatrix} b_1^T Y \\ b_2^T Y \\ \vdots \\ b_p^T Y \end{pmatrix}$

$Z^T Z = Y^T A^T A Y = b^T b = \sum_{i=2}^{p} (b_i^T Y)^2$ $(\triangle)$

$b_i^T Y \sim N(0, \underset{\overset{\shortparallel}{1}}{b_i^T b_i})$

$\Rightarrow (\triangle) \sim \chi^2_{p-1}(0)$

$\Rightarrow E(\triangle) = p-1$.

(b) $Y \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

$Q = Y^T \Sigma^{-1} Y - \frac{Y_1^2}{\sigma_1} \sim ?$

Sol. $Y^T \Big(\underbrace{\Sigma^{-1} - \begin{pmatrix} \frac{1}{\sigma_{11}} & 0 \\ 0 & 0 \end{pmatrix}}_{A} \Big) Y$

ETS $A^2 = A$ since $A^T = A$ by Fundamental Thm.

Verify $AV = AVAV$, $V = \Sigma$

Then $\mathrm{rank}(AV) = \mathrm{Tr}(AV) = 1$

$\Rightarrow Q \sim \chi^2_1(0)$

Q2

(a): $\hat{Y} = PY = (\sim) \begin{pmatrix} Y_1 \\ \vdots \\ Y_8 \end{pmatrix}$

$\hat{\sigma}^2 = \frac{400}{8-4} = 100$

$e = Y - \hat{Y} = (I-P)Y$

Sol. $\widehat{\mathrm{Var}\, e_2} = (1 - P_{22})\hat{\sigma}^2 = 62.6$

$\widehat{\mathrm{Var}\, e_3} = 68.2$

$\widehat{\mathrm{Var}\, e_1} = ???$ $(Q = I - P)$

$\widehat{\mathrm{Cov}(e_1, e_2)} = \hat{\sigma}^2 Q_{12} = \hat{\sigma}^2 Q_{21}$

$= 100(0 - (-0.242))$

$= 24.2$

$\widehat{\mathrm{Cov}(e_1, e_3)} = \hat{\sigma} Q_{31} = 6.1$

$\widehat{\mathrm{Cov}(e_2, e_3)} = -32.9$

(b): $P^2 = P$, $(P^2)_{ii} = (P)_{ii}$

$$P_{ii} = \sum_k P_{ik} P_{ki} = \sum_k P_{ik}^2 \quad (\Delta)$$

If r replicates

$$(\Delta) \geq r P_{ii}^2 \Rightarrow \frac{1}{r} \geq P_{ii}$$

$\frac{1}{3} \geq P_{22}$ if 2 other rows are the same as row 2.

So the answer is "NO"!

(c): (i): $P_{ij} = \Sigma_k P_{ik} P_{kj} \Rightarrow$

$$(P_{ij})^2 = (\Sigma_k P_{ik} P_{kj})^2 \quad (\text{C-S ineq.})$$

$$\leq \underbrace{(\Sigma_k P_{ik}^2)}_{= P_{ii}} \underbrace{(P_{kj}^2)}_{= P_{jj}}$$

(ii): $P^2 = P \Rightarrow (I-P)^2 = I-P$

Let $Q = I-P$, then:

$$q_{ij} \leq q_{ii} q_{jj} \quad \text{but}$$

$$q_{ij} = -P_{ij} \ \& \ q_{ii} = 1 - P_{ii},$$
$$q_{jj} = 1 - P_{jj}.$$

Q3: $y = \alpha 1 + X\beta + \varepsilon$, $\mathbb{E}\varepsilon = 0$

$\mathbb{E}\varepsilon\varepsilon^T = \mathrm{Cov}\,\varepsilon = V^{-1}$.

Find $\hat{\alpha}$!

Sol. $V^{\frac{1}{2}} y = \alpha \underbrace{V^{\frac{1}{2}} 1} + \underbrace{V^{\frac{1}{2}} X}\beta + V^{\frac{1}{2}}\varepsilon$

$$\tilde{y} = \alpha V^{\frac{1}{2}} 1 + (P+Q) V^{\frac{1}{2}} X\beta + V^{\frac{1}{2}}\varepsilon$$

$P = P_{C(V^{\frac{1}{2}}1)} = V^{\frac{1}{2}} 1 1^T V^{\frac{1}{2}} / 1^T V 1$,

$Q = I - P$.

$$\Rightarrow \mathbb{E}\tilde{y} = V^{\frac{1}{2}} 1 \overbrace{\left(\alpha + \frac{1^T V^{\frac{1}{2}} V^{\frac{1}{2}} X\beta}{1^T V 1}\right)}^{\gamma} + Q V^{\frac{1}{2}} X\beta$$

$$= \underbrace{V^{\frac{1}{2}} 1}_{\otimes} \gamma + \underbrace{Q V^{\frac{1}{2}} X}_{\odot}\beta \quad \odot \perp \otimes !$$

$$\hat{\gamma} = \frac{1^T V^{\frac{1}{2}}}{1^T V 1} V^{\frac{1}{2}} y = \frac{1^T V y}{1^T V 1}$$

$$= \hat{\alpha} + \frac{1^T V X \hat{\beta}}{1^T V 1}$$

$$\Rightarrow \hat{\alpha} = -\frac{1^T V X \hat{\beta}}{1^T V 1} + \frac{1^T V y}{1^T V 1}$$

$$\hat{\beta} = (X^T V^{\frac{1}{2}} Q V^{\frac{1}{2}} X)^{-1} X^T V^{\frac{1}{2}} Q V^{\frac{1}{2}} y$$

Verify

$$\hat{\alpha} = \frac{1^T (V - VX(X^T W^{-1} X)^{-1} X^T W^{-1}) y}{1^T V 1}.$$

Multiple Correlation Coefficient

$r = \text{Correlation}(Y, \hat{Y})$

$= \left(\sum_i (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})\right) / \sqrt{(\sum_i (Y_i - \bar{Y})^2)(\sum_i (\hat{Y}_i - \bar{\hat{Y}})^2}$

If $1 \in C(X)$, $\hat{Y} = PY \Rightarrow 1^T\hat{Y} = 1^T PY$

$\Rightarrow \bar{\hat{Y}} = \bar{Y}$ ①

$= (P1)^T Y$

$= 1^T Y$

$\bar{Y} 1 = \frac{1 1^T y}{n} = \underbrace{P_{C(1)}}_{11^T/n} Y \in C(X)$ ②

So $R^2 = r^2 = \frac{(\sum_i (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}))^2}{(\sum_i (Y_i-\bar{Y})^2)(\sum_i (\hat{Y}_i - \bar{Y})^2)}$ (Δ_1) (Δ_2)

$(\Delta_1) = (\sum_i (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y}))^2$

$= (\sum_i (\hat{Y}_i - \bar{Y})^2)^2$ Since

- $\sum_i (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \underbrace{e^T}_{\in C(X)^\perp} (\underbrace{PY - \bar{y}1}_{\in C(X)}) = 0$

$\Rightarrow R^2 = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2} = \frac{SSReg}{TSS}$.

Ex: Test $\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$.

$F = \frac{SSE_0 - SSE}{SSE} \frac{n-p}{p-1}$

$= \frac{n-p}{p-1} \frac{TSS - SSR_0 - (TSS - SSR)}{TSS - SSReg}$

$= \frac{1 - R_0^2 - (1 - R^2)}{1 - R^2} = \frac{R^2 - R_0^2}{1 - R^2} \frac{n-p}{p-1}$

$R_0^2 = \frac{\sum_i (\hat{Y}_i^0 - \bar{Y})^2}{TSS} = 0$ Since $\hat{Y}_i^0 = \bar{Y}\ \forall i$.

$\Rightarrow F = \frac{R^2}{1-R^2} \frac{n-p}{p-1}$

Test Statistic for H_0:

$F = \frac{R^2}{1-R^2} \frac{n-p}{p-1} \sim F_{p-1, n-p}$ Or

$R^2 = \frac{(p-1)F}{(n-p) + (p-1)F}$, note:

$X \sim F_{d_1, d_2} \Rightarrow \frac{d_1 X / d_2}{1 + d_1 X / d_2} \sim B(\frac{d_1}{2}, \frac{d_2}{2})$

$\Rightarrow R^2 \sim B(\frac{p-1}{2}, \frac{n-p}{2})$ &

$E R^2 = \frac{p-1}{n-1}$ (Exercise 4C. P113)

Geometry of LS coef.

$\theta = EY = X\beta$

$= \sum_{i=1}^{p} \beta_i X_i \in C(X)$

How to interpret β_i & $\hat{\beta}_i$ geometrically?

Focus on β_k & $\hat{\beta}_k$.

Let $V = \mathcal{L}(X_1, \cdots, X_{p-1})$

$\hat{X}_k = P(X_k | V)$

$X_k^\perp = X_k - \hat{X}_k$

$X_k^\perp$ provides info. other than those provided by $X_1, \cdots, X_{p-1} \in V$.

Observe that

$\langle \theta, X_p^\perp \rangle$

$= \langle \sum_{i=1}^{p} \beta_i X_i, X_p^\perp \rangle$

$= \beta_p \langle X_p^\perp, X_p^\perp \rangle \Rightarrow$

$\beta_p = \frac{\langle \theta, X_p^\perp \rangle}{\|X_p^\perp\|_2^2}$, $\hat{\beta}_p = \frac{\langle Y, X_p^\perp \rangle}{\|X_p^\perp\|_2^2}$

$$\mathrm{Var}\,\hat{\beta}_p = \frac{1}{\|x_p^\perp\|_2^2}\mathrm{Var}(x_p^{\perp\prime}Y) = \frac{\sigma^2}{\|x_p^\perp\|_2^2}$$

$\therefore$ $\mathrm{Var}\,\hat{\beta}_p$ Large if $\|x_p^\perp\|$ is small

i.e. $x_p - \sum_{i=1}^{p-1}\beta_i x_i \approx 0$

i.e. if x_p is nearly a linear combination of other x's. then $\hat{\beta}_p$ is very poorly estimated

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \mathrm{Cov}\left(\frac{\langle x_i^\perp, Y\rangle}{\|x_i^\perp\|_2^2}, \frac{\langle x_j^\perp, Y\rangle}{\|x_j^\perp\|_2^2}\right)$$

$$= \frac{\sigma^2\langle x_i^\perp, x_j^\perp\rangle}{\|x_i^\perp\|_2^2\|x_j^\perp\|_2^2}$$

$$= \cos(x_i^\perp, x_j^\perp)\frac{\sigma^2}{\|x_i\|_2\|x_j\|_2}$$

Thus, $\sigma^2(X^TX)^{-1}_{ij} = \frac{\sigma^2}{\|x_i\|_2\|x_j\|_2}\cos(x_i^\perp, x_j^\perp)$

Relation to R^2:

$$R^2 = SSReg/TSS$$

$$R^2_{k-1} = SSReg(X_1, \dots, X_{k-1})/TSS$$

$$R^2_k = SSReg(X_1, \dots, X_k)/TSS$$

Verify $\frac{R_k^2 - R_{k-1}^2}{1-R_{k-1}^2} = \frac{t^2}{n-k+t^2}$,

t = test stat for $\beta_k = 0$.

Partial Correlation

Let $v_1, v_2, x_1, x_2, \dots, x_k \in \mathbb{R}^n$,

$V = \mathcal{L}(x_1, \dots, x_k)$

$\hat{v}_i = P(v_i|V)$, $i=1,2$

$v_i^\perp = v_i - \hat{v}_i \in V^\perp$, $i=1,2$

The partial corr. coef. of V_1 & V_2 (with the linear effects of $x_1, \dots, x_k$ removed) is:

$$r_{v_1v_2\cdot x_1\cdots x_k} = \frac{\langle v_1^\perp, v_2^\perp\rangle}{\|v_1^\perp\|\|v_2^\perp\|}$$

$$\triangleq \cos_{x_1\cdots x_k}(v_1^\perp, v_2^\perp)$$

NOTE:

- $k=1$: $x_1 = 1$, $v_1^\perp = v_1 - P(v_1|1) = v_1 - \bar{v}_1 1$

$\Rightarrow$ $r_{v_1v_2\cdot 1}$ = raw corr. (pearson corr.)

Questions: Want to express pcc (partial corr. coef.) of order n in terms of pcc's of order $(n-1)$.

Ex $k=4$, x_1, x_2, x_3, x_4.

$$r_{14\cdot 23} = \frac{r_{14} - r_{13\cdot 2}r_{34\cdot 2}}{\sqrt{1-r_{13\cdot 2}^2}\sqrt{1-r_{34\cdot 2}^2}}$$

- Consider a general case:

$$x_1, x_2, \dots, x_{k-1}, x_k$$

& Let $V_J = \mathcal{L}(x_4, x_5, \dots, x_k)$

$\subseteq \mathcal{L}(x_3, x_4, \dots, x_k)$

$= V_I$, $I = \{3, \dots, k\}$, $J = \{4, \dots, k\}$

Let $\hat{x}_i = P(x_i|V_J)$ $i=1,2$

$x_i^\perp = x_i - \hat{x}_i \in V_J^\perp$, $i=1,2$

$x_3 = \hat{x}_3 + x_3^\perp = P(x_3|V_J) + (x_3 - P(x_3|V_J))$

For $i=1,2,\cdots$, WT find

$p(x_i|V_I)$. To do this,

recall $(Q_1 = I - P_{C(X_1)})$

$$P_{C(X_1|X_2)} = P_{C(X_1)} + Q_1X_2(X_2^TQ_1X_2)^{-1}X_2^TQ_1$$

(Do by orthogonization process:

$$EY = X_1\beta_1 + X_2\beta_2$$
$$= X_1(\beta_1 + (X_1^TX_1)^{-1}X_1^TX_2\beta_2) + Q_1X_2\beta_2\ )$$

$$w_i = x_i - p(x_i|V_I) \quad \text{for } i=1,2.$$
$$= x_i - (p(x_i|V_J) + p(x_i|x_3^\perp))$$
$$= x_i^\perp - p(\underbrace{x_i - \hat{x}_i}_{①} + \underbrace{\hat{x}_i}_{②} | \underbrace{x_3^\perp}_{③})$$
$$= x_i^\perp - p(x_i^\perp | x_3^\perp) \quad \text{since } ② \in V_J,\ ③ \in V_J^\perp$$

So $r_{12\cdot I}$ = pcc of x_1 & x_2 adjusting for $x_3, x_4, \cdots, x_k$

$$= \frac{\langle w_1, w_2\rangle}{\|w_1\|\|w_2\|},$$ depends only on $x_1^\perp, x_2^\perp$ & $x_3^\perp$.

$$\|w_1\|^2 = \langle w_1, w_1\rangle = \langle x_1^\perp - p(x_1^\perp|x_3^\perp),\ x_1^\perp - p(x_1^\perp|x_3^\perp)\rangle$$
$$= \left\langle x_1^\perp - \frac{\langle x_1^\perp, x_3^\perp\rangle}{\|x_3^\perp\|_2^2}x_3^\perp,\ x_1^\perp - \frac{\langle x_1^\perp, x_3^\perp\rangle}{\|x_3^\perp\|_2^2}x_3^\perp\right\rangle$$
$$= \langle x_1^\perp, x_1^\perp\rangle - \frac{\langle x_1^\perp, x_3^\perp\rangle^2}{\|x_3^\perp\|_2^2}$$

WLOG assume $\|x_i^\perp\|_2 = 1,\ i=1,2,3$

$$= 1 - \langle x_1^\perp, x_3^\perp\rangle^2 / \|x_1^\perp\|_2\|x_3^\perp\|_2$$
$$= 1 - r_{13\cdot J}^2.$$

Similarly,

- $\|w_2\|^2 = 1 - r_{23\cdot J}^2$
- $\langle w_1, w_2\rangle = \langle x_1^\perp - p(x_1^\perp|x_3^\perp),\ x_2^\perp - p(x_2^\perp|x_3^\perp)\rangle$

$$= r_{12\cdot J} - r_{23\cdot J}\, r_{13\cdot J}$$

$(\|x_i^\perp\|_2 = 1,\ i=1,2,3)$

In summary we showed

$$r_{12\cdot I} = r_{12\cdot 34\cdots k}$$
$$= \frac{r_{12\cdot J} - r_{13\cdot J}\, r_{23\cdot J}}{\sqrt{1-r_{13\cdot J}^2}\sqrt{1-r_{23\cdot J}^2}}$$

LHS: I & RHS: J. $\square$

Diagnostics

$$Y = X\beta + \varepsilon$$

$\mathbb{E}Y = X\beta$; $\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY$

$e = Y - \hat{Y} = (I-P)Y = QY$

- $\mathbb{E}e = (I-P)X\beta = QX\beta = 0$
- $\operatorname{Var} e = \sigma^2 Q$
- $\operatorname{Cov}(e, \hat{Y}) = \operatorname{Cov}(QY, PY) = 0$

Define ① $r_i = \dfrac{e_i}{\sqrt{\sigma^2(1-h_{ii})}}$ Internally Studentized Residual

② $t_i = \dfrac{e_i}{S_{(i)}\sqrt{1-h_{ii}}}$ where Externally Studentized residual

$S_{(i)}^2$ estimate σ^2 without the i^{th} case.

Q: Are they related & what are their distributions?

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{(X^TX)^{-1}X_i e_i}{1-h_{ii}}, \quad e_i = Y_i - X_i^T\hat{\beta}_{(i)}.$$

Use this to show relationship between

$$S_{(i)}^2 \ \& \ S^2$$

Verify: $(n-p-1)S_{(i)}^2 = (n-p)S^2 - \dfrac{e_i^2}{1-h_{ii}}$

$$t_i^2 \stackrel{d}{=} \frac{B}{1-B}(n-p-1), \quad B \sim \operatorname{Beta}\left(\tfrac{1}{2}, \tfrac{n-p-1}{2}\right)$$

Recall: $\left(\frac{\frac{a}{b}F}{1+\frac{a}{b}F} \sim B\left(\frac{a}{2}, \frac{b}{2}\right) \right)$

Exact Dist. of r_i^2 not available, but

$$\frac{r_i^2}{n-p} \sim B\left(\tfrac{1}{2}, \tfrac{n-p-1}{2}\right)$$

Test of outlier: Is the i^{th} case an outlier for the X-value.

$$Y = X\beta + \varepsilon$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1^T\beta \\ \vdots \\ X_n^T\beta \end{pmatrix} + \delta \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i^{th}$$

Show the test statistic for $\delta = 0$ is

$$F = t_i^2 \ \&$$

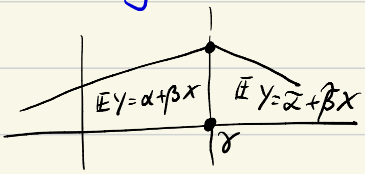
t_i externally studentized residual.

We need:

$$P_{C(X)} = X(X^TX)^{-1}X^T$$

$$\underbrace{P_{C(X|(0 \cdots 1 \cdots 0)^T)}}_{P_{C(X)} + \frac{(I-P_X)e_ie_i^T(I-P_X)}{e_i^T(I-P_X)e_i}} y = P_{C(X|(0 \cdots 1 \cdots 0)^T)} y$$

Two phase regression & Fieller's theorem



$$\mathbb{E}Y = \begin{cases}\alpha_1+\beta_1 x, & x\le\gamma\\ \alpha_2+\beta_2 x, & x\ge\gamma\end{cases}$$

Continuity assump.

$$\alpha_1+\beta_1\gamma = \alpha_2+\beta_2\gamma$$

$$\hat\gamma = \frac{\hat\alpha_1-\hat\alpha_2}{\hat\beta_2-\hat\beta_1}$$ point estimate

$y = X\beta+\varepsilon$ for a two phase reg.

Calibration: $Y, X, \hat\beta$ are known.

- Observe Y_0 but no info. on (x_0) was provided. Q: How to estimate X_0?

$$Y_0 = \hat\alpha + \hat\beta(X_0-\bar X)$$

$$\Rightarrow X_0 = \frac{Y_0-\hat\alpha}{\hat\beta}+\bar X = \frac{Y_0-\bar Y}{\hat\beta}+\bar X$$

Fieller's Theorem Suppose u, v are uncorrelated r.v.s,

$$u\sim N(\mu_u,\sigma_u^2),\ \sigma_u^2=\sigma^2 a_1$$

$$v\sim N(\mu_v,\sigma_v^2),\ \sigma_v^2=\sigma^2 a_2$$

$\hat\sigma_u = s a_1$. $\hat\sigma_v = s a_2$.

Define $g = t^2_{m,\alpha/2}\, s^2 a_2^2/v^2$, m: df associated with s^2 for estimating σ^2.

$$\sigma_u^2=\sigma^2 a_1,\quad \sigma_v^2 = \sigma^2 a_2$$

If $g<1$. a $100(1-\alpha)\%$ C.I. for $\frac{\mu_u}{\mu_v}$ is:

$$\frac{1}{1-g}\left(\frac{u}{v}\pm\frac{t_{m,\frac{\alpha}{2}}s}{v}\sqrt{(1-g)a_1^2+\frac{u^2a_2^2}{v^2}}\right)$$

Pf: $\theta\equiv\frac{\mu_u}{\mu_v}$ and $W=u-\theta V$

$$\mathbb{E}W = \mu_u-\theta\mu_v=0$$

$$\operatorname{Var}W = \sigma^2 a_1 + \theta^2\sigma^2 a_2$$

$$\widehat{\operatorname{Var}}W = s^2a_1+s^2a_2^{\underline{u}}$$

$$u-\theta v\sim N(0,\sigma^2a_1^2+\theta^2\sigma^2a_2^2)$$

Also $\frac{ms^2}{\sigma^2}\sim\chi^2_m$

$$\frac{(u-\theta v)/\sigma\sqrt{a_1^2+\theta^2a_2^2}}{\sqrt{\frac{ms^2}{\sigma^2}/m}}\sim t_m$$

$$(u-\theta v)^2 = t^2_{m,\frac{\alpha}{2}}\,s^2(a_1^2+\theta^2a_2^2)$$

Verify equation simplifies to

$$(1-g)\theta^2-\frac{2u}{v}\theta+\frac{u^2}{v^2}-g\frac{a_1^2}{a_2^2}=0$$

$$\theta=\frac{1}{2a}\left(-b\pm\sqrt{b^2-4ac}\right)$$

Verify

$$\theta\in\frac{1}{1-g}\left(\frac{u}{v}\pm\frac{t_{m,\frac{\alpha}{2}}s}{v}\sqrt{(1-g)a_1^2+\frac{u^2a_2^2}{v^2}}\right)$$

$u\leftarrow Y_0-\bar Y$; $v\leftarrow\hat\beta$

$$\operatorname{Var}u = \sigma^2+\frac{\sigma^2}{n} = \sigma^2\left(1+\frac1n\right)$$

$$\operatorname{Var}v = \sigma^2\left(\frac{1}{\sum_i(X_i-\bar X)^2}\right)$$

Verify $100(1-\alpha)\%$ C.I. for X_0 is

$$\bar X + \frac{Y_0-\bar Y}{\hat\beta(1-g)}\pm\frac{t_{n-2,\frac{\alpha}{2}}\,s}{\hat\beta(1-g)}\sqrt{\frac{1-g}{n}(1+n)+\frac{(Y_0-\bar Y)^2}{\hat\beta^2\sum_i(X_i-\bar X)^2}}$$

(Ex) $\mathbb{E}Y = \beta_0 + \beta_1(X - \bar{X})$

If $y=0$: $X_0 - \bar{X} = -\frac{\beta_0}{\beta_1}$

$\hat{X}_0 = \hat{\beta}_0 / \hat{\beta}_1$. Fieller's thm applies.

(Ex) Prognosting factors interaction assessment in a 2-trtment trial with trtmt T & C, along with a prognostic factor.

$$\mathbb{E} Y_? = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

- $X_1 = \begin{cases} 1 & \text{if trtmt } T \\ 0 & \text{if } C \end{cases}$
- X_2 some prognostic factor

$$\mathbb{E} Y_T = \beta_0 + \beta_1 + \beta_2 X_2 + \beta_3 X_2$$

$$\mathbb{E} Y_C = \beta_0 + \beta_2 X_2$$

- $\mathbb{E}(Y_T - Y_C) = \beta_1 + \beta_3 X_2$

Q: WTK X_2 for which $\mathbb{E}(Y_T - Y_C) > 0$.

i.e. WT estimate $X_2 = -\frac{\beta_1}{\beta_3}$

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}; \quad \mathrm{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$\tilde{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_3 \end{pmatrix}; \quad \mathrm{cov}(\tilde{\beta}) = V = \begin{pmatrix} V_{11} & V_{13} \\ V_{13} & V_{33} \end{pmatrix}$$

Let $w = \tilde{\beta}_1 + \theta \tilde{\beta}_3$, $\theta = -\frac{\beta_1}{\beta_3}$

then $\mathbb{E} w = 0$

$$\mathrm{Var}\, w = (V_{11} + \theta^2 V_{33} + 2\theta V_{13})\sigma^2$$

So $\frac{(n-4)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-4}$; $\frac{w}{\sqrt{\mathrm{Var}\, w}} \sim N(0,1)$

If σ^2 is known,

$$(\hat{\beta}_1 + \theta \hat{\beta}_3)^2 = Z^2_{\frac{\alpha}{2}} (V_{11} + \theta^2 V_{33} + 2\theta V_{13})\sigma^2$$

CI for θ are

$$\theta = \frac{-w_2 \pm \sqrt{w_2^2 - 4 w_1 w_2}}{2 w_2}.$$

$$w_1 = \hat{\beta}_1^2 - V_{11} Z^2_{\frac{\alpha}{2}}.$$

$$w_2 = 2\hat{\beta}_1 \hat{\beta}_3 - 2 V_{13} Z^2_{\frac{\alpha}{2}}$$

$$w_3 = \hat{\beta}_3^2 - V_{33} Z^2_{\frac{\alpha}{2}}$$

Note: $\beta_1 = 0$ iff $\frac{\hat{\beta}_1}{\sqrt{V_{11}}} > Z_{\alpha/2}$ iff $w_1 > 0$.

$\beta_3 = 0$ iff $\frac{\hat{\beta}_3}{\sqrt{V_{33}}} > Z_{\frac{\alpha}{2}}$ iff $w_3 > 0$

(Ex) Health economics.

Cost effectiveness ratio

$$R = \frac{\mu_{CT} - \mu_{CC}}{\mu_{ET} - \mu_{EC}}$$

$$\hat{R} = \frac{\bar{C}_T - \bar{C}_C}{\bar{E}_T - \bar{E}_C}$$

C.I. for R. Define

$$w = (\bar{C}_T - \bar{C}_C) - \theta(\bar{E}_T - \bar{E}_C)$$

Var w (estimate)

Extensions to correlated variables in the ratio.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \sigma^2 \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}\right)$$

WT find C.I. for $\theta = \mu_X / \mu_Y$

Let $W = X - \theta Y$

$\mathbb{E} W = 0$, $\operatorname{Var} W = \sigma^2 (V_{11} + V_{22}\theta^2 - 2\theta V_{12})$
$= \sigma^2 \cdot \sigma_w^2$

$$\frac{w}{\sqrt{\sigma^2 \cdot \sigma_w^2}} \sim \mathcal{N}(0,1)$$

$$\frac{r S_r^2}{\sigma^2} \sim \chi_r^2 .$$ S_r^2: unbiased estimator of σ^2

Then $$\frac{x - \theta y}{S_r \sqrt{V_{11} + \theta^2 V_{22} - 2\theta V_{12}}} \sim t_r$$

Verify the limits for θ are

$$\frac{1}{1-g}\left\{ \frac{X}{Y} + \frac{V_{12} g}{V_{22}} \pm \frac{t s}{Y}\sqrt{r} \right\}$$ where r is

$$V_{11} + \frac{2 X V_{12}}{Y} + \frac{X^2}{Y^2} V_{22} - g\left(V_{11} - \frac{V_{12}^2}{V_{22}}\right)$$

& $g = \dfrac{t^2 s^2 V_{22}}{Y^2}$ [Wiki...

Ex Two phase reg.

$\hat{\gamma}$ = estimated threshold when change occurs

(seber, p.161)

$$= \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}_1 - \hat{\beta}_2}$$

$W = (\hat{\alpha}_1 - \hat{\alpha}_2) + \gamma(\hat{\beta}_1 - \hat{\beta}_2) \Rightarrow$

$\mathbb{E} W = 0.$

• Violations

$Y = X\beta + \varepsilon$, $r(X) = p$, $\varepsilon \sim N_n(0, \sigma^2 I)$

• $\hat\beta = (X^TX)^{-1}X^Ty$. $Cov\,\hat\beta = \sigma^2(X^TX)^{-1}$

$\mathbb{E}\,Y = X\beta$ working model

$\mathbb{E}\,Y = X\beta + Z\gamma$ true model

$$\Rightarrow \mathbb{E}\hat\beta = (X^TX)^{-1}X^T(X\beta + Z\gamma)$$
$$= \beta + (X^TX)^{-1}X^TZ\gamma$$

unbiased if $X^TZ = 0$.

$$S^2 = \frac{1}{n-p}Y^TQY.$$

$$\mathbb{E}\,S^2 = \frac{1}{n-p}(\mathbb{E}Y^T\,Q\,\mathbb{E}Y + Tr\,Q(\sigma^2 I))$$
$$= \frac{1}{n-p}[\gamma^TZ^TQZ\gamma + \sigma^2(n-p)]$$
$$= \sigma^2 + \underbrace{\frac{1}{n-p}\gamma^TZ^TQZ\gamma}\ > 0$$

S^2 overestimates σ^2. unless $QZ = 0$

Fitted Model?

$\hat Y = X\hat\beta$.

$$\mathbb{E}\hat Y = X\beta + X(X^TX)^{-1}X^TZ\gamma$$
$$= X\beta + P_X Z\gamma$$

• Residuals $e = Y - \hat Y$

$$\mathbb{E}e = \mathbb{E}Y - \mathbb{E}X\hat\beta \qquad (L = (X^TX)^{-1}X^TZ)$$
$$= X\beta + Z\gamma - X(\beta + L\gamma)$$
$$= QZ\gamma \neq 0$$

So residual plots do not band around the $y=0$ line. What about Var e?

$$Var\ e = Var(QY)$$
$$= \sigma^2 Q \qquad \text{no change}$$

Verify as far as prediction at x_0

$\hat y = x_0^T\hat\beta$ under fitted model

True: $\hat y = x_0^T\hat\beta + z_0^T\hat\gamma$

$Var\ \hat y(x_0, z_0) \geq Var\ \hat y(x_0)$

Similar calculations for overfitting.

working model: $\mathbb{E}\,Y = X_1\beta_1 + X_2\beta_2$.

true model: $\mathbb{E}\,Y = X_1\beta_1$

$$\hat\beta = (X^TX)^{-1}X^TY$$

$$\mathbb{E}\hat\beta = (X^TX)^{-1}X^T(X_1\ X_2)\begin{pmatrix}\beta_1\\ 0\end{pmatrix}$$

$$\mathbb{E}\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} = \begin{pmatrix}\beta_1\\ 0\end{pmatrix} \Rightarrow$$

$\mathbb{E}\hat\beta_1 = \beta_1$ is still unbiased.

• Covariance is misspecified:

Assume $Cov\,\varepsilon = \sigma^2 I$

If $Cov\,\varepsilon = V\sigma^2$, then

$$\hat\beta_w = (X^TV^{-1}X)^{-1}X^TV^{-1}Y.$$

If $\mathbb{E}\,Y = X\beta$ holds

$$Cov\,\hat\beta_w = \sigma^2(X^TV^{-1}X)^{-1}.$$

Is $\hat\sigma^2 = \frac{1}{n-p}Y^TQY$ biased for σ^2 when $Cov\ \varepsilon = V\sigma^2$?

$\boxed{\text{claim}}$ If $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ are eigvals of V.

$$\frac{1}{n-p}\sum_{i=1}^{n-p}\mu_i \leq \mathbb{E}\frac{\hat\sigma^2}{\sigma^2} \leq \frac{1}{n-p}\sum_{i=n-p+1}^{n}\mu_i$$

$\boxed{\text{Lemma}}$ $H^T = H = H^2$ & $r(H) = r$. if $A^T = A$, then
$$\sum_{i=1}^{r} \lambda_i \le Tr(HA) \le \sum_{i=n-r+1}^{n} \lambda_i$$
where $\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$ are eig-vals of A.

Pf: $H^T = H = H^2 \Rightarrow \exists P$ s.t. $P^T H P = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = D_r$

$$Tr\, HA = Tr\, P D_r P^T A$$
$$= Tr\, (D_r^2 P^T A P)$$
$$= Tr\left[(PD_r)^T (APD_r)\right]$$

$\bullet\ PD_r = (P_1 \cdots P_n)(e_1 \cdots e_r\ O)$

$$= Tr \begin{pmatrix} P_1^T \\ P_2^T \\ \vdots \\ P_r^T \\ 0 \end{pmatrix} A (P_1 \cdots P_r\ 0) = \sum_{i=1}^{r} P_i^T A P_i$$

$$\Rightarrow Tr\, HA = \sum_{i=1}^{r} P_i^T A P_i, \quad \|P_i\|_2 = 1$$

Verify:
$$\sum_{i=1}^{r} \lambda_i \le Tr HA \le \sum_{i=n-r+1}^{n} \lambda_i$$

LACK OF FITNESS

$H_0: EY = X\beta$ vs $H_a: EY \ne X\beta$

Assumptions:

n_1 obs on y at $x_1^T = (x_{11}, x_{12}, \dots, x_{1k})$

n_2 obs on y at $x_2^T = (x_{21}, x_{22}, \dots, x_{2k})$

$\vdots$

n_g obs on y at $x_g^T = (x_{g1}, \dots, x_{gk})$

$n = \sum_{i=1}^{g} n_i$ (total # of obs)

Then:
$$y = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ \vdots \\ y_{n_2 2} \\ \vdots \\ y_{1g} \\ \vdots \\ y_{n_g g} \end{pmatrix}, \quad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_1^T \\ x_2^T \\ \vdots \\ x_2^T \\ \vdots \\ x_g^T \\ \vdots \\ x_g^T \end{pmatrix} \begin{matrix} \}\ n_1 \\ \\ \}\ n_2 \\ \vdots \\ \}\ n_g \end{matrix}$$

Reg y on $x_1, x_2, \dots, x_k$ &
$$SSE = Y^T Q Y.$$

Let
$$U = \begin{pmatrix} I - \frac{11^T}{n_1} & & O \\ & I - \frac{11^T}{n_2} & \\ O & & \ddots \\ & & & I - \frac{11^T}{n_g} \end{pmatrix}$$

(i.e. $\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$)

$$SSE = Y^T (Q - U + U) Y$$
$$= Y^T (Q - U) Y + Y^T U Y$$
$$= SSLoF + SSPE.$$

$\bullet\ (Q-U)(Q-U) = Q - UQ + U - QU$

why $X^T U = 0$? (verify it)

$\Rightarrow Y^T (Q-U) Y \perp\!\!\!\perp \underbrace{Y^T U Y}_{Y^T U Y}$

So $SSLOF = Y^T (Q-U) Y \perp\!\!\!\perp SSPE$

& $\frac{SSLOF}{\sigma^2} \sim \chi^2_{g-k}$

$\frac{SSPE}{\sigma^2} \sim \chi^2_{n-g}$

& so test statistic for
$$H_0: EY = X\beta$$
is $F = \frac{SSLOF/(g-k)}{SSPE/(n-g)} \sim F_{g-k, n-g}$

$\underbrace{df\, SSE}_{n-k} = (g-k) + (n-g).$

Scheffé's Method

$EY = X\beta$. $A \in \mathbb{R}^{q\times p}$, $r(A) = q$,

WT find an interval estimate for $A\beta$:

$$\hat\beta \equiv (X^TX)^{-1}X^T\beta.$$

$$A\hat\beta \sim \mathcal{N}_q(A\beta, \sigma^2 A(X^TX)^{-1}A^T)$$

$$\frac{1}{\sigma^2}(A\hat\beta - A\beta)^T [A(X^TX)^{-1}A^T]^{-1}(A\hat\beta - A\beta) \sim \chi^2_q$$

& $S^2 = \frac{\hat\sigma^2(n-p)}{\sigma^2} \sim \chi^2_{n-p}$, $p = \dim(\beta)$

so $$\frac{(A\hat\beta - A\beta)^T(A(X^TX)^{-1}A^T)^{-1}(A\hat\beta - A\beta)/q}{S^2} \sim F_{q,n-p}$$

so $1-\alpha = P(F_{q,n-p} \le F_{q,n-p,\alpha})$.

Let $\phi \equiv A\beta$. $L \equiv A(X^TX)^{-1}A^T$

Then

$$1-\alpha = P(\overbrace{(\hat\phi-\phi)}^{b}{}^T L^{-1}(\hat\phi-\phi) \le \overbrace{q\cdot S^2 F_{q,n-p,\alpha}}^{m})$$

$$= P(b^T L^{-1} b \le m)$$

$$= P\left(\max_{h\ne 0} \frac{(h^Tb)^2}{h^TLh} \le m\right) \text{ [KEY STEP]}$$

$$= P(\forall h_{h\ne 0},\ (h^Tb)^2 \le m\cdot h^TLh)$$

$$= P(\forall h,\ |h^Tb| \le \sqrt{m\, h^TLh})$$

$$= P(\forall h: h^T\phi \in h^T\hat\phi \pm \sqrt{m h^TLh})$$

Consider multiple comparison tests for one-way ANOVA

Def: Let $Z_1, \dots, Z_k$ & U are indept. RV with $Z_i \sim N(0,1)$ & $U \sim \chi^2_m(0)$.

define $$q \equiv \max_{i\ne j} \frac{|Z_i - Z_j|}{\sqrt{U/m}}$$

we call q has a studentized range dist. with k & m dfs & write $q \sim q_{k,m}$

Lemma: In a one-way ANOVA

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$j = 1, 2, \dots, n$; $i = 1, 2, \dots, k$

$$\bar Y_{i.} = \mu + \alpha_i + \bar\varepsilon_{i.};\quad \bar Y_{j.} = \mu + \alpha_j + \bar\varepsilon_{j.}$$

Verify $\max_{i\ne j} \frac{|\bar Y_{i.} - \bar Y_{j.} - (\alpha_i - \alpha_j)|}{\hat\sigma} \sim q_{k,\underbrace{k(n-1)}_{\text{df error for a 1-way ANOVA}}}$

It follows

$$P\left(\alpha_i - \alpha_j \in \bar Y_{i.} - \bar Y_{j.} \pm \frac{\hat\sigma}{\sqrt n} q_{k,k(n-1),\alpha} \text{ for } i\ne j\right)$$

$$= P\left(\frac{\sqrt n(\bar Y_{i.} - \bar Y_{j.} - (\alpha_i - \alpha_j))}{\hat\sigma} \le q_{k,k(n-1),\alpha} \text{ for } i\ne j\right)$$

$= 1-\alpha$. (Tukey's pair-comparison)

Q: How to construct sets of C.I. for contrasts in a 1-way ANOVA.

Lemma: Let $a_1, a_2, \dots, a_k$ be numbers.

Then $|a_i - a_j| \le b \ \forall i, j$

$$\iff \left|\sum_{i=1}^k c_i a_i\right| \le \frac{b}{2}\left(\sum_{i=1}^k |c_i|\right)$$

$\forall c$ s.t. $\sum_{i=1}^k c_i = 0$ (s.t. $c^T1 = 0$)

Goal: Construct CI for $\sum_i c_i\alpha_i$ $\forall c$

$$P\left(\sum_i c_i\alpha_i \in \sum_i c_i \bar Y_{i.} \pm \frac{\hat\sigma}{\sqrt n} q_{k,k(n-1),\alpha}\left(\sum_i \frac{|c_i|}{2}\right), \forall c\right)$$

$$= P\left(\left|\sum_i c_i(\bar Y_{i.} - \alpha_i)\right| \le \frac{\hat\sigma}{\sqrt n} q_{k,k(n-1),\alpha}\left(\sum_i \frac{|c_i|}{2}\right) \forall c\right)$$

$$\overset{\text{lemma}}{=} P\left(|\bar Y_{i.} - \alpha_i - (\bar Y_{j.} - \alpha_j)| \le \frac{\hat\sigma}{\sqrt n} q_{k,n(k-1),\alpha}\ \forall i,j\right)$$

$= 1-\alpha$. □

Transformations on X's or Y to meet model assumptions

$$\mathbb{E}Y = X\beta = \begin{pmatrix} x_1^T\beta \\ \vdots \\ x_n^T\beta \end{pmatrix}$$

- $\mathbb{E}e = 0 \quad \text{cov}\, Y = \sigma^2 I$
- Consider $\mathbb{E}Y = \alpha + \beta^T x$

Transform x by $g(x)$

$$\mathbb{E}Y = \alpha + \beta g(x)$$

Consider simple transformations like

$$g_\lambda(x) = \begin{cases} \ln x, & \lambda = 0 \\ \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}$$

- $\lim_{\lambda \to 0} \dfrac{x^\lambda - 1}{\lambda} = \ln x.$

In practice, run the regression with

$$\lambda = \{-2, -\tfrac{1}{2}, -\tfrac{1}{3}, 0, \tfrac{1}{3}, \tfrac{1}{2}, 1, 2, -1\}$$

Look at SSE_λ when we

- reg y on $g_\lambda(x)$ for each λ. Pick the λ with smallest SSE.

▸ Do same procedure on the response $(y>0)$ & ad-hoc.

▸ Alternatively, use variance stabilizing transformation

So that $V(y)$ is a constant

$$g(y) = g(\underset{\substack{\uparrow \\ \mathbb{E}Y}}{\mu}) + (y-\mu)g'(\mu) + o_p(y-\mu)$$

- $\text{Var}\, g(y) \approx \text{Var}\, Y [g'(\mu)]^2 \equiv \sigma^2 (g'(\mu))^2$

$$\Rightarrow \quad g'(\mu) = \sqrt{\frac{\sigma^2}{\text{Var}\, Y}}$$

$$g(\mu) = \int \frac{\sigma}{\sqrt{\text{Var}\, Y}}\, d\mu$$

if Poisson: $\sqrt{Y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

- In ANOVA, k groups may have σ_i^2, $i = 1, 2, \ldots, k$.

$H_0: \sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$. vs.

$H: \sigma_i \neq \sigma_j, \quad i \neq j$

Let $\hat{S}_i^2$ be the unbiased estimate of σ_i^2, $i = 1, 2, \ldots, k$. The ratio

Hartley's test $\quad \dfrac{\max_{1\le i\le k} \hat{S}_i^2}{\min_{1\le i\le k} \hat{S}_i^2}$

is close to 1 if H_0 holds, otherwise reject.

Levene's test

$Y_{11}, Y_{12}, \ldots, Y_{1n_1}$	$Z_{ij} = \lvert Y_{ij} - \bar{Y}_{i\cdot} \rvert$
$Y_{21}, Y_{22}, \ldots, Y_{2n_2}$	$\bar{Z}_{\cdot\cdot} = \frac{1}{n} \sum_i \sum_j Z_{ij}$
$\vdots$	
$Y_{k1}, Y_{k2}, \ldots, Y_{kn_k}$	$\bar{Z}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$

use

$$W = \frac{\sum_i n_i (\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2 / (k-1)}{\sum_i \sum_j (Z_{ij} - \bar{Z}_{i\cdot})^2 / (n-k)}$$

to test H_0 if $W > F_{k-1, n-k}$

If $\bar{Y}_{i\cdot}$ is replaced by $\tilde{Y}_{i\cdot}$: median in i^{th} group, then it is Brown-Forsythe test

Collinearity

Reg Y on $X_1 \cdots X_p$

$$\hat{Y} = X\hat{\beta} = (X_1 \cdots X_p)\begin{pmatrix}\hat\beta_1\\ \vdots \\ \hat\beta_p\end{pmatrix}$$

$$Var(\hat\beta_k) = \frac{\sigma^2}{\|X_k^{\perp}\|_2^2} = \sigma^2 \|X_k - P(X_k|V_{(-k)})\|_2^{-2}$$

$V_{(-k)} = \mathcal{L}(X_1, \dots, X_p$ but without $X_k)$

If $X_k \approx$ linear comb. of the other X's, then $\hat\beta_k$ will be estimated unreliably.

How to detect collinearity?

$R_j^2 \equiv R^2$ obtained when reg. X_j on $X_{(-j)}$.

$$VIF_j \equiv \frac{1}{1-R_j^2}$$ variance inflation factor

If VIF_j is large, say >10, then X_j is almost linearly related with the rest of X's.

What to do?

1. Omit the variable or variables with $VIF > 10$.
2. $EY = X\beta$

$= XR^{-1}R\beta$, R non-singular

where R is from the QR decomp. of X:

Householder matrix: $Hx = \begin{pmatrix}\|x\|_2\\ 0\end{pmatrix}$

$X = QR$

$$(X_1\ X_2 \cdots X_p) = \underbrace{(w_1\ w_2 \cdots w_p)}_{Q}\underbrace{\begin{pmatrix} \ddots & * \\ 0 & \ddots \end{pmatrix}}_{R}$$

$w_i^T w_j = 0$ & $w_i^T w_i = 1$.

$w_1 = \frac{X_1}{\|X_1\|_2}$, $c_{12}w_1 + c_{22}w_2 = X_2$.

$\gamma \equiv R\beta$.

$$\hat\gamma = \widehat{R\beta} = (Q^TQ)^{-1}Q^TY = Q^TY = R^{-T}X^TY$$

$$SSReg = Y^T P_{C(Q)} Y = YQQ^TY$$
$$= \sum_{i=1}^{p} (q_i^T Y)^2 \quad Q = [q_1 \cdots q_p]$$

HW: $EY = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$

I take obs at $X = 1, 2, \dots, 5$.

$$X = \begin{pmatrix} \mathbf{1} & \begin{matrix} 1 & 1 & 1 \\ 2 & 4 & 9 \\ 3 & 9 & 27 \\ 4 & 16 & 64 \\ 5 & 25 & 125 \end{matrix}\end{pmatrix} \quad \beta = \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix}$$

$$EY = X\beta = (XR^{-1})(R\beta)$$
$$= (w_1 \cdots w_4)\gamma$$

$$\Rightarrow 0 = \sum_{x=1}^{5} w_i(x) w_j(x)$$

$H_0: \beta_p = 0$ iff $H_0: \gamma_p = 0$.

Heteroscedasticity

(Cook's paper)

$Y_i = X_i^T\beta + \varepsilon_i$, $E\varepsilon_i = 0$, $Var\,\varepsilon_i = \sigma_i^2$

$EY = X\beta$, $r(X) = p$

Assume σ_i^2 depends on covariate

$\sigma_i^2 = f$(a q-vector covariate z) Can be one of the X's.

$= \exp\{z_i^T\lambda\}$ &

1st column of X is 1. Then $z = (1, z_1, \dots, z_k)$.

If $\lambda = \lambda_0 = (1, 0, 0, \dots, 0)$. Then testing

$H_0: \lambda = \lambda_0$, $\sigma_i^2 = e^{\lambda_0}$ = const

$$L(Y_i|\beta, \sigma_i^2) \to L = \prod_{i=1}^{n} L_i \to \left(\frac{\partial \log L}{\partial \beta}\right)\left(\frac{\partial \log L}{\partial \beta}\right)^T$$

PCA and Ridge Regression

$$\mathbb{E}Y = \underset{n\times p}{X}\beta = XUU^T\beta$$

where $UU^T = U^TU = I$.

$$\mathbb{E}Y = Z\gamma, \quad Z = XU \ \& \ \gamma = U^T\beta$$

$$(z_1, \cdots, z_p) = \left(\sum_{i=1}^{p} u_{i1}X_i, \cdots, \sum_{i=1}^{p} u_{ip}X_i\right)$$

Each z_i is a linear combination of X_i's. Pick u s.t.

$$u^TX^TXu = D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}$$

$$\underbrace{X^TX}_{p\times p} = UDU^T$$

Collinearity, $\lambda_{r+1}, \lambda_{r+2}, \cdots, \lambda_p = 0$.

Then $D = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$, $\Lambda_r = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{pmatrix}$

$$U = \{u_1 | u_2\}$$

$$\begin{pmatrix} u_1^T \\ u_2^T \end{pmatrix} X^TX (u_1 \ u_2) = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$$

$$u_1^TX^TXu_1 = \Lambda_r$$

$$(Xu_2)^TXu_2 = u_2^TX^TXu_2 = 0 \Rightarrow$$

$$Xu_2 = 0.$$

$$\mathbb{E}Y = X\beta = X(u_1 \ u_2)\begin{pmatrix} u_1^T\beta \\ u_2^T\beta \end{pmatrix}$$

$$= Xu_1\gamma_1$$

Xu_1 are the p.c. of X.

1^{st} p.c. of X:

$z_1 = Xu_1$ & $X^TXu_1 = \lambda_1 u_1$

$z_r = Xu_r$ & $X^TXu_r = \lambda_r u_r$

So Reg y on $z_1 \cdots z_r$

note $\quad z_i^Tz_j = u_i^TX^TXu_j$

$$= \begin{cases} \lambda_i & j = i \\ 0 & j \neq i \end{cases}$$

Screeplot: $\sum_{i=1}^{j} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i$

Ridge Regression

Recall $\mathrm{Var}\,\hat{\beta}_i = \dfrac{\sigma^2}{\|X_i\|_2^2}$. $\quad \hat{\beta}$ = LS estimate for β.

$$\hat{\beta}^k = (X^TX + kI)^{-1}X^Ty$$

Ridge Regression estimate for β

MSE of $\tilde{\beta}$ of β:

$$mse(\tilde{\beta}) = \mathbb{E}\underbrace{(\tilde{\beta} - \beta)}_{p\times 1}\underbrace{(\tilde{\beta} - \beta)^T}_{1\times p}$$

$$= \mathrm{Cov}(\tilde{\beta}) + bb^T, \quad b = \mathbb{E}(\tilde{\beta} - \beta).$$

Clearly $b(\hat{\beta}) = 0$. $\quad mse(\hat{\beta}) = \mathrm{Cov}(\hat{\beta}) = \sigma^2(X^TX)^{-1}$

$$b_k = (X^TX + kI)^{-1}X^Ty, \quad k > 0.$$

$$= G_kX^Ty$$

$$\mathbb{E}b_k = G_kX^TX\beta = (X^TX + kI)^{-1}X^TX\beta$$

$$= (X^TX + k(X^TX)(X^TX)^{-1})^{-1}X^TX\beta$$

$$= (I + k(X^TX)^{-1})^{-1}\beta$$

250A: $(I - AB)^{-1} = I - A(I - BA)^{-1}B$

$A = -kI$, $B = (X^TX)^{-1}$

$$\therefore \mathrm{bias}(b_k) = -kI(I + (X^TX)^{-1}kI)^{-1}(X^TX)^{-1}\beta.$$

$$= -k((X^TX)^{-1}(X^TX) + (X^TX)^{-1}kI)^{-1}(X^TX)^{-1}\beta$$
$$= -k(X^TX + kI)^{-1}\beta$$
$$= -k\, G_k\, \beta$$

$$Var\, b_k = Var\,(X^TX+kI)^{-1}X^Ty$$
$$= G_k X^T(\sigma^2 I)X G_k$$
$$= \sigma^2 G_k X^TX G_k$$

$$mse(b_k) = \sigma^2 G_k X^TX G_k + k^2 G_k \beta\beta^T G_k$$
$$= G_k(\sigma^2 X^TX + k^2\beta\beta^T)G_k$$

Show
$mse(\hat\beta) - mse(b_k) \geq 0$ for some k

Now show $Tr(mse(\hat\beta)) - Tr(mse(b_k)) \geq 0$ for some k.

$$[mse(b_k)]$$
$$= (G_k\beta\beta^T G_k k^2 + G_k X^TX G_k \sigma^2)$$
$$= (TDT^T+kI)^{-1}(TDT^T\sigma^2 + k^2\beta\beta^T)(TDT^T+kI)^{-1}$$

where $T^TX^TXT = D = \begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{pmatrix}$

$X^TX = TDT^T$ & $G_k = (X^TX+kI)^{-1}$

$$= T(D+kI)^{-1}(\sigma^2 D + \beta\beta^T k^2)(D+kI)T^T \quad (\triangle)$$

$$Tr(\triangle) = Tr\left\{\begin{pmatrix}\frac{1}{\lambda_1+k} & & \\ & \ddots & \\ & & \frac{1}{\lambda_p+k}\end{pmatrix}(D\sigma^2 + \beta\beta^T k^2)\begin{pmatrix}\frac{1}{\lambda_1+k} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\lambda_p+k}\end{pmatrix}\right\}$$

$$= Tr\begin{pmatrix}\frac{\lambda_1\sigma^2}{(\lambda_1+k)^2} & & \\ & \ddots & \\ & & \frac{\lambda_p\sigma^2}{(\lambda_p+k)^2}\end{pmatrix} + k^2\beta^T\begin{pmatrix}\frac{1}{(\lambda_1+k)^2} & & \\ & \ddots & \\ & & \frac{1}{(\lambda_p+k)^2}\end{pmatrix}\beta$$

$$= \sum_{i=1}^{p}\frac{\lambda_i\sigma^2 + k^2\beta_i^2}{(\lambda_i+k)^2}$$

$= g(k)$ if σ^2, X, β.

$$g(0) = \sum_{i=1}^{p}\frac{\sigma^2}{\lambda_i} = Tr(Cov(\hat\beta)) = Tr(mse(\hat\beta))$$

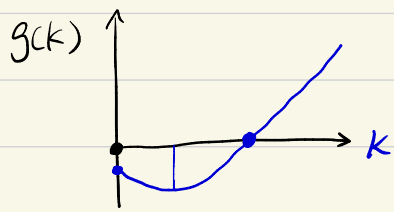
Let $g(k) - g(0) = Tr(mse(b_k)) - Tr(mse(\hat\beta))$

$$= \sum_{i=1}^{p}\frac{\lambda_i\sigma^2 + k^2\beta_i^2}{(\lambda_i+k)^2} - \sum_{i=1}^{p}\frac{\sigma^2}{\lambda_i^2}$$

Verify:

$$g'(k) = 0 = -\sum_{i=1}^{p}\frac{2\lambda_i(k\beta_i^2 - \sigma^2)}{(\lambda_i+k)^3} \quad !!!$$

Further, $g'(k) = \begin{cases}\geq 0 & \text{if } k \text{ small}\\ \leq 0 & \text{if } k \text{ Large}\end{cases}$



Can impose unrealistic assumptions:

$$k < \frac{2\sigma^2}{\beta^T\beta}$$

Use SVD then to view Ridge Reg est.

$$\underset{n\times p}{X} = \underset{n\times n}{U}\,\underset{n\times p}{D}\,\underset{p\times p}{V^T}$$

Verify

$$\hat\theta = X\hat\beta = X(X^TX)^{-1}X^Ty = \Sigma_i u_i u_i^T y$$
$$Xb_k = X(X^TX+kI)^{-1}X^Ty$$
$$= \Sigma_i \frac{\lambda_i^2}{\lambda_i^2+k}u_i u_i^T y$$

Since $\frac{\lambda_i^2}{\lambda_i^2+k} < 1$: Xb_k shrinks $X\hat\beta$.

James-Stein Estimator

See texts.

Lec 13

Bayesian Est. in LM

$\mathbb{E}Y = X\beta$. $Cov(e) = \sigma^2 I$

70-80s: Freq. vs. Bayesian

$f(y|\theta)$ density; $g(\theta)$ prior density

$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max} \ln L(y|\theta)$, $L(y|\theta) = \prod_{i=1}^n f(y_i|\theta)$

Posterior $p(\theta|y) = f(y,\theta)/f(y) \propto f(y|\theta)g(\theta)$

Ex Let $Y_1,\dots,Y_n \sim \mathcal{N}(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$.

$$L(y|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_i^n (y_i-\mu)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\{\sum_i (y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\}}$$

If σ is known,

$$L(y|\mu) \propto e^{-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2}$$

Assume $f(\mu) \propto e^{-\frac{1}{2\sigma_0}(\mu-\mu_0)^2}$

Posterior density is

$$p(\mu|y) \propto e^{-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2} e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}$$

$$= \exp\{-\tfrac{n}{2\sigma^2}(\bar{y}-\mu)^2 - \tfrac{1}{2\sigma_0^2}(\mu-\mu_0)^2\}$$

Let's see if $\exists\, \mu_n$ & σ_n^2 s.t. RHS is $\propto$

$$\exp\{-\tfrac{1}{2\sigma_n^2}(\mu-\mu_n)^2\}$$

Equate coefficients:

$$\mu^2: -\frac{1}{2\sigma_n^2} = -\frac{1}{2\sigma_0^2} - \frac{n}{2\sigma^2}$$

$$\sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

$$\mu: \frac{\mu_n}{\sigma_n^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}_n}{\sigma^2}$$

$$\Rightarrow \mu_n = \frac{\sigma^2}{\sigma^2+n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma^2+n\sigma_0^2}\bar{x}$$

posterior density: $\mu|y \propto \mathcal{N}(\mu_n, \sigma_n^2)$

Ex $Y \sim Bin(n,\theta)$, $\mathbb{P}(Y=y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$

Let $g(\theta)$ be a prior density for θ.

Posterior for $\theta \propto L(y|\theta)g(\theta)$

If $g(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$

Posterior is $\propto \underline{\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}$

Relative mean & variance of the Beta-Binomial r.v.

$\mathbb{E}Y = X\beta$, $Var(e) = \sigma^2 I$, $\theta^T = (\underset{p\times 1}{\beta^T}, \underset{1\times 1}{\sigma^2})$

$g(\theta) = g(\beta,\sigma^2) = g_1(\beta)g_2(\sigma^2)$

Preliminaries:

(a) $\int_0^\infty e^{-\frac{k}{x}} x^{-v-1}dx = \frac{1}{k^v\Gamma(v)}$

(b) $\int_0^\infty e^{-\frac{a}{x^2}} x^{-b-1}dx = \frac{\Gamma(b/2)}{2a^{b/2}}$

(c) $\|Y-X\beta\|_2^2 = (n-p)S^2 + \|X(\hat{\beta}-\beta)\|_2^2$

$$Y = X\beta + \varepsilon,\quad \varepsilon \sim \mathcal{N}(0,\sigma^2 I)$$

$f(\theta|y) \propto f(y|\theta)g(\theta)$

what is $g(\theta)$: $\underbrace{g_1(\beta)}_{\text{constant}}\underbrace{g_2(\sigma^2)}_{\propto \frac{1}{\sigma}}$

$$f(\theta|Y) \propto \frac{1}{\sigma}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\|Y-X\beta\|_2^2}$$

If interest is in β alone,

$$f(\beta|y) = \int_0^\infty \frac{1}{(\sqrt{2\pi})^n}\frac{1}{\sigma^{n+1}} e^{-\frac{1}{2\sigma^2}\|Y-X\beta\|_2^2} d\sigma$$

by (b)

$$= \frac{1}{2}\left(\frac{\|Y-X\beta\|_2^2}{2}\right)^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)$$

$$\Rightarrow f(\beta|y) \propto \|Y-X\beta\|^{-n}$$

by (c)

$$= \left((n-p)S^2 + \|X\hat{\beta}-X\beta\|_2^2\right)^{-\frac{n}{2}}$$

$$\propto \left(1+\frac{(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)}{(n-p)S^2}\right)^{-\frac{n}{2}}$$

Multivariate t-distribution

Page 475 (Lee & Seber)

Write $Y \sim t_m(v,\mu,\Sigma)$ if

$$f_Y(y) = \frac{\Gamma(\frac{v+m}{2})}{(\pi v)^{\frac{m}{2}}\Gamma(\frac{v}{2})}\frac{1}{|\Sigma|^{\frac{1}{2}}}\left(1+\frac{(y-\mu)^T\Sigma^{-1}(y-\mu)}{v}\right)^{-\frac{v+m}{2}}$$

$\mathbb{E}Y = \mu$, $\mathrm{Cov}\,Y = \Sigma$.

$\mathbb{E}(\mathbb{E}(Y|X)) = Y$; $\mathrm{Var}\,Y = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$

Before we worked with improper prior densities. Now work with proper priors

$$\theta^T = (\beta^T, \sigma^2)$$

$$f(\beta, \sigma^2) = \underbrace{f_1(\beta|\sigma^2)}_{N_p(m, \sigma^2 V)} \underbrace{f_2(\sigma^2)}_{\text{Inverted Gamma}}$$

$\boxed{IG}$

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$$

$$f(\theta|y) \propto f(y|\theta)g(\theta)$$

$$\propto (\sigma^2)^{-\frac{n+d+2+p}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2}\left[(y-X\beta)^T(y-X\beta) + (\beta-m)^T V^{-1}(\beta-m) + a\right]\right)$$

where $f(\sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{d+2}{2}}} e^{-a/2\sigma^2}$.

Let $Q = (y-X\beta)^T(y-X\beta) + (\beta-m)^T V^{-1}(\beta-m)$

$$f(\beta|y) \propto \int_0^\infty (\sigma^2)^{-\frac{n+d+2+p}{2}} e^{-\frac{1}{2\sigma^2}(Q+a)} d\sigma^2$$

$$\propto (Q+a)^{-\frac{d+n+p}{2}} \quad \text{(verify)}$$

$$\propto \left(1 + \frac{1}{n+d}(\beta-m_*)^T W_*^{-1}(\beta-m_*)\right)^{-\frac{d+n+p}{2}}$$

(Page 76)

$$\beta|y \sim t_p(n+d, m_*, W_*) \quad \square$$

LEC14 Linear mixed effect model

One-way Anova with Random Effects

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$
$$\varepsilon_{ij} \sim NID(0, \sigma^2)$$
$$\tau_i = \text{effect of trtmt } i.$$

$H_0: \tau_1 = \tau_2 = \cdots = \tau_k$

H_1: They are not the same.

When interest is in more than k trtmts, i.e., interest is in a population of trtmt, then sample k of them.

$\Rightarrow \tau_i \sim NID(0, \sigma_\tau^2)$ Random effect.

$\varepsilon_{ij} \sim NID(0, \sigma^2)$

& $\tau_i \perp \varepsilon_{ij} \ \forall\, i \,\&\, j$.

Hypothesis of interest is

$$H_0: \sigma_\tau^2 = 0 \,;\ H_1: \sigma_\tau^2 > 0.$$

we have data Y_{ij}, $i = 1, \cdots, k$, $j = 1, 2, \cdots, n$

$$\Sigma_i \Sigma_j (Y_{ij} - \bar{Y})^2 = \Sigma_i \Sigma_j (\bar{Y}_{i.} - \bar{Y} + Y_{ij} - \bar{Y}_{i.})^2$$
$$= \Sigma_i \Sigma_j (\bar{Y}_{i.} - \bar{Y})^2 + \Sigma_i \Sigma_j (Y_{ij} - \bar{Y}_{i.})^2$$
$$= SST_r + SSE$$

$\bar{Y}_{i.} = \mu + \alpha_i + \bar{\varepsilon}_{i.}$; $\bar{Y} = \mu + \bar{\alpha} + \bar{\varepsilon}$

(α & τ are the same thing)

$$SST_r = \sum_i^K \sum_j^n (\alpha_i - \bar{\alpha} + \bar{\varepsilon}_{i.} - \bar{\varepsilon})^2$$
$$\mathbb{E}\, SST_r = n\mathbb{E}\Big[\sum_i \big((\alpha_i - \bar{\alpha})^2 + (\bar{\varepsilon}_{i.} - \bar{\varepsilon})^2\big)\Big]$$
$$= n\sigma_\alpha^2 (k-1) \quad \left(Y_i \sim N(\mu, \sigma^2) \Rightarrow \frac{SSE}{\sigma^2} \sim \chi^2_{n-1} \right)$$
$$+ \sigma^2 (k-1)$$

$$\mathbb{E}\, SST_r/(k-1) = n\sigma_\alpha^2 + \sigma^2$$
$$\mathbb{E}(SSE) = \mathbb{E}\, n \sum_{i=1}^k (Y_{ij} - \bar{Y}_{i.})^2$$
$$= n\,\mathbb{E} \sum_{i=1}^k (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2$$
$$= n(k-1)\sigma^2$$

Since $\sum_{i=1}^k (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 \sim \chi^2_{k-1} \cdot \sigma^2$

$$\mathbb{E}\, MST_r = n\sigma_\tau^2 + \sigma^2$$
$$\Rightarrow \mathbb{E}\, \frac{MST_r - MSE}{n} = \sigma_\tau^2$$
$$\Rightarrow \hat{\sigma}_\tau^2 = \frac{MST_r - MSE}{n} \qquad \square.$$

C.I. for σ^2:

$$\frac{SSE}{\sigma^2} \sim \chi^2_{k(n-1)}$$

$\Rightarrow 100(1-\alpha)\%$ CI for σ^2

$$\frac{SSE}{\chi^2_{k(n-1),1-\frac{\alpha}{2}}} < \sigma^2 < \frac{SSE}{\chi^2_{k(n-1),\frac{\alpha}{2}}}$$

To find CI for $\sigma_\tau^2/\sigma^2 \equiv \theta$

$$\frac{(k-1)MST_r}{n\sigma_\tau^2 + \sigma^2} \sim \chi^2_{k(n-1)}$$

Since $MSE \perp MST_r$, we

$$\frac{MST_r}{MSE} \frac{\sigma^2}{n\sigma_\alpha^2 + \sigma^2} \sim F_{k,(n-1)k}$$

A $100(1-\alpha)\%$ C.I. for θ is

$$\left(\frac{1}{F}\frac{MST_r}{MSE} - 1\right)\frac{1}{n} \le \theta \le$$
$$\left(\frac{1}{F}\frac{MST_r}{MSE} - 1\right)\frac{1}{n}$$

upper $F_{\alpha/2}$

lower $F_{\alpha/2}$

Intraclass correlation Coef.

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma^2}$$

So $100(1-\alpha)\%$ C.I. for ρ is

$$\frac{L}{1+L} \le \rho \le \frac{U}{1+U}$$

where L & U are conf. limits for θ.

& noting that $\rho = \frac{\theta}{\theta+1}$

ANOVA – Two factor:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

Three factors:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

- $i=1,\dots,a$, $j=1,\dots,b$, $k=1,\dots,c$

l = # of replicates = $1,2,\dots,n$.

Fixed effects: $\sum_{i=1}^{a} \tau_i = 0$ for example.

Random effects: $\tau_i \sim NID(0, \sigma_\tau^2)$. Ex

Rule 1: fixed effects (say factor)

associate $\sum_i \tau_i^2 / a-1$ with its mean sq.

otherwise σ_τ^2 if factor is random.

$$y_{ijk} = \mu + \underbrace{\tau_i}_{A} + \underbrace{\beta_j}_{B} + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

Case (i): A, B both fixed.

Factor	a F i	b F j	n R k	EMS
τ_i	0	b	n	*
β_j	a	0	n	**
$(\tau\beta)_{ij}$	0	0	n	***
ε_{ijk}	1	1	1	Δ

(*) $E(MSA) = bn \frac{\sum_i \tau_i^2}{a-1} + \sigma^2$

(**) $E(MSB) = an \frac{\sum_j \beta_j^2}{b-1} + \sigma^2$

(***) $E(MSAB) = n \frac{\sum_i \sum_j (\tau\beta)_{ij}^2}{(a-1)(b-1)} + \sigma^2$

(Δ) $E(MSE) = \sigma^2$

Case (ii) A & B are random

Factor	a R i	b R j	n R k	EMS
τ_i	1	b	n	$bn\sigma_\tau^2 + n\sigma_{\tau\beta}^2 + \sigma^2$
β_j	a	1	n	$an\sigma_\beta^2 + n\sigma_{\tau\beta}^2 + \sigma^2$
$(\tau\beta)_{ij}$	1	1	n	$n\sigma_{\tau\beta}^2 + \sigma^2$
ε_{ijk}	1	1	1	σ^2

Test $\sigma_\tau^2 = 0 \Rightarrow \frac{MSA}{MSAB}$

$\sigma_\beta^2 = 0 \longrightarrow \frac{MSB}{MSAB}$

$\sigma_{\tau\beta}^2 = 0 \longrightarrow \frac{MSAB}{MSE}$

Case (iii) One R One F

	a	b	n	EMS
	F	R	R	
	i	j	k	
τ_i	0	b	n	$bn\frac{\sum_i \tau_i^2}{a-1} + n\sigma_{\tau\beta}^2 + \sigma^2$
β_j	a	1	n	$an\sigma_\beta^2 + \sigma^2$
$(\tau\beta)_{ij}$	1	1	n	$n\sigma_{\tau\beta}^2 + \sigma^2$
ε_{ijk}	1	1	1	σ^2

$H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_a : \frac{MSA}{MSAB}$

$H_{0B}: \sigma_\beta^2 = 0 : \frac{MSB}{MSE}$

$H_{0C}: \sigma_{\tau\beta}^2 = 0 : MSAB/MSE$

Satterwaithes' Approximations.

$\Rightarrow$ Approx. F-test.

Lec15 Satterwaite Approx

μ_1	μ_2
$\bar{x}_1$	$\bar{x}_2$
n_1	n_2
σ_1^2	σ_2^2

$\sigma_1 \neq \sigma_2$

Test $\mu_1 = \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{H_0}{\sim} t_b$$

$$b = \left(S_1^2/n_1 + S_2^2/n_2\right)^2 \Big/ \left[\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}\right]$$

Frequently, $\frac{r_i MSE_i}{\sigma^2} \sim \chi^2$,

Let $U_1, \ldots, U_k$ be independent χ^2 rvs with $r_1, \ldots, r_k$ df respectively.

Let $U = \sum_{i=1}^{k} a_i U_i$. Find constants ν such that $\nu U / \mathbb{E} U \sim \chi_b^2$

$$\mathbb{E} \frac{\nu U}{\mathbb{E} U} = b \Rightarrow \nu = b$$

Take variance on both sides.

$$2b = \mathrm{Var}\, \chi_b^2 = \mathrm{Var} \frac{\nu U}{\mathbb{E} U} = \frac{\nu^2}{(\mathbb{E} U)^2} \sum_i a_i^2 \mathrm{Var}\, U_i$$

$$= \frac{\nu^2}{(\mathbb{E} U)^2} \sum_i a_i^2 (2 r_i)$$

$$\Rightarrow b = \frac{\nu^2}{(\mathbb{E} U)^2} \sum_{i=1}^{k} (a_i \mathbb{E} U_i)^2 / r_i$$

$$\Rightarrow \hat{b} = \frac{(\sum_i a_i U_i)^2}{(\sum_{i=1}^{k} a_i^2 U_i^2)/r_i}$$

METHOD OF MOMENTS

- **Mixed linear model**
  - $Y = X\beta + Z\gamma + \varepsilon$

    $\mathbb{E}\varepsilon = 0$, $\mathrm{Cov}\,\varepsilon = R$
  - β: unknown vector of fixed effect.
  - γ: a vector of random effects

    $\mathbb{E}\gamma = 0$, $\mathrm{Cov}\,\gamma = D$,

    X, Z are known, $\mathrm{Cov}(\gamma, \varepsilon) = 0$.

$\mathbb{E}\,Y = X\beta$; $\mathrm{Cov}\,Y = ZDZ^T + R \equiv V$

To find MLEs, assume

$\gamma \sim \mathcal{N}(0, D)$, $\varepsilon \sim \mathcal{N}(0, R)$

Find MLEs for β & γ.

Write down likelihood function for y & γ in two steps.

$Y \mid \gamma \sim \mathcal{N}(X\beta + Z\gamma, R)$

$\gamma \sim \mathcal{N}(0, D)$

$$f_{Y,\gamma}(y, \gamma) = f_{Y|\gamma}(y) f_\gamma(\gamma)$$

$$\propto \frac{1}{|R|^{\frac{1}{2}}} e^{-\frac{1}{2}(y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma)} \frac{1}{|D|^{\frac{1}{2}}} e^{-\frac{1}{2}\gamma D^{-1}\gamma}$$

Dif. $\ln f_{Y,\gamma}(y, \gamma)$ w.r.t. β & γ

results

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & D^{-1} + Z^T R^{-1} Z \end{bmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

This is the set of Henderson's equations.

Can I Solve for $\hat\beta$ & $\hat\gamma$?

$$V = ZDZ^T + R = \mathrm{Cov}(y)$$

$$V^{-1} = (ZDZ^T + R)^{-1} \quad \text{Woodbury}$$

$$= R^{-1} - R^{-1} Z (D^{-1} + Z^T R^{-1} Z)^{-1} Z^T R^{-1}$$

$$\begin{pmatrix} \hat\beta \\ \hat\gamma \end{pmatrix} = \begin{pmatrix} \smile & \smile \\ \smile & \smile \end{pmatrix}^{-1} \begin{pmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{pmatrix}$$

2nd:

- $Z^T R^{-1} X \hat\beta + (D^{-1} + Z^T R^{-1} Z)\hat\gamma = Z^T R^{-1} y$

Express $\hat\gamma$ in terms of $(y - X\hat\beta)$

1st:

- $X^T R^{-1} X \hat\beta + X^T R^{-1} Z \hat\gamma = X^T R^{-1} y$.

Notice:

$$\hat\gamma = \left(D\{D^{-1} + Z^T R^{-1} Z\} - DZ^T R^{-1} Z\right) \times (D^{-1} + Z^T R^{-1} Z)^{-1} Z^T R^{-1} (y - X\hat\beta)$$

$$= \cdots = DZ^T V^{-1} (y - X\hat\beta)$$

Ex MLE's for dif. scenerios in the attached notes.

BLUP: best linear unbiased predictor (of u)

To be a BLUP, the following are requirements.

① $\hat{u}$ is a linear fct. of Y.

② $\hat{u}$ is unbiased for u:

$$\mathbb{E}(\hat{u}-u)=0.$$

③ $Var(\hat{u}-u) \le Var(v-u)$ & v is any linear unbiased predictor.

How does $\hat{u}$ look like?

$$\gamma \sim N(0, D)$$

$$Y \sim N(X\beta, V), \quad V = ZDZ^T + R$$

Then,

$$\begin{pmatrix} \gamma \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix}, \begin{pmatrix} D & DZ^T \\ ZD & V \end{pmatrix}\right)$$

$$Cov(\gamma, Y) = Cov(\gamma, X\beta + Z\gamma + \varepsilon)$$

$$= DZ^T$$

$$\mathbb{E}\gamma|Y = DZ^TV^{-1}(Y - X\beta)$$

$$\mathbb{E}(\mathbb{E}(\gamma|Y)) - \mathbb{E}\gamma = 0.$$