STAT3600 Ch. 2 Simple linear regression

**Assumptions:** $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Y random
X non-random

**Least squares method:** Find $\hat\beta_0, \hat\beta_1$ to minimize $\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$

why least-squares?
$\to a_0\hat\beta_0 + a_1\hat\beta_1$ is BLUE

$$\Rightarrow \hat\beta_0 = \bar Y - (S_{xy}/S_{xx})\bar x, \quad \hat\beta_1 = S_{xy}/S_{xx}$$

$\hat\beta_0$: expected value of Y when X is 0

$\hat\beta_1$: when X increases by 1 unit, expected value of Y increases by $\hat\beta_1$

$$S_{xx} = \sum (x_i - \bar x)^2 = \sum x_i^2 - n\bar x^2$$
$$S_{xy} = \sum (x_i - \bar x)(Y_i - \bar Y) = \sum (x_i - \bar x) Y_i = \sum (Y_i - \bar Y) x_i = \sum x_i Y_i - n\bar x \bar Y$$

**Inferrence about regression parameters:**

$$s^2 = \frac{\sum \hat\epsilon_i^2}{n-2} = \frac{SSE}{n-2} = MSE \sim (n-2)^{-1}\sigma^2\chi^2_{n-2},$$
$$E(s^2) = \sigma^2.$$

$(\hat\beta_0, \hat\beta_1)$ and $s^2$ independent
$\hat\beta_0, \hat\beta_1$ not independent

$$\hat\beta_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar x^2}{S_{xx}}\right)\sigma^2\right), \quad \hat\beta_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right),$$
$$Cov(\hat\beta_0, \hat\beta_1) = -\frac{\bar x \sigma^2}{S_{xx}}$$

**ANOVA table:**

| Source | S.S. | d.f. | M.S. |
|---|---|---|---|
| Regression | $SSR = S_{YY} - SSE = S_{xy}^2/S_{xx}$ | 1 | $MSR = SSR/1$ |
| Residual | $SSE = \sum \hat\epsilon_i^2 = S_{YY} - S_{xy}^2/S_{xx}$ | $n-2$ | $MSE = \frac{SSE}{n-2} = s^2$ |
| Total | $S_{YY} = \sum (Y_i - \bar Y)^2$ | $n-1$ | |

$S_{YY} = SSR + SSE$.

SSE: goodness of fit of model
SSR: effectiveness of X in explaining variation

$$F\text{-ratio} = \frac{MSR}{MSE}.$$

$H_0: \beta_1 = 0$
$H_1$: unrestricted

$\Rightarrow RR: \{F > F^{(\alpha)}_{1, n-2}\}$

small p-value $Pr(F_{1,n-2} > F)$
$\hookrightarrow$ significant linear relationship

$$R^2 = 1 - \frac{SSE}{S_{YY}} = \frac{SSR}{S_{YY}} = \frac{S_{xy}^2}{S_{xx}S_{YY}}$$

"coefficient of determination"

proportion of variation reduced by X.

big R $\to$ regression line effective in reducing total variation

**Linear combination** $\quad l = c_0\beta_0 + c_1\beta_1, \quad \hat{l} = c_0\hat{\beta}_0 + c_1\hat{\beta}_1$

$$s.e.(\hat{l}-l) = s\cdot\sqrt{c_0^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)+\frac{c_1^2}{S_{xx}}-2c_0c_1\left(\frac{\bar{x}}{S_{xx}}\right)} = s\sqrt{\frac{c_0^2}{n}+\frac{(c_0\bar{x}-c_1)^2}{S_{xx}}}$$

$s = \sqrt{MSE}$

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$$\boxed{\frac{\hat{l}-l}{s.e.(\hat{l}-l)} \sim t_{n-2} \Rightarrow \text{C.I. for } l:\ \hat{l} \pm s.e.(\hat{l}-l)\,t_{n-2}^{\alpha/2}}$$

$T^2 = F$. $\left(t_{n-2}^{(\alpha/2)}\right)^2 = F_{1,n-2}^{\alpha}$

$H_0: l = l_0 \quad H_1: l \neq l_0 \Rightarrow T = \left|\frac{\hat{l}-l_0}{s.e.(\hat{l}-l_0)}\right|$, 

p-value $= Pr(|t_{n-2}| > T)$
$= 2\cdot Pr(t_{n-2} > T)$

reject $H_0$ iff $|T| > t_{n-2}^{\alpha/2}$

$c = (1,0)^T$, $T = \dfrac{\hat{\beta}_0 - \beta_0}{s\sqrt{1/n + \bar{x}^2/S_{xx}}}$

$c = (0,1)^T$, $T = \dfrac{\hat{\beta}_1 - \beta_1}{s/\sqrt{S_{xx}}}$

$H_0: l \le l_0$ vs $H_1: l > l_0$
p-value $= p(t_{n-2} > T)$, RR $T > t_{n-2}^{\alpha}$

**Prediction**

$Y_j^* \sim N(\beta_0 + \beta_1 x_j^*, \sigma^2)$, predict $l = \sum_{j=1}^{m} a_j Y_j^*$

$l, \hat{l}$ independent (obs. vs future), $\hat{l} = \sum a_j(\hat{\beta}_0 + \hat{\beta}_1 x_j^*) = \hat{\beta}_0 \sum a_j + \hat{\beta}_1 \sum a_j x_j^*$

$l \sim N\left(\sum a_j(\beta_0 + \beta_1 x_j^*),\ \sigma^2 \sum a_j^2\right)$ ①

$\hat{l} \sim N\left(\sum a_j(\beta_0 + \beta_1 x_j^*),\ \sigma^2\left(\frac{(\sum a_j)^2}{n} + \frac{(\sum a_j(x_j^* - \bar{x}))^2}{S_{xx}}\right)\right)$ ②

$x^*$: new, incoming. $x$: build model

$100(1-\alpha)\%$ CI: $\hat{l} \pm s.e.(\hat{l}-l)\, t_{n-2}^{(\alpha/2)}$ → ① + ②

$\frac{\hat{l}-l}{s.e.(\hat{l}-l)} \sim t_{n-2}$

Special cases: CI.

$m=1, a=1$: $\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{n-2}^{(\alpha/2)}\, s\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{S_{xx}} + 1}$

$a_1 = a_2 = \cdots = a_m = \frac{1}{m}$: $\hat{\beta}_0 + \hat{\beta}_1 \bar{x}^* \pm t_{n-2}^{(\alpha/2)}\, s\sqrt{\frac{1}{n} + \frac{(\bar{x}^*-\bar{x})^2}{S_{xx}} + \frac{1}{m}}$

same mean, but CI of mean is shorter.

$m=2, a_1 = 1, a_2 = -1$: $\hat{\beta}_1(x_1^* - x_2^*) \pm t_{n-2}^{(\alpha/2)}\, s\sqrt{\frac{(x_1^*-x_2^*)^2}{S_{xx}} + 2}$

☆ prediction interval is wider than confidence interval (s.e. larger)

STAT3600 Ch. 3-4 Distribution theory, General linear model (1)

$E(AWB) = A \cdot E(W) \cdot B$

$\rightarrow E(AY+u) = A \cdot E(Y) + u$ $\quad Y \sim N_n(\mu, \Sigma),\ AY+u \sim N_k(A\mu, A\Sigma A^T)$

$\rightarrow Var(AY+u) = A \cdot Var(Y) \cdot A^T$ , $\rightarrow var(X+Y) = Var(X) + var(Y)$

$\rightarrow E(Y^T A Y) = tr(A\,var(Y)) + E(Y^T) A\, E(Y)$, $A$: square const. matrix

$Y = Y_1^2 + Y_2^2 + \cdots + Y_m^2 \iff Y \sim \chi^2_m,\ Y_i \overset{iid}{\sim} N(0,1)$

$Y_1 \sim \chi^2_m,\ Y_2\ \chi^2_n,\quad \dfrac{Y_1/m}{Y_2/n} \sim F_{m,n}$.

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$Z \sim N(0,1),\quad X \sim \chi^2_k,\quad Y = \dfrac{Z}{\sqrt{X/k}} \sim t_k$

$\rightarrow Y \sim N_n(\mu, \Sigma)$. then $A$ psd, symmetric, $B$, $A\Sigma B^T = 0 \Rightarrow Y^T A Y$ independent $BY$.

"Cochran's theorem"

$\rightarrow Y \sim N_n(0_n, \sigma^2 I_n)$. $A_1, \ldots A_m$ symmetric, $\sum A_i = I$, $A_i^2 = A_i$

then $Y^T A_1 Y, \ldots Y^T A_m Y$ independent, $Y^T A_i Y \sim \sigma^2 \chi^2_{tr(A_i)}$

---

$\boxed{\text{General linear model}}$ $Y = X\beta + \epsilon$. $\quad Y$: $n \times 1$, responses

$X$: $n \times d$, design matrix, rank $d < n$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{1p} \\ 1 & x_{21} & x_{2p} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$\beta$: $d \times 1$, $\epsilon$: $n \times 1$

Hat matrix: $H = X(X^TX)^{-1}X^T$ : $\boxed{HX = X \quad H^2 = H \quad H^T = H \quad (I-H)^2 = I-H \quad (I-H)^T = I-H}$ $\quad E(\epsilon) = 0,\ var(\epsilon) = \sigma^2 I$

LSE $\hat\beta = (X^TX)^{-1}X^TY$ . predicted $\hat Y = X\hat\beta = HY$

$S_{YY} = Y^TY - n\bar{Y}^2$ $\quad$ residuals $\hat\epsilon = Y - \hat Y = (I_n - H)Y$

$\boxed{\begin{array}{l} \epsilon \sim N_n(0_n, \sigma^2 I_n) \\ Y \sim N_n(X\beta, \sigma^2 I_n) \end{array}}$

under constraints $B\beta = c$: $\tilde\beta = \hat\beta + (X^TX)^{-1}B^T(B(X^TX)^{-1}B^T)^{-1}(c - B\hat\beta)$

---

$\boxed{\text{Inference}}$: $\hat\beta = (X^TX)^{-1}X^TY \sim N_d(\beta, \sigma^2(X^TX)^{-1})$ $\quad$ independent

$s^2 = (n-d)^{-1}\hat\epsilon^T\hat\epsilon \sim (n-d)^{-1}\sigma^2\chi^2_{n-d}$, $E(s^2) = \sigma^2$.

extract individual variance / covariance

**Reduced model vs Full model**

Full model $\Omega$: General linear model, constraints $A\beta = a$, $A: q\times d$, rank $q < r$, $a: \mathbb{R}^q$

Reduced model $w$: General linear model, constraints $A\beta = a$, and $B\beta = b$, $\begin{bmatrix} A \\ B \end{bmatrix}$ $r\times d$, rank $r \le d$, $b: \mathbb{R}^{r-q}$

$H_0: W$, vs $H_1: \Omega$. $q=0 \to$ unrestricted

| Source | s.s. | d.f. | m.s. |
|---|---|---|---|
| Regression fitting $w$ | $SSR_w = S_{YY} - SSE_w$ | $df(\text{total}) - df(\Omega) - df(\text{extra})$ | |
| Extra | $SSEXT = SSE_w - SSE_\Omega$ | # extra constraints | $MSEXT = \frac{SSEXT}{df}$ |
| Regression fitting $\Omega$ | $SSE_\Omega = \lVert Y - \hat{Y} \rVert^2 = (Y-HY)^T(Y-HY) = Y^TY - Y^THY$ | $n-d-$# original constraint | $MSE_\Omega = \frac{SSE_\Omega}{df}$ |
| Total | $S_{YY} = Y^TY - n\bar{Y}^2$ | $n-1$ | |

$d$: #col of $X$

$$F = \frac{MSEXT}{MSE_\Omega} \sim F_{k, n-d}.$$

$RR: \{F > F^{(\alpha)}_{k,n-d}\}$

$Pr(F_{k,n-d} > F) < \alpha. \Rightarrow$ effective in explaining variation

---

**Linear combination**

$c = [c_1, \dots, c_d]^T$, $\ell = c^T\beta$, $\hat{\ell} = c^T\hat{\beta}$ regression parameters

$\Rightarrow \hat{\ell} \sim N(\ell, \sigma^2 c^T(X^TX)^{-1}c)$, $s.e.(\hat{\ell} - \ell) = s\sqrt{c^T(X^TX)^{-1}c} = s.e.(\hat{\ell})$

$$\frac{\hat{\ell} - \ell}{s.e.(\hat{\ell} - \ell)} \sim t_{n-d}$$

Re: $\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$

C.I.: $\hat{\ell} \pm s.e.(\hat{\ell} - \ell)\, t^{\alpha/2}_{n-d}$,

Test $\ell = \ell_0$: $T = \left| \frac{\hat{\ell} - \ell_0}{s.e.(\hat{\ell} - \ell)} \right|$, $RR: \{T > t^{\alpha/2}_{n-d}\}$

equivalent to F-test for testing $\beta_j = 0$

**Prediction**

$Y_1^* \dots Y_m^*$ be independent future observations $a = [a_1 \dots a_m]^T$

$$Y^* = \begin{bmatrix} Y_1^* \\ \vdots \\ Y_m^* \end{bmatrix} \sim N_m(X^*\beta, \sigma^2 I_m)$$

$\ell = a^TY^*$, $\hat{\ell} = a^TX^*\hat{\beta}$

$$\Rightarrow s.e.(\hat{\ell} - \ell) = s\,(\underbrace{a^TX^*}_{\text{coeff}}\underbrace{(X^TX)^{-1}}_{\text{var}(\hat{\beta})}\underbrace{X^{*T}a}_{\text{coeff}} + \underbrace{a^Ta}_{\text{var}(Y^*)})^{1/2}$$

$$\Rightarrow C.I.: \hat{\ell} \pm s.e.(\hat{\ell} - \ell)\, t^{\alpha/2}_{n-d}$$

! Don't mix up $X$ vs $X^*$ !

$X$ ↑ previous model building

$X^*$ ↑ new, incoming

$\boxed{\text{Assumptions}}$ $Y_{ij} = \theta_i + \epsilon_{ij}$, $i = 1 \dots k$, $j = 1 \dots n_i$,

$E(\epsilon_{ij}) = 0$, $Var(\epsilon_{ij}) = \sigma^2$

$\boxed{\text{one-way: one factor, k levels}}$

reparametrize

$$Y = \mu + \tau_1 D_1 + \dots + \tau_{k-1} D_{k-1} + \epsilon$$

$D_i$: dummy variable

$D_i = 1$ and $D_{i'} = 0$ for $i' \neq i$

$$\mu = \theta_k, \quad \theta_i = \mu + \tau_i$$

why reparametrize?

→ use old stuff

→ both v1 ($\theta$) and v2 (reparametrized) use same # of free parameters

cell-means model: fitting

$\hat{\mu} = \bar{Y}_{k\cdot}$

$\hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}_{k\cdot}$

$\theta_i = \bar{Y}_{i\cdot}$

$\boxed{\text{two-way: interactive, (I, J) levels}}$

$$Y_{ijk} = \theta_{ij} + \epsilon_{ijk}$$

reparametrize

$$Y_{ijk} = \mu + \alpha_1 A_1 + \dots + \alpha_{I-1} A_{I-1} + \beta_1 B_1 + \dots + \beta_{J-1} B_{J-1}$$
$$+ (\alpha\beta)_{11} A_1 B_1 + \dots + (\alpha\beta)_{1,J-1} A_1 B_{J-1} + \dots + (\alpha\beta)_{I-1,J-1} A_{I-1} B_{J-1}$$

$$\mu = \theta_{IJ}, \quad \alpha_i = \theta_{iJ} - \theta_{IJ}, \quad \beta_j = \theta_{Ij} - \theta_{IJ}, \quad (\alpha\beta)_{ij} = \theta_{ij} - \theta_{iJ} - \theta_{Ij} + \theta_{IJ}$$

$\alpha_i$: main effects of A ; $\beta_j$: main effects of B. $\alpha\beta$: Interaction

cell-means model: fitting

$\hat{\mu} = \bar{Y}_{IJ\cdot}$

$\hat{\alpha}_i = \bar{Y}_{iJ\cdot} - \bar{Y}_{IJ\cdot}$

$\hat{\beta}_j = \bar{Y}_{Ij\cdot} - \bar{Y}_{IJ\cdot}$

$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{iJ\cdot} - \bar{Y}_{Ij\cdot} + \bar{Y}_{IJ\cdot}$

$\hat{\theta}_{ij} = \bar{Y}_{ij\cdot}$

$\boxed{\text{two-way: additive}}$

(no interactions)

$$Y_{ijk} = \mu + \alpha_1 A_1 + \dots + \alpha_{I-1} A_{I-1} + \beta_1 B_1 + \dots + \beta_{J-1} B_{J-1}$$

model fitting: $(\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j)$ from $(X^T X)^{-1} X^T Y$.

lse.

ANOVA one-way:

| Source | S.S. | d.f. | M.S. |
|---|---|---|---|
| Between-groups | $SST = S_{YY} - SSE$ | $k-1$ | $MST = SST/(k-1)$ |
| Within-groups | $SSE = \sum\sum(Y_{ij} - \bar{Y}_{i.})^2$ | $n-k$ | $MSE = SSE/n-k$ |
| Total | $S_{YY} = \sum\sum(Y_{ij} - \bar{Y}_{..})^2$ | $n-1$ | |

$$F = \frac{MST}{MSE} \sim F_{k-1, n-k}, \quad RR: \{F > F^{(\alpha)}_{k-1, n-k}\}$$

$$Pr(F_{k-1, n-k} > F) < \alpha.$$

$H_0$ rejected, factor has significant effects

two-way interactions

| Source | S.S. | df | MS |
|---|---|---|---|
| A, B | $S_{YY} - SSE_W$ | $I+J-2$ | |
| Interaction | $SSEXT_{AB} = SSE_W - SSE_\Omega$ | $(I-1)(J-1)$ | $MSEXT_{AB} = \frac{SSEXT_{AB}}{(I-1)(J-1)}$ |
| Error | $SSE_\Omega$ | $n-IJ$ | $MSE_\Omega = \frac{SSE_\Omega}{n-IJ}$ |
| Total | $S_{YY}$ | $n-1$ | |

d.f:
Total $= n-1$ $\quad F = \frac{MSEXT}{MSE}$
$A = I-1$
$B = J-1$
Interaction $= (I-1)(J-1)$

$$F = \frac{MSEXT}{MSE} \sim F_{(I-1)(J-1), n-IJ}$$

two-way main effects

| Source | S.S. | d.f. | M.S. |
|---|---|---|---|
| B, interactions | $S_{YY} - SSE_W$ | $I(J-1)$ | |
| A | $SSEXT_A$ | $I-1$ | $MSEXT_A = \frac{SSEXT_A}{I-1}$ |
| Error | $SSE_\Omega$ | $n-IJ$ | $MSE_\Omega = \frac{SSE_\Omega}{n-IJ}$ |
| Total | $S_{YY}$ | $n-1$ | |

$$F = \frac{MSEXT}{MSE} \sim F_{I-1, n-IJ}$$

"If there is strong evidence for the existence of interactions, studying the main effects may not be meaningful since these effects vary over different levels of the other factor. Averaging these effects over all levels of the other factor conceals the reality."

two-way main effects (additive)

| Source | S.S. | d.f. | M.S. | |
|---|---|---|---|---|
| B, μ | $S_{YY} - SSE_W$ | $J-1$ | | |
| A | $SSEXT_A = SSE_W - SSE_\Omega$ | $I-1$ | $MSEXT = \frac{SSEXT}{I-1}$ | $F = \frac{MSEXT}{MSE}$ |
| Error | $SSE_\Omega$ | $n-I-J+1$ | $MSE = \frac{SSE}{n-I-J+1}$ | |
| Total | $S_{YY}$ | $n-1$ | | |

**Analysis of treatment means**

$$L = a_1\psi_1 + a_2\psi_2 + \cdots + a_k\psi_k \qquad \psi_i: \text{ mean of treatment } i$$

$\sum a_i = 0 \to$ contrast

Re: CLT

$$\hat{L} = \sum_{r=1}^{k} a_r \bar{Y}_{r\cdot}$$

$$\bar{Y}_r \sim N\left(\psi_r, \frac{\sigma^2}{m_r}\right) \to \boxed{\hat{L} \sim N\left(L, \sigma^2 \sum_{r=1}^{k} \frac{a_r^2}{m_r}\right) \quad s.e.(\hat{L}) = s\left(\sum_{r=1}^{k} \frac{a_r^2}{m_r}\right)^{1/2}}$$

e.g. $C.I(L) = \hat{L} \pm s.e.(\hat{L})\, t_{n-k}^{\alpha/2}$

Test $L = L_0$: $T = \frac{\hat{L} - L_0}{s.e.(\hat{L})}$, $RR: |T| > t_{n-\underline{k}}^{\alpha/2} \to$ totally $K$ cells

Two-way classification model:

- Significant interactions: Study $\hat{L}$ at each fixed $j$ (other factor)

  e.g. $\bar{Y}_{1j\cdot} - \bar{Y}_{1j\cdot} \pm s\sqrt{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}\; t_{n-\underline{IJ}}^{(\alpha/2)} \to$ totally $IJ$ cells

- Insignificant interactions: lump together factors of other interval (re: meaningful)

  e.g. $\sum w_j(\theta_{1j} - \theta_{2j})$, $CI = \sum w_j(\bar{Y}_{1j\cdot} - \bar{Y}_{2j\cdot}) \pm s\sqrt{\sum w_j^2\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)}\, t_{n-IJ}^{\alpha/2}$

---

**Factorial design for balanced data**

Intuition: Originally, $\underbrace{Y_{ij}}_{K \text{ params}} = \underbrace{\mu + T_1 D_1 + \cdots + T_k D_k}_{1+K \text{ params}} \to$ constraint $D_k = 0$.

$$\underbrace{Y_{ij}}_{IJ} = \underbrace{\mu + \sum \alpha_i A_i + \sum \beta_j B_j + \sum \alpha\beta_{ij} A_i B_j}_{1 + I + J + IJ} \to \text{constraints (remove } I, J) \downarrow IJ \text{ params.}$$

alternateively, One-way: $\sum D_i = 0 \to K$ params

two-way: $\sum \alpha_i = 0 \to I-1$ params

$\sum \beta_j = 0 \to J-1$ params

$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0 \to I+J-1$ restrictions, $ab - a - b - 1 = (a-1)(b-1)$ params.

e.g. $K=2$. one-way.

$$Y_{ij} = \mu + T_1 D_1' + T_2 D_2' + \varepsilon_{ij}$$

$$= \mu + T_1 D_1 + \varepsilon_{ij} \qquad D_1 = D_1' - D_2'$$

$$= \begin{cases} \mu + T_1 + \varepsilon_{ij}, & i=1 \\ \mu - T_1 + \varepsilon_{ij}, & i=2 \end{cases}$$

$T_1' = T_1$

$T_2' = -T_1$

$T_1' + T_2' = 0.$

$$X = \begin{pmatrix} 1_{2n} & \begin{matrix} 1_n \\ -1_n \end{matrix} \end{pmatrix} \longleftrightarrow \begin{matrix} 1_{2n} : \mu \\ 1_n : T_1 \\ -1_n : T_2 \end{matrix}$$

$$\begin{pmatrix} \hat{\mu} \\ \hat{T}_1 \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 2n & 0 \\ 0 & 2n \end{pmatrix}^{-1} \begin{pmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} - Y_{2\cdot} \end{pmatrix}$$

$$= \frac{1}{2n} \begin{pmatrix} Y_{\cdot\cdot} \\ 2Y_{1\cdot} - Y_{\cdot\cdot} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{Y}_{\cdot\cdot} \\ \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot} \end{pmatrix}$$

$$\boxed{\hat{Y}_{ij} = \bar{Y}_{\cdot\cdot} + (\bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot}) D_1 - (\bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot}) D_2}$$

**Balanced data: Parameter estimators**

LSE is same as MLE (re: STAT2602)

One-way: $\hat{Y}_{ij} = \hat{\mu} + \hat{T}_1 D_1 + \dots + \hat{T}_K D_K$

$$\begin{cases} \hat{\mu} = \bar{Y}_{\cdot\cdot} \\ \hat{T}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} \end{cases}$$

Two-way: $\hat{Y}_{ij} = \hat{\mu} + \hat{\alpha}_1 A_1 + \dots + \hat{\alpha}_I A_I + \hat{\beta}_1 B_1 + \dots + \hat{\beta}_J B_J + \widehat{\alpha\beta}_{11} A_1 B_1 + \dots + \widehat{\alpha\beta}_{IJ} A_I B_J$

$$\begin{cases} \hat{\mu} = \bar{Y}_{\cdot\cdot\cdot} \\ \hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot} \\ \hat{\beta}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot} \\ \widehat{\alpha\beta}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot} \end{cases}$$ (inclusion-exclusion)

**Balanced data: ANOVA**

| Source | d.f. | S.S. | M.S. | F |
|---|---|---|---|---|
| Main$_A$ | $I-1$ | $SS_A = Jn \sum_{i=1}^I \hat{\alpha}_i^2$ | $MS_A$ | $MS_A/MS_E$ |
| Main$_B$ | $J-1$ | $SS_B = In \sum_{j=1}^J \hat{\beta}_j^2$ | $MS_B$ | $MS_B/MS_E$ |
| Interaction | $(I-1)(J-1)$ | $SS_{AB} = n \sum_{i=1}^I \sum_{j=1}^J \widehat{\alpha\beta}_{ij}^2$ | $MS_{AB}$ | $MS_{AB}/MS_E$ |
| Error | $(IJ)(n-1)$ | $SSE = S_{YY} - SS_{AB} - SS_A - SS_B$ | $MS_E$ | |
| Total | $(IJn)-1$ | $S_{YY} = \sum (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$ | | |

$n$: # of obs each cell

STAT3600 ch. 4, 7 General linear model (2), Regression diagnostics

Don't worry, this is all provided in formula sheet!

$\boxed{\text{Simultaneous Interval:}}$ Objective: Set of C.I. for multiple parameters s.t. $Pr(\text{all C.I. contain real value}) = 1-\alpha$.

1. Inference on regression parameters: $c^T \beta \in I(c)$

(a) Bonferroni: Finite $C = \{c_1 \dots c_g\}$

$$I(c_j) = c_j^T \hat\beta \pm s.e.(c_j^T \hat\beta)\, t_{n-d}^{(\alpha/2g)} = c_j^T \hat\beta \pm s\{c_j^T (X^T X)^{-1} c_j\}^{1/2}\, t_{n-d}^{\alpha/2g}$$

→ i.e. Old C.I. for inferrence, change $\alpha/2 \to \alpha/2g$ $g = \#$C.I.

(b) Scheffe: any $C$. $\tilde d$ = rank of $C$.

$$I(c) = c^T \hat\beta \pm s.e.(c^T \hat\beta) \sqrt{\tilde d F^{\alpha}_{\tilde d, n-d}} = c^T \hat\beta \pm s\{c^T (X^T X)^{-1} c\}^{1/2} \sqrt{\tilde d F^{\alpha}_{\tilde d, n-d}}$$

→ i.e. Old C.I. for inferrence, change $t_{n-d}^{\alpha/2} \to \sqrt{\tilde d F^{\alpha}_{\tilde d, n-d}}$

2. Prediction: $a^T Y^* \in J(a)$

(a) Bonferroni: Finite $A = \{a_1, \dots a_g\}$

$$J(a_i) = a_i^T X^* \hat\beta \pm s.e.(a_i^T X^* \hat\beta - a_i Y^*)\, t_{n-d}^{\alpha/2g}$$
$$= a_i^T X^* \hat\beta \pm s\{a_i^T X^* (X^T X)^{-1} X^{*T} a_i + a_i^T a_i\}^{1/2}\, t_{n-d}^{\alpha/2g}$$

(b) Scheffe: any $A$. $\tilde m$ = rank of $A$.

$$J(a) = a^T X^* \hat\beta \pm s.e.(a_i^T X^* \hat\beta - a_i Y^*) \sqrt{\tilde m F^{\alpha}_{\tilde m, n-d}}$$
$$= a^T X^* \hat\beta \pm s\{a_i^T X^* (X^T X)^{-1} X^{*T} a_i - a_i^T a_i\}^{1/2} \sqrt{\tilde m F^{\alpha}_{\tilde m, n-d}}$$

$\boxed{\text{Polynomial regression:}}$ $X_1 = X$, $X_2 = X^2$, $X_3 = X^3$, … $X_p = X^p$

⇒ Test degree of polynomial with ANOVA (reduced vs full)

$\boxed{\text{Interaction regression:}}$ Additive: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$

Interactive: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{12} X_{i1} X_{i2} + \epsilon_i$

multiply!

"effect of $X_1$ is unaffected by any changes in other explanatory variables"

Regression Diagnostics: violation of assumptions of linear models.

- Curvilinearity: Responses and explanatory variables are related in a more subtle way than linear → some curve in residual plot?
- Hetroscardacity: Responses have non-const. variance → inconsistent variance in residual e.g. increasing
- Dependency : non independence of error terms (e.g. related to time / X)
- Outliers / Influential observations

$(\hat{\varepsilon} = (I-H)Y \sim N(0, \sigma^2(I-H))$, correlated but neligible)

remoteness → Leverage: $x_i^T (X^T X)^{-1} x_i$ ($i^{th}$ diagonal element of $H = X(X^TX)^{-1}X^T$)

SLR: $h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2}$, $0 \le h_i \le 1$ $\quad var(\hat{\varepsilon}_i) = \sigma^2(1-h_i)$, $var(\hat{Y}_i) = \sigma^2 h_i$

"Measure of the remoteness of the $i^{th}$ observation from the remaining observations in space of explanatory variables"

$h_i \ge \frac{2d}{n}$ → high leverage.

Outliers → Studentised residual: $\hat{f}_i = \frac{\hat{\varepsilon}_i}{s\sqrt{1-h_i}}$, $s = MSE$

Studentised deleted residual: $\hat{r}_i = \frac{Y_i - \hat{Y}_{(i)}}{s_{(i)}/\sqrt{1-h_i}} = \hat{f}_i \sqrt{\frac{n-d-1}{n-d-\hat{f}_i^2}} \sim t_{n-d-1}$

$|\hat{r}_i| > 2$ → outlier

$\hat{Y}_{(i)} = x_i^T \beta_{(i)}$, prediction after removing $i^{th}$ observation

advantage: "if the linear model is correct, then all $\hat{f}_i / \hat{r}_i$ have same variance, comparable with one another" vs $\hat{\varepsilon}_i$

Influential observations → "If we exclude this observation we get very different analysis results"

Cook's distance: Use $\hat{\beta} \to \hat{\beta}_{(i)}$ (delete $i^{th}$ point)

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T X^T X (\hat{\beta} - \hat{\beta}_{(i)})}{d s^2} = \frac{\hat{f}_i^2}{d}\left(\frac{h_i}{1-h_i}\right)$$

$> F_{d,\,n-d}^{50\%}$ → influential

QQ plot (normal $\hat{\varepsilon}$ test)

1. Sort $\hat{\varepsilon}_i$ increasing
2. plot $\hat{\varepsilon}_i$ against $\Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$
3. linear → normal

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{(i)}}{s_{(i)}\sqrt{h_i}} = \hat{r}_i \sqrt{\frac{h_i}{1-h_i}}$$

$|DFFITS_i| > 2\sqrt{d/n}$ → influential

- Multicollinearity: $\tilde{X} = \left[\frac{c_1}{\|c_1\|}, \frac{c_2}{\|c_2\|}, \cdots \frac{c_d}{\|c_d\|}\right]$

linear dependence between explanatory variables

$\lambda_1 \cdots \lambda_d$ : eigenvalues of $\tilde{X}^T\tilde{X}$,
$r_1 \cdots r_d$ : eigenvectors, $\|r_i\| = 1$

$\eta_k = \left(\frac{\max_j \lambda_j}{\lambda_k}\right)^{1/2}$ "condition index", $\quad \phi_{kj} = \frac{\lambda_k^{-1} r_{jk}^2}{\sum_{l=1}^{d} \lambda_l^{-1} r_{jl}^2}$ "proportion of variance $\beta_j$ by $r_k$"

1. Find large condition index $\tilde{\beta}_k$
2. Find large values of $\phi_{kj}$ → linear combination