

TMA 4255 Dec. 2008

Problem 1 :

\hat{A} = average of Y when A is $+1$
 - average of Y when A is -1

$$= \frac{70.8 + 73.2 + 79.3 + 91.2}{4} - \frac{69.3 + 71.3 + 77.5 + 88.9}{4}$$

$$= \underline{\underline{1.8750}} \text{ (MINITAB)}$$

Alias-structure shows that this estimate is confounded with BD , CE , $ABCDE$

\hat{BC} = compute by contrast ± 1

$$= \frac{Y_1 + Y_2 - Y_3 - Y_4 - Y_5 - Y_6 + Y_7 + Y_8}{4} = \underline{\underline{4.7250}}$$

Confounded with DE , ABE , ACD .

Defining relation: From $D = AB$, $E = AC$ we get

$$I = ABD, I = ACE,$$

$$I = ABD \cdot ACE = BCDE$$

so defining relation is

$$\underline{\underline{I = ABD = ACE = BCDE}}$$

\Rightarrow Resolution is 3 (lowest number of letters)

Practical interpretation: Resolution $k \Rightarrow$
main effects are confounded with effects
of order $k-1$; second order interactions
are confounded with order $k-2$ etc.

b) It follows from alias-structure, by
setting all interactions with D or E
equal to 0; that the alias structure now is

I
A
B
C
D+AB
E+AC
BC
~~ABC~~ ABC

Thus A, B, C can be estimated unconfounded,
while D is confounded with AB,
E is confounded with AC
(their generators!)

$$\begin{aligned} \widehat{\text{Effect}} &= \frac{\sum_{i=1}^8 \pm y_i}{4} \Rightarrow \text{Var}(\widehat{\text{Effect}}) = \frac{1}{16} \cdot 8\sigma^2 \\ &= \underline{\underline{\frac{\sigma^2}{2}}} \end{aligned}$$

$$\sigma^2 = 2 \Rightarrow \text{Var}(\widehat{\text{Effect}}) = 1$$

Thus, say that an effect is significant if $|\widehat{\text{Effect}}| > z_{0.025} \sqrt{1} = \underline{1.96}$

Thus from MINITAB-result:

B, C, BC are the only significant effects.

$$A \text{ significant} \Leftrightarrow z_{\alpha/2} \leq 1.8750$$

$$\Leftrightarrow \alpha/2 \geq 0.0304 \Leftrightarrow \alpha \geq 0.0608$$

Thus: smallest sign level making A significant is $\alpha = 0.0608$ (p-value)

c) We have $m=7$ estimated effects.

(1) The median of the 7 numbers is the 4th smallest, which is 1.8750

Preliminary estimate is $1.5 \cdot 1.8750 = 2.8125$

(2) Throw out all effects with abs. value $\geq 2.5 \cdot 2.8125 = 7.0313$.

Then 6 remains, with median $\frac{0.2250 + 1.8750}{2} = 1.05$

so PSE = $1.5 \cdot 1.05 = \underline{1.575}$ g.e.d.

Problem 2:

a) Model $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$

where $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \delta_{ij} = \sum_j \delta_{ij} = 0$

$\varepsilon_{ijk} \sim N(0, \sigma^2)$, all independent.

Test statistic for interaction: $F_3 = \frac{S_3^2}{S^2} = \frac{\frac{SS(AB)}{(a-1)(b-1)}}{\frac{SSE}{ab(a-1)}} = \frac{\frac{SS(AB)}{6}}{\frac{SSE}{12}}$

(reject for large values)

Under H_0 is $F_3 \sim \text{Fisher}(6, 12)$

so critical value is $f_{6,12,0.05} = 3.00$

Observed F_3 is 2.29 so we do not reject.

Main effects:

A: $H_0: \alpha_i = 0$ for all i vs. $H_1: \text{not so}$

Test statistic:

$$F_1 = \frac{\frac{SS(A)}{a-1}}{\frac{SSE}{ab(n-1)}} = \frac{MSA}{MSE} = 93.52$$

$\sim \text{Fisher}(2, 12)$ under H_0

Reject at 5% level

(critical value 3.89)

B: $H_0: \beta_j = 0$ for all j vs. $H_1: \text{not so}$

$$F_2 = 1330.48 \quad (\sim \text{Fisher}(3, 12))$$

critical value 3.49.)

$$b) \cdot SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$SS(AB) = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$\frac{SSE}{\sigma^2} = \sum_i \sum_j \left(\underbrace{\frac{\sum_k (Y_{ijk} - \bar{Y}_{ij.})^2}{\sigma^2}} \right)$$

This is χ_{n-1}^2 by result!

also: Have ab terms in $\sum_i \sum_j$,

so $\frac{SSE}{\sigma^2}$ is $\chi_{ab(n-1)}^2$

Here: $ab(n-1) = 12$

Estimator: $S^2 = \frac{SSE}{12} = \text{MSE} = \underline{\underline{0.537}}$
↑
estimate

$$\frac{SS(AB)}{\sigma^2} \sim \chi_6^2 \Rightarrow \frac{SSE + SS(AB)}{\sigma^2} \sim \chi_{18}^2$$

$\Rightarrow \frac{SSE + SS(AB)}{18}$ is estimator of σ^2

Computed:
$$\frac{8.99 + 6.45}{18} = \underline{\underline{0.8578}}$$

Problem 3

a) The observations Y_{ijk} can be renumbered from 1, ..., 24 and for each of them is given a pair (x_1, x_2) as explanatory variables.

Must assume ϵ_{ijk} are independent and $N(0, \sigma^2)$

~~Good~~

Variation explained:
$$R^2 = \frac{SSR}{SST} = \frac{2242.6}{2261.4} = \underline{\underline{0.992}}$$

Significance is tested by $F = \frac{SSR/2}{SE/21} = 1251.40$ (99.2% explained)
 This is significant (Fisher (2, 21)) by any reasonable significance level.

$H_0: \beta_r = 0$ vs. $H_1: \beta_r \neq 0$ for $r = 1, 2$

Test statistic $T_r = \frac{\beta_r}{SD(\hat{\beta}_r)} \sim t_{21}$ under H_0

so reject at 1% if $|T_r| > t_{0.005, 21} = 2.831$

So both hypotheses are rejected.

The tests are essentially about the main effects, so the results of ~~of~~ ~~the~~ Problem 2 are confirmed.

~~Similarities~~ Similarities: Testing effects of length and wind speed

Differences:

The model in the present problem is more specified. Interaction is not taken care of in this problem.

b) Studentized residuals:

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}} \quad \text{where } e_i = y_i - \hat{y}_i$$

are supposed to behave like "N(0,1)" under the ~~the~~ model.

First plot (upper left) checks normality of the r_i (see theory in lectures and book). ~~Seems~~ ^{Seems} OK

~~Second~~ ^{upper right} plot (~~check~~) plots (\hat{y}_i, r_i)
Seems to be slightly \cap -shaped

Third plot (lower left): Seems to be larger variation for small x_1

Fourth plot: Seems to be \cap -shaped.

May get better fit by introducing explanatory variables $x_1 x_2, x_1^2, x_2^2$ (or some of these). This would mean to approximate the function of x_1, x_2 by its second order Taylor polynome.

Problem 4.

For the model (2) we compute

$$P(\bar{X} \in]0, 40]) = F(40) = 0.2212$$

$$P(\bar{X} \in]40, 60]) = F(60) - F(40) = 0.2090$$

$$P(\bar{X} \in]60, 80]) = F(80) - F(60) = 0.2019$$

$$P(\bar{X} \in]80, 100]) = F(100) - F(80) = 0.1583$$

$$P(\bar{X} \in]100, 120]) = F(120) - F(100) = 0.1042$$

$$P(\bar{X} \in]120, \infty) = 1 - F(120) = 0.1054$$

The expected values are found by multiplying these by 65.

Then:

$$\chi^2 = \frac{(7 - 65 \cdot 0.2212)^2}{65 \cdot 0.2212} + \frac{(14 - 65 \cdot 0.2090)^2}{65 \cdot 0.2090} + \dots = 10.15$$

Reject if $\chi^2 > \chi^2_{0.05, 6-1} = 11.070$

so do not reject at 5% (P-value is 0.071)

If parameters not given, we would use the maximum likelihood estimates and perform the same computation - but df would then be reduced by 2 (one for each estimated parameter).