STAT3015/4030/7030 GLMs 8/8/2017 (1)

Two-way ANOVA – the Trade-in data example

Model

Assumptions $\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$

$$Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$$

- Y_{ijk}: value
- μ: overall mean (intercept)
- α_i: age factor
- τ_j: gender factor
- γ_{ij}: age:gender interaction

$i = 1$ (Elderly), 2 (Middle), 3 (Young) levels of age

$j = 1$ (Female), 2 (Male) levels of gender / R default is alphabetical

$k = 1, 2, 3, 4, 5, 6$ observations $\forall\, i, j$ $(n = 3 \times 2 \times 6 = 36)$

Constraints (sum coding or 1/0/-1 indicator variables)

$$\sum_i \alpha_i = 0 \qquad \alpha_1 + \alpha_2 + \alpha_3 = 0 \qquad \alpha_3 = -(\alpha_1 + \alpha_2)$$

$$\sum_j \tau_j = 0 \qquad \tau_1 + \tau_2 = 0 \qquad \tau_2 = -\tau_1$$

$$\sum_i \sum_j \gamma_{ij} = 0 \Rightarrow \sum_i \gamma_{ij} = 0 \,,\; \sum_j \gamma_{ij} = 0$$

gender

age	1 = F	2 = M	
1 = F	γ_{11}	$-\gamma_{11}$	0
2 = M	γ_{21}	$-\gamma_{21}$	0
3 = Y	$-(\gamma_{11}+\gamma_{21})$	$(\gamma_{11}+\gamma_{21})$	0
	0	0	0

The R default for sum contrasts is to use the last category in alphabetic order as the "reference" category

cf Treatment contrasts (0/1 or dummy coding)
R default is the first category in alphabetic order

$$\alpha_1 = 0 \,, \quad \tau_1 = 0 = \tau_{\text{Female}}$$

$= \alpha_{\text{Elderly}}$

gender

age		
	0	0
	0	γ_{22}
	0	γ_{32}

Simple ANCOVA (Analysis of Covariance) Model

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 X_{ij} + \gamma_i X_{ij} + \varepsilon_{ij}$$

birthweight

Assumptions: $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

α_i represent the deviations due to the levels $i = 1, 2$ {"N", "Y"} of the factor smoker

X_{ij} represents the covariate gestation

γ_i like α_i represents the deviations (this time with the slope rather than the intercept) due to the levels of smoker

$j = 1, \ldots$ # "N" observations ; $1 \ldots$ # "Y" observations

Constraints: $\alpha_N = 0$; $\gamma_N = 0$

For Smoker = "N"

model is

$$\hat{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 X_{ij}$$

For Smoker = "Y" (using 0/1 dummy variables)

model is

$$\hat{Y}_{ij} = (\hat{\beta}_0 + \alpha_Y) + (\hat{\beta}_1 + \gamma_Y) X_{ij}$$