18 Some More Miscellaneous ANOVA Topics

18.1 Two-Way Analysis of Variance without Replicates

18.1 Note: Recall the 2-way Analysis ANOVA table in Note 17.11:

source	sum of squares	df	mean squares	test statistic
Factor A	SS_A	$a-1$	$\text{MS}_A = \text{SS}_A/(a-1)$	MS_A/MS_E
Factor B	SS_B	$b-1$	$\text{MS}_B = \text{SS}_B/(b-1)$	MS_B/MS_E
Interaction	SS_{AB}	$(a-1)(b-1)$	$\text{MS}_{AB} = \text{SS}_{AB}/((a-1)(b-1))$	$\text{MS}_{AB}/\text{MS}_E$
Error	SS_E	$ab(r-1)$	$\text{MS}_E = \text{SS}_E/(ab(r-1))$	
Total	SS_{TOTAL}	$abr-1$		

We assumed that we have r replicates per cell. However, for example in randomized block designs (see Example 15.5), this is usually not the case. If $r = 1$, then we have no degrees of freedom to estimate the error if we use an interaction term in the model! This makes sense, since the total degrees of freedom are $ab-1$, and we use $a-1+b-1+(a-1)(b-1) = ab-1$ degrees of freedom to estimate the main effects and the interaction. If the interaction is zero however, then the ANOVA table becomes:

source	sum of squares	df	mean squares	test statistic
Factor A	SS_A	$a-1$	$\text{MS}_A = \text{SS}_A/(a-1)$	MS_A/MS_E
Factor B	SS_B	$b-1$	$\text{MS}_B = \text{SS}_B/(b-1)$	MS_B/MS_E
Error	SS_E	$(a-1)(b-1)$	$\text{MS}_E = \text{SS}_E/((a-1)(b-1))$	
Total	SS_{TOTAL}	$ab-1$		

The expected mean squares for fixed, random and mixed effects (A fixed and B random) models are:

source	fixed effects	random effects	mixed effects
Factor A	$\sigma^2 + \frac{b}{a-1}\sum_i \alpha_i^2$	$\sigma^2 + b\sigma_A^2$	$\sigma^2 + \frac{b}{a-1}\sum_i \alpha_i^2$
Factor B	$\sigma^2 + \frac{a}{b-1}\sum_i \beta_i^2$	$\sigma^2 + a\sigma_B^2$	$\sigma^2 + a\sigma_B^2$
Error	σ^2	σ^2	σ^2

18.2 Example: When analyzing the data arising from the experiment that was run to compare the effects of in-patient and out-patient protocols on the lab measurements of resting metabolic rate in humans (Example 15.5), we assume the following model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ independent}.$$

Here α_i denotes the effect of the ith protocol and β_j denotes the effect for the jth subject. We are mostly interested in the effect for the three specific protocols in the experiment, and use the subject effect as blocking variable. Hence α is a fixed effect, and β is a random effect. Therefore we assume that $\sum_{i=1}^{a} \alpha_i = 0$, and $\beta_j \sim N(0, \sigma_B^2)$. Note that we also assumed that there is no protocol-patient interaction. This means that there is for example no patient for which the resting metabolic rate is different for one specific protocol only.

For the ANOVA table we get:

Source	Df	SS	MS	F	p
Protocol	2	0.024	0.012	0.140	0.870
Subject	8	23.243	2.905	34.740	< 0.001
Error	16	1.338	0.084		

We conclude that we can not detect a protocol effect, but that blocking was important. Judging from the Figure in Example 15.5, this makes a lot of sense!

18.3 Note: What if the assumption of no interaction is false? Substituting $r = 1$ into Table 17.12, we get:

source	fixed effects	random effects	mixed effects
Factor A	$\sigma^2 + \frac{b}{a-1}\sum_i \alpha_i^2$	$\sigma^2 + \sigma_{AB}^2 + b\sigma_A^2$	$\sigma^2 + \sigma_{AB}^2 + \frac{b}{a-1}\sum_i \alpha_i^2$
Factor B	$\sigma^2 + \frac{a}{b-1}\sum_i \beta_i^2$	$\sigma^2 + \sigma_{AB}^2 + a\sigma_B^2$	$\sigma^2 + a\sigma_B^2$
Interaction	$\sigma^2 + \frac{1}{(a-1)(b-1)}\sum_i \sum_j \gamma_{ij}^2$	$\sigma^2 + \sigma_{AB}^2$	$\sigma^2 + \sigma_{AB}^2$
Error	σ^2	σ^2	σ^2

Hence, even if we falsely assume that there is no interaction between A and B, we still get valid F-tests in the random effects model, and we get a valid F-test for the factor of interest in the mixed effects model!

18.2 Analysis of Covariance

The Analysis of Covariance (ANCOVA) model combines qualitative factors as used in an ANOVA with quantitative predictors as used in a standard linear regression analysis:

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

with $i = 1, \ldots, p,\ j = 1, \ldots, n_i$, and $\varepsilon_{ij} \sim N(0, \sigma^2)$ independent. This model is useful to adjust for a confounding variable (the covariate z) when making group comparisons in observational studies. It also allows for more precision in group comparisons in randomized studies.

18.4 Note: Hypothesis of interest are:

(a) $H_1 : \beta_i = \beta,\ \forall i.$ We first test if the relationship between Y and z is the same in all groups.

(b) $H_2 : \alpha_i = 0,\ \forall i.$ Assuming the slopes β_i are the same, we then test for group differences.

18.5 Note: Rewriting the model as

$$Y_{ij} = \mu + \alpha_i + \beta z_{ij} + \gamma_i z_{ij} + \varepsilon_{ij},$$

we see that H_1 implies no interaction between group and z (i. e. $\gamma_i = 0$), and if this is true, H_2 implies that there is no group effect (i. e. $\alpha_i = 0$).

18.6 Theorem: To obtain the least squares estimates, we do the following:

(a) For the full model, fit simple linear regression models separately across groups:

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})(z_{ij} - \bar{z}_{i\cdot})}{\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{i\cdot})^2} \qquad \text{and} \qquad \hat{\mu}_i = \bar{Y}_{i\cdot} - \hat{\beta}_i \bar{z}_{i\cdot} \qquad \text{for} \qquad i = 1, \ldots, p.$$

(b) Under H_1, the common slope estimate is obtained by pooling the sums of squares across groups:

$$\hat{\beta} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})(z_{ij} - \bar{z}_{i\cdot})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{i\cdot})^2} \qquad \text{and} \qquad \hat{\mu}_i = \bar{Y}_{i\cdot} - \hat{\beta} \bar{z}_{i\cdot} \qquad \text{for} \qquad i = 1, \ldots, p.$$

(c) Under H_2, the group means are replaced by the overall mean:

$$\hat{\beta} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})(z_{ij} - \bar{z}_{\cdot\cdot})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{\cdot\cdot})^2} \qquad \text{and} \qquad \hat{\mu} = \bar{Y}_{\cdot\cdot} - \hat{\beta} \bar{z}_{\cdot\cdot}.$$

18.7 Theorem: The residual sums of squares are:

(a) $RSS = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2 - \sum_{i=1}^{p}\hat{\beta}_i^2\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{i\cdot})^2$.

(b) $RSS_{H_1} = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2 - \hat{\beta}^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{i\cdot})^2$.

(c) $RSS_{H_2} = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 - \hat{\beta}^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{\cdot\cdot})^2$.

18.8 Theorem: Let $n = \sum_{i=1}^{p} n_i$.

(a) If H_1 is true, then

$$F_1 = \frac{(RSS_{H_1} - RSS)/(p-1)}{RSS/(n-2p)} \sim F_{p-1,\, n-2p}.$$

(b) If H_2 is true, then

$$F_2 = \frac{(RSS_{H_2} - RSS_{H_1})/(p-1)}{RSS_{H_1}/(n-p-1)} \sim F_{p-1,\, n-p-1}.$$

(c) If both H_1 and H_2 are true, then

$$F_2 = \frac{(RSS_{H_2} - RSS)/(2p-2)}{RSS/(n-2p)} \sim F_{2p-2,\, n-2p}.$$