Two serve and project - in more ways: two-way analysis of variance with random effects

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

March 13, 2024

Outline for today

- Two-way ANOVA using orthogonal projections

Two-way ANOVA

Consider two factors: T (treatment) and P (plot) with number of levels d_T and d_P. Moreover let $P \times T$ be the cross-factor (has $d_P d_T$ levels - one for each combination of levels of P and T).

Assume $P \times T$ is balanced with $n_{P \times T} = m$ observations for each level. Then P and T balanced too with numbers of observations $n_P = md_T$ and $n_T = md_P$ for each level.

Model with random P and $P \times T$ effects (e.g. to account for soil variation)

$$y_{ptr} = \xi + \beta_t + U_p + U_{pt} + \epsilon_{ptr}$$

$p = 1, \dots, d_P,\ t = 1, \dots, d_T,\ r = 1, \dots, m$

In vector form:

$$y = \mu + Z_P U_P + Z_{P \times T} U_{P \times T} + \epsilon$$

where $\mu \in L_T = \text{span} X_T$.

Similar to one-way anova define:

$$V_0 = L_0 \quad V_T = L_T \ominus V_0 \quad V_P = L_P \ominus V_0$$

Since $P \times T$ balanced it follows that $P_T P_P = P_P P_T = P_0$ which implies $Q_T Q_p = 0$ where $Q_T = P_T - P_0$ and $Q_P = P_P - P_0$. Hence V_T and V_P are orthogonal and $L_P + L_T = V_0 \oplus V_P \oplus V_T$.

Define further

$$V_{P\times T} = L_{P\times T} \ominus (L_P + L_T) \quad V_I = \mathbb{R}^n \ominus L_{P\times T}$$

Orthogonal decomposition:

$$\mathbb{R}^n = V_0 \oplus V_P \oplus V_T \oplus V_{P\times T} \oplus V_I$$

Dimensions of 'V' spaces: $f_0 = 1$, $f_P = d_P - 1$, $f_T = d_T - 1$, $f_{P\times T} = d_P d_T - d_P - d_T + 1 = (d_P - 1)(d_T - 1)$, $f_I = n - d_P d_T$.

Orthogonal projections on 'V' spaces: $Q_0 = P_0$, $Q_P = P_P - Q_0$, $Q_T = P_T - Q_0$, $Q_{P\times T} = P_{P\times T} - Q_P - Q_T - Q_0$ and $Q_I = I - P_{P\times T}$.

Covariance structure:

$$\begin{aligned}\mathbb{C}\mathrm{ov}\, Y &= \sigma_P^2 n_P P_P + \sigma_{P\times T}^2 n_{P\times T} P_{P\times T} + \sigma^2 I \\ &= \lambda_P \tilde{Q}_P + \lambda_{P\times T} \tilde{Q}_{P\times T} + \lambda_I \tilde{Q}_I\end{aligned}$$

where

$$\begin{aligned}\lambda_I &= \sigma^2 \\ \lambda_{P\times T} &= \sigma^2 + n_{P\times T}\sigma_{P\times T}^2 \\ \lambda_P &= \sigma^2 + n_{P\times T}\sigma_{P\times T}^2 + n_P\sigma_P^2\end{aligned}$$

and

$$\tilde{Q}_P = Q_0 + Q_P = P_P \quad \tilde{Q}_{P\times T} = Q_{P\times T} + Q_T \quad \tilde{Q}_I = Q_I$$

Interpretation of $V_{P\times T}$ for a fixed effects two-way ANOVA

Suppose that P and T are both systematic factors. Then the ordinary two-way ANOVA is

$$y_{ptr} = \xi + \alpha_p + \beta_t + \gamma_{pt} + \epsilon_{ptr}.$$

Then $L_{P\times T}$ is the space for the mean vector $\mu = (\xi + \alpha_p + \beta_t + \gamma_{pt})_{ptr}$.

Similarly, $L_P + L_T$ is the space for the mean vector $\mu = (\xi + \alpha_p + \beta_t)_{ptr}$ in case of the additive model where the $\gamma_{pt} = 0$ (no interaction).

Hence $V_{P\times T} = L_{P\times T} \ominus (L_P + L_T)$ is the 'interaction space'.

'$\tilde{Q}$'s are orthogonal projections on

$$\tilde{V}_P = V_0 + V_P \quad \tilde{V}_{P\times T} = V_{P\times T} + V_T \quad \tilde{V}_I = V_I$$

This corresponds to orthogonal decomposition

$$\mathbb{R}^n = \tilde{V}_P \oplus \tilde{V}_{P\times T} \oplus \tilde{V}_I$$

based on model without systematic effects ('V'-spaces aggregated into $\tilde{V}$-spaces).

Orthogonal decomposition of data vector:

$$
\begin{aligned}
Y &= Q_0 Y + Q_P Y + Q_T Y + Q_{P\times T} Y + Q_I Y \\
&= \tilde{Q}_P Y + \tilde{Q}_{P\times T} Y + \tilde{Q}_I Y
\end{aligned}
$$

Decomposition of mean vector $\mu \in L_T = V_0 \oplus V_T$:

$$
\mu = \tilde{Q}_P \mu + \tilde{Q}_{P\times T} \mu + \tilde{Q}_I \mu = Q_0 \mu + Q_T \mu + 0 = \mu_0 + \mu_T
$$

where $\mu_0 \in L_0$ and $\mu_T \in V_T$.

Note: one-to-one correspondence between μ and (μ_0,μ_T) (reparametrization)

As for one-way anova Y is decomposed into independent normal vectors

$$\tilde{Q}_P Y \sim N(\mu_0, \lambda_P \tilde{Q}_P) \quad \tilde{Q}_{P\times T} Y \sim N(\mu_T, \lambda_{P\times T} \tilde{Q}_{P\times T})$$
$$\tilde{Q}_I Y \sim N(0, \lambda_I \tilde{Q}_I)$$

Density for Y factorizes into product of 'densities' for $\tilde{Q}_P Y$, $\tilde{Q}_{P\times T} Y$ and $\tilde{Q}_I Y$.

Hence

$$\hat{\mu}_0 = P_0 \tilde{Q}_P Y = P_0 Y = \bar{Y}_{..} 1_n \quad \hat{\mu}_T = Q_T \tilde{Q}_{P\times T} Y = Q_T Y$$
$$\hat{\mu} = \hat{\mu}_0 + \hat{\mu}_T = P_T Y \quad \hat{\lambda}_P = \|\tilde{Q}_P Y - P_0 Y\|^2 / d_P = \|Q_P Y\|^2 / d_P$$
$$\hat{\lambda}_{P\times T} = \|\tilde{Q}_{P\times T} Y - Q_T Y\|^2 / (d_P d_T - d_P) = \|Q_{P\times T} Y\|^2 / (d_P d_T - d_P)$$
$$\hat{\lambda}_I = \hat{\sigma}^2 = \|Q_I Y\|^2 / (n - d_P d_T)$$

NB: since V_P and $V_{P\times T}$ have dimensions $f_P = d_P - 1$ and $f_{P\times T} = (d_P - 1)(d_T - 1)$ we often use these denominators instead of d_P and $d_P d_T - d_P$ for λ_P and $\lambda_{P\times T}$ (REML).

Note: MLE for μ is identical to MLE in the linear model with $\mu \in L_T$ and no random effects.

Strata

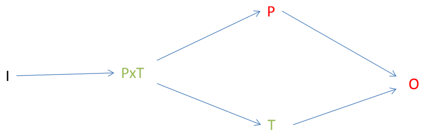
Crucial common point for one- and two-way anova models considered: we are able to decompose covariance matrix as linear combination of scaled projections on orthogonal '$\tilde{V}$'-subspaces where coefficients are in one-to-one correspondence with the variance components (the variances of random effects). This enables factorization of likelihood !

Fixed factors assigned to *strata* corresponding to '$\tilde{V}$' spaces for the random factors

Rule: fixed factor F belongs to B strata (B random factor) if B is the coarsest random factor which is finer than F (assuming that we work with designs where this rule gives unique allocation of each fixed factor to one random factor).

NB: F is finer than G (or G coarser than F) if levels of G can be obtained by merging levels of F. We then write $G \preceq F$.

Structure diagram



Arrow from F to G if G is coarser than F and there is no intermediate factor which is coarser than F and finer than G.

Note: three strata (black, green and red) corresponding to ‘random’ factors I, $P \times T$ and P.

NB: here we can not have both P and T random unless O random too (if O systematic no unique allocation to random factor strata). Also want to respect hierarchical principle so $P \times T$ can not be fixed if P random.

Relation to 'old school ANOVA'

By orthogonal decomposition

$$Y = Q_0 Y + Q_P Y + Q_T Y + Q_{P\times T} Y + Q_I Y$$

and Pythagoras,

$$\|Y - Q_0 Y\|^2 = \|Q_P Y\|^2 + \|Q_T Y\|^2 + \|Q_{P\times T} Y\|^2 + \|Q_I Y\|^2$$

With terminology SSF$=\|Q_F Y\|^2$ we get

$$\sum_{ptr} (Y_{ptr} - \bar{Y}_{\dots})^2 = SSP + SST + SSPT + SSI$$

Here, for example, SSP$=\sum_{ptr}(\bar{Y}_{p\cdot\cdot} - \bar{Y}_{\dots})^2$ and SSPT$=$ $\sum_{ptr}(\bar{Y}_{pt\cdot} - \bar{Y}_{p\cdot\cdot} - \bar{Y}_{\cdot t\cdot} + \bar{Y}_{\dots})^2$.

Sorry: in relation with one-way ANOVA I called left hand side SST (T for total) and SSI was SSE

Note also: $Q_T Y = (\bar{Y}_{\cdot t \cdot} - \bar{Y}_{\cdots})_{ptr}$.

This is the part of the mean vector that corresponds to sum to zero constraint: with

$$\hat{\beta}_t = \bar{Y}_{\cdot t \cdot} - \bar{Y}_{\cdots}$$

we have

$$\sum_t \hat{\beta}_t = 0$$

In other words: with $\mu_T = Q_T \mu \in V_T$ we have μ_T and $1_n \in Q_0$ orthogonal so

$$1_n^T \mu_T = 0 \quad \text{and} \quad 1_n^T \hat{\mu}_T = 0$$

Paper pulp example: two-way analysis of variance without treatment effect

P: card board with $d_P = 5$. T: position within cardboard (not a treatment) with $d_T = 4$. Number of replications $m = 4$.

Model:

$$y_{ptr} = \xi + U_p + U_{pt} + \epsilon_{ptr}$$

Variances σ^2_P, $\sigma^2_{P \times T}$ and σ^2

We can use exactly the same orthogonal decomposition as before. Only difference is that $Q_T \mu = Q_T 1_n \xi = 0$ which means $\tilde{Q}_{P \times T} Y \sim N(0, \lambda_{P \times T} \tilde{Q}_{P \times T})$ and REML and ML estimates for $\lambda_{P \times T}$ coincide and are equal to

$$\hat{\lambda}_{P \times T} = \|\tilde{Q}_{P \times T} Y\|^2 / (d_P d_T - d_P)$$

NB: in ANOVA table (next slide), SSP$=\|Q_P Y\|^2$, SST$=\|Q_T Y\|^2$, SSPT$=\|Q_{P \times T} Y\|^2$ and SSE$=\|Q_I Y\|^2$.

ANOVA table

```
 Analysis of Variance Table

Response: Reflektans
                       Df  Sum Sq  Mean Sq  F value    Pr(>F)
factor(Sted)            3 0.03600 0.011999  188.981 < 2.2e-16 ***
                                                            #SST
factor(Pap.nr.)         4 1.07520 0.268800 4233.472 < 2.2e-16 ***
                                                            #SSP
factor(Sted):factor(Pap.nr.)
                       12 0.02168 0.001807   28.452 < 2.2e-16 ***
                                                            #SSPT
Residuals              60 0.00381 0.000063
                                                            #SSE
```

$\hat{\sigma}^2 = 0.00006$ $\hat{\lambda}_{P\times T} = (0.036 + 0.02168)/15 = 0.00385$
$\hat{\lambda}_P = 1.0752/5 = 0.21504$ (or $\hat{\lambda}_P = 1.0752/4 = 0.0072 = 0.2688$).

By relations

$$\lambda_{P\times T} = \sigma^2 + n_{P\times T}\sigma^2_{P\times T}$$
$$\lambda_P = \sigma^2 + n_{P\times T}\sigma^2_{P\times T} + n_P\sigma^2_P$$

we obtain $\hat{\sigma}^2_P = (0.21504 - 0.00385)/16 = 0.0132$
$\hat{\sigma}^2_{P\times T} = (0.00385 - 0.00006)/4 = 0.0009475$ (or
$\hat{\sigma}^2_P = (0.2688 - 0.00385)/16 = 0.01655937$)

Recall: balanced design required - and difficult to remember rules for calculating variance components.

Using lmer and ML

```
> out2=lmer(Reflektans~(1|Sted:Pap.nr.)+(1|Pap.nr.),REML=F)
> summary(out2)

Random effects:
 Groups       Name        Variance  Std.Dev.
 Sted:Pap.nr. (Intercept) 9.454e-04 0.030747
 Pap.nr.      (Intercept) 1.320e-02 0.114890
 Residual                 6.349e-05 0.007968
Number of obs: 80, groups:  Sted:Pap.nr., 20; Pap.nr., 5
```

Using lmer and REML

```
> out2=lmer(Reflektans~(1|Sted:Pap.nr.)+(1|Pap.nr.))
> summary(out2)

Random effects:
 Groups       Name        Variance  Std.Dev.
 Sted:Pap.nr. (Intercept) 9.455e-04 0.030749
 Pap.nr.      (Intercept) 1.657e-02 0.128717
 Residual                 6.349e-05 0.007968
Number of obs: 80, groups:  Sted:Pap.nr., 20; Pap.nr., 5
```

Note REML and ML estimates for σ^2 and $\sigma^2_{P\times T}$ coincide (up to rounding error)

Explanation of `Reflektans~(1|Sted:Pap.nr.)+(1|Pap.nr.))`:

- no fixed formula: intercept always included as default
- `(1|Sted:Pap.nr.)` random intercepts for groups identified by variable `StedPap.nr.`
- `(1|Pap.nr.)` random intercepts for groups identified by variable `Pap.nr.`
- random effects specified by different terms independent.

A more complicated example: gene-expression

Gene (DNA string) composed of substrings (exons) which may be more or less expressed according to treatment.

Expression measured as intensities on micro-array (chip). One chip pr. patient-treatment.

Factors: E (exon 8 levels), P (patient, 10 levels), T (treatment, 2 levels)

Y: vector of intensities (how much is exon expressed).

Model:

$$y_{ept} = \xi + \alpha_e + \beta_t + \gamma_{et} + U_p + U_{pt} + \epsilon_{ept}$$

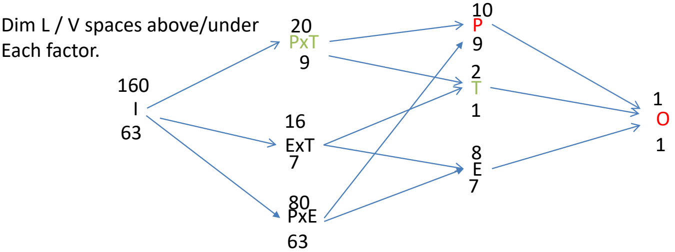
U_{pt} and U_p random chip and patient effects.

Main question: are exons differentially expressed - i.e. are $\gamma_{et} \neq 0$ or not (we will return to this question later)

Note: $P \times E \times T = I$ is balanced so every sub-factor (E, $E \times T$ etc.) is balanced. This implies orthogonal decomposition (see a following slide):

$$\mathbb{R}^{160} = V_0 \oplus V_P \oplus V_E \oplus V_T \oplus V_{P\times T} \oplus V_{E\times T} \oplus V_{P\times E} \oplus V_I$$

Structure diagram with random factors $P, P \times T, I$:



Decomposition with respect to random factors:

$$\mathbb{R}^{160} = \tilde{V}_P \oplus \tilde{V}_{P\times T} \oplus \tilde{V}_I$$

where $\tilde{V}_P = V_0 \oplus V_P$, $\tilde{V}_{P\times T} = V_T \oplus V_{P\times T}$ and $\tilde{V}_I = V_E \oplus V_{E\times T} \oplus V_{P\times E} \oplus V_I$.

Decomposition of $\mathbb{C}\mathrm{ov}\, Y$:

$$\mathbb{C}\mathrm{ov}\, Y = n_P \sigma_P^2 P_P + n_{P\times T} \sigma_{P\times T}^2 P_{P\times T} + \sigma^2 I = \lambda_P \tilde{Q}_P + \lambda_{P\times T} \tilde{Q}_{P\times T} + \lambda_I \tilde{Q}_I$$

Decomposition of μ:

$$\mu = \tilde{Q}_P \mu + \tilde{Q}_{P\times T} \mu + \tilde{Q}_I \mu = Q_0 \mu + Q_T \mu + Q_E \mu + Q_{E\times T} \mu$$

As before decomposition of Y into independent Gaussian vectors:

$$\tilde{Q}_P Y \sim N(Q_0 \mu, \lambda_P \tilde{Q}_P) \quad \tilde{Q}_{P\times T} Y \sim N(Q_T \mu, \lambda_{P\times T} \tilde{Q}_{P\times T})$$

$$\tilde{Q}_I Y \sim N\big((Q_E + Q_{E\times T})\mu, \lambda_I \tilde{Q}_I\big)$$

Orthogonality of 'V'-spaces

We use results of exercise 7 to deduce that

$$P_{P\times E}P_{E\times T} = P_E \quad Q_P P_{E\times T} = P_P P_{E\times T} - P_0 P_{E\times T} = P_0 - P_0 = 0$$

The second equality shows that V_P and $V_{E\times T}$ are orthogonal.

Further:

$$Q_{P\times E} = P_{P\times E} - Q_P - Q_E - Q_0 \quad Q_{E\times T} = P_{E\times T} - Q_E - Q_T - Q_0$$

Using the result in the upper equation and pairwise orthogonality of Q_P, Q_E, Q_T, Q_0 we get

$$Q_{P\times E}Q_{E\times T} = 0$$

Thus $V_{P\times E}$ and $V_{E\times T}$ are orthogonal too.

Orthogonality of $V_{P\times T}$, V_P and V_T was shown for two-way ANOVA.

Proceeding this way all 'V' spaces orthogonal.

Anova table:

```
> fit1=lm(intensity~treat*factor(exon)+factor(patient)+
                  factor(patient):treat,data=gene1)
> anova(fit1)
Analysis of Variance Table

Response: intensity
                      Df  Sum Sq Mean Sq  F value    Pr(>F
treat                 1    3.242   3.242  14.4796 0.0002199
factor(exon)          7  254.343  36.335 162.2852 < 2.2e-16
factor(patient)       9   15.405   1.712   7.6449 6.703e-09
treat:factor(exon)    7    2.238   0.320   1.4278 0.1998234
treat:factor(patient) 9    8.190   0.910   4.0643 0.0001345
Residuals             126 28.211   0.224
```

$\hat{\sigma}^2 = \hat{\lambda}_I = 0.224$ $\tilde{\lambda}_{P\times T} = 8.19/9 = 0.91$ $\tilde{\lambda}_P = 15.405/9 = 1.712.$

$\hat{\sigma}^2_{P\times T} = (0.91 - 0.224)/8 = 0.08575$
$\hat{\sigma}^2_P = (1.712 - 0.91)/16 = 0.050125$

With aov()

```
> fit1=aov(intensity~treatment*factor(exon)+
        Error(factor(patient)+factor(patient):treatment),data=ge
> summary(fit1)

Error: factor(patient)
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   15.4   1.712

Error: factor(patient):treatment
          Df Sum Sq Mean Sq F value Pr(>F)
treatment  1  3.242   3.242   3.563 0.0917 .
Residuals  9  8.190   0.910
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
                        Df Sum Sq Mean Sq F value Pr(>F)
factor(exon)             7 254.34   36.33 162.285 <2e-16 ***
treatment:factor(exon)   7   2.24    0.32   1.428    0.2
Residuals              126  28.21    0.22
```

Using lmer:

```
> fit1=lmer(intensity~treat*factor(exon)+
                   (1|patient)+(1|patient:treatment),data=g
> summary(fit1)
Random effects:
 Groups            Name        Variance Std.Dev.
 patient:treatment (Intercept) 0.08577  0.2929
 patient           (Intercept) 0.05011  0.2239
 Residual                      0.22389  0.4732
Number of obs: 160, groups:  patient:treatment, 20; patient
```

K-way ANOVA

Assume K factors $F_1, \ldots, F_K$ so that $F_1 \times \ldots \times F_K$ balanced.

Let $\mathcal{D}$ be the set of $0, I, F_1, \ldots, F_K$ and all cross-combinations of $F_1, \ldots, F_K$.

Then, in analogy with three-way (excs 7), all 'V' spaces orthogonal.

Let $\mathcal{B} \subseteq \mathcal{D}$ be the set of factors with random effects.

We then have

$$\Sigma = \sum_{B \in \mathcal{B}} \sigma_B^2 n_B P_B \quad \text{and} \quad P_B = \sum_{F \in \mathcal{D}: F \preceq B} Q_F$$

and we want

$$P_B = \sum_{B' \in \mathcal{B}: B' \preceq B} \tilde{Q}_{B'} \tag{1}$$

where $\tilde{Q}_B$ orthogonal projections on some spaces $\tilde{V}_B$ where $\mathbb{R}^n = \otimes_{B \in \mathcal{B}} \tilde{V}_B$.

K-way ANOVA continued

Given (1) we have required decomposition of Σ into sum of scaled orthogonal projections:

$$\Sigma = \sum_{B \in \mathcal{B}} \sigma_B^2 n_B P_B = \sum_{B' \in \mathcal{B}} \lambda_{B'} \tilde{Q}_{B'} \quad \text{where} \quad \lambda_{B'} = \sum_{B \in \mathcal{B} : B' \preceq B} n_B \sigma_B^2 .$$

Thus we can again obtain decomposition Y of into independent $\tilde{Y}_B = \tilde{Q}_B Y$, $B \in \mathcal{B}$.

Hence parameter estimates can easily be derived by analogy with results for linear normal model.

K-way ANOVA continued

A sufficient (and necessary) condition for (1) is that: for all $F \in \mathcal{D}$ there exists a $B \in \mathcal{B}$ such that $F \preceq B$ and $B \preceq B'$ for all other B' with $F \preceq B'$.

(this must be checked for a given model. Note this implies $I \in \mathcal{B}$ and that $B = B(F)$ is unique).

Then we define

$$\tilde{V}_B = \sum_{F \in \mathcal{D}: B(F) = B} V_F \quad \text{and} \quad \tilde{Q}_B = \sum_{F \in \mathcal{D}: B(F) = B} Q_F$$

Each $F \in \mathcal{D}$ belongs to precisely one $\tilde{V}_B$ so $\mathbb{R}^n = \otimes_{B \in \mathcal{B}} \tilde{V}_B$.

See also more details in note "Analysis of variance using orthogonal projections".

Even more general set-up can be found in Jesper Møller: Centrale statistiske modeller og likelihood baserede metoder.

Exercises

1. Check that $P_T P_P = P_0$.
2. Check that $L_T + L_P = V_0 \oplus V_P \oplus V_T$.
3. For a two-way balanced ANOVA, derive the decomposition of the covariance matrix of Y in terms of 'λ's and '$\tilde{Q}$'s.
4. Derive factorization of likelihood for balanced two-way.
5. Derive estimates of mean and variance parameters for balanced two-way.
6. Install the R-package `faraway` which contains the data set `penicillin`. The response variable is yield of penicillin for four different production processes (the 'treatment'). The raw material for the production comes in batches ('blends'). The four production processes were applied to each of the 5 blends. Fit anova models with production process as a fixed factor and blend as random factor. Try to use both the anova table and `lmer`.

7. In a balanced three-way design with factors F_1, F_2, and F_3 show that

$$P_{F_1 \times F_2} P_{F_2 \times F_3} = P_{F_2} \quad P_{F_1 \times F_2} P_{F_3} = P_0$$

Explain how this can be generalized to the following result:

$$P_{\prod_{i \in A} F_i} P_{\prod_{i \in B} F_i} = P_{\prod_{i \in A \cap B} F_i}$$

for a K-way balanced design with factors F_i, $i = 1, \ldots, K$ and $A, B \subseteq \{1, \ldots, K\}$ (taking $P_{\prod_{i \in \emptyset F_i}} = P_0$).