Lecture 8
Repeated Measures Analysis

DSA 8070 Multivariate Analysis
October 4 - October 8, 2021

Whitney Huang
Clemson University

Notes

Dog Experiment [Source: PSU STAT 505]

A completely randomized block design was carried out to determine the effects of 4 surgical treatments on coronary potassium in a group of 36 dogs. There are 9, 8, 9, and 10 dogs in each treatment group, respectively. Each dog was measured at four different time points (1, 5, 9, and 13 minutes) following one of four experimental treatments:

- Control - no surgical treatment is applied
- Extrinsic cardiac denervation immediately prior to treatment
- Bilateral thoracic sympathectomy and stellectomy 3 weeks prior to treatment
- Extrinsic cardiac denervation 3 weeks prior to treatment

We are looking at the treatment effect on the coronary sinus potassium levels

Notes

Notation of Approaches

Let Y_{ijk} be the potassium level for treatment i in dog j at time k:

- there are $a = 4$ treatments
- n_i dogs received treatment i (therefore, there are $n_1 + \cdots + n_a = 9 + 8 + 9 + 10 = 36$ dogs in total)
- $t = 4$, the number of observations over time

Approaches

- Split-plot ANOVA
- MANOVA
- Mixed Models

Notes

Approach 1: Split-plot ANOVA

Model: $Y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$,
where

- α_i: effect of treatment i
- $\delta_{j(i)}$: random effect of dog j receiving treatment i
- β_k: effect of time k
- $(\alpha\beta)_{ik}$: treatment by time interaction
- ε_{ijk}: random error

Assumptions:

- $\varepsilon_{ijk} \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma^2_\varepsilon)$
- $\delta_{j(i)} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2_\delta)$
- β_k does not depend on the dog $\Rightarrow$ no time by dog interaction

Notes

Split-plot ANOVA Table

Source	df	MS	F
Trt	$a-1$	$\text{MS}_{trt} = \frac{\text{SS}_{trt}}{a-1}$	$F = \frac{\text{MS}_{trt}}{\text{MS}_{error_1}}$
Error 1	$N-a$	$\text{MS}_{error_1} = \frac{\text{SS}_{error_1}}{N-a}$	
Time	$t-1$	$\text{MS}_{time} = \frac{\text{SS}_{time}}{t-1}$	$F = \frac{\text{MS}_{time}}{\text{MS}_{error_2}}$
Trt × Time	$(a-1)(t-1)$	$\text{MS}_{trt\times time} = \frac{\text{SS}_{trt\times time}}{(a-1)(t-1)}$	$F = \frac{\text{MS}_{trt\times time}}{\text{MS}_{error_2}}$
Error 2	$(N-a)(t-1)$	$\text{MS}_{error_2} = \frac{\text{SS}_{error_2}}{(N-a)(t-1)}$	
Total	$Nt-1$		

Notes

Dog Experiment Split-plot Analysis

```
> library(lmerTest)
> fit <- lmer(Response ~ Treatment * Time + (1 | Dog_id), data = dat)
> anova(fit)
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Treatment      3.3396 1.11319     3    32  6.0038  0.002297 **
Time           6.2043 2.06811     3    96 11.1540 2.404e-06 ***
Treatment:Time 3.4397 0.38219     9    96  2.0613  0.040573 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

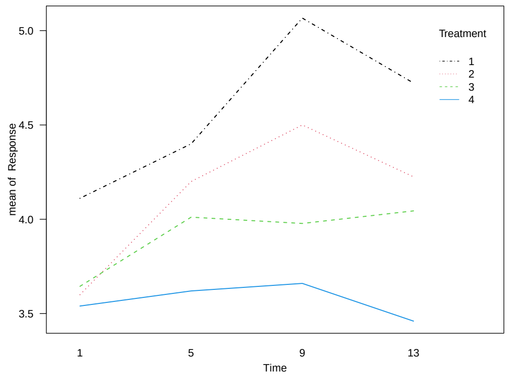
Hypothesis Tests:
We start with the interaction between treatment and time:

$$H_0 : (\alpha\beta)_{ik} = 0 \quad \forall i = 1, \cdots, a,\ k = 1, 2, \cdots, t$$

Result: We conclude the effect of treatment depends on time at $\alpha = 0.05$ level

Notes

Interaction Plot



Rejecting $H_0 : (\alpha\beta)_{ik} = 0$ means we reject the assumption of "parallelism"

Notes

Some Criticisms about the Split-ANOVA Approach

- The Split-plot ANOVA Approach assumes a constant correlation between any two observations from the same dog, that is, $\mathrm{Cor}(Y_{ijk}, Y_{ijk'}) = \frac{\sigma_\delta^2}{\sigma_\delta^2+\sigma_\varepsilon^2}$, this is the so-calle compound symmetry correlation structure
- This assumption is unlikely to be valid with repeated measurements over time as the correlation for two nearby time points is likely to be higher than the correlation for two far apart time points
- Next, we are going to take a multivariate approach (MANOVA) as an attempt to address this issue

Notes

Approach 2: MANOVA

Here we consider the observations over time from the same dog, dog j receiving treatment i as a single vector of interest

$$\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijt})^T,$$

and we will perform a one-way MANOVA

Assumptions:

- Dogs receiving treatment i have common mean vector $\boldsymbol{\mu}_i$
- All dogs have common covariance matrix $\boldsymbol{\Sigma}$
- Data from different dogs are independently sampled
- Data are multivariate normally distributed

Notes

Dog Experiment MANOVA Analysis

```
> dat <- read.table("dog1.txt")
> out <- manova(cbind(V3, V4, V5, V6) ~ as.factor(V1), data = dat)
> summary(out, test = "Wilks")
              Df   Wilks approx F num Df den Df  Pr(>F)
as.factor(V1)  3 0.48452    2.022     12 77.018 0.03316 *
Residuals     32
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: There are significant differences between at least one pair of treatments in at least one measurement of time

Criticism: MANOVA makes no assumptions regarding the temporal correlation structure, and hence, may be overparameterized leading to poor parameter estimates

Notes

Approach 3: Mixed Model Analysis

Main idea: Split-plot makes a too restrictive assumption while MANOVA makes no assumptions regarding the temporal correlation structure. The mixed model approach allows us to model the temporal correlation involving a limited number of parameters.

Model: $Y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$.

Assumptions:

- $\varepsilon_{j(ik)} \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma^2_\varepsilon)$
- $\delta_{j(i)} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2_\delta)$
- The correlation between the errors for the same dog depends only on the difference in observation time points: $|k - k'|$, e.g., $\mathrm{Cor}(Y_{ijk}, Y_{Y_{ijk'}}) = \rho^{|k-k'|}$ (Autoregressive with order 1)

Notes

Dog Experiment Mixed Model Analysis

```
> library(nlme)
> fit1 = gls(Response ~ Treatment * Time,
+           correlation = corCompSymm(form = ~ 1 | Dog_id), data = dat2)
> fit2 = gls(Response ~ Treatment * Time,
+           correlation = corAR1(form = ~ 1 | Dog_id), data = dat2)
> anova(fit1, fit2)
     Model df      AIC      BIC    logLik
fit1     1 18 275.8063 327.1429 -119.9032
fit2     2 18 277.5811 328.9177 -120.7906
```

Results:

- Based on both AIC/BIC, having an AR(1) does not necessarily improve the model fit (in this data)
- However, having the option of modeling repeated measurement error structure can be useful in general as it provides additional modeling choices

Notes