Universität Zürich[UZH]

Prof. Dr. Reinhard Furrer
Department Mathematical Modeling and Machine Learning (DM^{3}L)
Applied Statistics group
https://dm3l.uzh.ch/research

Dr. Bernadetta Tarigan
Office for Animal Welfare and 3Rs
Biostatistics Consulting
www.tierschutz.uzh.ch

Contact: www.math.uzh.ch/animalwelfare

Sample Size Calculation for one-way ANOVA with Repeated Measures in Time

B. Tarigan & K. Cherneva & G. Asthakov

Compiled on May 20, 2025 (v2.3)

This document is a companion to our statistical methods consultation service, mainly –but not limited– for animal research. Thus, the described methods/approaches here can be used and generalized for clinical research and all other fields of research.

Contents

1 Introduction **2**

2 The model **2**

3 Simulation for sample size calculation **3**

4 Statistical analysis methods **4**

5 Generalized Estimating Equation (GEE) **5**

1 Introduction

We consider an experimental study with repeated measures, where the subjects are the experimental units of a Completely Randomized (CR) design with one treatment factor and we take multiple measurements (of the continuous primary outcome of interest) on each subject over a period of time. The objective of this type of study is typically to compare not only the effect of the treatment groups but also the trend over time. CR design with one treatment factor means that all the treatment levels (*complete*) are assigned to the experimental units at *random*. Thus each of the unit has equal probability of receiving any treatment level and is independent of the other units.

Our primary interest is to calculate sample size in such design with power analysis. The sample size is the number of experimental units that are needed in each of the treatment level (the number of homogeneous replicates). For this purpose we first view such design formally within the framework of Split-Plot ANOVA. It is not because ANOVA is the best way to analyze data resulting from such design, but because ANOVA is the best way to think of your data and to plan your designs (see Casella, 2010, page 1). Using this Split-Plot design framework we can identify the parameters involved in the model and the condition for the model to hold. The sample size calculation is based in simulating data under the model of Split-Plot design, given some estimates of the parameters and the condition.

We also discuss some methods to analyze such data of repeated measurements, apart from Split-Plot ANOVA. In the last section we mention another method of calculating the sample size based on Generalized Estimating Equation (GEE) approach.

2 The model

We view such design within the framework of ANOVA. That is, as a variant of **Split Plot** model

$$y_{ijk} = \mu + \tau_i + \epsilon_{ij} + \gamma_k + (\tau\gamma)_{ik} + \delta_{ijk} \tag{1}$$

where τ_i's are the treatment effects, γ_k's are the time effects and $(\tau\gamma)_{ik}$'s are the interaction effects. There are two sources of error of the subject j. The first source is ϵ_{ij}, the subject error due to the treatment level i (the whole plot error). The second source is δ_{ijk}, the subject error due to the treatment level i and the time level k (the split plot error). Both errors are assumed normal distributed: $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and $\delta_{ijk} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\delta^2)$. That is, the variability across the treatment level i and the variability across the time point k are assumed to be constant and identical, respectively. Moreover, both random effects are assumed to be mutually independent. Our notation and exposition follows Casella, 2010, Section 5.2, page 175.

Let Σ denote the variance-covariance matrix of size txt of the time factor with t levels. Let μ_i denote the mean vector of size t of the treatment factor for each treatment level i, with $i = 1, \cdots, g$. Thus, the model assumptions are equivalent to assuming that for each treatment level i, over the time level $k = 1, \cdots, t$, the subject j follows multivariate normal distribution with mean vector μ_i and covariance matrix Σ. That is,

$$Y_{ij} \overset{iid}{\sim} \mathcal{N}(\mu_i, \Sigma) \ ,$$

for each subject $j = 1, \cdots, n$ where n is the sample size per treatment level. Note that here we consider a balanced design. That is, the sample size of each treatment level is the same, $n_i = n$, for all $i = 1, \cdots, g$.

We now look closer at the structure of the variance-covariance matrix Σ of the time factor, which is actually quite simple. From the model assumptions it follows that the covariance between two different time points k and k' of the same subject j within the same treatment level i is

$$\mathrm{Cov}(Y_{ijk}, Y_{ijk'}) = \mathrm{E}[(\epsilon_{ij} + \delta_{ijk})(\epsilon_{ij} + \delta_{ijk'})] = \sigma_\epsilon^2 \ ,$$

and that

$$\mathrm{Var}(Y_{ijk}) = \mathrm{Var}(\epsilon_{ij} + \delta_{ijk}) = \sigma_\epsilon^2 + \sigma_\delta^2 \ .$$

Thus, *the variances are identical* (the diagonal components) with value $\sigma_\epsilon^2 + \sigma_\delta^2$ and *the covariances are also identical* (the off-diagonal components) with value σ_ϵ^2. This property is called **Compound Symmetry** (CS) condition.

As a consequent of CS condition, clearly measurements of the same subject over the time are no longer independent but correlated. The correlation of the same subject in the same treatment level over any two different time points is strictly positive and identical with the value

$$\text{Corr}(Y_{ijk}, Y_{ijk'}) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_\delta^2} > 0 \ .$$

Let ρ a short hand for the above correlation value, $\rho := \sigma_\epsilon^2/(\sigma_\epsilon^2 + \sigma_\delta^2)$. Let R denote the correlation matrix of size txt of the time factor with t levels. Thus, under the condition, the correlation matrix is

$$R = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ & 1 & \rho & \dots & \rho \\ & & \ddots & \dots & \vdots \\ & & & 1 & \rho \\ & & & & 1 \end{bmatrix} \tag{2}$$

which is completely determined by the correlation value ρ and this is typically called as *exchangeable* correlation matrix.

The CS is a sufficient condition for conducting ANOVA (with the corresponding correct F tests therein), but it is not a necessary condition.

As a side note, we notice that the model (1) is also known as **effects model**. This model is equivalent to the so-called **cell means model**

$$y_{ijk} = \mu_{ik} + \varepsilon_{ijk} \tag{3}$$

with a different mean μ_{ik} for each combined level of the treatment factor and the time factor, where the experimental errors ε_{ijk} are assumed to be multivariate normal distributed over the time and independent due to the complete randomization of the experimental units to the treatment levels.

In ANOVA terminology, the model (1) is sometimes called as **mixed ANOVA** model, as it is a mixed of between-subject factor (the treatment factor with CRD) and the within-subject factor (the repeated times on the subjects). This model can also be seen, in regression terminology, as a **linear mixed effect model** where we have the combination of the fixed effects of the treatment factor, time factor and their interaction and the random effects of the two sources of error.

3 Simulation for sample size calculation

The calculation of sample size (n per treatment level) is based on simulated power in Split-Plot ANOVA on simulated measurements.

Measurements (i.e., primary outcome or response variable) is simulated under the multivariate normal distribution $Y_{ij} \overset{iid}{\sim} \mathcal{N}(\mu_i, \Sigma)$, for each treatment level i. Thus, the input are:

(i) the number of group g (i.e., the number of levels of the treatment factor),
(ii) the number of time repeated measurements t (i.e, the number of levels of the time factor),
(iii) the expected mean vectors μ_i for each group over the time (μ_i is a vector of size t for each i, $i = 1, \cdots, g$),
(iv) the variance vector of the time factor (of size t),
(v) the correlation value ρ of the exchangeable correlation matrix R of the time factor as in (2).

Given the correlation matrix R and the variance vector of the time repeated measurements, we can calculate the variance-covariance matrix Σ. Then, for a given sample size n, we simulate Y_{ij}, $j = 1, \cdots, n$ following the multivariate normal distribution $\mathcal{N}(\mu_i, \Sigma)$, for each treatment level i.

On this data set we conduct Split Plot ANOVA and extract the p-value for the treatment factor. We repeat the data set simulation many times, say at least $B = 1000$ times, to obtained B p-values. The simulated power for the given sample size n is defined as the proportion of how many p-values that are smaller than the desired significant level α. We repeat the whole procedure for different sample size n. Thus, for each sample size we simulated the power and then draw a power curve.

This procedure is available on our shiny R interactive application SampleSizeR (see Astakhov *et al.*, 2025).

As an **example**, consider a fictitious study investigating 3 diets on steers during 12 weeks growth trial. Body weights (kg) will be measured repeatedly every 3 weeks (at week 3, week 6, week 9 and week 12). Thus we have 4 time levels. We consider a balance CRD design. That is, the 3 diets are randomly assigned to the 3 groups with the same number of replicates. We assume that the steers have similar initial weight.

We further assume that *the correlation of the same steer decreases over time* is fixed with $\rho = 0.4$. and that *the variation in body weight is increasing* as weight increases over time. We consider $(36, 64, 100, 144)$ are the variance vector for the four time levels. As for the estimated means of the three diets over the four time points, we use

$$\mu = \begin{bmatrix} 250 & 280 & 300 & 330 \\ 260 & 295 & 315 & 345 \\ 255 & 295 & 320 & 355 \end{bmatrix} .$$

We run the simulation a few times, each time with these parameter values and significant level of $\alpha = 0.05$. We see that we need 7 steers per treatment group, to achieve a power about 0.80.

4 Statistical analysis methods

There are three possible ways to analyze such repeated-measures-in-time data:

(i) F tests for Split Plot Design that requires the CS condition,
(ii) F tests for CRD on the summary statistics of the time repeated measurements (e.g., the slopes over time),
(iii) linear mixed effects models.

For the first approach (i), in assessing the model assumption the CS condition can be relaxed into the so-called *sphericity* condition (aka Huynh-Feldt condition). Instead of assuming that the variances and covariances across the time levels are respectively identical, we assume that the variance of the *difference* between any two time levels is identical. That is, $\text{Var}(Y_{ijk} - Y_{ijk'})$ is a constant.

The CS condition implies the sphericity condition:

$$\begin{aligned} \text{Var}(Y_{ijk} - Y_{ijk'}) &= \text{Var}(Y_{ijk}) + \text{Var}(Y_{ijk'}) - 2\ \text{Cov}(Y_{ijk}, Y_{ijk'}) \\ &= (\sigma_\epsilon^2 + \sigma_\delta^2) + (\sigma_\epsilon^2 + \sigma_\delta^2) - 2\sigma_\epsilon^2 \\ &= 2\sigma_\delta^2 \ , \end{aligned}$$

for any two of time levels $k \neq k'$ with k and k' from $1, \cdots, t$. The correlation matrix under the sphericity condition is similar as the matrix R of the CS condition, that the correlation between any two time levels is positive and identical. The sphericity condition is typically tested with the Mauchly's test, which is provided in most of statistics software. We refer to Lawson, 2015, Section 9.5, page 366, for more details.

As we have mentioned in Introduction, the mixed ANOVA model is not the best way to analyze data resulting from such repeated measures in time. The second approach (ii) is another way to analyze such data. The repeated measures can be summarized into one measurement, such as an average or a slope of a regression line. Since there is now only one measurement per subject, the correlation problem disappears. A detailed example is shown in Casella, 2010, Example 5.17, page 217.

The third approach (iii), linear mixed effects models, is probably the most family of models that are often used in analyzing such data. Without discussing details, we note that this larger family is more advantageous over

the other two approaches. With this family, we can (a) analyze *unbalanced* design (e.g., different number of time repeats, different number of subject in each treatment level), (b) handle *missing data*, (c) treat the *time as a continuous variable* instead of categorical therefor more informative, (d) incorporate more clusters other than the time cluster and (e) assume *non-normal* distributions on the (random) effects of the two sources of variability in the model.

We recommend Lazic, 2016, Section 4.6, page 181, for further reading on the limitations of analyzing repeated measurement using ANOVA and on the advantage of using linear mixed effects.

5 Generalized Estimating Equation (GEE)

As we saw in Section 4, the CS condition can be relaxed to the sphericity condition. However, this sphericity condition is still limiting. It is sometimes more reasonable to assume that the correlation decreases over time as the observations are further apart in time. A plausible correlation model is

$$\text{Corr}(Y_{ijk}, Y_{ijk'}) = \rho^{|k-k'|} \ ,$$

which is known as an $AR(1)$ correlation structure (Auto Regressive of order 1). In such situation, we can use the GEE approach to calculate sample size. The details are given in Astakhov (2024).

References

Astakhov, G. (2024). Sample size calculation:implementation of different scenarios for the online tool samplesizer. https://www.math.uzh.ch/consulting/fileadmin/user_upload/Astakhov_Georgy.pdf.

Astakhov, G., Cherneva, K., Furrer, R., and Tarigan, B. (2025). Samplesizer: calculate sample sizes within completely randomized design. http://shiny.math.uzh.ch/git/reinhard.furrer/SampleSizeR/.

Casella, G. (2010). *Statistical Design*. Springer.

Lawson, J. (2015). *Design and Analysis of Experiments with R*. CRC Press.

Lazic, S. (2016). *Experimental Design for Laboratorist Biologists*. Cambridge University Press.

Version Information

v1.1	04.02.2021	
v1.2	19.01.2022	change reference style
v2.1	25.04.2024	revise contents and the example (strict CS condition), change D-Math to DM3L
v2.2	11.07.2024	link to the master thesis
v2.3	May 20, 2025	minor revision, layout, uzh-logo

```
print( sessionInfo(), locale=FALSE)


## R version 4.3.2 (2023-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] knitr_1.45
##
## loaded via a namespace (and not attached):
## [1] compiler_4.3.2 tools_4.3.2    highr_0.10     xfun_0.42      evaluate_0.23
```