Random Effects ANOVA in Uncertainty Evaluation

B. Arendacká, A. Täubner, S. Eichstädt, T. Bruns, C. Elster

Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany,
Email: barbora.arendacka@ptb.de

Abstract. *Random effects ANOVA models are well established and implemented in statistical software, they provide enough freedom to model dependencies between repeated measurements (e.g. a stronger dependency between measurements obtained within one day as compared to measurements obtained on different days) and they offer a closer look at the structure of the uncertainty (by splitting the variability due to different sources). The paper illustrates these concepts and advocates the use of random effects ANOVA models for analysis of long-term repeated experiments conducted to assess repeatability of a measurement. Variability revealed by such an experiment is important when creating an uncertainty budget. The concepts are illustrated with the help of data coming from calibration of accelerometers.*

Keywords: Random Effects, ANOVA, Type A Uncertainty

1. Introduction

To assess repeatability of a measurement, a long-term repeated experiment can be carried out. For example, in the context of calibration of accelerometers, an accelerometer may be repeatedly mounted into the measurement setup and then the frequency response function (FRF) may be measured repeatedly in a usual manner. Such a long-term experiment reveals variability that may not be observed when an accelerometer is mounted into the setup only once or twice, which is what would be done in a routine calibration. Thus the variability observed in the long-term experiment is important when creating an uncertainty budget. In order to determine this uncertainty, the simplest way one may think of, is to look at the mean values (and their variability) for each of the different mountings of the accelerometer. However, this would give us a valid result only if the numbers of repeated measurements per mounting, as well as the variability of measurements within each mounting, are always the same. Moreover, this approach does not exploit the data to the fullest and leaves much of potentially useful information untouched. A finer view can be achieved by employing random effects ANOVA(= Analysis Of VAriance) models. We describe them in detail in the next section. We then illustrate the method using data obtained within the EMRP project IND09 'Traceable dynamic measurement of mechanical quantities'. There, torque, force and pressure are of main interest. However, since acceleration is a fundamental quantity for dynamic measurements and a traceable primary calibration of accelerometers has been realised in PTB, see e.g. [1], calibration of accelerometers provides a good basis for gaining experience with implementation of the approach that can be later directly transferred to the other mechanical quantities the project focuses on.

2. Subject and Methods

Data

To illustrate the capabilities of random effects ANOVA models, we will use a part of a large dataset obtained in an experiment conducted to assess repeatability of FRF measurements for calibration of an accelerometer. The measurement campaign involved two different accelerometers that were repeatedly mounted into the measurement setup used for calibration of accelerometers in PTB (for details see [1, 2]). However, to make the presentation concise, we will consider only data resulting from a repeated mounting of a back-to-back accelerometer, type 8305 (Brüel & Kjaer) and excitation of the system by a sinusoidal excitation at frequency 4000 Hz. In addition, we will focus only at the amplitude of the FRF and show the values in arbitrary units. The input acceleration was determined from measurements by a laser interferometer pointing its rays at the top of the mounted accelerometer. The accelerometer was mounted repeatedly 20 times within a period of roughly 5 weeks. Each time the frequency response was determined repeatedly from a sinusoidal fit to the measured signals; 10 times with laser interferometer in 'position 0°' (i.e. with laser rays pointing at 2 points on the top of the accelerometer lying on a line having 0° angle with a certain reference surface), 10 times in 'position 90°'. Thus, our dataset consists of sets of 10 and 10 measured amplitudes (position 0° and 90°) obtained for 20 different mountings (see Fig. 1). The main interest is to assess the variability of an average of amplitudes obtained for a single mounting, which would be the value, reported for the amplitude of the FRF in a routine calibration. However, as we will see, employing random effects ANOVA models we can obtain a finer picture of the structure of the uncertainty and answer also such questions as e.g. what improvement in uncertainty the averaging over positions 0° and 90° brings.

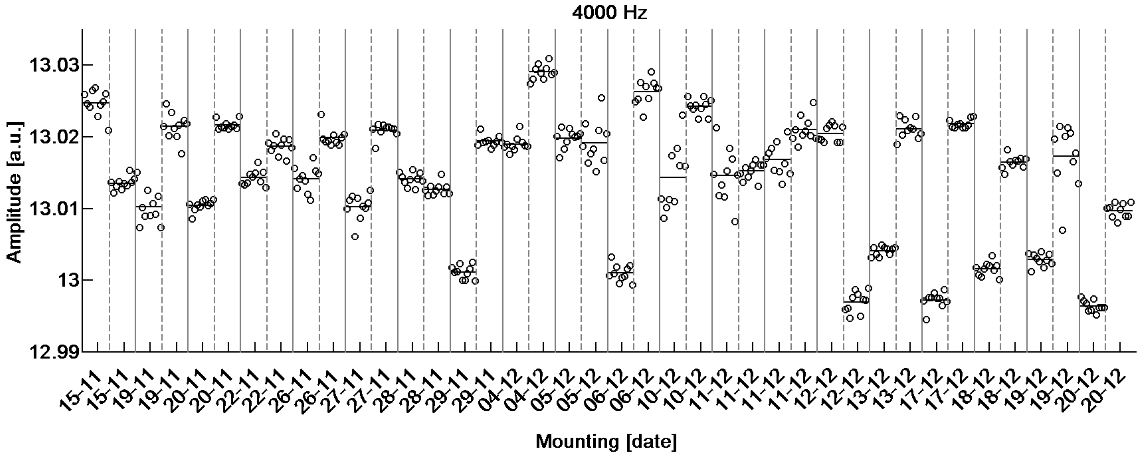

Fig. 1. Amplitude of the frequency response (a.u.) for 20 different mountings of the accelerometer, each time for two different adjustments of the laser interferometer. The horizontal lines show respective averages.

Random Effects ANOVA

The ANOVA model in our case can be of the form:

$$y_{mpr} = a + b_p + A_m + B_{mp} + E_{mpr}\,, \tag{1}$$

where y_{mpr} denotes the rth (r=1,..,10) amplitude measured for mounting m (m=1,...,20) at position p (p=1,2), a denotes the common mean (our measurand), b_p is the fixed effect of position p , $A_m \sim N(0,\ \sigma_M^2)$, $B_{mp} \sim N(0,\ \sigma_{MP}^2)$ are mutually independent random effects and $E_{mpr} \sim N(0,\ \sigma^2)$ are mutually independent random errors. $N(.,.)$ denotes a normal distribution. To make it clearer, observe the data in Fig. 1. It seems that for each combination of mounting and position, the measured values fluctuate around a certain level. These fluctuations are modelled by the random errors. The levels around which the measured values fluctuate are moreover shifted up and down around some common value (not shown) in a random manner. These shifts are modelled by the random effects. The effect due to mounting, A_m, brings a shift common for measurements obtained within 1 mounting, the effect due to the combination of mounting-position, B_{mp}, refines further the shift due to mounting by adjusting it with a random value for each mounting-position combination. Since these effects are random, we are interested in their possible size captured by the variances σ_M^2, σ_{MP}^2. In contrast, the fixed effects b_p account for a possible systematic effect of the two different positions, i.e. their contribution to the shift is constant for a given position, and thus we are interested in their values directly. To make the model identifiable, we assume that the fixed effects $b_1 + b_2$ sum to 0, so that the shifts from the common level *a* have zero mean. This also ensures that the average of amplitudes obtained within one mounting, $\bar{y}_{m..}$, is an unbiased estimator of our measurand *a*. Before fitting the model in Eq. 1 to our data, we may look once more at Fig. 1 and observe closely the fluctuations of the measured values around the depicted levels. It seems that the variability of these fluctuations varies, being rather large for some combinations of mounting-position. Thus instead of our original assumption $E_{mpr} \sim N(0,\sigma^2)$, we may rather assume a different variance for the random error within each mounting–position combination; $E_{mpr} \sim N(0,\ \sigma_{mp}^2)$.

Implied (Co)variance Structure

The model in Eq. 1 implies a certain covariance structure of the measurements. For a given mounting, the single measured value, y_{mpr}, the average of values at position *p*, $\bar{y}_{mp.}$ and the average of all values within the mounting, $\bar{y}_{m..}$, have the following distributions:

$$y_{mpr} \sim N(a + b_p, \sigma_M^2 + \sigma_{MP}^2 + \sigma_{mp}^2) \tag{2}$$

$$\bar{y}_{mp.} \sim N(a + b_p, \sigma_M^2 + \sigma_{MP}^2 + \sigma_{mp}^2 / 10) \tag{3}$$

$$\bar{y}_{m..} \sim N(a, \sigma_M^2 + \sigma_{MP}^2 / 2 + \frac{1}{4}[\sigma_{m1}^2 / 10 + \sigma_{m2}^2 / 10]) \tag{4}$$

Moreover, measurements within one mounting-position have covariance $\sigma_M^2 + \sigma_{MP}^2 \geq 0$, while measurements obtained within one mounting but at different positions have covariance $\sigma_M^2 \geq 0$. It is easy to see that if all random errors have the same variance, i.e. $\sigma_{mp}^2 = \sigma^2$, the variance in Eq. 4 turns into the expected $\sigma_M^2 + \sigma_{MP}^2 / 2 + \sigma^2 / 20$. The type A relative standard uncertainty of $\bar{y}_{m..}$ can be expressed via relative standard uncertainties due to mounting, mounting-position interaction and random errors as $\sqrt{u_{rM}^2 + u_{rMP}^2 + u_{rE}^2}$, where $u_{rM} = \sigma_M / a$, $u_{rMP} = \sigma_{MP} / (\sqrt{2}a)$, see Eq. 4. When the accelerometer is mounted anew, u_{rE}^2 can be determined from the sampling variances of the repeated measurements at the two positions. The other two contributions $u_{rM}^2 + u_{rMP}^2$ must be taken from a (past) long-term experiment.

3. Results

For fitting the model in Eq. 1 to our data (Fig. 1), we used function *lme* in the library *nlme* of R [3] and estimated the parameters with the restricted maximum likelihood method. σ_M was estimated as $1.6{*}10^{-6}$, σ_{MP} as 0.0075687. Moreover the 95% confidence interval for σ_M (obtained with function *intervals()*) contained 0. This suggests that the effect of mounting is negligible and supports the current practice of mounting an accelerometer applying a torque of 2 N.m, which was established in order to prevent any effects on the measured values. Thus, refitting the data to a model in Eq. 1 with A_m (the mounting effect) omitted (i.e. σ_M^2 fixed at 0), yields the final estimates: a=13.014408, σ_{MP}=0.00757 and all the different σ_{mp} between 0.00062 and 0.00452. Taking the minimum (maximum) of these in place of both the σ_{m1}, σ_{m2} in Eq. 4 and dividing the square root of the resulting variance by a, we obtain that the type A relative standard uncertainty of $\bar{y}_{m..}$ ranges roughly from 0.0411% to 0.0418%. Similarly, using Eq. 3 we can see that the relative standard uncertainty for average over only one position ranges from 0.0582% to 0.0592%. Correcting the average for the systematic bias b_p and including its uncertainty changes these numbers only slightly. Thus the improvement factor due to taking an average of 2 positions is approximately $1/\sqrt{2}$: clearly, in the variances in Eq. 3 and Eq. 4 with σ_M^2=0, σ_{MP}^2 is the dominant part. However, were σ_M^2 non-zero, the improvement factor would be closer to 1 (1 means no improvement). In the extreme case of σ_M^2 being dominant, averaging over two positions would not lead to a real improvement.

4. Discussion

Functions for fitting random effects models can be found in other statistical software as well. If one assumes common variance for random errors, MATLAB and *nlmefit* can be used. In our case, we could then use even textbook formulas from the so called ANOVA table. However, more general approaches allow accommodating further features, e.g. a possible trend observed within the mounting-position repeated measurements, and one can then verify whether the trend is negligible with respect to the overall uncertainty as desired.

Of course, before drawing conclusions from the estimates, one should check the goodness of the fit of the chosen model, see e.g. [3]. In our case, in model with common error variance a plot of standardized residuals versus fitted values would suggest going for the more complex model (fitted above) with different variances for the mounting-position groups.

All in all, with the help of statistical software fitting random effects ANOVA models to similar long-term experiments is not difficult and does justice to the effort invested in the experiment by enabling deeper insights into and verification of different assumptions about the measurement process.

Acknowledgements

The research leading to these results has received funding from the European Union on the basis of Decision No. 912/2009/EC.

References

[1] von Martens HJ et al. Recent advances in vibration and shock measurements and calibrations using laser interferometry. *SPIE* 5503: 1-19, 2004.

[2] International Standard ISO 16063-11, Methods for the calibration of vibration and shock transducers – Part 11: Primary vibration calibration by laser interferometry. International Organization for Standardization, Geneva, 1999.

[3] Pinheiro J, Bates D. Mixed Effects Models in S and S-PLUS. Springer, New York, 2000.