

# Analysis of means: Examples using package **ANOM**

Philip Pallmann

August 12, 2015

## Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduction</b>                          | <b>1</b> |
| <b>2</b> | <b>ANOM in a two-way layout</b>              | <b>2</b> |
| <b>3</b> | <b>ANOM with (overdispersed) count data</b>  | <b>4</b> |
| <b>4</b> | <b>ANOM with linear mixed-effects models</b> | <b>6</b> |
| 4.1      | Using <code>nlme</code> . . . . .            | 6        |
| 4.2      | Using <code>lme4</code> . . . . .            | 6        |
| 4.3      | Ignoring the repeated structure . . . . .    | 8        |

## 1 Introduction

The analysis of means (ANOM) is a graphical method for presenting multiple group comparisons with an overall mean (“grand mean”). Ever since Ott published his pioneering paper (Ott 1967), ANOM has enjoyed great popularity in quality control, and piles of extensions and applications have been discussed. A nice SAS-based overview of ANOM is given in the book by Nelson et al. (2005).

The purpose of this vignette is to illustrate how R’s broad functionality can be exploited for ANOM-type data analysis. We have already elaborated on various real-world data applications of ANOM in the tutorial publication accompanying this package (Pallmann and Hothorn 2015). In the following we want to delve into some further non-trivial data scenarios:

- ANOM with more than one factor e.g., in a two-way layout (Section 2),
- ANOM with non-normal data e.g., (overdispersed) Poisson count data using generalized linear models (Section 3),
- ANOM with correlated data in clustered, nested, hierarchical, or multilevel structures e.g., repeated measures with multiple raters using linear mixed-effects models (Section 4).

All comparisons are performed at a familywise type I error level of 5%.

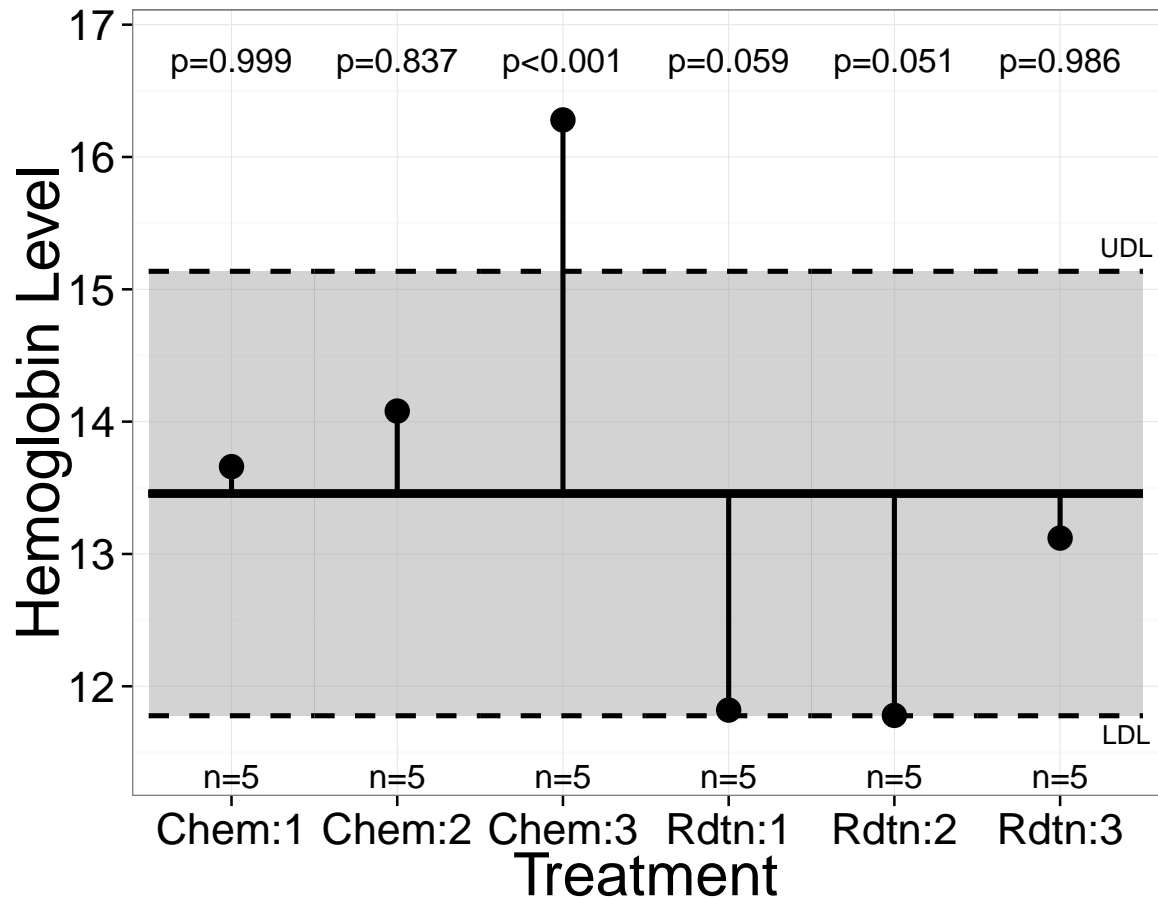


Figure 1: ANOM decision chart for the hemoglobin data based on a pseudo-one-way analysis.

## 2 ANOM in a two-way layout

Nelson et al. (2005) describe a trial on 30 male cancer patients in a balanced complete two-way layout. Each patient was treated with one therapy (chemo or radiation) and one out of three drugs (named 1, 2, and 3), and the level of hemoglobin was measured afterwards.

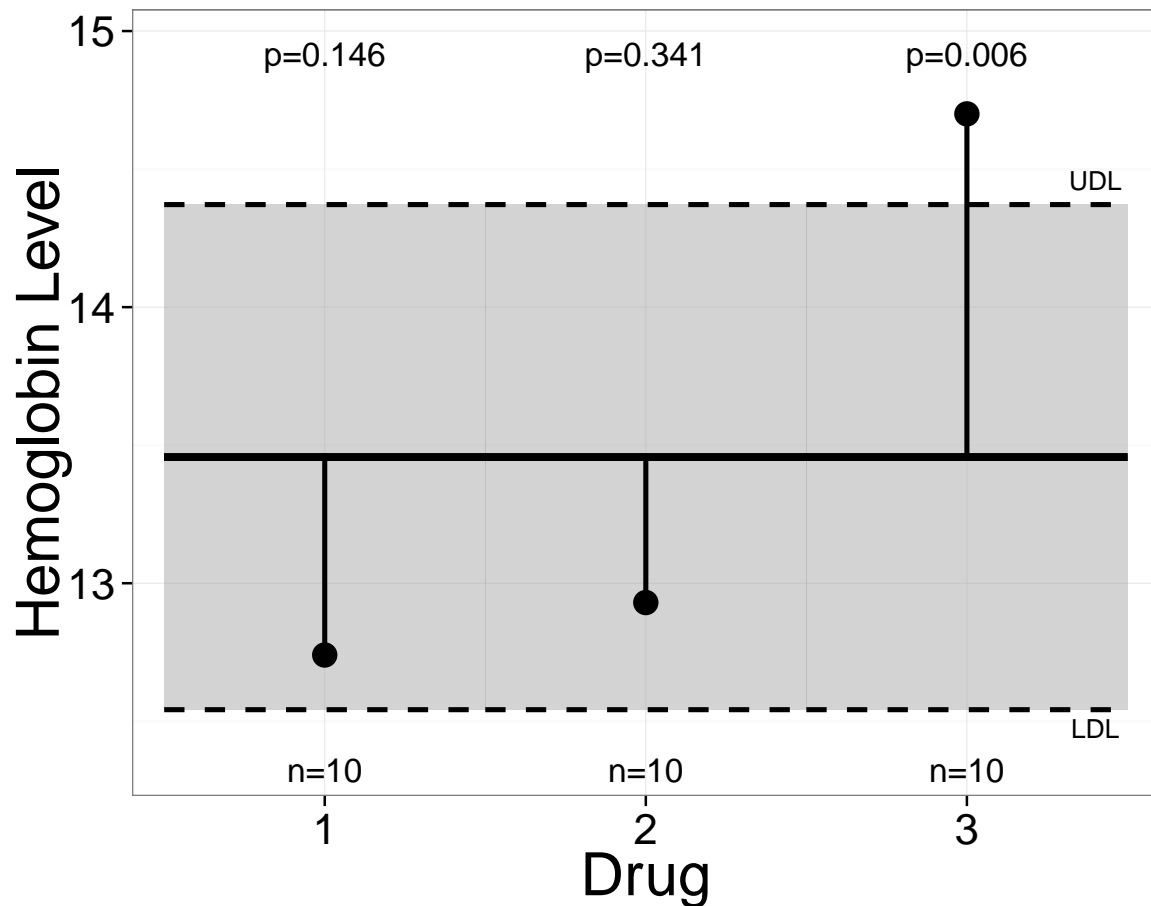
```
library(ANOM)
data(hemoglobin)
```

The six drug-therapy combinations may be compared in a simplified pseudo-one-way layout where the new factor's levels are the combinations of therapy and drug.

```
hemoglobin$the <- as.factor(abbreviate(hemoglobin$therapy))
hemoglobin$td <- with(hemoglobin, the:drug)
hemodel <- lm(level ~ td, hemoglobin)
he <- glht(hemodel, mcp(td="GrandMean"), alternative="two.sided")
ANOM(he, xlabel="Treatment", ylabel="Hemoglobin Level")
```

We find that chemotherapy together with drug 3 leads to hemoglobin levels that are significantly above the grand mean (Figure 1). On the contrary, if radiation therapy is combined with either drug 1 or 2, the levels of hemoglobin are very close to the lower decision limit.

Acting on the assumption that the two treatment factors “therapy” and “drug” do not interact, we may also investigate the marginal drug effects across both therapies. An analysis of variance suggests that the interaction effect is negligible.



**Figure 2:** ANOM decision chart for the hemoglobin data based on a two-way analysis.

```
hemodel2 <- lm(level ~ drug * therapy, hemoglobin)
anova(hemodel2)

## Analysis of Variance Table
##
## Response: level
##          Df Sum Sq Mean Sq F value    Pr(>F)
## drug       2 23.369   11.684   5.5385 0.0105262 *
## therapy    1 44.408   44.408  21.0499 0.0001184 ***
## drug:therapy 2  2.245    1.122   0.5320 0.5941993
## Residuals 24 50.632    2.110
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This result clears the way for making reasonable inferences for drugs pooled over both types of therapy.

```
hemodel3 <- lm(level ~ drug + therapy, hemoglobin)
he3 <- glht(hemodel3, mcp(drug="GrandMean"), alternative="two.sided")
ANOM(he3, xlabel="Drug", ylabel="Hemoglobin Level")
```

Figure 2 shows that drug 3 raises hemoglobin levels significantly compared to the grand mean. Keep in mind that this type of pooled analysis would be inept in the presence of a therapy-drug interaction; here one should rather perform separate drug comparisons for radiation and for chemotherapy.

### 3 ANOM with (overdispersed) count data

Beall (1942) reports a study on the effectiveness of six insect sprays to control the tomato hornworm *Manduca quinquemaculata* (formerly called *Phlegethontius quinquemaculatus*), which is a common pest of solanaceae. The experiment was carried out in a randomized complete block design: each spray was applied to plants in each of six blocks, and the number of insects found after treatment was counted and recorded (taking two random samples per block and spray).

The data is stored in R's basic `datasets` package, but this version is incomplete as the block structure of the experiment was missed out for some reason. We include a block variable according to Table 7 in Beall's original publication.

```
data(InsectSprays)
InsectSprays$block <- as.factor(rep(1:6, each=2))
```

A common distributional assumption for count data is Poisson, so it sounds like a good idea to fit a Poisson generalized linear model with a logarithmic link function to the data. However, recall that the variance of a Poisson random variable equals its expected value. This basic assumption is more often than not violated with real-world data, and we usually include a variance inflation factor  $\phi$  in the model to cope with this overdispersion (or extra-Poisson variability).

```
insmodel1 <- glm(count ~ spray + block, data=InsectSprays,
                 family=quasipoisson(link="log"))
summary(insmodel1)$dispersion
## [1] 1.198022
```

In our case the estimated variance inflation of 19.8% is negligible, hence we will most likely not harm our analysis when using a simple Poisson GLM without an overdispersion parameter. An analysis of deviance reveals that both treatment and block bring about highly significant effects.

```
insmodel2 <- glm(count ~ spray + block, data=InsectSprays,
                 family=poisson(link="log"))
anova(insmodel2, test="Chisq")

## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: count
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                71      409.04
## spray    5       310.71      66      98.33 < 2.2e-16 ***
## block    5        25.34      61       72.99 0.0001198 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We perform an ANOM to find out which insect sprays lead to lower or higher counts of insects compared to average. The logarithmic transformation of the counts is automatically reversed to simplify interpretation of the resulting graph.

```
ins <- glht(insmodel2, mcp(spray="GrandMean"))
ANOM(ins)

## Error in ANOMintern(mu = m * cit[, 1], n = ss, gm = m, lo = m * cit[, : object 'back'
## not found
```

Figure 3 shows that all six treatments are significantly different from the grand mean: more insects were counted after applying sprays A, B, and F, and fewer with sprays C through E. The varying widths of

the gray band (upped minus lower decision limit) just reflect the dependence of the Poisson variance on the expected value.

## 4 ANOM with linear mixed-effects models

Pinheiro and Bates (2000) describe data from an investigation of ergonomic stools performed by Wretenberg et al. (1993). They had nine people assess the physical effort to rise from each of four stool types: a stool of ordinary height, a low ordinary stool, a one-legged stool, and a pneumatic stool. Each subject tested each stool type once and rated the perceived exertion on the so-called Borg scale, which takes values from 6 to 20. We wish to detect types of stools that are significantly more or less difficult to rise from in comparison to average.

```
library(nlme)
data(ergoStool)
```

This is a complete block design; each of the four stools under study is evaluated nine times. One challenge is that the scoring is clearly subjective e.g., a person suffering from chronic back pain is more likely to assign lower ratings than a healthy individual. To account for this variability between raters, a linear mixed-effects model is our method of choice.

The model for the  $i$ th individual has the general form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

where  $\mathbf{y}_i$  are the outcomes,  $\boldsymbol{\beta}$  the fixed and  $\mathbf{b}_i$  the random effects,  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  the design matrices for the fixed and random effects, and  $\boldsymbol{\epsilon}_i$  the residuals. We assume that  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$  and  $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$ , and additionally  $\text{cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}$ .

In our ergonomic stools example, we model fixed stool effects and treat the test persons as random. We hereby make the assumption that the testers are representative of all people who could sit on these stools and that their scorings are normally distributed with between-rater variance  $\sigma_b$ .

There are several concurrent and (in part) complementary frameworks for mixed-effects modeling which coexist in the R world. Here we illustrate ANOM using two of the most popular and best-established packages, `nlme` (Pinheiro et al. 2013) and `lme4` (Bates et al. 2013).

### 4.1 Using nlme

We model the perceived effort to rise (on the Borg scale) as a function of the type of stool and include a random subject effect to acknowledge between-rater variability.

```
library(nlme)
esmodel1 <- lme(effort ~ Type, random=~1|Subject, data=ergoStool)
es1 <- glht(esmodel1, mcp(Type="GrandMean"), alternative="two.sided")
ANOM(es1, xlabel="Stool Type", ylabel="Exertion (Borg Scale)")
```

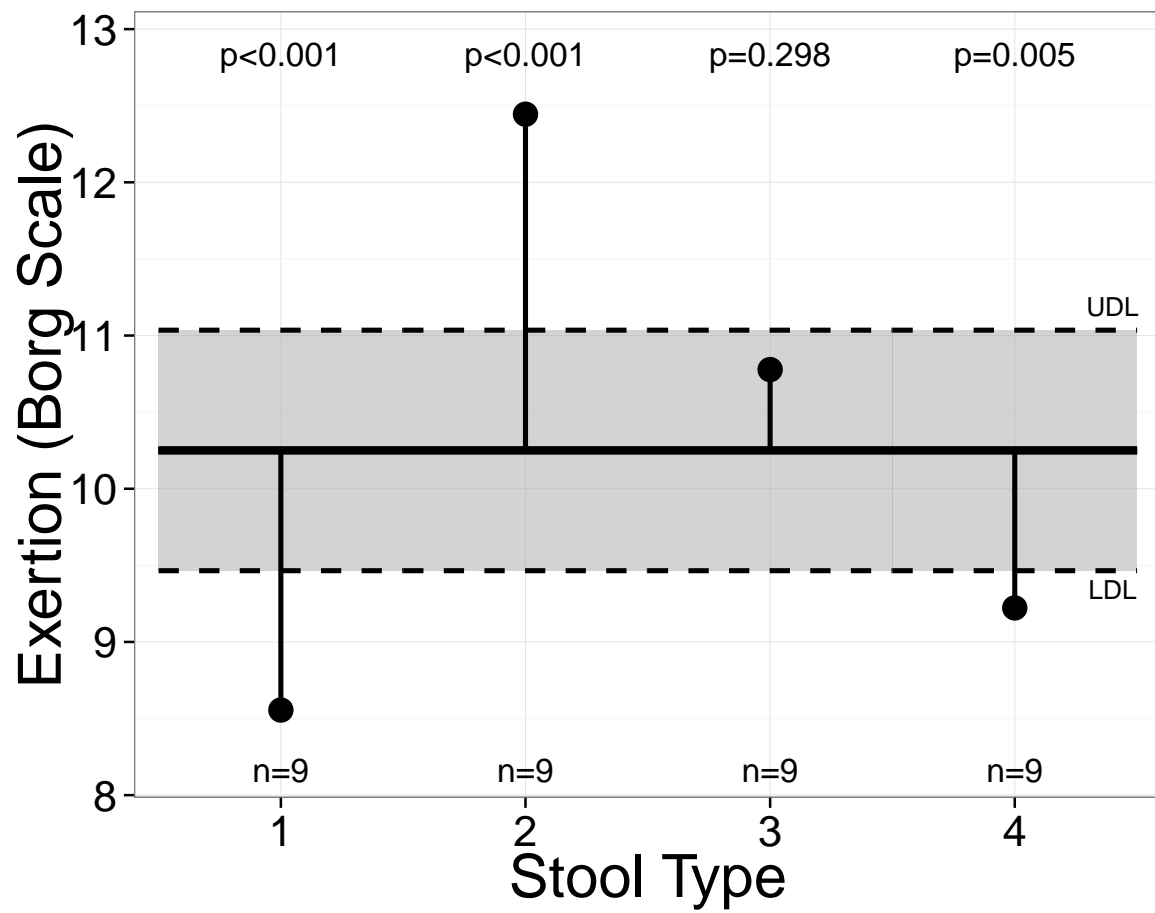
Figure 4 displays that it is significantly more difficult to rise from stools of type 2. On the contrary, the perceived exertion is significantly below the grand mean for stool types 1 and 4.

### 4.2 Using lme4

An identical graphic can be generated with the functionality in `lme4`.

```
library(lme4)
esmodel2 <- lmer(effort ~ Type + (1|Subject), data=ergoStool)
es2 <- glht(esmodel2, mcp(Type="GrandMean"), alternative="two.sided")
ANOM(es2, xlabel="Stool Type", ylabel="Exertion (Borg Scale)")
```

Note that there are instances when the two mixed-effects model packages cannot be used interchangeably. Only `nlme` supports imposing a pattern like compound symmetry or AR(1) on the covariance matrix of the random effects and/or residuals whereas `lme4`'s strong points are sophisticated random effects hierarchies and nested structures, and it also fits generalized linear mixed-effects models.



**Figure 3:** ANOM decision chart for the ergonomic stool data based on a linear mixed-effects model.

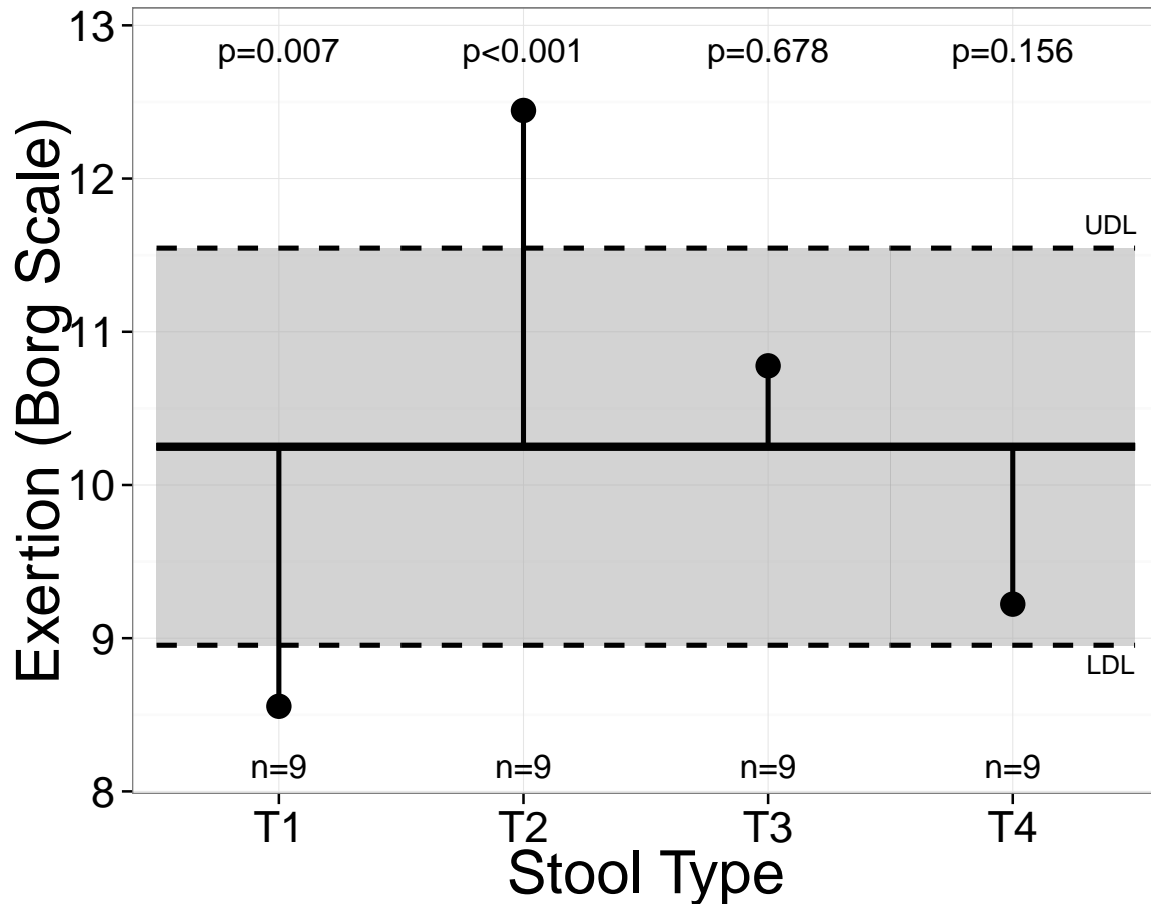


Figure 4: ANOM decision chart for the ergonomic stool data based on a standard linear model.

### 4.3 Ignoring the repeated structure

One may wonder what would happen if we did not account for the fact that each chair was tested by several individuals. We fit a simple linear model and observe that the decision limits drift apart so that stool type 4 is no longer significant (Figure 5).

```
esmodel3 <- lm(effort ~ Type, ergoStool)
es3 <- glht(esmodel3, mcp(Type="GrandMean"), alternative="two.sided")
ANOM(es3, xlabel="Stool Type", ylabel="Exertion (Borg Scale)")
```

The reason is quite simple: when fitting a mixed-effects model, we partition the total variance of the data into a component explained by the variability among the nine raters and a residual component. If we ignore the clustered structure of the data and fit a standard linear model, all variability is summed up in the residual variance, thus making it harder to detect significant stool effects.



## References

- D. Bates, M. Maechler, B. Bolker, and S. Walker. *lme4: Linear mixed-effects models using Eigen and S4*, 2013. URL <http://CRAN.R-project.org/package=lme4>. R package version 1.0-5.
- G. Beall. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, 32(3-4):243–262, 1942.
- P. R. Nelson, P. S. Wludyka, and K. A. F. Copeland. *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA, and ASA, Alexandria, VA, 2005. ISBN 0-89871-592-X.
- E. R. Ott. Analysis of means - a graphical procedure. *Industrial Quality Control*, 24:101–109, 1967.
- P. Pallmann and L. A. Hothorn. Analysis of means (ANOM): A generalized approach using R. *To appear in Journal of Applied Statistics*, 2015.
- J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and nonlinear mixed effects models*, 2013. URL <http://CRAN.R-project.org/package=nlme>. R package version 3.1-111.
- J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY, 2000. ISBN 0-387-98957-9.
- P. Wretenberg, U. P. Arborelius, and F. Lindberg. The effects of a pneumatic stool and a one-legged stool on lower limb joint load and muscular activity during sitting and rising. *Ergonomics*, 36(5): 519–535, 1993.