

METHODOLOGY

Open Access



# A recurring error in evaluating the effects of different pesticides, pollutants and fertilizers with a zero level

Gaétan Moreau\* 

## Abstract

**Background:** The Quenouille-Addelman solution has been proposed to properly analyze linear models with a crossed or factorial arrangement of treatments that includes a qualitative/categorical and a quantitative factor with a zero level, a situation particularly prevalent in ecotoxicological studies. However, a review of the recent literature reveals that this solution isn't used, perhaps due to a lack of recognition that zero-level factors can produce incomplete factorial arrangements.

**Results:** Using practical examples, I demonstrate that the conclusions of a study can be substantially altered if the Quenouille-Addelman solution is not used when warranted.

**Conclusions:** Suspecting that the lack of a detailed method may have contributed to the underutilization of the solution, I describe how to apply the solution using current statistical software packages and discuss how the solution can be adapted to address some experimental situations not previously considered.

**Keywords:** Factorial arrangement, Incomplete factorial arrangement, Qualitative factor, Zero level

## Introduction

Analyzing a study with a crossed or factorial arrangement of treatments that includes a zero level is an underestimated challenge because often, a zero amount of each level of a qualitative/categorical factor is essentially the same treatment. Consider a study in which  $p$  different pesticides are applied at  $r$  different rates, with one of those rates being zero. Since it doesn't matter which pesticide is applied at zero rate, there are  $\{p(r-1) + 1\}$  single treatment combinations rather than  $p \times r$ . Thus, an analysis with a two-way linear model cannot be carried out because the sums of squares (SS) and degrees of freedom (df) of the analysis require adjustment to account for the incomplete factorial arrangement of treatments. The situation described above is common in the literature

examining applications (e.g., pesticides, adhesive dental cements), enrichment (e.g., fertilization, isotope enrichment study), inoculations (e.g., growth hormones, vaccines), pollutants, length of storage, etc.

Quenouille (1953) and Addelman (1974) independently proposed a solution, hereinafter referred to as the Quenouille-Addelman (QA) solution for linear models to deal with the issue discussed above. In short, the QA solution involves amalgamating the SS of two models to obtain a single model, without increasing the Type I and Type II error (see below). However, despite the existence of this solution, it is practically never applied when necessary. To support this claim, a literature search of recent publications with a qualitative factor and a quantitative factor including a zero level is presented in Additional file 1: Review of the frequency of use of the Quenouille-Addelman solution in the literature. This search showed that none of the reviewed studies applied the QA solution. Instead, 11.4% of the studies used an erroneous factorial

\*Correspondence: gaetan.moreau@umoncton.ca

Université de Moncton, Moncton, NB E1A 3E9, Canada



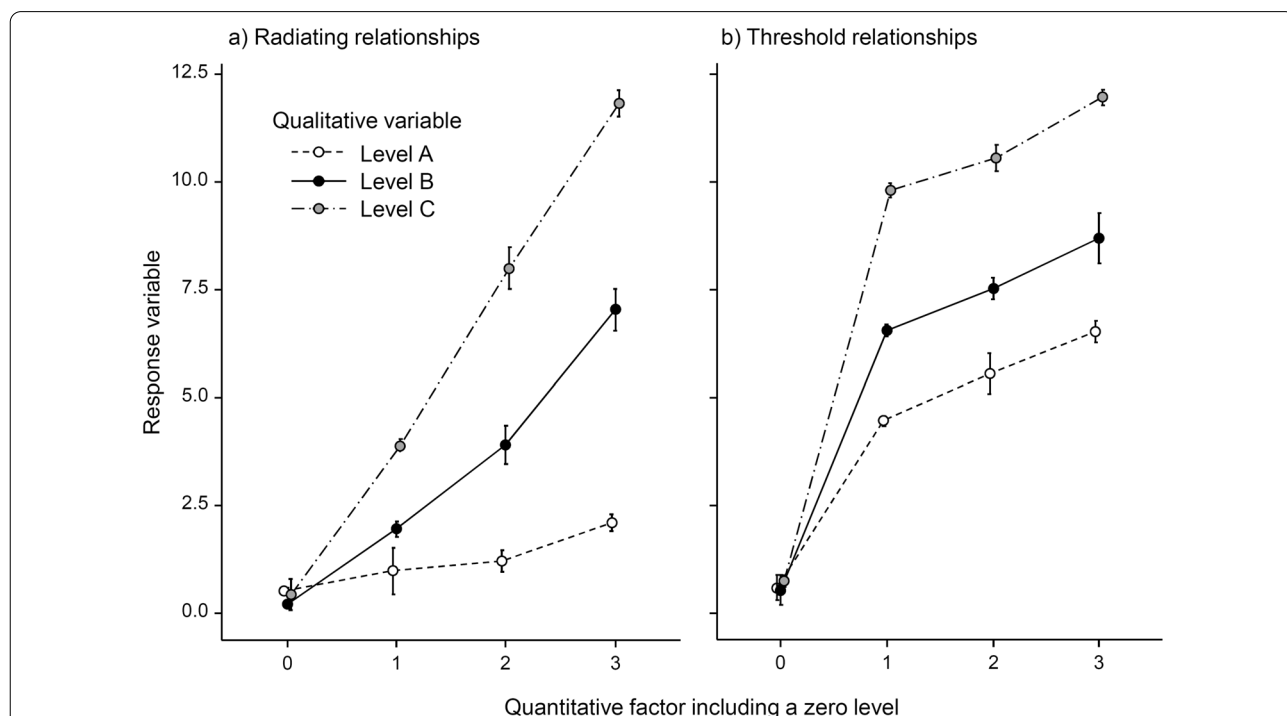
© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

linear model, 2.5% excluded control treatment data from analysis, 36.7% performed a one-way test on a variable combining the qualitative and quantitative factors, and 49.4% used other inadequate approaches such as leaving out the comparison between qualitative treatments (Additional file 1: Review of the frequency of use of the Quenouille-Addelman solution in the literature). All aforementioned approaches are either biased, or contribute to information loss, as explained below. The resulting corollary is that the QA solution has been largely forgotten, perhaps due to a misunderstanding of the impact of zero levels in factorial arrangements. An update on the subject appears overdue. Herein, I (1) demonstrate how noncompliance with the QA solution alters the conclusions of a study, (2) describe how to achieve the solution using current statistical packages, and (3) examine how the solution can be adapted to solve situations not considered by Quenouille (1953) and Addelman (1974).

**The Quenouille-Addelman solution and substitute (flawed) approaches**

To date, the adverse effects of not using the QA solution when warranted have not been demonstrated. In a review paper, Gates (1991) discusses the solution using an example where the adjustment imparted is subtle, which does not do justice to the effect this solution has in most

publications where it was used (Quenouille 1953; Green et al. 1976, 1977; Conrad et al. 1993; Lu and Nielsen 1993; Cushman et al. 1998; Olivier et al. 2000; Gong et al. 2001; Moreau and Bauce 2001, 2003). Using simulated data (available in Additional file 2: Simulated data used to produce Fig. 1 and Table 1) inspired by the aforementioned studies and literature review, I determined that the trends followed by the different levels of the qualitative variable as the quantitative variable increases can predict the effects of the solution. In the uncommon situation where the relationships with the quantitative variable of all individual qualitative treatments extend linearly from level zero (e.g., Fig. 1a), the adjustment provided by the solution is at its lowest. The interpretation of the results is slightly modified although the result of a linear model fit can substantially change (Table 1). Only one of all the published studies using the QA solution (i.e., Gong et al. 2001) reported such data. For all other situations, significant differences between the QA solution and an unadjusted model occur if the solution is not used (i.e., Quenouille 1953; Green et al. 1976, 1977; Conrad et al. 1993; Lu and Nielsen 1993; Cushman et al. 1998; Olivier et al. 2000; Moreau and Bauce 2001, 2003). If the relationships with and without the zero level are different (e.g., Fig. 1b), the adjustment conferred by the solution is at its highest (Table 1). In the latter example, the



**Fig. 1** Theoretical examples of factorial arrangements of treatments involving one qualitative factor with three levels and one quantitative factor with four levels that includes a zero amount. In **a** the relation radiates linearly from the zero level while in **b** the relation does not radiate in the same way

**Table 1** Inadequate and Quenouille-Addelman solutions for the data presented in Fig. 1 a, b

Inadequate							Quenouille-Addelman solution						
Source of variation	df	SS	MS	F	P		Source of variation	df	SS	MS	F	P	
Fixed-effect solution for a completely randomized design based on the data presented in Fig. 1a													
Qualitative variable (A)	2	101.77	50.89	208.40	>0.01		Qualitative variable (A)	2	136.60	68.30	326.35	>0.01	
Quantitative variable (Z)	3	156.66	52.22	213.87	>0.01		Quantitative variable (Z)	3	156.66	52.22	249.52	>0.01	
A × Z	6	60.59	10.10	41.36	>0.01		A × Z	4	25.76	6.44	30.77	>0.01	
Error	12	2.93	0.24				Error	14	2.93	0.21			
Total	23	321.95					Total	23	321.95				
Fixed effect solutions for a completely randomized design based on the data presented in Fig. 1b													
Qualitative variable (A)	2	64.48	32.24	179.94	>0.01		Qualitative variable (A)	2	84.20	42.10	274.14	>0.01	
Quantitative variable (Z)	3	255.79	85.26	475.89	>0.01		Quantitative variable (Z)	3	255.79	85.26	555.20	>0.01	
A × Z	6	19.87	3.31	18.48	>0.01		A × Z	4	0.10	0.03	0.16	0.95	
Error	12	2.15	0.18				Error	14	2.15	0.15			
Total	23	342.29					Total	23	342.24				
Polynomial fixed effect solution for a completely randomized design based on the data presented in Fig. 1b													
Qualitative variable (A)	2	64.48	32.24	179.94	>0.01		Qualitative variable (A)	2	84.20	42.10	274.14	>0.01	
Quantitative variable (Z)	1	206.66	206.66	1153.45	>0.01		Quantitative variable (Z)	1	206.66	206.66	1345.69	>0.01	
Linear	1	39.58	39.58	220.91	>0.01		Linear	1	39.58	39.58	257.73	>0.01	
Quadratic	1	9.54	9.54	53.25	>0.01		Quadratic	1	9.54	9.54	62.12	>0.01	
Cubic	1						Cubic	1					
A × Z	2	12.16	6.08	33.93	>0.01		A × Z	2	0.01	0.01	0.03	0.97	
Linear	2	5.71	2.86	15.93	>0.01		Linear	2	0.10	0.05	0.33	0.73	
Quadratic	2	2.00	1.00	5.58	0.02		Quadratic	2	2.15	1.08	0.15		
Cubic	2						Cubic	2					
Error	12	2.15	0.18				Error	14	2.15	0.15			
Total	23	342.28					Total	23	342.24				

unadjusted model and the QA solution offer contrasting results, one indicating a strong interaction and the other not. Changing the scale (i.e., applying data transformation) does not help. Thus, in most cases, the QA solution increases the SS associated with the qualitative variable at the expense of the interaction term. Unadjusted models have an inflated Type II error rate when evaluating the main effect of the qualitative variable and an inflated Type I error rate when evaluating interactions.

Other approaches have often been used instead of the QA solution (Additional file 1: Review of the frequency of use of the Quenouille-Addelman solution in the literature). For instance, some authors omitted control treatment data from the analysis. While this can approximate the QA solution in some situations, it can also alter the results because the main effect of the quantitative variable is not evaluated over its entire range. Other authors repeatedly used the control treatment in a series of analyses with the other treatments, which results in inflated Type I error rate. In many cases, authors combined the qualitative and quantitative variables into a single variable and performed a one-way test followed by multiple comparison (post hoc) tests, a procedure comparing each treatment to a single control (e.g., Dunnett or Williams test) or orthogonal contrasts. For example, a dose of 0, 25 and 100 ml of a given pesticide could be classified as control, low dose and high dose, respectively. This, however, does not change the fact that a control dose of two different pesticides is the same treatment. In addition, this approach means that the interaction between the factors cannot be examined and trend analysis (see below) is impossible. The same would apply with an ANCOVA or a regression model. A Dunnett or Williams test is also considerably less informative than the QA solution. For example, using the data in Fig. 1a, a Dunnett test only reveals that the zero level is different from all but one of the treatment combinations (i.e., Level A at the value 1 of the quantitative factor). The Dunnett or Williams test also precludes subsequent tests without the zero level because this inflates the Type I error rate. It is possible that contrasts could be used to derive main effect and interaction test statistics approximating the QA solution, but to our knowledge, no one has yet investigated this avenue.

#### The Quenouille-Addelman solution for fixed-effects two-way linear models

Quenouille (1953) and Addelman (1974) presented a hand calculation solution for a two-way fixed-effects linear model. Although some steps of the solution can be performed using statistical packages (Gates 1991), the solution is generally hard to perform in one execution

(Hocking 2013). Suspecting that the lack of a detailed example may have contributed to the underutilization of the QA solution, I describe below a step-by-step approach to achieving the solution with most packages.

1. Using the whole dataset, calculate the unadjusted SS and df for all sources of variation using a two-way linear model.
2. Remove the zero level from the dataset and run the same analysis as in step 1.
3. Create a table of the SS and by combining the two linear models. Take the SS and df of the quantitative variable, error and total obtained from the first model. The SS and df of the qualitative factor and interaction are obtained from the second model.
4. Increase the number of df associated with the error term to incorporate the degrees lost by the interaction term because the difference between the qualitative factor for the zero level can only be chance differences (Quenouille 1953). The sum of the df of the treatments ( $A + Z + [A \times Z]$ ; Table 1) is now equal to the number of distinct treatments minus one. Of course, if there is any reason to suspect that there is a difference between the qualitative factors for the zero level, their SS can also be calculated separately according to the method presented by Quenouille (1953). The unadjusted two-way linear model and the QA solution both yield the same total SS if the design is balanced [i.e., no missing data; see Addelman (1974) and Gates (1991)].
5. Calculate the mean squares ( $MS = SS \div df$ ), F-values ( $F = MS \div MS_{\text{Error}}$ ) and P-values (tabulated using functions included in spreadsheets or probability tables) of the adjusted model. For examples of this last step, the reader is invited to refer to a statistical textbook or to the QA solution presented on the right side of Table 1.

#### The Quenouille-Addelman solution in other experimental situations

Below, I identify solutions, if possible, for analytical and experimental situations that cannot be solved using the calculations provided by Quenouille (1953) and Addelman (1974) and have not been previously addressed in the literature.

#### Polynomial contrasts

Because a linear model does not identify which of the pairs of means are different when there are more than two levels, additional tests are often required. Instead of post-hoc tests, when quantitative variables with fixed intervals are used, an effective approach is to perform a

trend analysis using polynomial contrasts (Keppel 1982). For example, data in Fig. 1b allow for a third-order polynomial contrast model presented at the bottom of Table 1. Note that the polynomial model contains one level less for the interaction term than for the main effect of the continuous variable due to the adjustment associated with the QA solution (Table 1).

#### **Mixed models, maximum likelihood and REML**

Models with both fixed and random effects are nowadays analyzed using mixed model procedures, Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) (Zuur et al. 2009). Most software programs performing mixed model analyses now incorporate REML estimation as a default option (Gurka 2006). While several statistical packages do not display a complete table of SS when performing these analyses, the MS and df can be obtained and allow for the inverse calculation of the error term. For example, a REML of the data from Gates (1991) developed with the lmer function in the lme4 package of R (R Core Team 2021) can be used to calculate the MS of the error term by dividing the MS of any fixed effect by its F-value. Once these terms are secured through cross-multiplications, the QA solution presented above can be applied.

#### **Unbalanced designs**

An unbalanced dataset with missing data presents a challenge because the SS cannot be estimated independently and do not add up with the error term(s) to the total SS as they would in a balanced design. This non-orthogonality means that the Type I SS is affected by the order in which the terms are included in the model. One way to deal with this situation is to remove missing cells, randomly remove samples from the dataset until equilibrium is reached and apply the QA solution. Because most researchers are unwilling to throw data away, another approach is to fit the missing cells using imputation techniques (reviewed in van Ginkel et al. 2007), and then apply the QA solution. A third method is to apply the QA solution using Type III SS but if two treatments exhibit different levels of imbalance (e.g., if the control has fewer missing data than the other treatments), this leads to biases in SS estimations and result in a different total SS for the unadjusted linear model and the QA solution. Ultimately, the choice between strategies to deal with missing data should depend upon the situation at hand (see review by Graham 2009).

#### **Three-way ANOVAs and higher-order models**

The methodology presented by Quenouille (1953) and Addelman (1974) does not apply to higher-order models such as three-, four- or five-way linear models. Although

the potential for an inflated rate of Type I error increases as the order of a model increases (Cohen 2001), these models are frequently applied and need to be addressed.

In the case of a three-way linear model with one zero-level quantitative variable and two qualitative variables, the solution is similar to the two-way fixed-effect linear model presented above. The SS of the error term and the quantitative variable are retrieved as usual while the SS of the two qualitative variables and all interactions are only calculated for non-zero quantities of the quantitative factor. The degrees of freedom associated with the main effects are not modified but the degrees of freedom associated with all four interaction terms are reduced and transferred to the error term. An example of a 3-way ANOVA calculation is shown in Additional file 3: Data and solutions for a 3-way analysis. Fitting a four-way and higher-order models with a single quantitative variable follows the same procedure.

A higher-order linear model with at least one qualitative variable could also include more than one quantitative variable with a zero level. An example of this situation would be a study of the effect of tillage (i.e., qualitative variable), nitrogen fertilization (i.e., quantitative variable that includes a zero level), and pesticide application (i.e., a quantitative variable that includes a zero level) on the biomass of a given crop. However, the mathematical solution has not been developed to our knowledge for this situation and cannot be solved using the QA solution as discussed here.

#### **GLMs, GAMs and Bayesian models**

The QA solution has not been developed for generalized, additive and Bayesian models. Considering the usefulness of these approaches, I stress the need of developing an equivalent of the QA solution for these models in the near future. However, it is important to note that in the literature search presented in Additional file 1: Review of the frequency of use of the Quenouille-Addelman solution in the literature, no study used any of these approaches to handle the data and thus, that the solution for a linear model herein is still relevant.

#### **Discussion/conclusion**

In this article, the emphasis has been placed primarily on hypothesis testing for the QA solution but many contemporary analyses focus instead on estimating the variability associated with the mean or median. As the sum of the degrees of freedom differs when the QA solution is applied, the eventual calculations of confidence intervals or error estimates will be impacted. For fixed-effects models, these calculations can be easily adjusted by following the methods described in standard statistical textbooks. On the other hand, more complex models (e.g.,

REML) require a mathematical solution that is beyond the scope of this article.

Statistical errors are generally not intentional. In most cases where the QA solution was not used when needed, the study authors were probably unaware that they were making a mistake. It is likely that a lack of statistical literacy also contributes to this situation. Likewise, poor statistical literacy among editors probably exacerbates this problem. As a peer reviewer, I have suggested to some authors to employ the QA solution. However, the suggested changes have never been enforced by the editorial board, perhaps due to a lack of awareness that noncompliance with the QA solution inflates type I and type II error rates. In their defense, the QA solution has practically fallen into oblivion since 2010 as a single review article (Moreau et al. 2015) has cited Addelman (1974). Quenouille (1953) was cited 27 times in this same period but not for an application of the solution discussed herein. My aspiration with this article is to rectify this situation and reduce the incidence of this recurring error in future publications.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43170-022-00128-0>.

**Additional file 1.** Review of the frequency of use of the Quenouille-Addelman solution in the literature.

**Additional file 2.** Simulated data used to produce Fig.1 and Table 1.

**Additional file 3.** Data and solutions for a 3-way analysis.

## Acknowledgements

The author thanks F. Horgan for stimulating discussions that led to the writing of this manuscript, as well as K. LeBlanc, D. Boudreau, L. Tousignant, N. Hammami and two anonymous reviewers for comments on an earlier version of this manuscript.

## Author contributions

Not applicable.

## Funding

Not applicable.

## Availability of data and materials

All data generated or analysed during this study are included in this published article [and its Additional files information files].

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 3 May 2022 Accepted: 19 August 2022

Published online: 05 September 2022

## References

- Addelman S. Computing the analysis of variance table for experiments involving qualitative factors and zero amounts of quantitative factors. *Am Stat.* 1974;28:21–2.
- Cohen BH. Explaining psychological statistics. 2nd ed. Wiley; 2001.
- Conrad KM, Mast MG, Mac Neil JH, Ball HR. Composition and gel-forming properties of vacuum-evaporated liquid egg white. *J Food Sci.* 1993;58:1013–6.
- Cushman LC, Pemberton HB, Miller JC, Kelly JW. Interactions of flower stage, cultivar, and shipping temperature and duration affect pot rose performance. *HortScience.* 1998;33:736–40.
- Gates CE. A user's guide to misanalysing planned experiments. *HortScience.* 1991;26:1261–5.
- Gong H, Lawrence AL, Gatlin DM, Jiang DH, Zhang F. Comparison of different types and levels of commercial soybean lecithin supplemented in semipurified diets for juvenile *Litopenaeus vannamei* Boone. *Aquac Nutr.* 2001;7:11–7.
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol.* 2009;60:549–76.
- Green JR, Lawhon JT, Cater CM, Mattil KF. Protein fortification of corn tortillas with oilseed flours. *J Food Sci.* 1976;41:656–60.
- Green JR, Lawhon JT, Cater CM, Mattil KF. Utilization of hole undefatted glandless cottonseed kernels and soybeans to protein-fortify corn tortillas. *J Food Sci.* 1977;42:790–4.
- Gurka MJ. Selecting the best linear mixed model under REML. *Am Stat.* 2006;60:19–26.
- Hocking RR. Methods and applications of linear models: regression and the analysis of variance. Hoboken: Wiley; 2013.
- Keppel G. Design and analysis: a researcher's handbook. 2nd ed. Englewood Cliff: Prentice-Hall; 1982.
- Lu DD, Nielsen SS. Heat inactivation of native plasminogen activators in bovine milk. *J Food Sci.* 1993;58:1010–2.
- Moreau G, Bauce É. Developmental polymorphism: a major factor for understanding sublethal effects of *Bacillus thuringiensis*. *Entomol Exp Appl.* 2001;98:133–40.
- Moreau G, Bauce É. Feeding behavior of spruce budworm (Lepidoptera: Tortricidae) larvae subjected to multiple exposures of *Bacillus thuringiensis* variety *kurstaki*. *Ann Entomol Soc Am.* 2003;96:231–6.
- Moreau G, Michaud J-P, Schoenly KG. Experimental design, inferential statistics, and computer modeling. In: Tomberlin JK, Benbow ME, editors. Forensic entomology: international dimensions and frontiers. CRC Press, Taylor & Francis Group. 2015; pp. 205–230.
- Olivier F, Tremblay R, Bourget E, Rittschof D. Barnacle settlement: field experiments on the influence of larval supply, tidal level, biofilm quality and age on *Balanus amphitrite* cyprids. *Mar Ecol Progr Ser.* 2000;199:185–204.
- Quenouille MH. The design and analysis of experiment. Griffin; 1953.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2021.
- van Ginkel JR, van der Ark LA, Sijtsma K. Multiple imputation for item scores when test data are factorially complex. *Br J Math Stat Psychol.* 2007;60:315–37.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. Springer; 2009.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.