

Multiplicative Models for Combining Information from Several Sensory Experiments: a Bayesian Analysis

Bareng A. S. Nonyane^{a,*} and Chris M. Theobald^b

School of Mathematics, University of Edinburgh
and Biomathematics & Statistics Scotland,
The King's Buildings, Edinburgh, Scotland EH9 3JZ, U.K.

^a *aletta@schoolph.umass.edu*

^b *cmt@bioss.ac.uk*

November 7, 2006

* *Correspondence* B.A.S Nonyane, Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, 416 Arnold House, 715 North Pleasant Street, Amherst, MA, 01002, USA. Fax: 1 413 545 1645

Abstract

If data are available from a series of quantitative sensory experiments on the same type of product, and the assessors in these experiments are drawn from a common pool, then it is possible to combine information from the experiments on the relative biases and variability of individual assessors, and to examine the influence of possible temporal effects over the series. Such a combination of information is illustrated using a series of apple-tasting experiments conducted with the main aim of monitoring assessor performance over time. Models which include random effects and multiplicative interaction terms have been used for modelling heterogeneous interaction between assessors and products in individual sensory experiments. Such models are extended here to analyse data from series of experiments. A Bayesian approach is used that allows for adjustment for missing observations and for the use of information on assessors' previous performance when analysing future experiments. This use of previous information leads to a reduction in the average variance of product differences.

Key Words: Assessor performance; Bayesian inference; Multiplicative model; Sensory studies.

1 Introduction

Sensory experiments are widely used by the food industry for reasons such as establishing consumer preferences for different food and drink varieties on the market and in product development. Usually, series of tasting experiments are conducted with a panel of trained assessors who give their scores or rankings on food products. The performance of the assessors may be evaluated regularly over time, and one might omit from the panel those assessors whose performance is not consistent. The literature has a wide range of methods for evaluating assessor performance: see for example, Naes (1998) , Rossi (2001) and McEwan, Hunter, van Gemmert and Leah (2002).

This paper considers the analysis of data from quantitative sensory studies in which the performance of assessors over a sequence of experiments can be examined. The models and methods proposed are illustrated on a series of apple-tasting experiments, which are described in Section 2. Heterogeneity of assessors' responses to products, and of error variance between assessors, are modelled with a Bayesian multiplicative model, and the individual-specific parameters in this model are used to measure assessor performance relative to others in the panel. Multiplicative models are discussed in Section 3, and extended to series of experiments in Section 4. The apple data are analysed in Section 5. Section 6 discusses the analysis of a future experiment given information about assessor performance from previous experiments, and this is followed by a general discussion.

2 Apple-tasting data

The data studied here were obtained from the Hannah Research Institute in Ayr, Scotland. A series of apple-tasting experiments was conducted over several years with the aim of monitoring assessor performance over time. Data sets from 45 of the experiments spanning the period March 1996 to April 1998 were available for analysis. For these studies, 14 trained female assessors were involved: the number of assessors attending individual experiments varied from 9 to 14 per experiment, and the number of experiments attended from 11 to 43.

In each experiment, several apple varieties were tasted, the choice of varieties depending on the availability in the local market at the time of the experiment. Varieties were sometimes tasted in several versions, with different countries of origin or methods of packaging, and there were some

0119 inconsistencies in recording this information. Thus for consistency we treat the combinations of
0120 varieties and versions as unique to every experiment, and refer to them as *products*. This results
0121 in between 4 and 16 products per experiment (most often 8), corresponding to between 4 and
0122 11 varieties. The order followed for presenting products to assessors within a tasting session
0123 was based on the rows of Williams Latin-square designs (Williams (1949), MacFie, Bratchel,
0124 Greenhoff and Vallis (1989)). A Williams Latin square is such that every product occurs equally
0125 often in every position, and is preceded by every other product equally often; also each assessor
0126 receives every product the same number of times. The last of these conditions is satisfied even
0127 when the number of assessors is not a multiple of the number of products, as in most of the
0128 experiments considered here. The allocation was replicated 3 times for each assessor (with a few
0129 exceptions), and the replicates were separated by break periods the lengths of which were not
0130 reported.

0141 Each assessor was allocated a private booth in which to do the tasting, and she entered the
0142 scores by dragging a mouse pointer along a bar marked from 0 to 100 on a computer screen.
0143 Each product was scored on 11 attributes: fruitiness, sweetness, acidity, bitterness, presence
0144 of perfumed smell, floral sensation, after-taste, persistence, hardness, crunchiness and overall
0145 acceptability of the product. We present analyses only for the sweetness scores which appear to
0146 be roughly normally distributed: some of the other attributes had scores with high proportions
0147 of zeros, for which an assumption of normality would not be appropriate.

0148 A combined analysis of the sweetness attribute from the 45 experiments was carried out in order
0149 to

- 0150 • model assessor performance across experiments
- 0151 • adjust mean product effects within experiments for missing assessors
- 0152 • incorporate information on assessors' use of the sweetness scale to improve precision in the
0153 analysis of future experiments which include the same assessors.

3 Multiplicative models for assessor performance in individual experiments

Yates and Cochran (1938) and Finlay and Wilkinson (1963) introduced multiplicative models which are widely used in agricultural trials to model heterogeneity of variety variance between growing environments. Brockhoff and Skovgaard (1994) drew an analogy between variety trials and sensory experiments in which a panel of assessors evaluate products, and proposed the use of multiplicative models for sensory studies.

Consider first unreplicated sensory data in which y_{ij} denotes the attribute score of assessor i for product j , and there is no standard measure of attribute intensity to which the scores can be compared. It is often found that there is heterogeneity of product and error variances between assessors. When that is the case, a model which takes this heterogeneity into account by including a multiplicative term may be used. This is given by

$$y_{ij} = \alpha_i + \beta_i \theta_j + \epsilon_{ij}, \quad (1)$$

where α_i is the additive effect of assessor i , θ_j is the effect of product j , β_i is a coefficient for assessor i reflecting her use of the measurement scale and ϵ_{ij} is a random error term, assumed to be normally distributed with zero mean. Thus, the interaction between assessor and product is modelled by a regression on the unknown product effects with coefficients β_i . This model can allow for unequal residual variances for assessors, so that $\text{var}(\epsilon_{ij}) = \sigma_i^2$. Theobald and Mallinson (1978) considered model (1), with normally distributed θ_j and unequal residual variances, in the context of comparative calibration of measuring instruments. They noted a connection with factor analysis models having a single common factor, and emphasised precision, defined as $\pi_i = \beta_i/\sigma_i$, as a measure of the performance of instrument i .

The β_i were interpreted by Brockhoff and Skovgaard (1994) as measures of assessors' discriminating ability, and called *assessor expansiveness* by Mead and Gay (1995). If assessor i is good at discriminating she will have low residual standard deviation σ_i relative to β_i , and thus a large value of π_i .

Brockhoff and Skovgaard (1994) also suggested how this multiplicative model may be generalised to account for the whole design structure when each product is tasted more than once, that is,

0237 in a replicated experiment. Thus, when replication effects are given index r , their model has
0238
0239 the form

$$0240 \quad y_{ijr} = \alpha_i + \beta_i(\theta_j + \gamma_r) + \epsilon_{ijr}, \quad (2)$$

0241
0242
0243
0244
0245 where γ_r is a replication effect and this, like the product effect, is realized through the assessors'
0246 use of the scale.
0247

0248
0249
0250 Frequentist methods for fitting such multiplicative models were explored by Digby (1979), Oman
0251 (1991), Gogel, Cullis and Verbyla (1995) and Nabugoomu, Kempton and Talbot (1999).

0252
0253
0254
0255 Smith, Cullis, Brockhoff and Thompson (2003) extended model (2) to allow for more general
0256 designs and unbalanced data. Their models may include fixed effects in addition to assessor
0257 effects, and random effects other than those for products. The single multiplicative term in (1)
0258 is generalised to a sum of such terms, corresponding to a factor analysis model with several
0259 common factors. Model fitting is by REML. When the error variances σ_i^2 of assessors are
0260 allowed to differ, methods such as REML and maximum likelihood have the disadvantage that
0261 some variance estimates may be zero. For example, the data presented in Smith et al. (2003)
0262 lead to many zero estimates and to very dissimilar estimates of assessor expansiveness.
0263
0264
0265
0266
0267
0268
0269
0270

0271 Our interest is in extending model (2) to apply to series of experiments on products of the same
0272 type, and then to future experiments. We seek also to avoid unstable estimates of error variances
0273 and assessor expansiveness.
0274
0275
0276
0277

0278 **4 A multiplicative model for series of experiments**

0279
0280
0281 Plots of mean sweetness scores over replicates versus products for four of the experiments are
0282 shown in Figure 1. The different lines do not correspond to the same assessors in all four plots
0283 because attendance was not the same for all experiments. The plots show general consistency
0284 between assessors within experiments in terms of the ordering of products on the scale from 0
0285 to 100, but there appear to be differences in mean scores over products and in expansiveness.
0286
0287
0288
0289
0290 Within experiments, a model with multiplicative interaction effects, like (1) or (2), is therefore
0291 more suitable than an additive model. A combined model for all experiments may also be
0292 considered, with common values over experiments for some parameters, such as error variances.
0293
0294
0295

We want to make inferences about product effects in future experiments, so it is necessary to treat these effects as random, as in Smith et al. (2003). We go further and adopt the Bayesian approach which considers *all* model parameters as randomly sampled from probability distributions. This approach is intended to

- avoid unreasonable estimates of error variances and expansiveness
- avoid the use of large-sample approximations for inferences
- permit the inclusion of expert knowledge into the analysis.

Expert knowledge of similar sensory studies is incorporated by specifying prior distributions for variance components, and for the overall expectation of the responses.

In order to extend model (1) to a multiplicative model for a series of experiments, we allow the product effects and the additive and multiplicative effects of assessors to depend on the experiment. Thus model (1) is generalized to

$$y_{ijk} = \alpha_{ik} + \beta_{ik} \theta_{j(k)} + \epsilon_{ijk}, \quad (3)$$

where α_{ik} is the effect of assessor i in experiment k , β_{ik} is assessor i 's expansiveness in experiment k , $\theta_{j(k)}$ is the effect of product j within that experiment and the error terms ϵ_{ijk} are assumed to be distributed as $N(0, \sigma_i^2)$ independently. The effects α_{ik} for the i th assessor are assumed to be normally distributed about her mean effect α_i , and similarly the β_{ik} for any i are taken to be normal with mean β_i . Product effects are nested within experiments, so for any k the $\theta_{j(k)}$ are assumed to be normally distributed about a mean product effect ϕ_k for experiment k . The experiment effects are thus realized via the assessor's expansiveness, like the replicate effects γ_r in (2).

An alternative notation might express α_{ik} as the sum of α_i and an interaction effect δ_{ik} , say, of mean zero, with similar decompositions for the β_{ik} and $\theta_{j(k)}$. The parametrization chosen for (3) is an example of *hierarchical centring* (Gelfand, Sahu and Carlin (1995)), which has been shown to hasten convergence when fitting Bayesian linear models.

Table 1 shows the structure assumed for the parameters in model (3), and also includes the prior distributions. Greek and Roman letters indicate respectively unknown parameters and values

0355 specifying prior distributions. The subscripts A, B, P, E refer to the populations of assessor ef-
0356 fects, assessor expansiveness, product effects and experiment effects, while \sim means ‘distributed
0357 as’ or ‘distributed independently as’, according to the context. The unknown parameters in
0358 the model are assumed to be statistically independent *a priori* unless the dependence is made
0359 explicit in Table 1.
0360
0361
0362
0363
0364

0365 The first two entries in the top row of Table 1 express the α_{ik} and α_i as normally distributed
0366 random effects with expectations α_i and μ_A respectively and variance components σ_{AE}^2 and σ_A^2 .
0367 The third entry defines the prior distribution for μ_A as normal, and the entries in the second row
0368 define those for σ_{AE}^2 and σ_A^2 . Various choices could be made for the latter distributions, but the
0369 most convenient mathematically is that the reciprocals of all the variance components are given
0370 gamma or equivalently scaled- χ^2 distributions. Following Theobald, Talbot and Nabugoomu
0371 (2002), these distributions are conveniently specified using a prior estimate of the variance and
0372 a corresponding degrees-of-freedom parameter, larger values of this parameter indicating greater
0373 confidence in the estimate. For example, σ_A^2 is given a prior estimate denoted by s_A^2 with degrees
0374 of freedom d_A , and the notation $\sigma_A^2 \sim IC(s_A^2, d_A)$ means that $d_A s_A^2 \sigma_A^{-2}$ has the distribution
0375 $\chi^2(d_A)$, so that s_A^{-2} is interpretable as the expectation of σ_A^{-2} in the prior distribution.
0376
0377
0378
0379
0380
0381
0382
0383
0384
0385

0386 The third and fourth rows in Table 1 relate to the expansiveness parameters, and are similar
0387 to the first two rows, except that the β_i are centred around a value of 1. This is necessary in
0388 order to fix the scale of these parameters, since in models (1) and (3) multiplying all the β s by
0389 a positive constant and dividing the θ s by the same constant would leave the right hand side
0390 unchanged. The product effects are similarly made identifiable by centring their means ϕ_k on
0391 zero. The final entry in Table 1 specifies a common prior distribution for the assessors’ residual
0392 variances.
0393
0394
0395
0396
0397
0398
0399
0400

0401 **5 Combined analysis of apple-tasting experiments**

0402 The data on the sweetness of apples from the 45 experiments were combined using means over
0403 replicates within combinations of assessor and product rather than using raw scores directly.
0404 This simplification causes a slight loss of efficiency because of missing values, but it is the usual
0405 procedure for combining data for meta-analysis, such as in analysis of variety trials in the UK.
0406 We note, though, that Smith, Cullis and Thompson (2001) advocate analysing individual plot
0407
0408
0409
0410
0411
0412
0413

0414 yields for such trials. From now on, we treat these means as our responses.
0415
0416

0417 To apply the model of Section 3 to these data, we specify the prior mean m_A and variance r_A^2 of
0418 μ_A in Table 1 as 33 and 30. The prior estimates of the variance components and their degrees of
0419 freedom are given in Table 2. These values were obtained from an expert who had been involved
0420 in the programme of sensory studies carried out at the Hannah Research Institute (E A Hunter,
0421 personal communication). The few degrees of freedom for some of the variances indicate his lack
0422 of experience with analysing series of experiments rather than individual ones.
0423
0424
0425
0426
0427
0428

0429 The Bayesian analysis was carried out using the WinBUGS program, which is freely avail-
0430 able from <http://www.mrc-bsu.cam.ac.uk/bugs>. Using an iterative process known as *Markov*
0431 *Chain Monte Carlo* (Gilks, Richardson and Spiegelhalter (1996)), it generates long sequences
0432 of values drawn from the joint posterior distribution of the unknown parameters, that is their
0433 distribution given the prior distribution and the information in the data. These sequences can
0434 then be used to approximate the posterior distributions of individual parameters.
0435
0436
0437
0438
0439
0440

0441 Summaries of the posterior distribution may include expectations and standard errors for in-
0442 dividual parameters: such summaries are given for the parameters specific to the assessors in
0443 Model (3) in Table 3. The posterior expectations for the mean effects α_i of assessors have a
0444 range of about 50, suggesting large differences in the perception of sweetness. Thus the absence
0445 of certain assessors from a particular experiment could substantially influence the mean scores
0446 for products. The large standard error for assessor 12 is due to the fact that she left the panel
0447 of assessors after the first 16 experiments. The estimates of precision for assessors 1, 2, 5, 6,
0448 7, 8 and 13 suggest that they have high discriminating abilities. Using the estimates of expan-
0449 siveness or standard deviation alone leads to different judgements about which assessors are the
0450 best discriminators.
0451
0452
0453
0454
0455
0456
0457
0458
0459
0460

0461 Systematic effects of experiment number might arise from changes in the source or composition
0462 of products. Such effects for apple sweetness may be investigated by plotting the posterior
0463 expectations of the mean product effects within experiments ϕ_k against the experiment numbers
0464 k , which are in date order. Considering these effects rather than the simple mean scores for
0465 experiments corrects for the absence of particular assessors from some of them: the posterior
0466 expectations of the individual product effects $\theta_{j(k)}$ are corrected in the same way. The plot is
0467
0468
0469
0470
0471
0472

0473 given in Figure 2. A downward trend from the first experiment is followed by a sharp increase
0474 around the twentieth, and then a slight decline to the last. No firm explanation can now be
0475 found for the sudden increase, but it occurs at the start of a set of seven experiments in each
0476 of which only four products were tasted. A partial auto-correlation plot revealed no significant
0477 auto-correlation between the posterior expectations of the ϕ_k .
0478
0479
0480
0481
0482
0483

0484 **6 Analysis of future experiments**

0486 By using the knowledge gained about the assessors from a combined analysis of data from
0487 several experiments, as described in Sections 4 and 5, it should be possible to make more precise
0488 comparisons of product effects in future experiments which use the same product type and pool
0489 of assessors, and to correct estimated product effects for missing assessors. To adapt model (3)
0490 for future experiments, the product effects are assumed to have the same structure as those for
0491 the past series, with mean product effects drawn from the same population as the past ones.
0492 We use f to replace the index k identifying the experiment and e to replace j as the product
0493 index, so that the future responses by assessor i , denoted by y_{ief} , satisfy
0494
0495
0496
0497
0498
0499
0500

$$0501 \quad y_{ief} = \alpha_{if} + \beta_{if} \theta_{e(f)} + \epsilon_{ief}. \quad (4)$$

0502
0503 By analogy with the fifth row of Table 1, future product effects $\theta_{e(f)}$ are then assumed to be
0504 drawn from a Normal population $N(\phi_f, \sigma_{PE}^2)$, where ϕ_f is itself drawn from $N(0, \sigma_E^2)$.
0505
0506
0507

0508 It would also be possible to extend this model further to a future experiment that included
0509 some assessors who did not take part in the past experiments, so long as their mean effects,
0510 expansiveness and variances could be assumed to be drawn from the same distributions as for
0511 the existing assessors.
0512
0513
0514
0515
0516
0517
0518

0519 To illustrate the possible benefit from including information provided by the analysis of previ-
0520 ous experiments with the same assessors, we compare analyses of data from a single experiment
0521 (treated as a future one) first ignoring information on model parameters from the other exper-
0522 iments and then including this information. Since the main interest in sensory experiments is
0523 likely to be to establish differences between products, we measure the reduction in the average
0524 variance (over all pairs of products) of product differences in the posterior distribution. If a
0525
0526
0527
0528
0529
0530
0531

and b denote any two of p products assessed in the future experiment then the average posterior variance of the $p(p - 1)$ product differences is given by

$$\frac{1}{p(p - 1)} \sum_{a \neq b} \text{var}(\theta_{a(f)} - \theta_{b(f)}), \quad (5)$$

where the variance relates to the posterior distribution. Theobald et al. (2002) show that (5) may be simplified to

$$\frac{2}{p - 1} \left\{ \sum_{e=1}^p \text{var}(\theta_{e(f)}) - p \text{var}(\bar{\theta}_f) \right\}, \quad (6)$$

where $\bar{\theta}_f$ denotes the mean of the $\theta_{e(f)}$. Thus the inclusion of past information on assessors is measured by examining the reduction in (6).

For the data on the sweetness of apples, the effect of including past information is illustrated by using the entire series as past data and one experiment which all 14 of the assessors happened to attend as the future experiment. The data from this experiment are thus used twice, but their influence on the posterior distribution from the combined analysis is slight since it arises from only one of 45 experiments.

The comparison under the heading ‘Expert prior distribution’ in Table 4 was made assuming the prior distribution specified in Section 5. In order to examine the robustness of this comparison to changes in the prior distribution, a second comparison was made after dividing by 5 the degrees of freedom corresponding to all the variance components: this is shown under the heading ‘Diffuse prior distribution’. As expected, the latter prior distribution gives a higher average posterior variance of product differences, but in both cases including information from other experiments results in a reduction of over 50% in the average variance of the product differences.

7 Discussion

We consider a situation in which quantitative sensory data are available from several experiments on similar products with overlapping sets of assessors. In such a situation, it should be possible to extend the usual analysis of individual experiments by combining the information they contain on assessor effects and variance components. Such extensions include

0591
0592
0593
0594
0595
0596
0597
0598
0599
0600
0601
0602
0603
0604
0605
0606
0607
0608
0609
0610
0611
0612
0613
0614
0615
0616
0617
0618
0619
0620
0621
0622
0623
0624
0625
0626
0627
0628
0629
0630
0631
0632
0633
0634
0635
0636
0637
0638
0639
0640
0641
0642
0643
0644
0645
0646
0647
0648
0649

- examining any pattern in the scores over time
- adjusting estimates to account for missing assessors
- using the information gained on assessors in the past to obtain more precise analyses of future experiments.

We have shown how these extensions can be achieved in the context of a multiplicative model for normally distributed scores. Such models can be expected to fit better than additive ones when the assessors differ in how much of the recording scale they use, but fitting multiplicative models in the usual frequentist framework can lead to zero estimates of variance components and unstable estimates of assessor expansiveness. The Bayesian framework offers a solution to these problems while incorporating expert knowledge into the analysis, and it deals with missing data and prediction for future experiments.

The effects of missing assessors on product means can be corrected by using assessors' responses in the experiments that they did attend. Incorporating information from past analyses into future ones is also seen to increase precision by reducing the average variance of product differences.

One objection to Bayesian modelling is that choosing appropriate prior distributions may be difficult. In cases where prior information seems unreliable, it may be given a low weight by using small degrees of freedom for the variance parameters, and this is better than ignoring prior information completely.

Acknowledgements

This work was done as part of the first author's PhD study which was funded by the Cecil Renaud Charitable and Educational Trust Fund, South Africa. This work was also supported by funds from the Scottish Executive Environment and Rural Affairs Department. The apple-tasting data were kindly provided by the Hannah Research institute in Ayr, Scotland. We are grateful to the late Rob Kempton for suggesting this study and to E A Hunter for his advice on the analysis of apple-tasting data.

References

0709
0710
0711
0712 Brockhoff, P. M. and Skovgaard, M. (1994). Modelling individual differences between assessors
0713 in sensory evaluations. *Food Quality and Preference*, 5, 215-224.
0714

0715
0716
0717 Digby, P. G. N. (1979). Modified joint regression analysis for incomplete variety by environment
0718 data. *Journal of Agricultural Science*, 93, 81-86.
0719

0720
0721
0722 Finlay, K. W. & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding pro-
0723 gramme. *Australian Journal of Agricultural Research*, 14, 742-754.
0724

0725
0726
0727 Gelfand, A. E., Sahu, S. K. & Carlin, B. P. (1995). Efficient parametrizations for normal linear
0728 mixed models. *Biometrika*, 82,479-488.
0729

0730
0731
0732 Gilks, W.R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Prac-*
0733 *tice*. London: Chapman and Hall.
0734

0735
0736
0737 Gogel, B. J., Cullis, B. R. & Verbyla, A. P. (1995). REML estimation of multiplicative effects
0738 in multi-environment variety trials. *Biometrics*, 51, 744-749.
0739

0740
0741
0742 MacFie, H.J.H, Bratchell, N., Greenhoff, K. & Vallis, L. V. (1989). Designs to balance the order
0743 of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, 4,
0744 129-148.
0745

0746
0747
0748 McEwan, J. A., Hunter, E. A., van Gemert, L. J. & Leah, P. (2002). Proficiency testing for
0749 sensory profile panels: measuring panel performance. *Food Quality and Preference*, 13, 181-190.
0750

0751
0752
0753 Mead, R. & Gay, C. (1995). Sequential design of sensory trials. *Food Quality and Preference*,
0754 6, 271-280.
0755

0756
0757
0758 Nabugoomu, F., Kempton, R. A. & Talbot, M. (1999). Analysis of series of trials where va-
0759 rieties differ in sensitivity to locations. *Journal of Agricultural, Biological and Environmental*
0760
0761
0762
0763
0764
0765
0766
0767

0768 Statistics, 4(3), 310-325.
0769
0770
0771
0772 Naes, T. (1998). Detecting individual differences among assessors and differences among repli-
0773 cates in sensory profiling. *Food Quality and Preference*, 9(3), 107-110
0774
0775
0776
0777 Oman, S. D. (1991). Multiplicative effects in mixed model analysis of variance. *Biometrika*,
0778 78(4), 729-739.
0779
0780
0781
0782
0783 Rossi, F. (2001). Assessing sensory panelist performance using repeatability and reproducibility
0784 measures. *Food Quality and Preference*, 12, 467-479.
0785
0786
0787
0788 Smith, A. B., Cullis, B. R., & Thompson, R. (2001). Analyzing variety by environment data
0789 using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57, 1138-
0790 1147.
0791
0792
0793
0794
0795 Smith, A., Cullis, B., Brockhoff, P. & Thompson, R. (2003). Multiplicative mixed models for
0796 the analysis of sensory evaluation data. *Food Quality and Preference*, 14, 387-395.
0797
0798
0799
0800
0801 Theobald, C. M. & Mallinson, J. R. (1978). Comparative calibration, linear structural relation-
0802 ships and congeneric measurements. *Biometrics*, 34, 39-45.
0803
0804
0805
0806
0807 Theobald, C. M., Talbot, M. & Nabugoomu, F. (2002). A Bayesian approach to regional and
0808 local-area prediction from crop variety trials. *Journal of Agricultural, Biological and Environ-*
0809 *mental Statistics*, 7(3), 403-419.
0810
0811
0812
0813
0814 Williams, E. J. (1949). Experimental Designs Balanced for the Estimation of Residual Effects
0815 of Treatments. *Australian Journal of Scientific Research*, 2(2), 149-168.
0816
0817
0818
0819
0820 Yates, F. & Cochran, W.G. (1993). The Analysis of Groups of Experiments. *Journal of Agri-*
0821 *cultural Science*, 28, 556-580.
0822
0823
0824
0825
0826

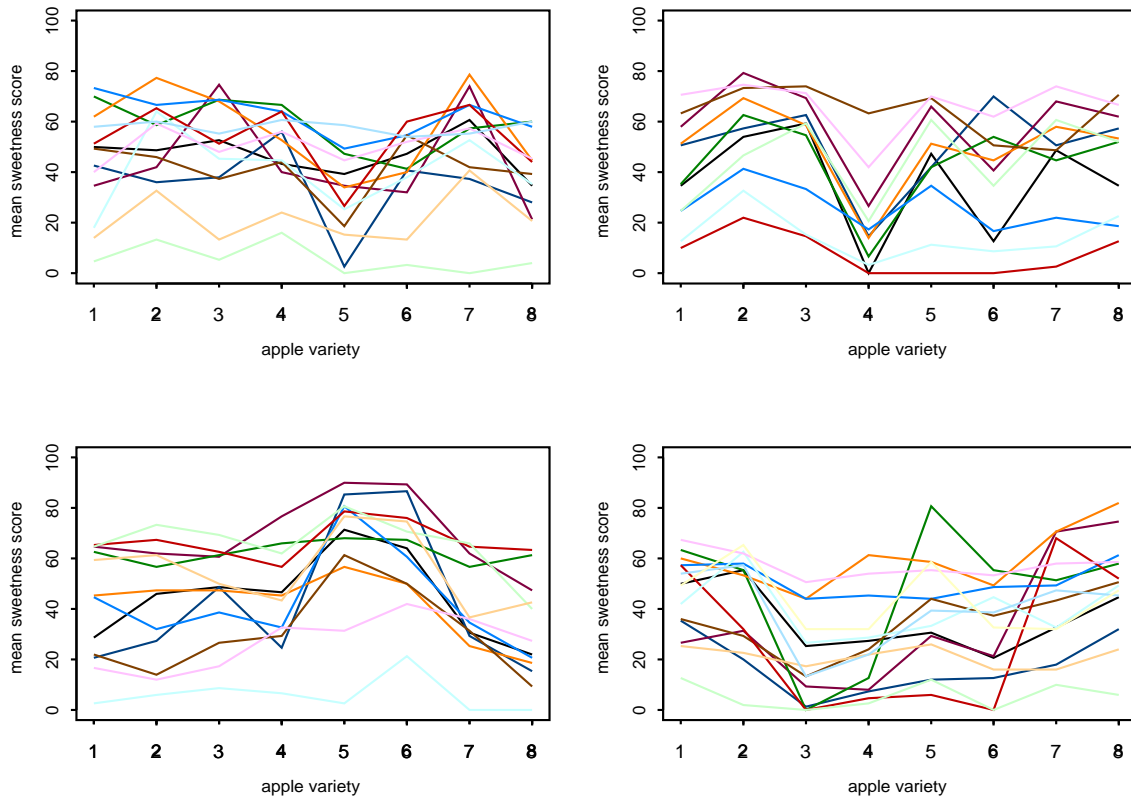
Figure captions

Figure 1: Plots of mean sweetness scores against apple varieties for four experiments

Figure 2: Plot of posterior expectations of mean product effects within experiments against experiment number

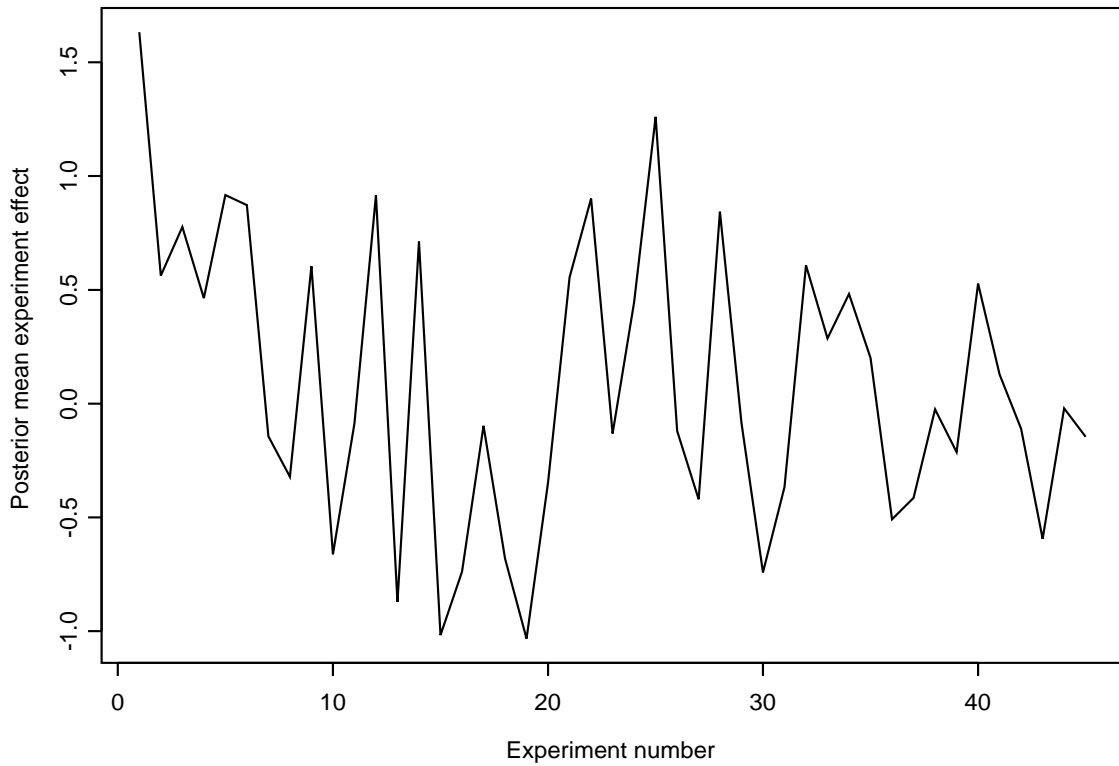
0886
0887
0888
0889
0890
0891
0892
0893
0894
0895
0896
0897
0898
0899
0900
0901
0902
0903
0904
0905
0906
0907
0908
0909
0910
0911
0912
0913
0914
0915
0916
0917
0918
0919
0920
0921
0922
0923
0924
0925
0926
0927
0928
0929
0930
0931
0932
0933
0934
0935
0936
0937
0938
0939
0940
0941
0942
0943
0944

Figure 1: Plots of mean sweetness scores against apple varieties for four experiments



0945
0946
0947
0948
0949
0950
0951
0952
0953
0954
0955
0956
0957
0958
0959
0960
0961
0962
0963
0964
0965
0966
0967
0968
0969
0970
0971
0972
0973
0974
0975
0976
0977
0978
0979
0980
0981
0982
0983
0984
0985
0986
0987
0988
0989
0990
0991
0992
0993
0994
0995
0996
0997
0998
0999
1000
1001
1002
1003

Figure 2: Plot of posterior expectations of mean product effects within experiments against experiment number



Tables

Table 1: Structure of the parameters in multiplicative model (3) for combined data. The notation $\sigma^2 \sim IC(s^2, d)$ means that $d s^2 \sigma^{-2}$ has the distribution $\chi^2(d)$.

$\alpha_{ik} \sim N(\alpha_i, \sigma_{AE}^2)$	$\alpha_i \sim N(\mu_A, \sigma_A^2)$	$\mu_A \sim N(m_A, r_A^2)$
$\sigma_{AE}^2 \sim IC(s_{AE}^2, d_{AE})$	$\sigma_A^2 \sim IC(s_A^2, d_A)$	
$\beta_{ik} \sim N(\beta_i, \sigma_{BE}^2)$	$\beta_i \sim N(1, \sigma_B^2)$	
$\sigma_{BE}^2 \sim IC(s_{BE}^2, d_{BE})$	$\sigma_B^2 \sim IC(s_B^2, d_B)$	
$\theta_{j(k)} \sim N(\phi_k, \sigma_{PE}^2)$	$\phi_k \sim N(0, \sigma_E^2)$	
$\sigma_{PE}^2 \sim IC(s_{PE}^2, d_{PE})$	$\sigma_E^2 \sim IC(s_E^2, d_E)$	$\sigma_i^2 \sim IC(s^2, d)$

1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121

Table 2: Values defining the prior distributions of variance components for model (3) applied to the apple-tasting data.

Variance	Estimate	Degrees of freedom
σ_A^2	264	7
σ_B^2	0.14	9
σ_E^2	30	5
σ_{AE}^2	66	4
σ_{BE}^2	0.0225	5
σ_{PE}^2	40	15
σ_i^2	50	30

Table 3: Expectations and standard errors of the posterior distributions for the parameters of model (3) that relate to individual assessors in the apple-tasting data.

Assessor	Mean effect	Expansiveness	SD	Precision
i	α_i (SE)	β_i (SE)	σ_i (SE)	π_i (SE)
1	48.6 (2.29)	1.26 (0.15)	9.92 (0.46)	0.127 (0.017)
2	26.6 (2.14)	1.01 (0.14)	7.18 (0.38)	0.141 (0.020)
3	30.1 (2.39)	1.31 (0.16)	13.48 (0.64)	0.097 (0.013)
4	47.8 (2.13)	1.31 (0.16)	12.46 (0.56)	0.106 (0.014)
5	65.1 (2.32)	1.19 (0.15)	8.18 (0.43)	0.146 (0.021)
6	57.2 (2.15)	0.78 (0.14)	5.41 (0.28)	0.146 (0.027)
7	40.4 (2.44)	1.03 (0.15)	7.69 (0.43)	0.135 (0.022)
8	35.8 (2.10)	1.44 (0.16)	10.17 (0.51)	0.142 (0.015)
9	28.1 (2.22)	0.80 (0.14)	9.66 (0.43)	0.083 (0.015)
10	60.8 (2.16)	0.65 (0.13)	6.49 (0.29)	0.101 (0.021)
11	15.6 (2.05)	0.82 (0.14)	9.58 (0.49)	0.085 (0.016)
12	22.2 (4.16)	0.54 (0.21)	6.68 (0.53)	0.081 (0.032)
13	53.2 (2.04)	0.81 (0.12)	5.19 (0.26)	0.156 (0.025)
14	47.9 (2.23)	0.93 (0.15)	9.19 (0.45)	0.102 (0.018)

1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239

Table 4: Average posterior variances of product differences from analysing a single apple-tasting experiment ignoring or including data from the other experiments: two prior distributions are assumed

	Expert prior distribution	Diffuse prior distribution
Analysis ignoring other experiments	9.51	12.41
Analysis including data on other experiments	4.49	6.09