UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

# A Bayesian-Frequentists Approach for detecting Outliers in a One-way Variance Components Model

By

A.J. van der Merwe, P.C.N. Groenewald and M.R. Sjölander

Department of Mathematical Statistics and Actuarial Science, University of the Free State, Republic of South Africa

**Abstract**

The most common Bayesian approach for detecting outliers is to assume that outliers are observations which have been generated by contaminating models. An alternative idea was used by Zellner (1975) and Chaloner (1994). They studied the properties of realized regression error terms. Posterior distributions for individual realized errors and for linear and quadratic combinations of them were derived. In this note the theory and results derived by Chaloner (1994) are extended. Since it is not clear to us what the frequentist properties of the Bayesian procedures of Chaloner and Zellner are (i.e. what the size of the type I error or the power of their tests are) a Bayesian-frequentist approach is used for detecting outliers in a one-way variance components model. For illustration purposes, the Sharples (1990) contaminated data are used. It is concluded that the Bayesian frequentist approach seems to be more conservative than Chaloner's method.

**Keywords:**

Random effects model; posterior distributions; outlying observations; Bayesian-frequentist approach; predictive distribution; control limits

## 1. Introduction

The most common Bayesian approach for detecting outliers is to assume that outliers are observations which have been generated by models that are contaminated and which are different from the model that has generated the rest of the data. These contaminating models are usually considered to be either mean-shift or inflated-variance models. Previous research along these lines are given in Box and Tiao (1968), Freeman (1980), Pettit and Smith (1985), Verdinelli and Wasserman (1991) and Hoeting, Raftery and Madigan (1996).

For approaches of Bayesian model checking, see for example Dey, Gelfand, Swartz and Vlachos (1998), O'Hagan (2003), Marshall and Spiegelhalter (2003) and Bayarri and Castellanos (2007).

An alternative idea was used by Zellner (1975). He studied the properties of realized regression error terms. Posterior distributions for individual realized errors and for linear and quadratic combinations of them were derived. Zellner and Moulton (1985) on the other hand used the posterior distributions of the realized error terms to construct a residual plot. The approach used by Chaloner and Brant (1988) is an extention of the ideas applied by Zellner (1975) and Zellner and Moulton (1985). They defined an outlier to be an observation with a large realized error which has been generated by the model under consideration. Chaloner and Brant (1988) calculated the exact posterior probability of an observation being an outlier as well as the joint posterior probability of any two observations being outliers. Let $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \ldots, m,$ and independently of each other. The $\varepsilon_1, \ldots, \varepsilon_m$ are realized errors or residuals. An outlier is defined as any observation with $|\varepsilon_i| > \tilde{k}$ for a suitable value of $\tilde{k}$. If

$$\tilde{k} = \sigma\Phi^{-1}\left\{0.5 + \tfrac{1}{2}\left(0.95^{\frac{1}{m}}\right)\right\} \tag{1.1}$$

then the prior probability of no outlier is 0.95. Given the data, the posterior probabilities can be calculated. According to Chaloner (1994), any observation with a posterior probability, $pr(|\varepsilon_i| > \tilde{k}|data)$ larger than the prior probability $2\Phi(-\tilde{k})$ would be a possible outlier.

In this note the theory and results derived by Chaloner (1994) for the balanced one-way random effects model will be extended. The data that will be used are the Sharples (1990) contaminated data and the possible outliers are indicated by asterisks in Table 1.1.

## 2. The Model and the Example

The balanced one-way random effects model is defined as

$$Y_{ij} = \theta + r_i + e_{ij} \quad (i = 1, \ldots, I,\ j = 1, \ldots, J) \tag{1.2}$$

where

$e_{ij} \sim N(0, \sigma_1^2)$ ; $r_i \sim N(0, \sigma_2^2)$, independently of each other.

Also

$v_1 = I(J-1)$ ; $v_2 = I-1$ ; $Y_{i.} = \sum_{j=1}^{J} Y_{ij}$ ;

$\bar{Y}_{i.} = \frac{1}{J} Y_{i.}$ ; $Y_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}$ ; $\bar{Y}_{..} = \frac{1}{IJ} Y_{..}$ ;

$v_1 m_1 = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(Y_{ij} - \bar{Y}_{i.}\right)^2$ ; $v_2 m_2 = J \sum_{i=1}^{I} (\bar{Y}_{i.} - \bar{Y}_{..})^2$

where $v_1 m_1$ and $v_2 m_2$ are the within and between groups sum of squares respectively.

Example 1.1

Table 1.1: Sharples Generated Data with Possible Outliers indicated by an Asterisk

| Group | Measurements | | | | | | $\bar{Y}_{i.}$ |
|---|---|---|---|---|---|---|---|
| 1 | 24.80 | 26.90 | 26.65 | 30.93 | 33.77 | **63.31*** | 34.39 |
| 2 | 23.96 | 28.92 | 28.19 | 26.16 | 21.34 | 29.46 | 26.34 |
| 3 | 18.30 | 23.67 | 14.47 | 24.45 | 24.89 | 28.95 | 22.46 |
| 4 | **51.42*** | 27.97 | 24.76 | 26.67 | 17.58 | 24.29 | 28.78 |
| 5 | 34.12 | 46.87 | **58.59*** | 38.11 | 47.59 | 44.67 | 44.99 |
| | | | | | | | $\bar{Y}_{..}$ = 31.39 |

$I = 5$ , $J = 6$ , $v_1 = I(J-1) = 25$ , $v_2 = I - 1 = 4$ ,

$v_1 m_1 = 2282.0893$ , $v_2 m_2 = 1837.0937$

$\hat{\sigma}_1^2 = 91.2836$ , $\hat{\sigma}_2^2 = 61.3316$

## 3. Prior and Posterior Distributions

As prior the Jeffreys' independent prior

$p(\theta, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2} (\sigma_1^2 + J\sigma_2^2)^{-1}$

will be used. See for example Box and Tiao (1973, Ch.5). It can easily be shown that, given the variance components, the posterior distribution of $e_{ij} = Y_{ij} - \theta - r_i$ is normal with

$$E\left[e_{ij} \middle| \underline{Y}, \sigma_1^2, \sigma_2^2\right] = Y_{ij} - \frac{J\sigma_2^2 \bar{Y}_{i.} + \sigma_1^2 \bar{Y}_{..}}{\sigma_1^2 + J\sigma_2^2} \text{ and } Var\left[e_{ij} \middle| \underline{Y}, \sigma_1^2, \sigma_2^2\right] = \frac{\sigma_1^2}{IJ} \left\{ \frac{\sigma_1^2 + IJ\sigma_2^2}{\sigma_1^2 + J\sigma_2^2} \right\} \quad (3.1)$$

Also, given the variance components, the posterior distribution of $r_i$ is normal with

$$E\left[r_i \middle| \underline{Y}, \sigma_1^2, \sigma_2^2\right] = \frac{J\sigma_2^2}{\sigma_1^2 + J\sigma_2^2} (\bar{Y}_{i.} - \bar{Y}_{..}) \text{ and } Var\left[r_i \middle| \underline{Y}, \sigma_1^2, \sigma_2^2\right] = \frac{\sigma_2^2 (I\sigma_1^2 + J\sigma_2^2)}{I(\sigma_1^2 + J\sigma_2^2)} \quad (3.2)$$

The posterior distribution of the variance components is given by

$$p(\sigma_1^2, \sigma_2^2 | \underline{Y}) \propto (\sigma_1^2)^{-\frac{1}{2}(v_1+2)} (\sigma_1^2 + J\sigma_2^2)^{-\frac{1}{2}(v_2+2)} exp\left\{-\frac{1}{2}\left[\frac{v_1 m_1}{\sigma_1^2} + \frac{v_2 m_2}{\sigma_1^2 + J\sigma_2^2}\right]\right\}$$

$$\sigma_1^2 > 0 \ ; \ \sigma_2^2 > 0 \text{ and } \sigma_1^2 + J\sigma_2^2 = \sigma_{12}^2 > \sigma_1^2. \quad (3.3)$$

The posterior distribution of the variance components as well as that of $\frac{e_{ij}}{\sigma_1}$ and $\frac{r_i}{\sigma_2}$ are given in Chaloner (1994).

Simulation of the variance components follow easily from (3.3) as follows:

Since $\frac{v_1 m_1}{\sigma_1^2} \sim \chi_{v_1}^2$ and $\frac{v_2 m_2}{\sigma_{12}^2} \sim \chi_{v_2}^2$, $\sigma_1^2$, $\sigma_{12}^2$ and $\sigma_2^2 = \frac{\sigma_{12}^2 - \sigma_1^2}{J}$ can easily be simulated. If a negative value of $\sigma_2^2$ is obtained, disregard this value as well as the corresponding $\sigma_1^2$ value. It is our opinion that this is the best method to simulate the variance components if the percentage of negative $\sigma_2^2$ values are zero or very small. The simulated values can then be substituted in the conditional normal distributions (for example equations (3.1) and (3.2)). Repeat the above steps until $\tilde{l}$ permissible values are obtained. For the Sharples data $\tilde{l}$ will be taken as 100 000. By averaging the $\tilde{l}$ conditional distributions (Rao-Blackwell method), the unconditional posterior distributions can be obtained.

## 4. The Bayesian Method of Chaloner

Chaloner (1994) defined two sets of residuals:

The within-group residuals $\frac{e_{ij}}{\sigma_1}$ and the $\frac{r_i}{\sigma_2}$ between-group residuals. The within-group residuals measure how far $Y_{ij}$ is from its mean and the between-group residuals measure how far $r_i$ is from zero.

The unconditional posterior distributions of these residuals can easily be obtained by simulating the variance components from equation (3.3). Using equation (1.1) for a sample of size 30 gives $\tilde{k} = 3.14$. The prior probability of an observation being an outlier is therefore 0.0017. The posterior probabilities that observations $Y_{16}$, $Y_{41}$ and $Y_{53}$ are outliers are

(a) $P\left(\left|\frac{e_{16}}{\sigma_1}\right| > \tilde{k} | \underline{Y}\right) = 0.43810$, (b) $P\left(\left|\frac{e_{41}}{\sigma_1}\right| > \tilde{k} | \underline{Y}\right) = 0.04963$ and

(c) $P\left(\left|\frac{e_{53}}{\sigma_1}\right| > \tilde{k} | \underline{Y}\right) = 0.00059$. The corresponding Bayes factors are

(i) $\widetilde{BF}_{16} = \frac{0.43810}{0.0017} = 257.71$, (ii) $\widetilde{BF}_{41} = \frac{0.04963}{0.0017} = 29.19$ and (iii) $\widetilde{BF}_{53} = \frac{0.00059}{0.0017} = 0.35$.

In the following table, Jeffreys' interpretation of Bayes factors are given (Jeffreys (1961)).

**Table 4.1: Jeffreys' Scale of Evidence for Bayes Factor $BF_{12}$**

| **Bayes Factor** | **Interpretation** |
|---|---|
| $BF_{12}<1$ | Negative support for Model $M_1$ |
| $1<BF_{12}<3$ | Barely worth mentioning support for $M_1$ |
| $3<BF_{12}<10$ | Substantial evidence for $M_1$ |
| $10<BF_{12}<30$ | Strong evidence for $M_1$ |
| $30<BF_{12}<100$ | Very strong evidence for $M_1$ |
| $BF_{12}>100$ | Decisive evidence for $M_1$ |

The posterior probabilities of observations $Y_{16}$ and $Y_{41}$ are larger than 0.0017. Observations $Y_{16}$ and $Y_{41}$ are therefore possible outliers. The Bayes factors in Table 4.1 also give strong and decisive evidence that this is the case.

Similar results can be obtained for the between-group residuals. Since there are only five groups, it follows from equation (1.1) that for $m = 5$, $\tilde{k} = 2.57$ the prior probability that any one of the groups will be outlying is 0.0102. The posterior probabilities are (a) $P\left(\left|\frac{r_1}{\sigma_2}\right| > \tilde{k}|\underline{Y}\right) = 0.000431$, (b) $P\left(\left|\frac{r_3}{\sigma_2}\right| > \tilde{k}|\underline{Y}\right) = 0.00614$ and

(c) $P\left(\left|\frac{r_5}{\sigma_2}\right| > \tilde{k}|\underline{Y}\right) = 0.03393$. It therefore seems that Group 5 is a possible outlier because its posterior probability is larger than the prior probability of 0.0102. The corresponding Bayes factors are

(i) $\widetilde{BF}_1 = \frac{0.000431}{0.0102} = 0.042$, (ii) $\widetilde{BF}_2 = \frac{0.00614}{0.0102} = 0.60$ and (iii) $\widetilde{BF}_3 = \frac{0.03393}{0.0102} = 3.33$.

Since the Bayes factor of Group 5 is between three and ten, then according to Table 4.1 there is substantial evidence that Group 5 is outlying.

## 5. A Bayesian-Frequentist Approach for Outlier Detection

### 5.1. The Known-Variance Case

Since it is not clear to us what the frequentist properties of the Bayesian procedures of Chaloner and Zellner are (i.e. what the size of the Type I error or the power of their tests are) a Bayesian-frequentist approach will be used for detecting outliers in a one-way random effects model. According to Bayarri and Berger (2004), statisticians should readily use both Bayesians and frequentist ideas. Objective Bayesians and frequentist methods often give similar results for normal linear models. See for example the results in Tables 5.1 and 5.2. Reid and Cox (2014) on the other hand mentioned that "A hybrid method of inference that uses Bayesian reasoning with impersonal priors, if the results are well calibrated in the frequency sense, may be ideal, but to date the construction of these priors is elusive." It is a very desirable situation if the resulting Bayesian procedure will also have good frequentist properties.

From equation (3.2) it follows that the standardised random effect $r_i^* = \frac{r_i}{\sqrt{Var[r_i|\sigma_1^2,\sigma_2^2,data]}}$

is also normally distributed with $E[r_i^*|\sigma_1^2,\sigma_2^2,data] = \frac{J\sigma_2\sqrt{I}}{\sqrt{(\sigma_1^2+J\sigma_2^2)(I\sigma_1^2+J\sigma_2^2)}}(\bar{Y}_{i.}-\bar{Y}_{..})$ and $Var[r_i^*|\sigma_1^2,\sigma_2^2,data] = 1$.

In Table 5.1 the Bayesian results for the random effects are illustrated and in Table 5.2 the SAS printout for the Sharples data are given. Since the true parameter values are not known, the point estimates $\hat{\sigma}_1^2 = 91.2836$ and $\hat{\sigma}_2^2 = 61.3316$ are used in the formulae.

**Table 5.1: Solution for Random Effects - Bayes Procedure - Variances Known**

| Group | $E[r_i\|\sigma_1^2,\sigma_2^2,data]$ | $\sqrt{Var[r_i\|\sigma_1^2,\sigma_2^2,data]}$ | $E[r_i^*\|\sigma_1^2,\sigma_2^2,data]$ |
|---|---|---|---|
| 1 | 2.4048 | 4.6924 | 0.5125 |
| 2 | -4.0492 | 4.6924 | -0.8629 |
| 3 | -7.1607 | 4.6924 | -1.5260 |
| 4 | -2.0915 | 4.6924 | -0.4457 |
| 5 | 10.8967 | 4.6924 | 2.3222 |

Table 5.2: Solution for Random Effects – SAS – Satterthwaite Procedure

| Group | Estimate | Std Err Pred | DF | t Value | Pr>\|t\| |
|---|---|---|---|---|---|
| 1 | 2.4048 | 4.6924 | 6.11 | 0.51 | 0.6263 |
| 2 | -4.0492 | 4.6924 | 6.11 | -0.86 | 0.4207 |
| 3 | -7.1607 | 4.6924 | 6.11 | -1.53 | 0.1769 |
| 4 | -2.0915 | 4.6924 | 6.11 | -0.45 | 0.6711 |
| 5 | 10.8966 | 4.6924 | 6.11 | 2.32 | 0.0585 |

A comparison between the two tables show that $\sqrt{Var[r_i|\sigma_1^2,\sigma_2^2,data]}$ is equal to the standard error of a predictor (Std Err Pred) and the expected value of the standardized random effect $E[r_i^*|\sigma_1^2,\sigma_2^2,data]$ is equal to the t Value in the SAS Printout.

As mentioned the expected values of the standardised residuals will be used as measures for detecting possible outliers. If in a certain data set

$|E[r_i^*|\sigma_1^2,\sigma_2^2,data]| > \tilde{k}$ then $r_i$ will be considered an outlier. The values of $\tilde{k}$ will be obtained from the predictive distribution of $E[r_i^*|\sigma_1^2,\sigma_2^2]$.

### 5.2. The Predictive Distribution of $\boldsymbol{E\left[r_i^* \middle| \sigma_1^2, \sigma_2^2\right] = Y_i^*}$

Let $Y_i^* = \dfrac{J\sigma_2 \sqrt{I}(\overline{Y}_{i.} - \overline{Y}_{..})}{\sqrt{(\sigma_1^2 + J\sigma_2^2)(I\sigma_1^2 + J\sigma_2^2)}}$

where $\overline{Y}_{i.}$ $(i = 1, \dots, 5)$ and $\overline{Y}_{..}$ are considered to be random variables. The predictive distribution of $Y_i^*$ will tell us what possible values $E[r_i^* | \sigma_1^2, \sigma_2^2]$ might take on in future experiments. The following theorem can now be proved:

<u>Theorem 5.1</u>

If $H_0$ is true (i.e if the model is correct) then $\underline{Y}^* = [Y_1^* \quad Y_2^* \quad \dots \quad Y_I^*]'$ given the variance components is multivariate normally distributed with mean $E[Y_i^*] = 0$ and variance $Var[Y_i^*] = \frac{J(I-1)\sigma_2^2}{(I\sigma_1^2 + J\sigma_2^2)}$ $(i = 1, \dots, I)$. The correlation coefficient between $Y_l^*$ and $Y_m^* = \rho_{l,m} = \frac{-1}{I-1}$, $l = 1, \dots, I,\ m = 1, \dots, I, l \neq m$.

The proof is given in Appendix A.

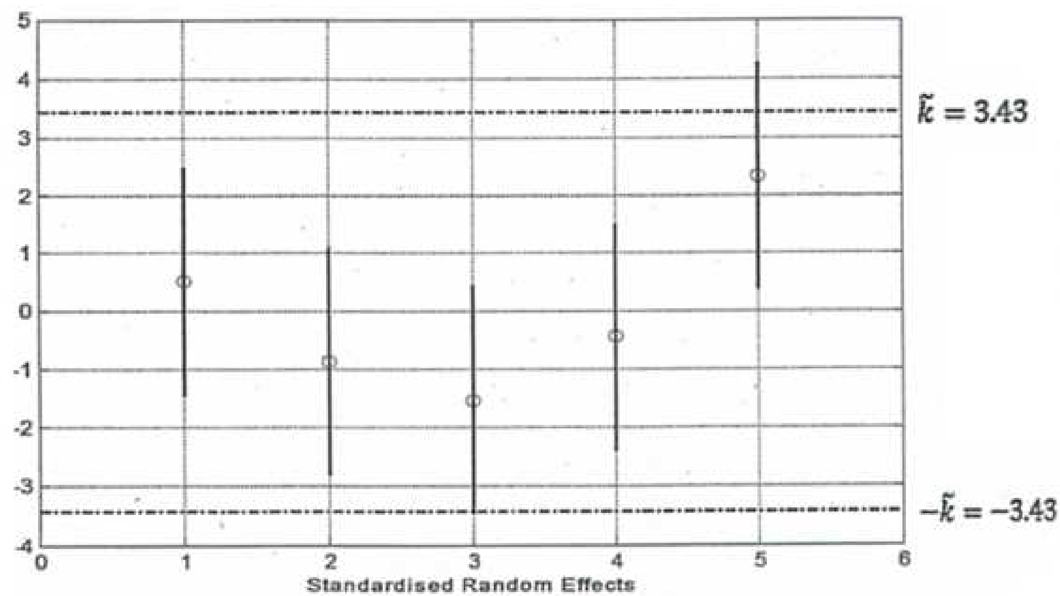
From Theorem 5.1 it follows that if $I = 5$,

$$\tilde{k} = 2.57\sqrt{Var[Y_i^*]} = 2.57\sqrt{\frac{(I-1)J\sigma_2^2}{(I\sigma_1^2 + J\sigma_2^2)}} = 2.57\sqrt{\frac{(4)(6)(61.3316)}{5(91.2836) + 6(61.3316)}} = 3.4341$$

if the point estimates are substituted for the parameter values.

This means that in 95% of future experiments all the $E[r_i^* | \sigma_1^2, \sigma_2^2, data]$, $i = 1, \dots, 5$ will fall between -3.4341 and 3.4341. In Figure 5.1 the control limits for the Sharples data are illustrated.

**Figure 5.1: Means and 95% Intervals for $r_i^*$, $i = 1, \dots, 5$**




According to our method, Group 5 is not an outlying group, because $P(|r_5^*| > 3.4341) = 0.1441$. The mean of Group 5 is 2.32 which is smaller than 3.45.

### 5.3. Unknown Variances

In the unknown variance case, $\tilde{k}$ can be obtained by simulation. It was mentioned in Section 3 that:

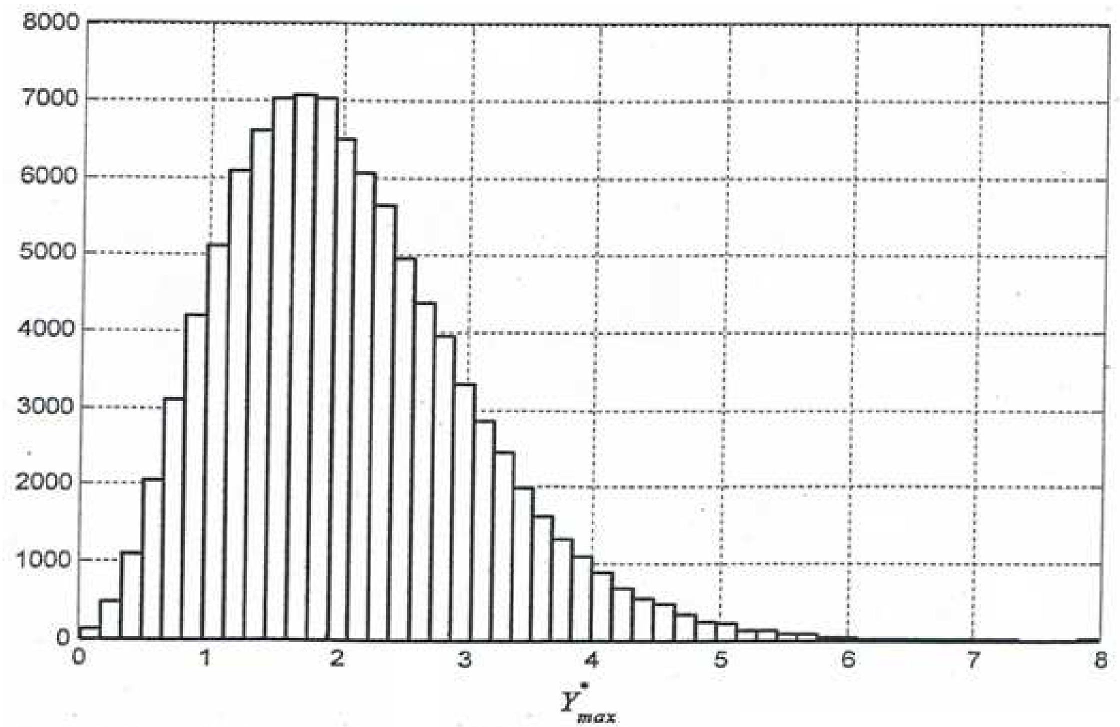
(i) Since $\frac{v_1 m_1}{\sigma_1^2} \sim \chi^2_{v_1}$ and $\frac{v_2 m_2}{\sigma_{12}^2} \sim \chi^2_{v_2}$, $\sigma_1^2$, $\sigma_{12}^2$ and $\sigma_2^2 = \frac{\sigma_{12}^2 - \sigma_1^2}{J}$ can easily be simulated.

(ii) For each pair of simulated variance components $(\sigma_1^2, \sigma_2^2)$ draw $\underline{Y}^*$ from the multivariate normal distribution given in Theorem 5.1. Since $\underline{Y}^*$ is a singular normal distribution, only $I - 1$ random variables can be drawn. For the Sharples data draw $Y_1^*, Y_2^*, Y_3^*, Y_4^*$ and calculate $Y_5^* = 0 - \sum_{l=1}^{4} Y_l^*$.

(iii) Calculate $Y_{max}^* = \max(|Y_i^*|; i = 1,2, \dots, 5)$.

(iv) Repeat steps (i) – (iii) 100 000 times and draw a histogram of $Y_{max}^*$.

**Figure 5.2: Histogram of $Y_{max}^{*}$**




$\boldsymbol{Y_{max}^{*0.95} = 3.8383 = \widetilde{k}}$ **(100 000 simulations)**

The control limits as well as the means and 95% Bayesian confidence intervals for $r_i^*$ are given in Figure 5.3, and in Figure 5.4 the unconditional posterior distributions of $r_i^*$ are displayed. The unconditional posterior distributions are obtained by averaging the conditional posterior distributions, i.e. the Rao-Blackwell method.

**Figure 5.3: Means and 95% Intervals for $r_i^*$, $i = 1, \dots, 5$, Unknown Variances**

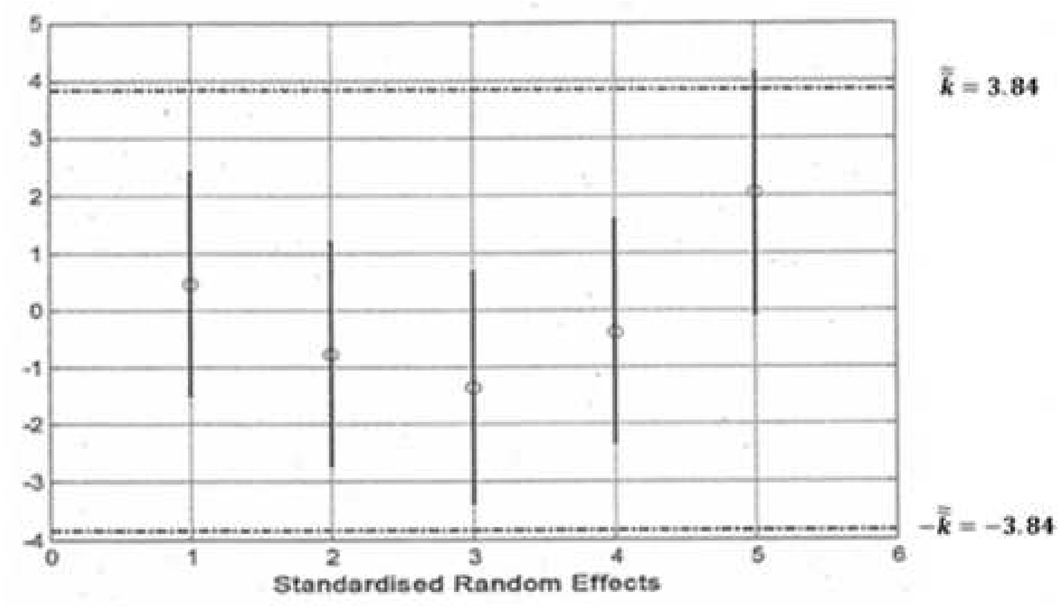
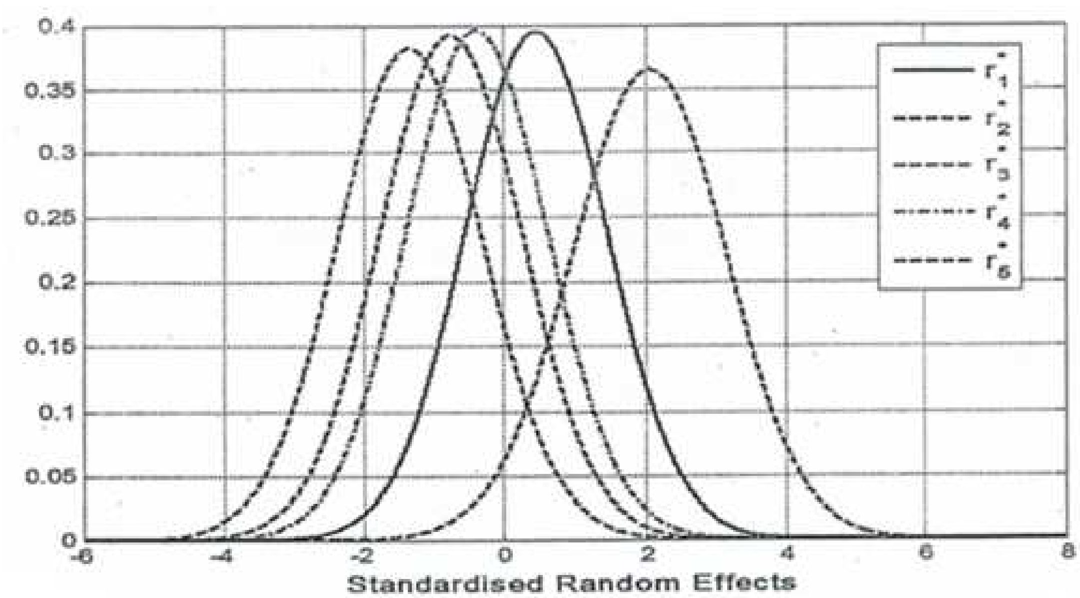



**Figure 5.4: Unconditional Posteriors of $r_i^*$, $i = 1, \dots, 5$**




In Table 5.3 the means, variances and 95% intervals for $r_i^*$ are given, and in the last column the outlying probabilities of the five groups are given.

**Table 5.3: Probabilities that Groups are outlying**

| $r_i^*$ | $Mean(r_i^*)$ | $Var(r_i^*)$ | 95% Interval | $P(Outlier)$ |
|---|---|---|---|---|
| 1 | 0.4507 | 1.0091 | -1.518 – 2.420 | 0.0004 |
| 2 | -0.7589 | 1.0257 | -2.744 – 1.266 | 0.0012 |
| 3 | 1.3421 | 1.0804 | -3.379 – 0.695 | 0.0083 |
| 4 | -0.3920 | 1.0069 | -2.359 – 1.575 | 0.0003 |
| 5 | 2.0424 | 1.1862 | -0.092 – 4.177 | 0.0500 |

Inspection of Table 5.3 shows that the probability that Group 5 is outlying is now 0.0500 which is smaller than 0.1441, the probability for the known variance case. The estimation of the variance components using Monte Carlo simulation leads to more uncertainty and that is the reason for the smaller probability. According to the Bayesian-frequentist procedure there is no reason to believe that any one of the groups are outlying.

A possible criticism of the Bayesian-Frequentist procedure so far could have been the apparent double use of the data. The same data are used for obtaining the posterior distribution of the $r_i^*$ and the predictive distribution of the $Y_i^*$. As mentioned by Bayarri and Castellanos (2007): "This can result in severe conservatism incapable of detecting clearly inappropriate models." See also Bayarri and Berger (2000) and Bayarri and Morales (2003). One way to avoid this double use of the data are to use part of the data for computing the posterior distribution and the rest for prediction. This is what we aim to do in the last part of this section and in the next section. All the data will be used for obtaining the posterior distributions of the standardised random effects and residuals but for prediction purposes the possible outliers will be deleted from the data. By deleting $Y_{16}$, $Y_{41}$ and Group 5 from the data set it was found that $\tilde{k} = 3.44$. The value given in Figure 5.1 where the point estimates are substituted for the variance components is $\tilde{k} = 3.43$. These two values are for all practical purposes the same. It can therefore be concluded that group 5 is not an outlying group. For the unbalanced random effects model the Monte Carlo simulation procedure is somewhat more complicated and will be discussed in the next section.

## 6. <u>Outliers in the Case of Individual Observations</u>

From equation (3.1) it is clear that the standardised residual $e_{ij}^* = \frac{e_{ij}}{\sqrt{Var\left[e_{ij}|\underline{Y},\sigma_1^2,\sigma_2^2\right]}}$ is normally distributed with mean $E\left[e_{ij}^*|\underline{Y},\sigma_1^2,\sigma_2^2\right] = \frac{Y_{ij}-\frac{J\sigma_2^2}{\sigma_1^2+J\sigma_2^2}\bar{Y}_{i.}-\frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\bar{Y}_{..}}{\sqrt{\frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\left(\sigma_2^2+\frac{\sigma_1^2}{IJ}\right)}}$

and variance $Var\left[e_{ij}^*|\underline{Y},\sigma_1^2,\sigma_2^2\right] = 1, i = 1,\dots,I,\ j = 1,\dots,J.$

If for a certain data set $\left|E\left[e_{ij}^*\middle|\sigma_1^2,\sigma_2^2,\underline{Y}\right]\right| > k^*$ then $e_{ij}$ will be considered an outlier. As before the value of $k^*$ will be obtained from the predictive distribution of $E\left[e_{ij}^*\middle|\sigma_1^2,\sigma_2^2\right]$.

### 6.1. The Predictive Distribution of $E\left[e_{ij}^*\middle|\sigma_1^2,\sigma_2^2\right] = Y_{ij}^*$

Let $Y_{ij}^* = \dfrac{Y_{ij} - \frac{J\sigma_2^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{i.} - \frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{..}}{\sqrt{\frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\left(\sigma_2^2+\frac{\sigma_1^2}{IJ}\right)}}$

where $Y_{ij}$, $\overline{Y}_{i.}$ and $\overline{Y}_{..}$ are considered to be future observations, i.e. random variables. The predictive distribution of $Y_{ij}^*$ $(i = 1,\dots,I,\ j = 1,\dots,J)$ will be an indication of what possible values $E\left[e_{ij}^*\middle|\sigma_1^2,\sigma_2^2\right]$ might take on in future experiments. The following Theorem can now be proved.

<u>Theorem 6.1</u>

If the data are generated by the model given in equation (1.2) then $\underline{Y}^* = [Y_{11}^* \quad Y_{12}^* \quad \dots \quad Y_{IJ}^*]'$ will be multivariate normally distributed with mean

$E\left[Y_{ij}^*\middle|\sigma_1^2,\sigma_2^2\right] = 0 \quad (i = 1,\dots,I,\ j = 1,\dots,J)$

$Var\left[Y_{ij}^*\middle|\sigma_1^2,\sigma_2^2\right] = \frac{1}{\sigma_1^2+IJ\sigma_2^2}\{I(J-1)(\sigma_1^2+J\sigma_2^2)+\sigma_1^2(I-1)\}$

$Cov\left[Y_{ij}^*,Y_{lm}^*\middle|\sigma_1^2,\sigma_2^2\right] = \frac{-\sigma_1^2}{\sigma_1^2+IJ\sigma_2^2}$

$Cov\left[Y_{ij}^*,Y_{lj}^*\middle|\sigma_1^2,\sigma_2^2\right] = \frac{-\sigma_1^2}{\sigma_1^2+IJ\sigma_2^2}$

and

$Cov\left[Y_{ij}^*,Y_{im}^*\middle|\sigma_1^2,\sigma_2^2\right] = -1$

The proof is given in Appendix B.

To avoid the double use of the data in the simulation procedure of the variance components, the possible outliers $Y_{16}$, $Y_{41}$ and $Y_{53}$ are deleted from the data set. Group 5 will however not be deleted because it was shown in Section 5 that it is not an outlying group. From this it follows that $n_1 = 5, n_2 = 6, n_3 = 6, n_4 = 5$ and $n_5 = 5$. To simplify the simulation procedure (since the posterior can then be expressed in hierarchical form.) $\gamma = \frac{\sigma_2^2}{\sigma_1^2}$ is first simulated and then $\sigma_1^2$. $\sigma_2^2$ follows from the product of $\gamma$ and $\sigma_1^2$.

Two objective priors will be considered, namely the Probability-matching and Reference priors. These priors often lead to procedures with good frequentist properties. An in-depth discussion of the nature and merits of the Reference and Probability-matching priors lie outside the scope of this article, but the interested reader should consult Berger and Bernardo (1992) as well as Datta and Ghosh (1995).

The Probability-matching prior for the parameters $(\theta, \gamma, \sigma_1^2)$ is given by:

$$\Pr_1(\theta, \gamma, \sigma_1^2) \propto \sigma_1^{-2} \left\{ \sum_{i=1}^{I} \frac{n_i^2}{(1+\gamma n_i)^2} - \frac{1}{n} \left( \sum_{i=1}^{I} \frac{n_i}{1+\gamma n_i} \right)^2 \right\}^{\frac{1}{2}} = \sigma_1^{-2} \Pr_1(\gamma) \tag{6.1}$$

which is also a Reference prior for the parameter groupings $(\theta, \gamma, \sigma_1^2)$, $(\gamma, \theta, \sigma_1^2)$ and $(\gamma, \sigma_1^2, \theta)$.

The Reference prior for the parameter groupings $(\theta, \sigma_1^2, \gamma)$, $(\sigma_1^2, \gamma, \theta)$ and $(\sigma_1^2, \theta, \gamma)$ is given by:

$$\Pr_2(\theta, \sigma_1^2, \gamma) \propto \sigma_1^{-2} \left\{ \sum_{i=1}^{I} \frac{n_i^2}{(1+\gamma n_i)^2} \right\}^{\frac{1}{2}} = \sigma_1^{-2} \Pr_2(\gamma) \tag{6.2}$$

where $\sum_{i=1}^{I} n_i = n$

If $n_1 = n_2 = \cdots = n_i = J$ both these priors simplify to Jeffreys' independent prior $\mathrm{P}(\theta, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2}(\sigma_1^2 + J\sigma_2^2)^{-1}$.

From equations (6.1) and (6.2) it follows that the posterior distributions $\Pr_1(\gamma|\underline{Y})$ and $\Pr_2(\gamma|\underline{Y})$ are given by:

$$\Pr_1(\gamma|\underline{Y}) \propto \Pr_1(\gamma) \prod_{i=1}^{I} \left( \frac{1}{1+\gamma n_i} \right)^{\frac{1}{2}} \left( \sum_{i=1}^{I} \frac{n_i}{1+\gamma n_i} \right)^{-\frac{1}{2}} \left[ v_1 m_1 + \sum_{i=1}^{I} \frac{n_i (\overline{Y}_{i.} - \hat{\theta})^2}{1+\gamma n_i} \right]^{-\frac{1}{2}(n-1)} \tag{6.3}$$

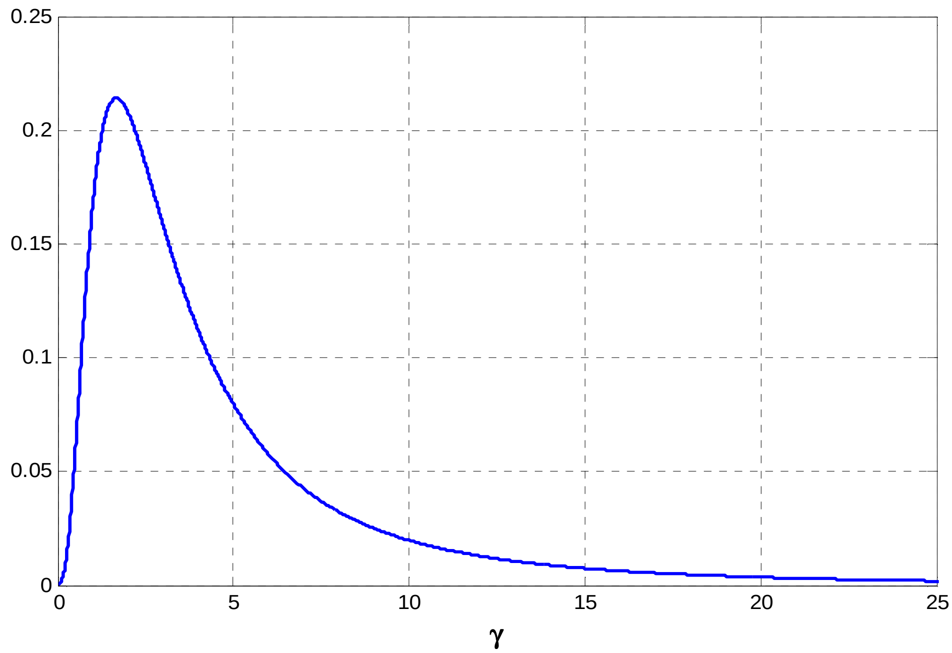
and

$$\Pr_2(\gamma|\underline{Y}) \propto \Pr_2(\gamma) \prod_{i=1}^{I} \left( \frac{1}{1+\gamma n_i} \right)^{\frac{1}{2}} \left( \sum_{i=1}^{I} \frac{n_i}{1+\gamma n_i} \right)^{-\frac{1}{2}} \left[ v_1 m_1 + \sum_{i=1}^{I} \frac{n_i (\overline{Y}_{i.} - \hat{\theta})^2}{1+\gamma n_i} \right]^{-\frac{1}{2}(n-1)} \tag{6.4}$$

$\mathrm{v}_1 = n - I$, $\mathrm{v}_1 m_1 = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left( Y_{ij} - \overline{Y}_{i.} \right)^2$, and $\hat{\theta} = \dfrac{\sum_{i=1}^{I} \overline{Y}_{i.} \frac{n_i}{1+\gamma n_i}}{\sum_{i=1}^{I} \frac{n_i}{1+\gamma n_i}}$

In Figure 6.1, the two posterior distributions $\Pr_1(\gamma)$ and $\Pr_2(\gamma)$ are displayed.

**Figure 6.1: Posterior Distribution of $\mathbf{Pr_1}(\gamma|\underline{Y})$ and $\mathbf{Pr_2}(\gamma|\underline{Y})$**




$$Mean(\gamma) = 5.0846 \quad , \quad Median(\gamma) = 3.318 \quad , \quad Mode(\gamma) = 1.67 \quad , \quad Var(\gamma) = 28.404$$

From Figure 6.1 it is clear that the posterior distributions $\mathrm{Pr}_1(\gamma|\underline{Y})$ and $\mathrm{Pr}_2(\gamma|\underline{Y})$ are for all practical purposes the same. In the Monte Carlo simulation procedure only $\mathrm{Pr}_1(\gamma|\underline{Y})$ will therefore be used.

The join posterior distribution can be written as

$$\mathrm{p}(\theta, \sigma_1^2, \gamma|\underline{Y}) = \mathrm{p}(\theta|\sigma_1^2, \gamma, \underline{Y})\mathrm{p}(\sigma_1^2|\gamma, \underline{Y})\mathrm{Pr}(\gamma|\underline{Y})$$

and the simulation method to obtain $k^*$ is as follows:

(i) By using a rejection method simulate $\gamma$ from $\mathrm{Pr}(\gamma|\underline{Y})$.

(ii) Given $\gamma$, $\sigma_1^2$ has an Inverse Gamma distribution and a simulated value of $\sigma_1^2$ can be obtained from the equation:

$$\sigma_1^2 = \frac{1}{\chi_{n-1}^2}\left[\mathrm{v}_1 m_1 + \sum_{i=1}^{I} \frac{n_i(\overline{Y}_{i.} - \hat{\theta})^2}{1 + \gamma n_i}\right]$$

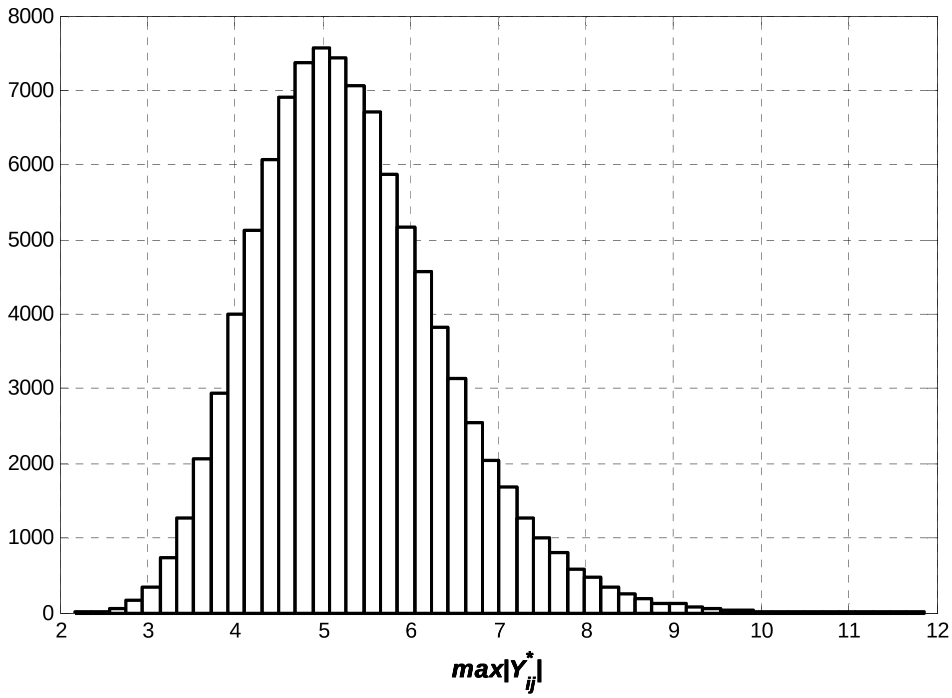
(iii) From (i) and (ii) it follows that $\sigma_2^2 = \sigma_1^2\gamma$.

(iv) For each pair of simulated variance components $(\sigma_1^2, \sigma_2^2)$ draw $\underline{\underline{Y}}^* = [Y_{11}^* \quad Y_{12}^* \quad \dots \quad Y_{IJ}^*]'$ from the multivariate normal distribution defined in Theorem 6.1

(v) Calculate the maximum of the absolute values that have been drawn and call it $\max\left|Y_{ij}^*\right| \quad i=1,\dots,I,\ j=1,\dots,J.$

(vi) Repeat steps (i) – (v) 100 000 times and draw a histogram of the simulated $\max\left|Y_{ij}^*\right|$ values

**Figure 6.2: Histogram of $\max\left|Y_{ij}^*\right|$**




$$\max\left|Y_{ij}^*\right|^{0.95} = 7.2584 = k^*$$

$$\text{Mean} = 5.3196\ , \qquad \text{Median} = 5.2119\ , \qquad \text{Mode} = 5\ , \qquad \text{Variance} = 1.1506$$

If the prior $p(\theta, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2}\sigma_2^{-1}$ is used (see for example Bayarri and Castellanos (2007)) then $k^* = 7.2649$**.**

The control limits as well as the means and 95% Bayesian confidence intervals for $e_{ij}^*$ are given in Figure 6.3, and in Figure 6.4, the unconditional posterior distributions are displayed. As mentioned, the unconditional posterior distributions are obtained using the Rao-Blackwell method.

**Figure 6.3: Means and 95% Intervals for $e^*_{ij}, i = 1, \ldots, 5,\ j = 1, \ldots, 6$**


**Standardised Errors**

**Figure 6.4: Unconditional Posteriors of $e^*_{41}, e^*_{53}, e^*_{16}$**

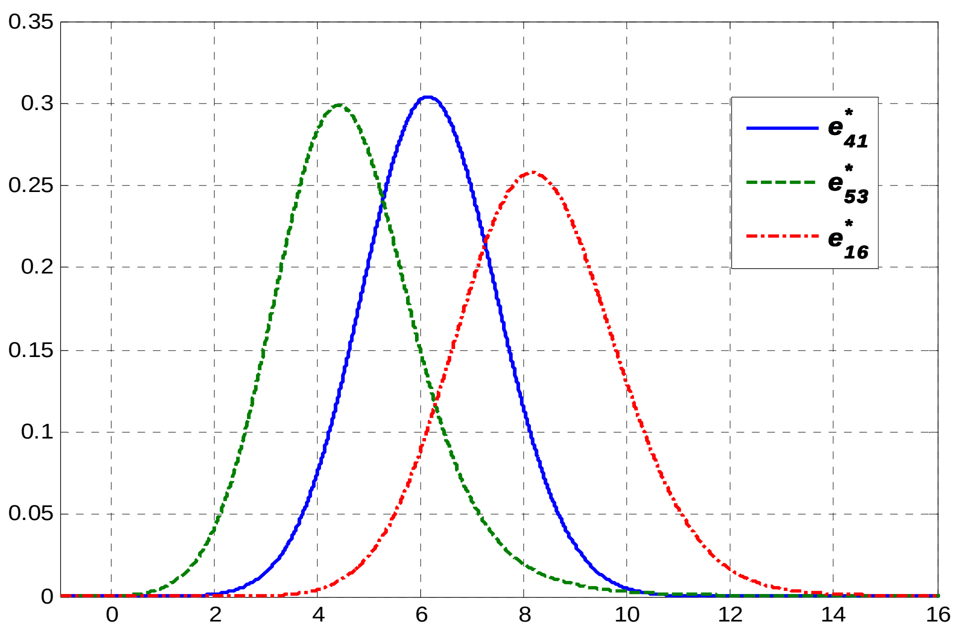


**Table 6.1 Probabilities that Individual Observations are Outlying**

| $e_{ij}^*$ | $Mean(e_{ij}^*)$ | $Var(e_{ij}^*)$ | 95% Interval | $P(\lvert e_{ij}^*\rvert > 7.26)$ |
|---|---|---|---|---|
| 41 | 6.1981 | 1.7257 | 3.623 – 8.773 | 0.2094 |
| 53 | 4.6637 | 2.2356 | 1.773 – 7.954 | 0.0412 |
| 16 | 8.3100 | 2.5608 | 5.174 – 11.447 | 0.7451 |

Since the mean of $e_{16}^*$ is larger than $k^* = 7.26$, it can be concluded that $Y_{16}$ is an outlying observation. This is also clear from Table 6.1 where $P(|e_{16}^*| > 7.26) = 0.7451$.

## 7. Conclusion

In this note the Bayesian procedures of Zellner (1975) and Chaloner (1994) are extended for the balanced one-way random effects model. Since it is not clear to us what the frequentist properties of their methods are (i.e. what the size of the type I error or the power of their tests are), a Bayesian-frequentist approach is used for detecting outliers. The Sharples generated data (Sharples (1990) and Chaloner (1994) ) are used for illustration purposes.

Chaloner (1994) concluded that observations $Y_{16}$ and $Y_{41}$ as well as Group 5 are possible outliers, since their posterior probabilities are larger than the prior probability of an individual observation or group being an outlier.

Inspection of Tables 5.3 and 6.1, as well as Figures 5.3 and 6.2 show that only observation $Y_{16}$ is an outlier.

The Bayesian-frequentist approach therefore seems to be more conservative than Chaloner’s method.

# Appendix A

Proof of Theorem 5.1

$$Y_i^* = \frac{J\sigma_2\sqrt{I}(\bar{Y}_{i.}-\bar{Y}_{..})}{\sqrt{(\sigma_1^2+J\sigma_2^2)(I\sigma_1^2+J\sigma_2^2)}} \quad i = 1, \dots, I$$

where

$$\bar{Y}_{i.} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij} = \frac{1}{J}\sum_{j=1}^{J}(\theta + r_i + e_{ij}) = \theta + r_i + \bar{e}_{i.}$$

and

$$\bar{Y}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta + r_i + e_{ij}) = \theta + \bar{r}_{.} + \bar{e}_{..}$$

Therefore

$$E(\bar{Y}_{i.} - \bar{Y}_{..}) = \theta - \theta = 0$$

Since $r_i$ and $e_{ij}$ are uncorrelated it follows that

$$\begin{aligned} Var(\bar{Y}_{i.} - \bar{Y}_{..}) &= Var(r_i - \bar{r}_{.}) + Var(\bar{e}_{i.} - \bar{e}_{..}) \\ &= \sigma_2^2\frac{(I-1)}{I} + \sigma_1^2\frac{(I-1)}{IJ} \\ &= \frac{(I-1)}{IJ}(\sigma_1^2 + J\sigma_2^2) \quad i = 1, \dots, I \end{aligned}$$

and therefore

$$Var(Y_i^*) = \frac{J(I-1)\sigma_2^2}{(I\sigma_1^2+J\sigma_2^2)} \qquad \text{(A.1)}$$

Also

$$\begin{aligned} Cov(\bar{Y}_{l.} - \bar{Y}_{..})(\bar{Y}_{m.} - \bar{Y}_{..}) &= E(\bar{Y}_{l.} - \bar{Y}_{..})(\bar{Y}_{m.} - \bar{Y}_{..}) - E(\bar{Y}_{l.} - \bar{Y}_{..})E(\bar{Y}_{m.} - \bar{Y}_{..}) \\ &= E(\bar{Y}_{l.} - \bar{Y}_{..})(\bar{Y}_{m.} - \bar{Y}_{..}) \\ &= E\{[(r_l - \bar{r}_{.}) + (\bar{e}_{l.} - \bar{e}_{..})][(r_m - \bar{r}_{.}) + (\bar{e}_{m.} - \bar{e}_{..})]\} \\ &= E[(r_l - \bar{r}_{.})(r_m - \bar{r}_{.})] + E[(\bar{e}_{l.} - \bar{e}_{..})(\bar{e}_{m.} - \bar{e}_{..})] \end{aligned}$$

since the expected values of the cross products are zero.

Now $E[(r_l - \overline{r}_.)(r_m - \overline{r}_.)] = E[r_l r_m] - E[r_l \overline{r}_.] - E[\overline{r}_. r_m] + E[\overline{r}_.^2]$

$$= 0 - \frac{\sigma_2^2}{I} - \frac{\sigma_2^2}{I} + \frac{\sigma_2^2}{I}$$

$$= \frac{-\sigma_2^2}{I}$$

and

$E[(\overline{e}_{l.} - \overline{e}_{..})(\overline{e}_{m.} - \overline{e}_{..})] = E[\overline{e}_{l.}\overline{e}_{m.}] - E[\overline{e}_{l.}\overline{e}_{..}] - E[\overline{e}_{..}\overline{e}_{m.}] + E[\overline{e}_{..}^2]$

$$= 0 - \frac{\sigma_1^2}{IJ} - \frac{\sigma_1^2}{IJ} + \frac{\sigma_1^2}{IJ}$$

$$= \frac{-\sigma_1^2}{IJ}$$

Therefore

$Cov(\overline{Y}_{l.} - \overline{Y}_{..})(\overline{Y}_{m.} - \overline{Y}_{..}) = \frac{-1}{IJ}(\sigma_1^2 + J\sigma_2^2)$

and

$Cov(Y_l^*, Y_m^*) = \frac{-J\sigma_2^2}{(I\sigma_1^2 + J\sigma_2^2)}, l = 1, \dots, I,\ m = 1, \dots, I, l \neq m.$

From this it follows that the correlation coefficient between

$Y_l^*$ and $Y_m^* = \rho_{l,m} = \frac{-1}{I-1}$ (A.2)

**Appendix B**

$$Y_{ij}^* = \frac{Y_{ij} - \frac{J\sigma_2^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{i.} - \frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{..}}{\sqrt{\frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\left(\sigma_2^2+\frac{\sigma_1^2}{IJ}\right)}},\ i = 1, \dots, I,\ j = 1, \dots, J$$

Let $\tilde{Y}_{ij} = Y_{ij} - \frac{J\sigma_2^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{i.} - \frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\overline{Y}_{..}$

$$= Y_{ij} - a\overline{Y}_{i.} - b\overline{Y}_{..}$$

where $a = \frac{J\sigma_2^2}{\sigma_1^2+J\sigma_2^2}$, $b = \frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}$ and $a + b = 1$

The model is

$$Y_{ij} = \theta + r_i + e_{ij}$$

where $e_{ij} \sim N(0, \sigma_1^2)$ and $r_i \sim N(0, \sigma_2^2)$

(a) we will first derive $Var(Y_{ij}^*)$

Now $Var(\tilde{Y}_{ij}) = Var(r_i - ar_i - b\overline{r}_{.}) + Var(e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..})$

since $r_i$ and $e_{ij}$ are independently distributed of each other

Further

$$Var(r_i - ar_i - b\overline{r}_{.}) = Var[b(r_i - \overline{r}_{.})] = b^2\sigma_2^2\frac{(I-1)}{I}$$

and

$$Var(e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..}) = \frac{\sigma_1^2}{J}\left[(J-1) + b^2\left(\frac{I-1}{I}\right)\right]$$

Therefore

$$Var(\tilde{Y}_{ij}) = \frac{\sigma_1^2}{J}\left\{(J-1) + \frac{\sigma_1^2}{\sigma_1^2+J\sigma_2^2}\left(\frac{I-1}{I}\right)\right\}$$

and

$$Var(Y_{ij}^*) = \frac{1}{\sigma_1^2+IJ\sigma_2^2}\{I(J-1)(\sigma_1^2 + J\sigma_2^2) + \sigma_1^2(I-1)\} \tag{B.1}$$

(b) $Cov(\tilde{Y}_{ij}, \tilde{Y}_{lm}) = E(\tilde{Y}_{ij}\tilde{Y}_{lm})$ because $E(\tilde{Y}_{ij}) = E(\tilde{Y}_{lm}) = 0$

$(i = 1, \dots, I;\ j = 1, \dots, J; l = 1, \dots, I;\ m = 1, \dots, J; i \neq l \text{ and } j \neq m)$

Therefore

$$\begin{aligned} Cov(\tilde{Y}_{ij}, \tilde{Y}_{lm}) &= E(Y_{ij} - a\overline{Y}_{i.} - b\overline{Y}_{..})(Y_{lm} - a\overline{Y}_{l.} - b\overline{Y}_{..}) \\ &= E\{b(r_i - \overline{r}_{.}) + (e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..})\}\{b(r_l - \overline{r}_{.}) + (e_{lm} - a\overline{e}_{l.} - b\overline{e}_{..})\} \\ &= b^2E(r_i - \overline{r}_{.})(r_l - \overline{r}_{.}) + E(e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..})(e_{lm} - a\overline{e}_{l.} - b\overline{e}_{..}) \end{aligned}$$

Now

$$b^2E(r_i - \overline{r}_{.})(r_l - \overline{r}_{.}) = \frac{-b^2\sigma_2^2}{I}$$

and

$$E\left(e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..}\right)(e_{lm} - a\overline{e}_{l.} - b\overline{e}_{..}) = \frac{-b^2\sigma_1^2}{IJ}$$

From this it follows that

$$Cov\left(\tilde{Y}_{ij}, \tilde{Y}_{lm}\right) = \frac{-b^2}{IJ}(\sigma_1^2 + J\sigma_2^2) = \frac{-(\sigma_1^2)^2}{IJ\left(\sigma_1^2 + J\sigma_2^2\right)}$$

and

$$Cov(Y_{ij}^*, Y_{lm}^*) = \frac{-\sigma_1^2}{\sigma_1^2 + IJ\sigma_2^2} \qquad \text{(B.2)}$$

(c) In a similar way it can be shown that

$$Cov(Y_{ij}^*, Y_{lj}^*) = \frac{-\sigma_1^2}{\sigma_1^2 + IJ\sigma_2^2} \qquad \text{(B.3)}$$

(d)
$$\begin{aligned} Cov\left(\tilde{Y}_{ij}, \tilde{Y}_{im}\right) &= b^2 E(r_i - \overline{r}_{.})^2 + E\left(e_{ij} - a\overline{e}_{i.} - b\overline{e}_{..}\right)(e_{im} - a\overline{e}_{i.} - b\overline{e}_{..}) \\ &= b^2\sigma_2^2\frac{(I-1)}{I} - \frac{\sigma_1^2}{J}\left\{(1 - b^2) + \frac{b^2}{I}\right\} = -\frac{\sigma_1^2}{J}\left\{\frac{\sigma_1^2 + IJ\sigma_2^2}{I(\sigma_1^2 + J\sigma_2^2)}\right\} \end{aligned}$$

and

$$Cov(Y_{ij}^*, Y_{im}^*) = -1 \qquad \text{(B.4)}$$

**References**


1) Bayarri, M. J. and Berger, J.O. (2000). *p*-Values for Composite Null Models (with Discussion). J. Amer. Statist. Assoc. 95: 1127–1142, 1157–1170.
2) Bayarri, M.J. and Berger, J. (2004). The Interplay between Bayesian and Frequentist Analysis. Statistical Science, 19, 1, 58-80.
3) Berger, J. O. and Bernardo, J. M. (1992a). On the Development of Reference Priors. In *Bayesian Statistics 4, 35–60* (eds. Bernardo J.M., Berger J.O., Dawid, A. P., and Smith, A. F. M.) Oxford: University Press.
4) Bayarri, M. J. and Castellanos, M.E. (2007). Bayesian Checking of the Second Levels of Hierarchical Models. Statistical Science Vol. 22. No. 3: 222–243.
5) Bayarri, M. J. and Morales, J. (2003). Bayesian Measures of Surprise for Outlier Detection. J. Statist. Plann. Inference 111: 3–22.
6) Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis, Reading. MA, Addison-Wesley.
7) Box, G.E.P., Tiao, G.C., (1968). A Bayesian Approach to some Outlier Problems. Biometrika. 55, 119-129.
8) Chaloner, K., (1994). Residual Analysis and Outliers in Bayesian Hierarchical Models. In: Smith, A., Freeman, P. (Eds.), Aspects of Uncertainty. Wiley, Chichester, UK.
9) Chaloner, K. and Brant, R. (1988). A Bayesian Approach to Outlier Detection and Residual Analysis, Biometrika. 75, 651-9.
10) Datta, G. S. and Ghosh, J. K. (1995). On Priors providing Frequentist Validity of Bayesian Inference. *Biometrika* 82: 37–45.
11) Dey, D.K., Gelfand, A.E., Swartz, T.B. and Vlachos, A.K. (1998). A Simulation-intensive Approach for Checking Hierarchical Models. Test 7, 325-346.
12) Freeman, P.R., (1980). On the number of Outliers in Data from a Linear Model. In: Bernardo, J.M., DeGroot, M.H., Smith, A.F.M., Lindley, D.V. (Eds.), Bayesian Statistics. Valencia University Press, Valencia, pp. 349-365.
13) Hoeting, J., Raftery, A.E., Madigan, D., (1996). A Model for Simultaneous Variable Selection and Outlier Identification in Linear Regression. Comput. Statist. Data Anal. 22, 251-270.
14) Jeffreys, H. (1961). Theory of Probability. Oxford: Oxford University Press.
15) Marshall, E.C. and Spiegelhalter, D.J. (2003). Approximate Cross-validatory Predictive Checks in Disease Mapping Models. Stat. Med. 22, 1649-1660
16) O'Hagan, A. (2003). HSSS Model Criticism (with Discussion). In Highly Structured Stchastic Systems (P.J. Green, N.L. Hjort and S.T. Richardson, eds.) 423-445. Oxford Univ. Press. MR2082403
17) Pettit, L.I., Smith, A.F.M., (1985). Outliers and Influential Observations in Linear Models. In: Bernardo, J.M., DeGroot, M.H., Smith, A.F.M., Lindley, D.V. (Eds.), Bayesian Statistics, Vol. 2. North-Holland, Amsterdam, pp. 473-494.

18) Reid, N. and Cox, D.R. (2014). On Some Principles of Statistical Influence, International Statistical Review: 0,0; 1-16.
19) Sharples, L.D. (1990). Identification and Accommodation of Outliers in General Hierarchical Models, Biometrika, 77, 445-53.
20) Verdinelli, I., Wasserman, L., (1991). Bayesian Analysis of outlier problems using the Gibbs sampler. Statist. Comput. 1, 105-117.
21) Zellner, A. (1975). Bayesian Analysis of Regression Error terms, J. Am. Statist. Assoc., 70, 138-144
22) Zellner, A. and Moulton, B.R. (1985) Bayesian Regression Diagnostics with Applications to International Consumption and Income Data, J. Econometrics, 29, 187-211.