Levene's mean-based test: exact and approximate distributions

Technical report

Mick O'Neill

Biometry Unit
Faculty of Agriculture
University of Sydney
Australia 2006

Introduction

So much interest has been shown in Levene's (1960) test of variance homogeneity that it is becoming standard output in statistical packages, at least for one-way designs. The original suggestion was extremely simple. Take the residuals (R_{ij}) from a one-way ANOVA, form absolute values (D_{ij}) of these residuals, and re-analyse the (D_{ij}) via a one-way ANOVA. Since D_{ij} has expected value proportional to σ_i, the standard deviation of the ith treatment, it was argued that the follow-up ANOVA would test for equality of treatment variances. Furthermore, since the ANOVA F-test is known to be robust, Levene's test should out-perform Bartlett's test, which is known to be sensitive to departures from normality.

Recently, Keyes and Levy (1997) showed that this procedure is valid *only when the design is balanced*. This should have appeared obvious, since, if n_i is the number of replicates in the ith treatment, the exact mean for D_{ij} is

$$E(D_{ij}) = \sqrt{\frac{2}{\pi}\left(1-\frac{1}{n_i}\right)}\sigma_i, \qquad i = 1, ..., t.$$

Thus, the F-test is *not* testing the hypothesis H_0: $\sigma_1 = \sigma_2 = \ldots = \sigma_t = \sigma(\text{say})$ whenever the designed is unbalanced; statistical packages are therefore producing incorrect tests in these cases. Keyes and Levy suggested a simple remedy: analyse

$$U_{ij} = \left(1-\frac{1}{n_i}\right)^{-\frac{1}{2}} D_{ij}$$

instead. This has no effect on the F-test and p-values produced by the packages whenever the design is balanced, and clearly rectifies the test otherwise. Others have realised the implication of design imbalance on the F-statistic (for example, Alan Miller, 1973, personal communication), but to our knowledge this is the first time it has been pointed out in print.

Another useful modification is the following. Since

$$\mathrm{var}(D_{ij}) = \left(1-\frac{2}{\pi}\right)\left(1-\frac{1}{n_i}\right)\sigma_i^2, \qquad i = 1, ..., t \tag{1}$$

scaling the U_{ij} to have variance σ_i^2 would allow for a more intuitively appealing interpretation of the ANOVA mean squares. Thus, in this paper we suggest an analysis of

$$U_{ij} = \left(1-\frac{2}{\pi}\right)^{-\frac{1}{2}}\left(1-\frac{1}{n_i}\right)^{-\frac{1}{2}}\left|Y_{ij} - \bar{y}_i\right| \tag{2}$$

in order to detect variance heterogeneity. Again, this scaling makes no difference to the F-statistic when the design is balanced.

There remain several equally serious worries over the use of Levene's test in its current format. Take for example the following *count* data, which are non-normal and appear to have unequal means and unequal variances:

		Treatment		
1	2	3	4	5
35	61	12	0	36
22	40	10	2	36
11	45	0	4	33
6	40	6	1	18

SPSS (Release 6.1) returns a p-value of 0.088 for Levene's test, whereas the output from Minitab (Release 10) appears as follows:

```
Bartlett's Test (normal distribution)
Test Statistic:    8.214
p-value:           0.084

Levene's Test (any continuous distribution)
Test Statistic:    1.246
p-value:           0.334
```

The huge difference in Levene's p-values (0.088 versus 0.334) is due to Minitab's use of the *median*, rather than *mean*, in calculating the D_{ij}. This is done in the belief that the resultant test statistic is more robust to non-normality (hence the claim "any continuous distribution" in Minitab's output). However, to the author's knowledge, the theoretical properties of this statistic have not been explored.

The behaviour of Levene's test statistic, based on the mean, is very strange even for normal data. Several people have observed that, when the hypothesis of equal variances is true, Levene's test has inflated power. For example, for t=4, n_i=4, Yitnosumarto and O'Neill (1986) observed an empirical significance level of 9.2% for a 5% test, using 2000 simulations; Keyes and Levy (1997) observed 9.8% for the same case, based on 10,000 simulations: this rose to 17.7% for the unbalanced design with (2,3,4,5) replicates for the four treatments (using their modified statistic). What is more surprising is that the power of the test can *fall* to (sometimes well) below the empirical significance level for some alternatives. Thus, for the latter example, Keyes and Levy (1997) observed an empirical power of only 11.4% when the true variances are (1,2,3,4) respectively. All this has led to the conservative advice that the hypothesis of equal variances should be rejected only for small p-values (for example < 0.01) (Milliken and Johnson, 1984).

This advice seems quite unsatisfactory. The problem is simply that Levene's F-statistic does not follow an F-distribution with (t-1) numerator and $\sum_i \left(n_i - 1\right)$ =(N-t) denominator degrees of freedom, even for data distributed normally.

To illustrate this further, consider the results of Levene's test from 75,000 simulations, based on independent, normally distributed N(0,1) data from 4 treatments each with 4 replicates, using the modification suggested in Eq (2).

Table 1 presents summary statistics, and Figure 1 presents the empirical distributions, of the Treatment Sum of Squares (TSS), Residual Sum of Squares (RSS) and F-statistic based on a one-way analysis of the scaled absolute residuals U_{ij}. By way of comparison, corresponding theoretical values are included, based on the distributions assumed on performing an ANOVA: χ^2_{t-1}, χ^2_{N-t} and $F_{t-1, N-t.}$ respectively.

Table 1 Summary statistics for Bartlett's and Levene's adjusted tests based on Eq (4) using the *median* (as in Minitab) and the *mean* (as in SPSS), based on 75,000 simulations of data assumed N(0,1); t=4, n_i=4

statistic	Bartlett's test	Levene's test								
		TSS			RSS			F		
		Sample		theoretical	Sample		theoretical	Sample		theoretical
		based on *median*	based on *mean*	based on χ^2_3	based on *median*	based on *mean*	based on χ^2_{12}	based on *median*	based on *mean*	based on $F_{3,12}$
Mean	2.993	3.787	3.892	3.000	15.980	10.797	12.000	1.290	1.772	1.200
s.d.	2.416	3.313	3.320	2.449	8.984	5.456	4.899	1.841	2.336	1.249
Q1	1.224	1.461	1.537	1.213	9.420	6.797	8.438	0.395	0.640	0.408
Q2	2.378	2.883	3.017	2.366	14.250	9.884	11.340	0.799	1.227	0.835
Q3	4.097	5.108	5.270	4.108	20.621	13.822	14.845	1.540	2.137	1.561

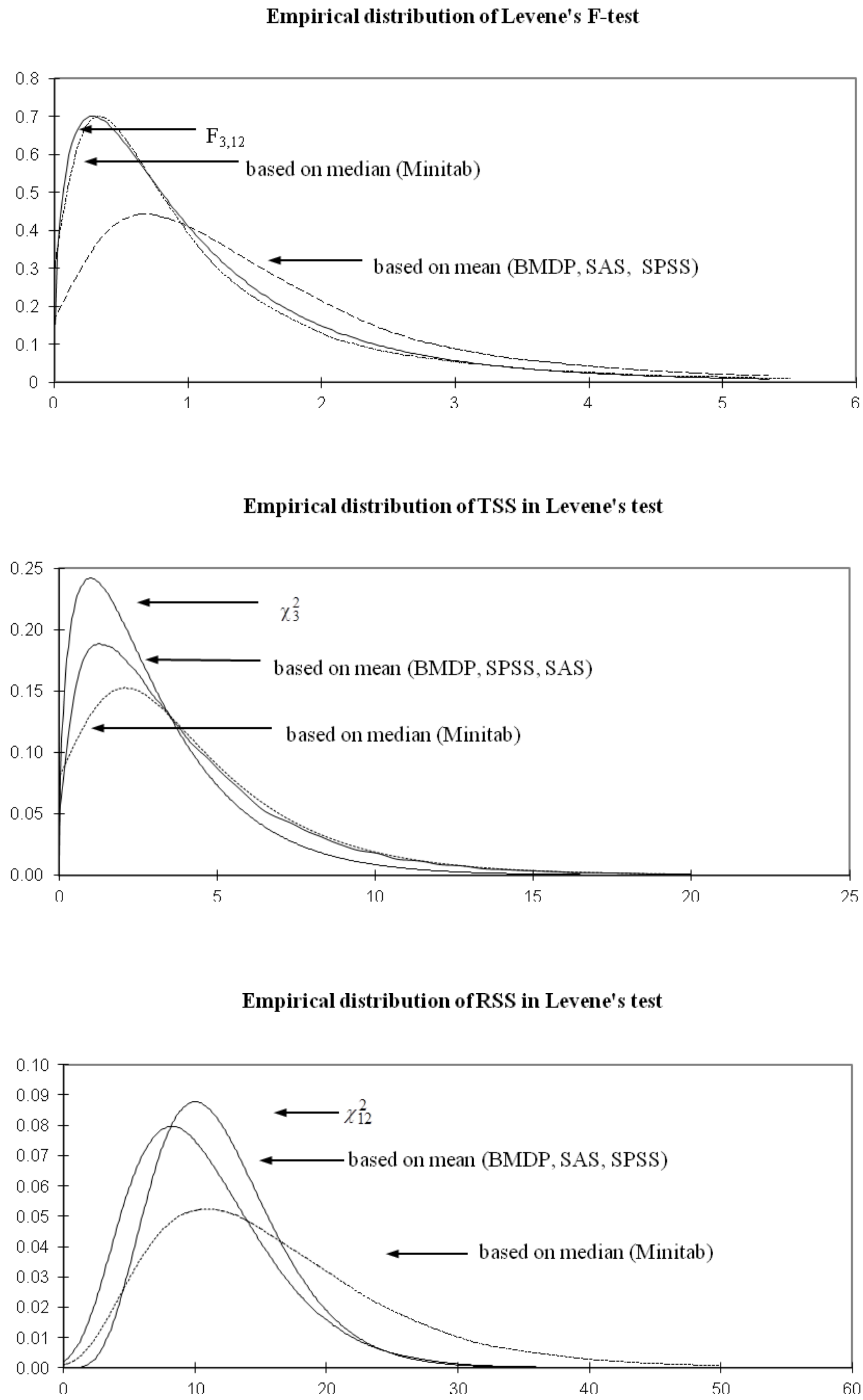

Figure 1 **Empirical and assumed distributions of Levene's F statistic (top), TSS (middle)**

It is clear that the distributions underlying the ANOVA are inappropriate. In particular, the empirical F-distribution indicates that the correct distribution for Levene's F-test should be squatter and longer-tailed, and different for the different tests (based on *mean* and *median*); consequently, the p-values calculated assuming an $F_{t-1,\, N-t.}$ distribution will be too small in both cases. It is surprising, however, that the test based on the *median* does so well, given that the component distributions are so unlike the relevant χ^2 distributions.. The distribution of the RSS in this case is much larger than χ^2_{12}; this compensates for the slightly larger TSS in the ratio, with the result that the distribution of the F-statistic *almost by accident* follows more closely the standard F-distribution.

It is surprising also that the change from *mean* to *median* in the formation of Levene's test makes so much difference. In the simulations, the median-based test was almost always (about 75%) smaller than the mean-based test. By way of illustration only, a regression of the simulated data resulted in

$$\textit{median-based } F = -0.0589 + 0.761 \textit{ mean-based } F$$

with $R^2 = 93.2\%$. A random selection of 4,000 of the 75,000 data points is plotted in Figure 2. Moreover, only once in 75,000 times was the median-based test "significant" at 5% when the mean-based test was not. The full empirical percentages are shown in Table 2.

Table 2 Empirical 5% significance levels of Levene's test (t = 4, n = 4), based on 75,000 simulations

Mean-based F-test	Median-based F-test (Minitab)		Mean-based F-test
(BMDP, SAS, SPSS)	<5%	>5%	
<5%	67,103	1	67,104
>5%	3,051	4,845	7,896 (10.5%)
Median-based F-test	70,154	4,846 (6.5%)	75,000

The purpose of this paper, then, is to study the exact distribution of Levene's test (as modified in Eq (2) and using the *mean*), and to determine whether a better test can be recommended in general.

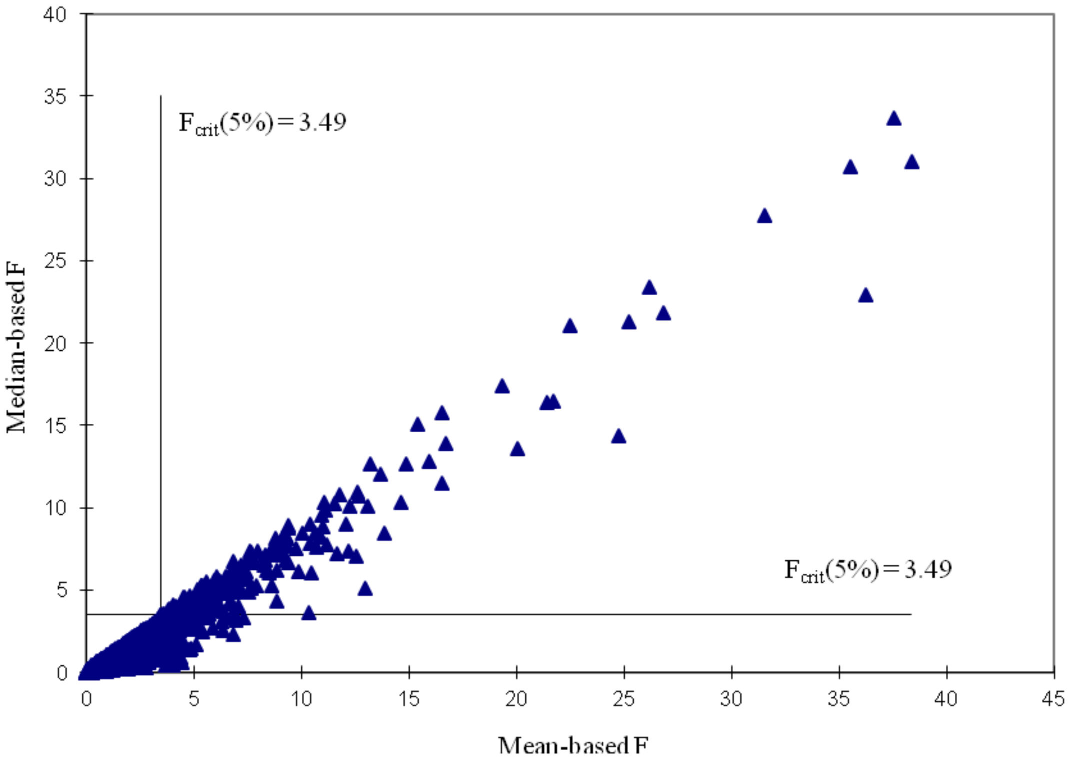

Figure 2 **Plot of a random selection of mean-based and median-based F values**

Exact Distribution of Levene's test

Let Y_{ij} represent the jth replicate (j=1, ..., n$_i$) in the ith treatment (i=1, ..., t) of a oneway design, and σ_i^2 the ith population variance. Let $\bar{y}_{i.}$ represent the ith treatment mean. Then the residuals, defined as

$$R_{ij} = Y_{ij} - \bar{y}_{i.},$$

are multivariate normal with zero means. They are independent across treatments, but for the ith treatment the covariance matrix of the n$_i$ residuals which form the vector $\mathbf{R}_i$ is

$$\text{cov}\left(\mathbf{R}_i\right) = \sigma_i^2\left[\mathbf{I}_{n_i} - \frac{1}{\text{n}_i}\mathbf{J}_{n_i}\right] \tag{3}$$

where $\mathbf{I}_{n_i}$ is an (n$_i$ x n$_i$) identity matrix and $\mathbf{J}_{n_i}$ an (n$_i$ x n$_i$) matrix of ones.

The U_{ij}, defined in Eq (2), are therefore "semi-normal" or "folded-normal" variables. They are certainly not normal; their probability density function (pdf) and covariance structure are quite complicated (see Appendix 1). The repercussion of these facts on Levene's test is quite dramatic.

Keyes and Levy's (1996) modified Levene's test statistic is simply the ratio of the Treatment Mean Square to the Residual Mean Square of a oneway ANOVA of the U_{ij}, defined as

$$F = \frac{\text{N}-\text{t}}{\text{t}-1}\,\frac{\sum_{i=1}^{\text{t}} \text{n}_i(\bar{u}_{i.} - \bar{u})^2}{\sum_{i=1}^{\text{t}}\sum_{j=1}^{\text{n}_i}(U_{ij} - \bar{u}_{i.})^2} \tag{4}$$

where $\bar{u}_{i.}$ is the ith mean of the U_{ij} and $\bar{u}$ is the overall mean. Note that $\bar{u}_{i.}$ is proportional to the well-known *mean deviation* for the ith treatment. The mean deviation was studied extensively some fifty years ago, especially by Godwin (1945, 1946); the probability integral was tabulated by Hartley (1946).

Results for Levene's (modified) statistic

1. The Residual SS can be expressed as

$$\text{RSS} = \sum_{i=1}^{t}\sum_{j=1}^{n_i-1} V_{ij}^2 ,$$

where the V_{ij} are uncorrelated (but not independent, except across treatments); they are contrasts among folded-normal variables, have zero means and variances given by

$$\operatorname{var}(V_{ij}) = \frac{\left(1-\frac{2}{\pi}s_{n_i}\right)}{\left(1-\frac{2}{\pi}\right)}\sigma_i^2 ,$$

where

$$s_{n_i} = \frac{1}{n_i-1}\left(\sqrt{n_i(n_i-2)} + \sin^{-1}\left(\frac{1}{n_i-1}\right)\right). \quad (5)$$

Yitnosumarto and O'Neill (1986) expanded s_n as

$$s_n = 1 + \frac{1}{2(n-1)^2} + \frac{1}{24(n-1)^4} + \ldots \quad (6)$$

in studying the first-order moments of Levene's (unmodified) statistic.

2. The Treatment SS can be expressed as

$$\text{TSS} = \sum_{i=1}^{t-1} Z_i^2 ,$$

where the Z_i are correlated (except for balanced designs, in which they are uncorrelated, but not independent); they are contrasts among the scaled treatment mean deviations, and have zero means only if H_0: $\sigma_1 = \sigma_2 = \ldots = \sigma_t = \sigma$ is true; their variances and covariances are complex, except under H_0 for balanced designs (n_i = n for all i). In this case,

$$\operatorname{var}(Z_j) = \left[1 + \frac{2}{\pi-2}(n-1)(s_n-1)\right]\sigma^2$$

3. The V_{ij} and Z_i are uncorrelated, but not independent.

The proof of these results is given in Appendix 2.

The exact distribution of Levene's F-statistic, then, is quite complex. Since the V_{ij} and Z_i are not normal, neither the RSS nor the TSS is exactly proportional to χ^2. Nor is the RSS independent of the TSS, although the V_{ij} and Z_i are uncorrelated.

It is, however, possible to obtain the means, variances and covariances of the RSS and TSS. The results are proved in Appendix 3. The means were presented in Yitnosumarto and O'Neill (1986) and Keyes and Levy (1996). The results that follow are valid for any n>1, although some terms may drop out for n<5, because they are derived from cross-product moments of terms which will not exist for these cases. The terms that drop out are obvious in each case.

Exact mean and variance of RSS

$$E(RSS) = \sum_{i=1}^{t}\left(n_i - 1\right)\left[1 - \frac{2}{\pi - 2}\left(s_{n_i} - 1\right)\right]\sigma_i^2 \tag{7}$$

$$\begin{aligned}\operatorname{var}(RSS) = \left(\frac{2}{\pi-2}\right)^2 \sum_{i=1}^{t}\Bigg\{ & \frac{(n_i-2)(n_i-3)}{n_i-1}\sqrt{n_i(n_i-4)} - n_i(n_i-2) + \frac{\pi^2\left(n_i^3 - n_i^2 - n_i + 3\right)}{2n_i(n_i-1)} \\ & - \frac{4\pi(n_i^2 - 3n_i + 3)}{n_i\left(n_i-1\right)^2}\sqrt{n_i(n_i-2)} - \left[\pi\frac{(3n_i^2 + n_i - 6)}{2n_i\left(n_i-1\right)} + 2\sqrt{n_i(n_i-2)} + \sin^{-1}\frac{1}{n_i-1}\right]\sin^{-1}\frac{1}{n_i-1} \\ & + \frac{3\left(n_i-2\right)\left(n_i-3\right)}{2n_i(n_i-1)}\left[\pi + \frac{4(n_i-3)}{n_i-1}\sqrt{n_i(n_i-2)}\right]\sin^{-1}\frac{1}{n_i-3} \\ & + \frac{3\left(n_i-2\right)\left(n_i-3\right)}{n_i(n_i-1)}\left[5\sin^{-1}\frac{1}{n_i-1}\sin^{-1}\frac{1}{n_i-3} - 4I_{n_i}\right]\Bigg\}\sigma_i^4 \end{aligned} \tag{8}$$

where I_n is a transcendental function defined for n>3 by

$$I_n = \int_0^{\sin^{-1}\frac{1}{n-3}} \tan^{-1}\sqrt{\frac{2-(n-1)(n-4)\tan^2(\varphi)}{n(n-3)}}\,d\varphi\,. \tag{9}$$

I_n must be evaluated numerically, and so an extremely good approximation (for n>4) is developed in Appendix 4 (see Table 3); for n=4 we have $I_4 = \frac{\pi}{2}\tan^{-1}\frac{1}{\sqrt{2}}$. We find that the contribution of I_n to the variance is relatively small, except for small n.

Table 3 Exact and approximate values of I_n obtained from expansions in Appendix 4

n	Exact value	Approximate value
4	0.966 793	-
5	0.196 684	0.192 630
6	0.097 553	0.097 313
7	0.058 979	0.058 948
8	0.039 665	0.039 659
9	0.028 554	0.028 552
10	0.021 557	0.021 557

Exact mean and variance of TSS

$$E(TSS) = \sum_{i=1}^{t}\left(1-\frac{\mathrm{n}_i}{\mathrm{N}}\right)\left[1+\frac{2}{\pi-2}\left(\mathrm{n}_i-1\right)\left(s_{n_i}-1\right)\right]\sigma_i^2+\frac{2}{\pi-2}\sum_{i=1}^{t}\mathrm{n}_i\left(\sigma_i-\bar{\sigma}\right)^2$$

$$=(\mathrm{t}-1)+\frac{2}{\pi-2}\sum_{i=1}^{t}\left(1-\frac{\mathrm{n}_i}{\mathrm{N}}\right)\left(\mathrm{n}_i-1\right)\left(s_{n_i}-1\right)\sigma_i^2+\frac{2}{\pi-2}\sum_{i=1}^{t}\mathrm{n}_i\left(\sigma_i-\bar{\sigma}\right)^2 \qquad (10)$$

where $\bar{\sigma}$ is a (weighted) average of the σ_i. The latter term in Eq (10) disappears under the hypothesis of equal treatment variances.

The exact variance of *TSS* for the general case is given in Appendix 3: it involves fourth-order moments of the scaled treatment mean deviations. For the remainder of this paper, however, we concentrate on the distribution under the hypothesis of equal treatment variances; for this case

$$\mathrm{var}(TSS) = \sum_{i=1}^{t}\mathrm{n}_i^2\left(1-\frac{\mathrm{n}_i}{\mathrm{N}}\right)^2\left(\mu_{4i}-\mu_{2i}^2\right)+\frac{2}{\mathrm{N}^2}\left[\left(\sum_{i=1}^{t}\mathrm{n}_i^2\mu_{2i}\right)^2-\sum_{i=1}^{t}\mathrm{n}_i^4\mu_{2i}^2\right] \qquad (11)$$

where μ_{ki} is the k$^{\mathrm{th}}$ central moment of the i$^{\mathrm{th}}$ scaled treatment mean deviation $\bar{u}_{i.}$. Specifically,

$$\mu_{2i} = \mathrm{var}\left(\bar{u}_{i.}\right) = \frac{\sigma^2}{n_i}\left[1 + \frac{2}{\pi - 2}(n_i - 1)\left(s_{n_i} - 1\right)\right] \tag{12}$$

and

$$\mu_{4i} = \left(\frac{2}{\pi-2}\right)^2 \frac{\sigma^4}{n_i^2}\left\{\left(6n_i + \pi\frac{3n_i^3 - 8n_i^2 + 13n_i - 12}{n_i(n_i-1)^2}\right)\sqrt{n_i(n_i-2)} - 3n_i^2 - \frac{4n_i(n_i-2)}{n_i-1}\sqrt{(n_i-1)(n_i-3)}\right.$$

$$+\frac{(n_i-2)(n_i-3)}{n_i-1}\sqrt{n_i(n_i-4)} - \pi\frac{3n_i^2 - 5n_i + 8}{n_i - 1} + \pi^2\frac{3(n_i^2 - n_i + 2)}{4n_i(n_i-1)} - \frac{12(n_i-2)^2}{n_i-1}\sin^{-1}\frac{1}{n_i-2}$$

$$+3\left(2n_i + \pi\frac{n_i^2 - n_i + 2}{2n_i(n_i-1)}\right)\sin^{-1}\frac{1}{n_i-1} + \frac{3(n_i-2)(n_i-3)}{2n_i(n_i-1)}\left[\pi + 4\frac{n_i-3}{n_i-1}\sqrt{n_i(n_i-2)}\right]\sin^{-1}\frac{1}{n_i-3}$$

$$\left.+\frac{3(n_i-2)(n_i-3)}{n_i(n_i-1)}\left(\sin^{-1}\frac{1}{n_i-1}\sin^{-1}\frac{1}{n_i-3} - 4I_{n_i}\right)\right\}. \tag{13}$$

Note that for balanced designs with equal variances, the formulae simplify somewhat. Here, let μ_k denote the k^{th} central moment of each of the scaled treatment mean deviations. Then

$$E(TSS) = (t-1)n\mu_2 \tag{14}$$

$$\mathrm{var}(TSS) = \frac{(t-1)n^2}{t}\left((t-1)\mu_4 - (t-3)\mu_2^2\right) \tag{15}$$

This formula produces as a special case the well known result for the exact variance of a simple sample variance, taking $n_i = 1$, N = t: see, for example, Kendall and Stuart (1977, p260; our *t* is their *n*).

The theoretical mean and standard deviation corresponding to the sample moments in Table 1 can now be obtained exactly. These are given in Table 4 for the simulation considered in Table 1. Table 5 gives a selection of the exact moments for four equally replicated treatments (assuming without loss of generality $\sigma = 1$), the complete set being plotted in Figure 3. By way of comparison, the ratios of exact variance to exact mean are also plotted; for χ^2 distributions this ratio should be 2.

Table 4 Means and standard deviations for Levene's modified test based on Eq (4) using the *mean* (as in SPSS), based on 75,000 simulations of data assumed N(0,1); t=4, n_i=4

	TSS		RSS	
statistic	sample	population	sample	population
Mean	3.892	3.884	10.797	10.821
s.d.	3.320	3.318	5.456	5.482

Table 5 A selection of exact means, variances and variance/mean ratios for TSS and RSS in Levene's modified test for 4 treatments; t(n-1) is shown for reference

	TSS			RSS			
n	Mean	Variance	ratio	t(n-1)	Mean	Variance	ratio
2	6.000	31.823	5.304	4	0	0	-
3	4.344	13.929	3.207	8	6.208	12.095	1.948
4	3.884	11.010	2.834	12	10.821	30.058	2.778
5	3.660	9.592	2.620	16	15.119	44.061	2.914
6	3.527	8.796	2.494	20	19.297	56.949	2.951
7	3.439	8.287	2.410	24	23.415	69.327	2.961
8	3.376	7.934	2.350	28	27.499	81.429	2.961
9	3.329	7.675	2.306	32	31.561	93.365	2.958
10	3.292	7.477	2.271	36	35.610	105.192	2.954
15	3.188	6.928	2.173	56	55.750	163.549	2.934
20	3.138	6.677	2.127	76	75.816	221.380	2.920
25	3.110	6.532	2.101	96	95.854	279.020	2.911
30	3.091	6.439	2.083	116	115.879	336.569	2.904
35	3.077	6.373	2.071	136	135.897	394.069	2.900
40	3.067	6.325	2.062	156	155.910	451.539	2.896
45	3.060	6.287	2.055	176	175.920	508.989	2.893
50	3.054	6.258	2.049	196	195.928	566.425	2.891
60	3.045	6.213	2.041	236	235.941	681.270	2.887
70	3.038	6.182	2.035	276	275.949	796.093	2.885
80	3.033	6.159	2.030	316	315.956	910.901	2.883
90	3.030	6.141	2.027	356	355.961	1025.700	2.881
100	3.027	6.127	2.024	396	395.965	1140.493	2.880

Notice that for the simulation of Table 1, the ratio of exact variance to exact mean is approximately 2.8 for both TSS and RSS, quite different from the theoretical ratio of 2.0 for χ^2 distributions. Furthermore, it appears that the limiting ratio is 2 for TSS (as one would expect from the Central Limit Theorem), but larger than 2 for the RSS. We therefore examine the large-sample properties of these moments.

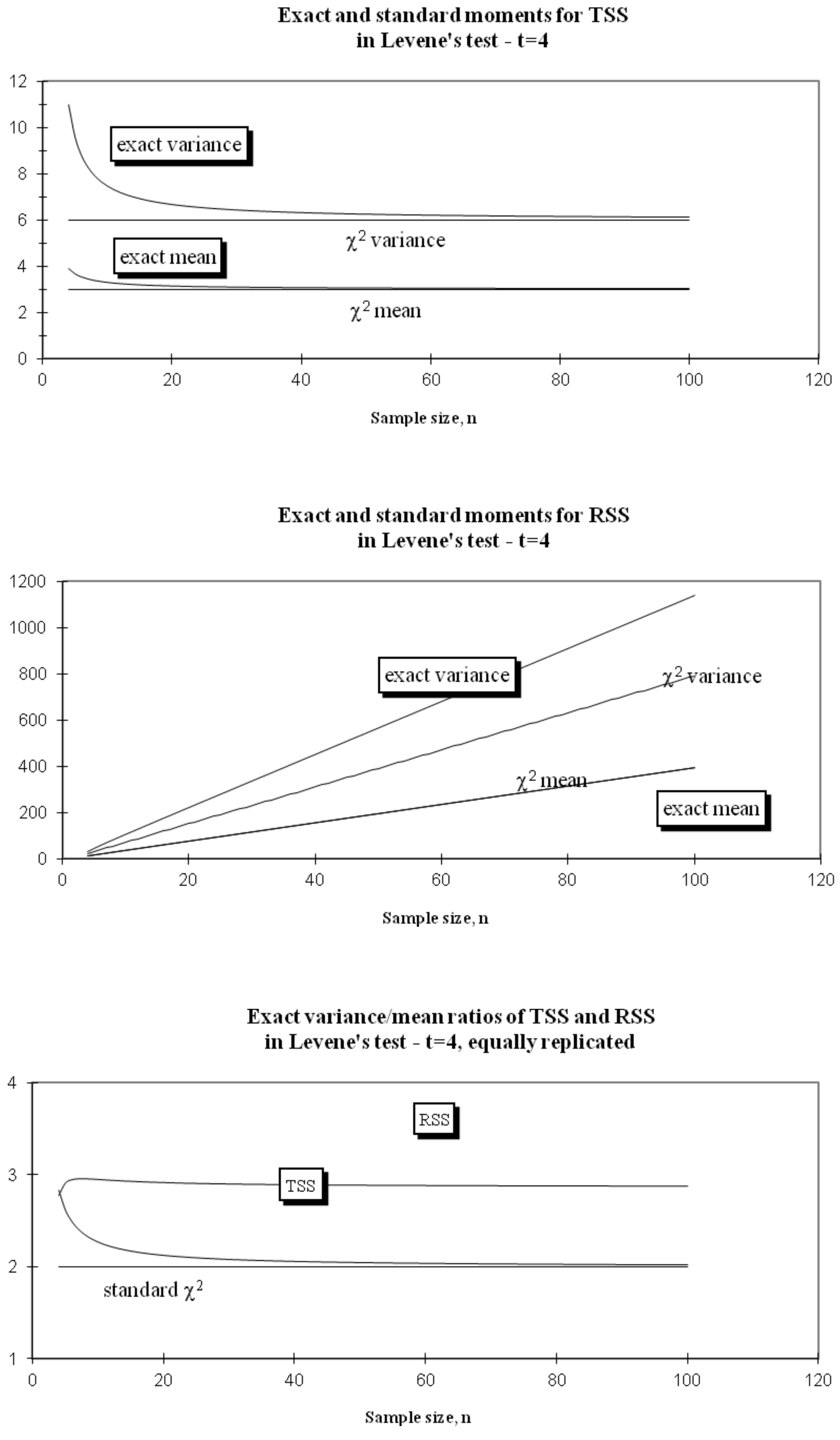

Figure 3 Exact and standard moments of TSS (top) and RSS (middle) and their ratio (bottom)

Exact covariance and correlation between TSS and RSS

Not unexpectedly, the covariance between RSS and TSS turns out to involve the moments of the individual scaled absolute residuals as well as those of the scaled treatment mean deviations. The exact result is

$$\operatorname{cov}(TSS, RSS) = \sum_{i=1}^{t}\left(1-\frac{\mathrm{n}_i}{\mathrm{N}}\right)\left(\lambda_{4i} + (\mathrm{n}_i - 1)\left[\lambda_{22i} + 2\lambda_{31i} + (\mathrm{n}_i - 2)\lambda_{211i}\right] - \mathrm{n}_i^2\left[\mu_{4i} - \mu_{2i}^2 + \mu_{2i}\sigma^2\right]\right).$$

The λs are given in Appendix 3. For 4 equally replicated treatments, the covariances and correlations are given in Table 6.

Table 6 Exact covariances and correlations between TSS and RSS in Levene's adjusted test for 4 equally replicated treatments

n	covariance	correlation
3	5.9980	0.4621
4	4.3979	0.2418
5	3.9012	0.1898
6	3.6225	0.1619
7	3.4432	0.1437
8	3.3178	0.1305
9	3.2252	0.1205
10	3.1539	0.1125
11	3.0974	0.1059
12	3.0515	0.1003
13	3.0135	0.0955
14	2.9815	0.0914
15	2.9541	0.0878
16	2.9305	0.0845
17	2.9099	0.0816
18	2.8917	0.0790
19	2.8757	0.0766
20	2.8613	0.0744
21	2.8484	0.0724
22	2.8367	0.0706
23	2.8262	0.0689
24	2.8165	0.0673
25	2.8077	0.0658

In the simulation exercise (with 4 replicates), the correlation turned out to be 0.246.

Asymptotic distribution of Levene's adjusted test

We now expand the exact means and variances of TSS and RSS in the case of equally replicated designs (under the hypothesis of equal treatment variances) to examine the limiting behaviour of Levene's test. Since exact formulae are available, we are interested in terms up to order n^{-1} only.

$$E(TSS) = (t-1) + O(n^{-1}) \tag{16}$$

$$\mathrm{var}(TSS) = 2(t-1) + O(n^{-1}) \tag{17}$$

$$E(RSS) = t(n-1) + O(n^{-1}) \tag{18}$$

$$\mathrm{var}(RSS) = \frac{2t}{(\pi-2)^2}\left[(n-1)\left(\pi^2-8\right)+2(\pi-8)^2\right] + O(n^{-1}) \tag{19}$$

Hence the limit of the ratio of variance to mean is 2 for TSS but $2(\pi^2-8)/(\pi-2)^2 = 2.869$ for RSS - and hence Figure 2. The RSS is severely overdispersed, and will *never* follow a χ^2 distribution, even asymptotically. Levene's test can therefore *never* follow an F distribution with standard degrees of freedom, even asymptotically.

Approximate distribution of Levene's adjusted test

A standard approximation to the distribution of Levene's adjusted test can be investigated as follows. We replace the TSS by a variable distributed as $c_1\sigma^2\chi^2_{\nu_1}$, where c_1 and ν_1 are chosen so that the first *two* central moments of this distribution coincide with $E(TSS)$ and $\mathrm{var}(TSS)$: Yitnosumarto and O'Neill's (1986) approximation was based on equating means only. Similarly, we replace the RSS by a variable distributed as $c_2\sigma^2\chi^2_{\nu_2}$, where c_2 and ν_2 are chosen so that the first *two* central moments of this distribution coincide with $E(RSS)$ and $\mathrm{var}(RSS)$. Thus, under the hypothesis of equal treatment variances and assuming for the moment that the χ^2 distributions are independent, Levene's adjusted test can be approximated by a cF_{ν_1,ν_2} distribution, where

$$c = \frac{c_1}{c_2}\frac{\nu_1}{t-1}\frac{N-t}{\nu_2}, \quad c_1 = \frac{\mathrm{var}(TSS)}{2E(TSS)}, \quad c_2 = \frac{\mathrm{var}(RSS)}{2E(RSS)}, \quad \nu_1 = \frac{2[E(TSS)]^2}{\mathrm{var}(TSS)}, \quad \nu_2 = \frac{2[E(RSS)]^2}{\mathrm{var}(RSS)}.$$

Notice that c_1 and c_2 are simply *half* the ratios given in Table 4; c_1 therefore tends to unity for large sample sizes, whereas c_2 tends to $(\pi^2-8)/(\pi-2)^2 = 1.435$. This is an important result, because strictly speaking the individual treatment components of RSS each have this approximate distribution. That the multiplier c_2 is almost constant, especially for n>4, gives weight to the χ^2 approximation for the overall RSS: we do not end up with a sum of approximate χ^2 distributions, each with different multipliers and different degrees of freedom. Finally, notice that $c_1 \nu_1 = E(TSS)$ and $c_2 \nu_2 = E(RSS)$, so that, from Eq (16)-Eq(18), c tends to unity, .ν_1 tends to (t-1), whereas ν_2 tends to $(\pi-2)^2/(\pi^2-8)$ t(n-1) = 0.697 t(n-1).

To illustrate, for a design with 4 replicates in each of 4 treatments we use an approximate 1.436 $F_{2.741,7.791}$ distribution rather than an $F_{3,12}$ distribution. For a large number of replicates in each of 4 treatments, we use an approximate $F_{3,\ 2.788(n-1)}$ distribution rather than an $F_{3,\ 4(n-1)}$ distribution. For the example being considered, the approximate distribution is not particularly good, as is shown in Figure 4. The reason is that for small samples the TSS and RSS are correlated, and hence the initial assumption of approximating these by *independent* variables which are proportional to χ^2 distributions is inappropriate, at least for small samples.

We are currently investigating an approximation which better fits the tail probabilities.

Acknowledgements

I wish to express my appreciation to the staff at Reading University, UK, where much of this work was carried out whilst the author was on leave. Special thanks go to David Porter for many suggestions on the difficult integrals involved in Appendix 3, and to Marilyn Collins for obtaining the SPSS output. Thanks also to Phil Leppard from the University of Adelaide for obtaining the BMDP output.

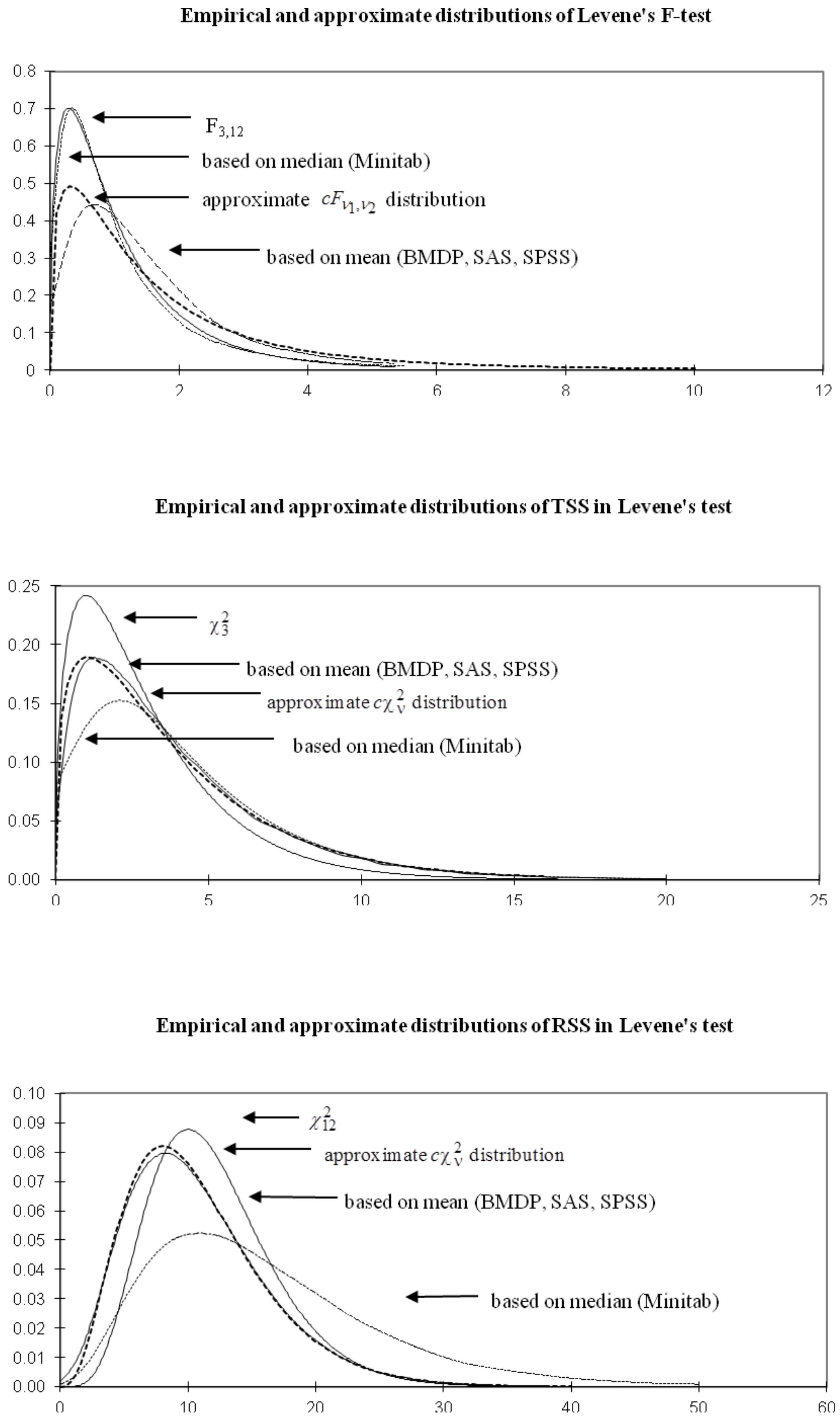

Figure 4 Empirical and assumed distributions involved in Levene's modified F statistic; the approximate F-distribution is obtained by ignoring the covariance between TSS and RSS

References

Fisher, R.A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Not. Roy. Astron. Soc.* **80**, 758-70.

Godwin, H.J. (1945). On the distribution of the estimate of mean deviation obtained from samples from a normal population. Biometrika, **33**, 254-6.

Godwin, H.J. (1946). A further note on the mean deviation. Biometrika, **35**, 304-9.

Gradshteyn, I.S. and Ryzhik, I.M. (1994). *Table of Integrals, Series and Products*. Translated from the Russian by Scripta Technica, Inc., and edited by Alan Jeffrey. Academic Press, New York.

Hartley, H.O. (1945). Note on the calculation of the distribution of the estimate of mean deviation in normal samples. Biometrika, **33**, 257-65.

Kendall, M. And Stuart, A. (1977). *The Advanced Theory of Statistics; Volume 1 (Fourth Edition).* Charles Griffin, London.

Keyes, T.K. and Levy, M.S. (1997). Analysis of Levene's test under design imbalance. J. Educ. and Behav. Statistics, *in press*.

Milliken, G. and Johnson, D. (1984). *Analysis of Messy Data, Volume 1: Designed Experiments*. Van Nostrand Reinhold Company, New York.

Yitnosumarto, S. and O'Neill, M.E. (1986). On Levene's tests of variance homogeneity. *Austral. J. Statist.*, **28**, 230-241.

Appendix 1

Distribution and second-order moments of the U_{ij}

Let $\phi(\quad ,\ldots, \quad ;\Sigma)$ denote the pdf of a multinormal variable which has mean **0** and covariance matrix Σ.

Firstly, let $\{ Y_j$, j=1, ... ,n$\}$ be NID(μ, σ^2). Then $\mathbf{R} = \{R_j\} = \{Y_j - \bar{y}\}$ is multinormal with mean **0** and covariance matrix $\boldsymbol{\Sigma} = \sigma^2\left[\mathbf{I}_n - \frac{1}{\mathrm{n}}\mathbf{J}_n\right]$. Marginally,

$$D_j = |R_j| = |Y_j - \bar{y}| \tag{A1.1}$$

is a semi-normal variable with pdf $\left[\phi\left(-D_j; \frac{\mathrm{n}-1}{\mathrm{n}}\sigma^2\right) + \phi\left(D_j; \frac{\mathrm{n}-1}{\mathrm{n}}\sigma^2\right)\right]$, where $D_j \geq 0$. It is well known that

$$E(D_j) = \sqrt{\frac{2}{\pi}\left(1 - \frac{1}{\mathrm{n}}\right)\sigma^2} \tag{A1.2}$$

and

$$\mathrm{var}(D_j) = \left(1 - \frac{2}{\pi}\right)\left(1 - \frac{1}{\mathrm{n}}\right)\sigma^2 . \tag{A1.3}$$

More generally, the pdf of a subset of (say the first) *k* of the $\{D_j\}$ is the sum of 2^k components:

$$\left[\phi\left(-D_1, -D_2, \ldots, -D_k; \boldsymbol{\Sigma}^*_{k,n}\right) + \phi\left(D_1, -D_2, \ldots, -D_k; \boldsymbol{\Sigma}^*_{k,n}\right) + \ldots + \phi\left(D_1, D_2, \ldots, D_k; \boldsymbol{\Sigma}^*_{k,n}\right)\right] \tag{A1.4}$$

where $\boldsymbol{\Sigma}^*_{k,n} = \sigma^2\left[\mathbf{I}_k - \frac{1}{\mathrm{n}}\mathbf{J}_k\right]$. This is equivalent to "folding" a multinormal distribution k times, and hence the description "folded-normal"; the resultant pdf is non-zero only over the positive region $\{D_1 \geq 0, \ldots, D_k \geq 0\}$. Fisher (1920) considered the bivariate distribution (although there is a slight misprint in the constant of his pdf). In particular, he gave the (correct) result

$$E(D_1 D_2) = \frac{2}{\pi}\frac{1}{\mathrm{n}}\left(\sqrt{\mathrm{n}(\mathrm{n}-2)} + \sin^{-1}\left(\frac{1}{\mathrm{n}-1}\right)\right)\sigma^2 \tag{A1.5}$$

from which

$$\mathrm{cov}\left(D_1, D_2\right) = \frac{2}{\pi}\left(1 - \frac{1}{\mathrm{n}}\right)\left(\mathrm{s}_n - 1\right)\sigma^2 \tag{A1.6}$$

and

$$\mathrm{cov}\left(U_1, U_2\right) = \frac{2}{\pi - 2}\left(\mathrm{s}_n - 1\right)\sigma^2 \tag{A1.7}$$

where

$$U_j = \left(1 - \frac{2}{\pi}\right)^{-1/2}\left(1 - \frac{1}{\mathrm{n}}\right)^{-1/2} D_j \,. \tag{A1.8}$$

The matrix $\mathbf{\Sigma}^*_{k,n}$ has a simple structure. In particular,

$$\left|\mathbf{\Sigma}^*_{k,n}\right| = \frac{\mathrm{n} - \mathrm{k}}{\mathrm{n}}\sigma^{2\mathrm{k}} \tag{A1.9}$$

and

$$\mathbf{\Sigma}^{*-1}_{k,n} = \sigma^{-2}\left[\mathbf{I}_k + \frac{1}{\mathrm{n} - \mathrm{k}}\mathbf{J}_k\right] \tag{A1.10}$$

The variables U_j, defined in Eq (A1.8), are scaled to have variances σ^2 and a covariance structure defined in Eq (A1.7). The pdf of a subset of (say the first) k of the $\left\{U_j\right\}$ is the sum of 2^{k} components:

$$\left[\phi\left(-U_1, -U_2, \ldots, -U_k; \mathbf{\Sigma}_{k,n}\right) + \phi\left(U_1, -U_2, \ldots, -U_k; \mathbf{\Sigma}_{k,n}\right) + \ldots + \phi\left(U_1, U_2, \ldots, U_k; \mathbf{\Sigma}_{k,n}\right)\right] \tag{A1.11}$$

where

$$\mathbf{\Sigma}_{k,n} = \frac{\pi}{\pi - 2}\frac{\mathrm{n}}{\mathrm{n} - 1}\sigma^2\left[\mathbf{I}_k - \frac{1}{\mathrm{n}}\mathbf{J}_k\right] \tag{A1.12}$$

$$\left|\mathbf{\Sigma}_{k,n}\right| = \left(\frac{\pi}{\pi - 2}\frac{\mathrm{n}}{\mathrm{n} - 1}\right)^{\mathrm{k}}\frac{\mathrm{n} - \mathrm{k}}{\mathrm{n}}\sigma^{2\mathrm{k}} \tag{A1.13}$$

and

$$\mathbf{\Sigma}^{-1}_{k,n} = \left(\frac{\pi}{\pi - 2}\frac{\mathrm{n}}{\mathrm{n} - 1}\sigma^2\right)^{-1}\left[\mathbf{I}_k + \frac{1}{\mathrm{n} - \mathrm{k}}\mathbf{J}_k\right] \tag{A1.14}$$

Appendix 2

Proof of Results 1-3

Let $\mathbf{R}$ represent the vector of residuals (in the order R_{11}, R_{12}, ...) from a oneway ANOVA of normal data, Y_{ij}, assumed NID(μ_i, σ_i^2). Then $\mathbf{R}$ is multinormal with mean **0** and covariance matrix

$$\text{var}\left(\mathbf{R}\right) = BlockDiagonal\left[\sigma_i^2\left(\mathbf{I}_{n_i} - \frac{1}{\text{n}_\text{i}}\mathbf{J}_{n_i}\right)\right]$$

where with this notation we indicate only the i^{th} of t similarly-defined block matrices. Using the results of Appendix 1, the covariance matrix of the corresponding vector $\mathbf{U}$ is

$$\text{var}\left(\mathbf{U}\right) = BlockDiagonal\left[\frac{\pi}{\pi - 2}\sigma_i^2\left(\alpha_i \mathbf{I}_{n_i} + \beta_i \mathbf{J}_{n_i}\right)\right]$$

where

$$\alpha_i = 1 - \frac{2}{\pi}\text{s}_{n_i},$$

$$\beta_i = \frac{2}{\pi}\left(\text{s}_{n_i} - 1\right).$$

Now make an orthogonal transformation $\mathbf{T} = \mathbf{P}'\mathbf{U}$, where the columns of **P** are constructed in the following way. The first $\left(\text{n}_1 - 1\right)$ columns are contrasts among the n_1 folded-normal *U*-variables in treatment 1; the next $\left(\text{n}_2 - 1\right)$ columns are contrasts among the n_2 folded-normal *U* variables in treatment 2; and so on, completing $\mathbf{P}_1$ the first (N-t) columns of **P**. The next (t-1) columns are contrasts among the t treatment mean deviations; these form $\mathbf{P}_2$. The final column, $\mathbf{P}_3$, is proportional to the unit vector. An example of **P** based on Helmert matrices, in the case of t=3 with (2,3,4) replicates is

$$\mathbf{P} = \left[\begin{array}{c|cc|ccc|cc|c}
\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & \frac{3}{\sqrt{30}} & \frac{4}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
-\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & \frac{3}{\sqrt{30}} & \frac{4}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
\hline
0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 & 0 & 0 & -\frac{2}{\sqrt{30}} & \frac{4}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 & 0 & 0 & -\frac{2}{\sqrt{30}} & \frac{4}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
0 & 0 & -\frac{2}{\sqrt{6}} & 0 & 0 & 0 & -\frac{2}{\sqrt{30}} & \frac{4}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
\hline
0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} & 0 & -\frac{5}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} & 0 & -\frac{5}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
0 & 0 & 0 & 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} & 0 & -\frac{5}{\sqrt{180}} & \frac{1}{\sqrt{9}} \\
0 & 0 & 0 & 0 & 0 & -\frac{3}{\sqrt{12}} & 0 & -\frac{5}{\sqrt{180}} & \frac{1}{\sqrt{9}}
\end{array}\right]$$

For simplicity write $\mathbf{T}' = \left\{V_{11}, \ldots, V_{t,n_t-1} \;\middle|\; Z_1, \ldots, Z_{t-1} \;\middle|\; G\right\} = \left\{\mathbf{V}' | \mathbf{Z}' | \mathbf{G}\right\}$; G is proportional to the overall mean $\bar{u}$. Then the covariance matrix of $\mathbf{T}$ is

$$\operatorname{var}(\mathbf{T}) \;=\; \mathbf{P}' \; \mathit{BlockDiagonal}\left[\frac{\pi}{\pi-2}\sigma_i^2\left(\alpha_i \mathbf{I}_{n_i} + \beta_i \mathbf{J}_{n_i}\right)\right] \; \mathbf{P}.$$

This matrix can be partitioned to conform with the partitioning of $\mathbf{T}$. The orthogonality of the columns of $\mathbf{P}$ ensures the following.

$$\operatorname{var}(\mathbf{V}) \;=\; \mathit{BlockDiagonal}\left[\frac{\pi}{\pi-2}\sigma_i^2\left(\alpha_i \mathbf{I}_{n_i-1}\right)\right]$$

$$\operatorname{cov}(\mathbf{V},\mathbf{Z}) \;=\; \mathbf{0}_{N-t,t-1}$$

$$\operatorname{cov}(\mathbf{V},\mathbf{G}) \;=\; \mathbf{0}_{N-t,1}$$

$$\operatorname{var}(\mathbf{Z}) \;=\; \mathbf{P}_2' \; \mathit{BlockDiagonal}\left[\frac{\pi}{\pi-2}\sigma_i^2\left(\alpha_i \mathbf{I}_{n_i} + \beta_i \mathbf{J}_{n_i}\right)\right] \; \mathbf{P}_2$$

$$= \frac{\pi}{\pi-2}(\alpha + \mathrm{n}\beta)\sigma^2 \; \mathbf{I}_{t-1} \text{ for equal replication and equal treatment variances}$$

$$\operatorname{cov}(\mathbf{Z},\mathbf{G}) \;=\; \mathbf{P}_2' \; \mathit{BlockDiagonal}\left[\frac{\pi}{\pi-2}\sigma_i^2\left(\alpha_i \mathbf{I}_{n_i} + \beta_i \mathbf{J}_{n_i}\right)\right] \; \mathbf{P}_3$$

$$= \mathbf{0}_{t-1,1} \qquad \text{for equal replication and equal treatment variances}$$

$$\mathrm{var}(\mathbf{G}) = \frac{\pi}{\pi-2}\frac{1}{N}\sum n_i(\alpha_i + n_i\beta_i)\sigma_i^2$$

$$= \frac{\pi}{\pi-2}(\alpha + n\beta)\sigma^2 \qquad \text{for equal replication and equal treatment variances.}$$

Appendix 3

Moments of RSS and TSS

3.1 Preamble

Consider first the Residual SS for a single treatment, based on the U_j as defined in Eq (A1.7):

$$RSS = \sum_{j=1}^{n}\left(U_j - \bar{u}\right)^2 = \frac{\mathrm{n}-1}{\mathrm{n}}\Sigma U_j^2 - \frac{2}{\mathrm{n}}\Sigma^* U_j U_k \tag{A3.1}$$

where Σ^* represents summation over all ordered subscripts (in this case j and k such that j<k). Now the U_j are identically distributed, and hence application of the second-order moments given in Appendix 1 gives rise to *E*(RSS).

Next we expand RSS^2. Gathering similar terms, we finally obtain

$$RSS^2 = \frac{1}{\mathrm{n}^2}\Big[(\mathrm{n}-1)^2\Sigma U_j^4 + 2(\mathrm{n}^2 - 2\mathrm{n} + 3)\Sigma^* U_j^2 U_k^2 - 4(\mathrm{n}-3)\Sigma^* U_j^2 U_k U_\ell$$

$$- 4(\mathrm{n}-1)\Sigma^* U_j^3 U_k + 24\Sigma^* U_j U_k U_\ell U_m\Big]. \tag{A3.2}$$

Again, the U_j are identically distributed, and so the $E(RSS^2)$ can be evaluated in terms of the fourth-order moments of the folded-normal variables. We obtain

$$E(RSS^2) = \frac{\mathrm{n}-1}{\mathrm{n}}\Big[(\mathrm{n}-1)\lambda'_4 + (\mathrm{n}^2 - 2\mathrm{n} + 3)\lambda'_{22} - 2(\mathrm{n}-2)(\mathrm{n}-3)\lambda'_{211}$$

$$- 4(\mathrm{n}-1)\lambda'_{31} + (\mathrm{n}-2)(\mathrm{n}-3)\lambda'_{1111}\Big] \tag{A3.3}$$

where as many subscripts are used as are necessary for the non-central moments

$$\lambda'_{abcd} = E(U_j^a U_k^b U_\ell^c U_m^d). \tag{A3.4}$$

These moments can be found from multinormal theory (in the case of λ'_4 and λ'_{22}) or by direct integration. The integral that gives λ'_{1111}, however, is extremely complex, and involves a transcendental function. We can obtain this more directly.

3.2 Moments of the mean deviation

Consider the mean deviation

$$\bar{d} = \frac{1}{n}\Sigma D_i = \frac{1}{n}\Sigma\left|Y_i - \bar{y}\right| .$$

Godwin (1946) obtained the s[th] moment of the distribution of $\bar{d}$ (when sampling Y_i from a standardised normal distribution) as

$$E\left(\bar{d}^{s}\right) = n^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}(2/n)^{s}\, I(n,0,s) \tag{A3.5}$$

where $I(n,0,s)$ can be calculated from a recurrence relation involving certain integrals, namely

$$I(n,r,s) = \frac{r(n-r)(s-1)}{n} I(n,r,s-2) + \frac{(n-2r)(n-2r-1)(s-1)}{n} I(n,r+1,s-2)$$
$$+ \frac{(n-2r)(n-2r-1)}{n} I(n-1,r,s-1) + \frac{r(n-r)}{n} I(n-1,r-1,s-1) \tag{A3.6}$$

A knowledge of $I(n,0,0)$, $I(n,1,0)$ and $I(n,2,0)$ is sufficient to give the first five moments of $\bar{d}$. Using direct integration, Godwin obtained the following starting values for Eq (A3.5):

$$I(n,0,0) = \frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}$$

$$I(n,1,0) = \frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}\left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\frac{1}{n-1}\right]$$

and

$$I(n,2,0)=\frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}\left[\frac{1}{16}+\frac{1}{8\pi}\left(\sin^{-1}\frac{1}{n-1}+\sin^{-1}\frac{1}{n-3}\right)+\frac{5}{4\pi^2}\sin^{-1}\frac{1}{n-1}\sin^{-1}\frac{1}{n-3}-\frac{1}{\pi^2}I_n\right] \tag{A3.7}$$

where

$$I_n = \int_0^{\sin^{-1}\frac{1}{n-3}} \tan^{-1}\sqrt{\frac{2-(n-1)(n-4)\tan^2(\varphi)}{n(n-3)}}\,d\varphi\,. \tag{A3.8}$$

Now adjusting back to a normal distribution with standard deviation σ, we obtain from Eq (A3.5) and Eq (A3.6) the following moments:

First moment

$$E\left(\bar{d}\right)=\frac{\sqrt{n}}{(2\pi)^{\frac{1}{2}(n-1)}}\frac{2}{n}(n-1)I(n-1,0,0)\sigma=\sqrt{\frac{2}{\pi}}\sqrt{\frac{n-1}{n}}\sigma \tag{A3.9}$$

Second moment

$$E\left(\bar{d}^2\right)=\frac{\sqrt{n}}{(2\pi)^{\frac{1}{2}(n-1)}}\frac{4}{n^2}(n-1)\left[I(n,1,0)+I(n-1,0,1)\right]\sigma^2$$

$$=\frac{\sqrt{n}}{(2\pi)^{\frac{1}{2}(n-1)}}\frac{4}{n^2}(n-1)\left[I(n,0,0)\left(\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\frac{1}{n-1}\right)+(n-2)I(n-2,0,0)\right]\sigma^2$$

$$=\frac{2}{\pi}\frac{n-1}{n^2}\left[\frac{\pi}{2}+\sqrt{n(n-2)}+\sin^{-1}\frac{1}{n-1}\right]\sigma^2 \tag{A3.10}$$

and hence

$$\mathrm{var}\left(\bar{d}\right)=\frac{2}{\pi}\frac{n-1}{n^2}\left[\frac{\pi}{2}-n+\sqrt{n(n-2)}+\sin^{-1}\frac{1}{n-1}\right]\sigma^2 \tag{A3.11}$$

Third moment

$$E\left(\overline{d}^{3}\right)=\frac{\sqrt{n}}{(2\pi)^{\frac{1}{2}(n-1)}}\frac{8}{n^{3}}(n-1)\left[2I(n,1,1)+I(n-1,0,2)\right]\sigma^{3}.$$

From Eq (A3.6),

$$I(n,1,1)=\frac{(n-2)(n-3)}{n}I(n-1,1,0)+\frac{(n-1)}{n}I(n-1,0,0)$$

$$=\frac{1}{n}\left(\frac{(n-2)(n-3)}{4}+\frac{(n-2)(n-3)}{2\pi}\sin^{-1}\frac{1}{n-2}+(n-1)\right)I(n-1,0,0)$$

so that, after gathering similar terms,

$$E\left(\overline{d}^{3}\right)=\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\frac{1}{n^{4}}\left[2\pi\sqrt{n(n-1)}\left(\frac{3n^{2}-4n+4}{4}+\frac{3(n-2)^{2}}{2\pi}\sin^{-1}\frac{1}{n-2}\right)\right.$$

$$\left.+n(n-1)(n-2)\sqrt{n(n-3)}\right]\sigma^{3} \qquad \text{(A3.12)}$$

Fourth moment

$$E\left(\overline{d}^{4}\right)=\frac{\sqrt{n}}{(2\pi)^{\frac{1}{2}(n-1)}}\frac{16}{n^{4}}(n-1)\left[3I(n,1,2)+I(n-1,0,3)\right]\sigma^{4}.$$

From Eq (A3.6),

$$I(n,1,2)=\frac{(n-1)}{n}\left[I(n,1,0)+I(n-1,1,0)\right]+\frac{(n-2)(n-3)}{n}\left[I(n,2,0)+I(n-1,1,1)\right]$$

After some simplification, we obtain

$$I(n,1,2)=\frac{(2\pi)^{\frac{1}{2}(n-5)}}{n^{\frac{3}{2}}}\left\{\frac{\pi^{2}}{4}\left(n^{2}-n+2\right)+\frac{\pi}{2}\frac{n^{3}-2n^{2}+5n-8}{n-1}\sqrt{n(n-2)}\right.$$

$$+\frac{\pi}{2}\left(n^{2}-n+2\right)\sin^{-1}\frac{1}{n-1}+\frac{n-3}{2(n-1)}\left[\pi(n-1)(n-2)+2(n-3)(n-4)\sqrt{n(n-2)}\right]\sin^{-1}\frac{1}{n-3}$$

$$+5(\mathrm{n}-2)(\mathrm{n}-3)\sin^{-1}\frac{1}{\mathrm{n}-1}\sin^{-1}\frac{1}{\mathrm{n}-3}-4(\mathrm{n}-2)(\mathrm{n}-3)I_n\Bigg\}$$

Combining this with *I(n-1,0,3)* gives

$$E\left(\bar{d}^4\right)=\left(\frac{2}{\pi}\right)^2\frac{\mathrm{n}-1}{\mathrm{n}^5}\Bigg\{\frac{3\pi^2}{4}\left(\mathrm{n}^2-\mathrm{n}+2\right)+\mathrm{n}(\mathrm{n}-2)(\mathrm{n}-3)\sqrt{\mathrm{n}(\mathrm{n}-4)}$$

$$+\frac{\pi}{\mathrm{n}-1}\left(3\mathrm{n}^3-8\mathrm{n}^2+13\mathrm{n}-12\right)\sqrt{\mathrm{n}(\mathrm{n}-2)}+\frac{3\pi}{2}\left(\mathrm{n}^2-\mathrm{n}+2\right)\sin^{-1}\frac{1}{\mathrm{n}-1}$$

$$+\frac{3(n-2)(n-3)}{2(\mathrm{n}-1)}\left[\pi(\mathrm{n}-1)+4(\mathrm{n}-3)\sqrt{\mathrm{n}(\mathrm{n}-2)}\right]\sin^{-1}\frac{1}{\mathrm{n}-3}$$

$$+15(\mathrm{n}-2)(\mathrm{n}-3)\sin^{-1}\frac{1}{\mathrm{n}-1}\sin^{-1}\frac{1}{\mathrm{n}-3}-12(\mathrm{n}-2)(\mathrm{n}-3)I_n\Bigg\}\sigma_i^4 \qquad \text{(A3.13)}$$

3.3 Fourth-order moments of the (scaled) absolute residuals $\{U_j\}$

Recall that $U_j = \left(1-\frac{2}{\pi}\right)^{-1/2}\left(1-\frac{1}{n}\right)^{-1/2}|R_j|$, where $\{R_j\}$ is multinormal with mean **0** and covariance $\Sigma = \sigma^2\left[\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right]$. Hence, from multinormal theory,

$$\lambda'_4 = E\left(U_j^4\right) = 3\left(\frac{\pi}{\pi-2}\right)^2 \sigma^4 \qquad \text{(A3.14)}$$

and

$$\lambda'_{22} = E\left(U_j^2 U_k^2\right) = \left(\frac{\pi}{\pi-2}\right)^2 \frac{n^2-2n+3}{(n-1)^2}\sigma^4 \qquad \text{(A3.15)}$$

The derivations of the next five results are quite long, and the proofs are available upon request.

Result A3.1

$$\lambda'_{31} = E\left(U_j^3 U_k\right) = \left(\frac{\pi}{\pi-2}\right)^2 \frac{2}{\pi}\left[\frac{2n^2-4n+3}{(n-1)^3}\sqrt{n(n-2)} + \frac{3}{n-1}\sin^{-1}\frac{1}{n-1}\right]\sigma^4$$

Result A3.2

$$\lambda'_{211} = E\left(U_j^2 U_k U_\ell\right) = \left(\frac{\pi}{\pi-2}\right)^2 \frac{2}{\pi}\left[\frac{n^2-2n+3}{(n-1)^3}\sqrt{n(n-2)} + \frac{n-3}{(n-1)^2}\sin^{-1}\frac{1}{n-1}\right]\sigma^4$$

Result A3.3

$$\lambda'_{1111} = E\left(U_j U_k U_\ell U_m\right) = \left(\frac{2}{\pi-2}\right)^2\left[\frac{n}{(n-1)^2}\sqrt{n(n-4)} + \frac{3\pi}{2(n-1)^2}\left(\sin^{-1}\frac{1}{n-3} - \sin^{-1}\frac{1}{n-1}\right)\right.$$

$$\left. + \frac{6(n-3)}{(n-1)^3}\sqrt{n(n-2)}\sin^{-1}\frac{1}{n-3} + \frac{15}{(n-1)^2}\sin^{-1}\frac{1}{n-1}\sin^{-1}\frac{1}{n-3} - \frac{12}{(n-1)^2}I_n\right]\sigma^4$$

Result A3.4

$$\lambda'_{21} = E\left(U_j^2 U_k\right) = \left(\frac{\pi}{\pi-2}\right)^{\frac{3}{2}} \sqrt{\frac{2}{\pi}}\left[\frac{n^2-2n+2)}{(n-1)^2}\right]\sigma^3$$

Result A3.5

$$\lambda'_{111} = E\left(U_j U_k U_\ell\right) = \left(\frac{2}{\pi-2}\right)^{\frac{3}{2}}\left[\frac{n}{n-1}\sqrt{\frac{n-3}{n-1}} + \frac{3(n-2)}{(n-1)^2}\sin^{-1}\frac{1}{n-2}\right]\sigma^3$$

3.4 Mean and variance of RSS

Using Eq (A3.1) and Eq (A1.7) we have, for a single treatment,

$$E(\text{RSS}) = (n-1)\left[\text{var}\left(U_j\right) - \text{cov}\left(U_j, U_k\right)\right]$$

$$= (n-1)\left[1 - \frac{2}{\pi-2}\left(s_n - 1\right)\right]\sigma^2 \qquad \text{(A3.16)}$$

or $$(n-1)\frac{2}{\pi-2}\left[\frac{\pi}{2} - s_n\right]\sigma^2$$

Now we use Eq (A3.3) and the results above, obtaining, for a *single* treatment

$$\text{var}(RSS) = \left(\frac{2}{\pi-2}\right)^2 \left\{ \frac{(n-2)(n-3)}{n-1}\sqrt{n(n-4)} - n(n-2) + \frac{\pi^2\left(n^3-n^2-n+3\right)}{2n(n-1)} \right.$$

$$- \frac{4\pi\left(n^2-3n+3\right)}{n(n-1)^2}\sqrt{n(n-2)} - \frac{1}{2}\left[\pi\frac{3n^2+n-6}{n(n-1)} + 4\sqrt{n(n-2)} + 2\sin^{-1}\frac{1}{n-1}\right]\sin^{-1}\frac{1}{n-1}$$

$$+ \frac{3(n-2)(n-3)}{2n(n-1)}\left[\pi + \frac{4(n-3)}{n-1}\sqrt{n(n-2)}\right]\sin^{-1}\frac{1}{n-3}$$

$$\left. + \frac{3(n-2)(n-3)}{n(n-1)}\left[5\sin^{-1}\frac{1}{n-1}\sin^{-1}\frac{1}{n-3} - 4I_n\right]\right\}\sigma^4 \qquad \text{(A3.17)}$$

These two formulae are easily generalised to cover the case of *t* treatments.

3.5 Mean and variance of TSS

The Treatment SS is simply a weighted sum of squares of the (independent) treatment mean deviations. Thus,

$$TSS = \sum_{i=1}^{t} n_i \left(\bar{u}_{i.} - \bar{u}\right)^2 = \sum n_i \left(1 - \frac{n_i}{N}\right)\bar{u}_{i.}^2 - \frac{2}{N}\sum{}^{*} n_i n_j \bar{u}_{i.}\bar{u}_{j.} \qquad \text{(A3.18)}$$

From Eq (A3.9) to Eq (A3.12) we have

$$E\left(\bar{u}_{i.}\right) = \sqrt{\frac{2}{\pi - 2}}\sigma_i \qquad \text{(A3.19)}$$

$$E\left(\bar{u}_{i.}^2\right) = \frac{2}{\pi - 2}\frac{1}{n_i}\left[\frac{\pi}{2} + \sqrt{n_i(n_i - 2)} + \sin^{-1}\frac{1}{n_i - 1}\right]\sigma_i^2$$

$$= \frac{2}{\pi - 2}\frac{1}{n_i}\left[\frac{\pi}{2} + (n_i - 1)s_{n_i}\right]\sigma_i^2 \qquad \text{(A3.20)}$$

and hence the *second* central moment of the i$^{\text{th}}$ adjusted mean deviation is

$$\mu_{2i} = \text{var}\left(\bar{u}_{i.}\right) = \frac{\sigma_i^2}{n_i}\left[1 + \frac{2}{\pi - 2}(n_i - 1)\left(s_{n_i} - 1\right)\right] \qquad \text{(A3.21)}$$

where μ_{ki} represents the k$^{\text{th}}$ centred moment of $\bar{u}_{i.}$. Next,

$$E\left(\bar{u}_{i.}^3\right) = \left(\frac{2}{\pi - 2}\right)^{\frac{3}{2}}\frac{1}{n_i}\left[\frac{\pi}{2}\frac{3n_i^2 - 4n_i + 4}{n_i(n_i - 1)} + \frac{3(n_i - 2)^2}{n_i(n_i - 1)}\sin^{-1}\frac{1}{n_i - 2} + \left(n_i - 2\right)\sqrt{\frac{n_i - 3}{n_i - 1}}\right]\sigma_i^3 \qquad \text{(A3.22)}$$

and hence the *third* central moment of the i$^{\text{th}}$ adjusted mean deviation is

$$\mu_{3i} = \left(\frac{2}{\pi - 2}\right)^{\frac{3}{2}}\frac{\sigma_i^3}{n_i}\left\{2n_i - 3\sqrt{n_i(n_i - 2)} + \left(n_i - 2\right)\sqrt{\frac{n_i - 3}{n_i - 1}} - \frac{\pi}{2}\frac{n_i - 4}{n_i(n_i - 1)}\right.$$

$$\left. - 3\sin^{-1}\frac{1}{n_i - 1} + \frac{3(n_i - 2)^2}{n_i(n_i - 1)}\sin^{-1}\frac{1}{n_i - 2}\right\} \qquad \text{(A3.23)}$$

Finally,

$$E\left(\bar{u}_{i.}^{4}\right)=\left(\frac{2}{\pi-2}\right)^{2}\frac{1}{n_i^3\left(n_i-1\right)}\left\{\frac{3\pi^2}{4}\left(n_i^2-n_i+2\right)+n_i(n_i-2)(n_i-3)\sqrt{n_i(n_i-4)}\right.$$

$$+\frac{\pi}{n_i-1}\left(3n_i^3-8n_i^2+13n_i-12\right)\sqrt{n_i(n_i-2)}+\frac{3\pi}{2}\left(n_i^2-n_i+2\right)\sin^{-1}\frac{1}{n_i-1}$$

$$+\frac{3(n_i-2)(n_i-3)}{2(n_i-1)}\left[\pi(n_i-1)+4(n_i-3)\sqrt{n_i(n_i-2)}\right]\sin^{-1}\frac{1}{n_i-3}$$

$$\left.+15\left(n_i-2\right)\left(n_i-3\right)\sin^{-1}\frac{1}{n_i-1}\sin^{-1}\frac{1}{n_i-3}-12\left(n_i-2\right)\left(n_i-3\right)I_{n_i}\right\}\sigma_i^4 \qquad \text{(A3.24)}$$

and hence the *fourth* central moment of the i^th adjusted mean deviation is

$$\mu_{4i}=\left(\frac{2}{\pi-2}\right)^{2}\frac{\sigma_i^4}{n_i}\left\{\left(6n_i+\pi\frac{3n_i^3-8n_i^2+13n_i-12}{n_i(n_i-1)^2}\right)\sqrt{\frac{n_i-2}{n_i}}-3n_i-4\left(n_i-2\right)\sqrt{\frac{n_i-3}{n_i-1}}\right.$$

$$+\frac{\left(n_i-2\right)\left(n_i-3\right)}{n_i-1}\sqrt{\frac{n_i-4}{n_i}}-\pi\frac{3n_i^2-5n_i+8}{n_i(n_i-1)}+\frac{3\pi^2}{4}\frac{n_i^2-n_i+2}{n_i^2(n_i-1)}$$

$$-\frac{12\left(n_i-2\right)^2}{n_i(n_i-1)}\sin^{-1}\frac{1}{n_i-2}+\frac{3}{2}\left(\pi\frac{n_i^2-n_i+2}{n_i^2(n_i-1)}+4\right)\sin^{-1}\frac{1}{n_i-1}$$

$$+\frac{3(n_i-2)(n_i-3)}{2n_i^2(n_i-1)}\left[\pi+4\frac{n_i-3}{n_i-1}\sqrt{n_i(n_i-2)}\right]\sin^{-1}\frac{1}{n_i-3}$$

$$\left.+\frac{3\left(n_i-2\right)\left(n_i-3\right)}{n_i^2\left(n_i-1\right)}\left(5\sin^{-1}\frac{1}{n_i-1}\sin^{-1}\frac{1}{n_i-3}-4I_{n_i}\right)\right\} \qquad \text{(A3.25)}$$

Thus, because of the independence of the $\bar{u}_{i.}$,

$$E(TSS)=\sum n_i\left(1-\frac{n_i}{N}\right)E\left(\bar{u}_{i.}^{2}\right)-\frac{2}{N}\sum{}^{*}n_i n_j E\left(\bar{u}_{i.}\right)E\left(\bar{u}_{j.}\right)$$

$$=\frac{2}{\pi-2}\left\{\sum\left(1-\frac{n_i}{N}\right)\left[\frac{\pi}{2}+\left(n_i-1\right)s_{n_i}\right]\sigma_i^2-\frac{2}{N}\sum{}^{*}n_i n_j\sigma_i\sigma_j\right\}$$

$$= \frac{2}{\pi - 2}\left\{\sum\left(1 - \frac{n_i}{N}\right)\left[\frac{\pi}{2} + (n_i - 1)s_{n_i} - n_i\right]\sigma_i^2 + \sum n_i\left(\sigma_i - \bar{\sigma}\right)^2\right\}$$

$$= \frac{2}{\pi - 2}\left\{\sum\left(1 - \frac{n_i}{N}\right)\left[\frac{\pi - 2}{2} + (n_i - 1)\left(s_{n_i} - 1\right)\right]\sigma_i^2 + \sum n_i\left(\sigma_i - \bar{\sigma}\right)^2\right\}$$

$$= \sum\left(1 - \frac{n_i}{N}\right)\left[1 + \frac{2}{\pi - 2}(n_i - 1)\left(s_{n_i} - 1\right)\right]\sigma_i^2 + \frac{2}{\pi - 2}\sum n_i\left(\sigma_i - \bar{\sigma}\right)^2 \qquad \text{(A3.26)}$$

where $\bar{\sigma}$ is a (weighted) average of the σ_i. The latter term in Eq (A3.23) disappears under the hypothesis of equal treatment variances.

In a manner similar to Eq (A3.2), we have

$$TSS^2 = \sum n_i^2\left(1 - \frac{n_i}{N}\right)^2 \bar{u}_{i.}^4 + 2\sum{}^* n_i n_j\left(1 - \frac{n_i}{N} - \frac{n_j}{N} + \frac{3n_i n_j}{N^2}\right)\bar{u}_{i.}^2\bar{u}_{j.}^2$$

$$-\frac{4}{N}\sum_{i \neq j,k}\sum_{j<k} n_i\left(1 - \frac{3n_i}{N}\right)n_j n_k \bar{u}_{i.}^2\bar{u}_{j.}\bar{u}_{k.} - \frac{4}{N}\sum_{i \neq j} n_i^2\left(1 - \frac{n_i}{N}\right)n_j \bar{u}_{i.}^3\bar{u}_{j.}$$

$$+\frac{24}{N^2}\sum{}^* n_i n_j n_k n_\ell \bar{u}_{i.}\bar{u}_{j.}\bar{u}_{k.}\bar{u}_{\ell.} \qquad \text{(A3.27)}$$

Since the treatment mean deviations are independent, their non-central moments can be inserted into Eq (A3.27) and var(*TSS*) can be calculated for the general case. Under the hypothesis that the treatment variances are all equal (to σ^2 say), the variance can be obtained more directly. Under this hypothesis, Eq (A3.18) and Eq (A3.27) can both be thought of as being expanded about *centred* mean deviations, and hence

$$E(TSS) = \sum n_i\left(1 - \frac{n_i}{N}\right)\mu_{2i} \qquad \text{(A3.28)}$$

$$E\left(TSS^2\right) = \sum n_i^2\left(1 - \frac{n_i}{N}\right)^2\mu_{4i} + 2\sum{}^* n_i n_j\left(1 - \frac{n_i}{N} - \frac{n_j}{N} + \frac{3n_i n_j}{N^2}\right)\mu_{2i}\mu_{2j} \qquad \text{(A3.29)}$$

and hence

$$\mathrm{var}(TSS) = \sum \mathrm{n}_i^2 \left(1 - \frac{\mathrm{n}_i}{\mathrm{N}}\right)^2 \left(\mu_{4i} - \mu_{2i}^2\right) + \frac{4}{\mathrm{N}^2} \sum {}^* \mathrm{n}_i^2 \mathrm{n}_j^2 \mu_{2i} \mu_{2j}$$

$$= \sum \mathrm{n}_i^2 \left(1 - \frac{\mathrm{n}_i}{\mathrm{N}}\right)^2 \left(\mu_{4i} - \mu_{2i}^2\right) + \frac{2}{\mathrm{N}^2} \left[\left(\sum \mathrm{n}_i^2 \mu_{2i}\right)^2 - \sum \mathrm{n}_i^4 \mu_{2i}^2 \right] \qquad \text{(A3.30)}$$

Table A3.1 presents up to fourth-order moments for the adjusted absolute residuals U_{ij}, for up to 25 replicates. Likewise, Table A3.2 presents up to fourth-order moments for the adjusted treatment mean deviations, $\bar{u}_{i.}$.

Table A3.1 Moments (non-centred, top; centred, bottom) of the adjusted absolute residuals, U_{ij}

n	λ'_{1}	λ'_{11}	λ'_{2}	λ'_{111}	λ'_{21}	λ'_{3}	λ'_{1111}	λ'_{22}	λ'_{211}	λ'_{31}	λ'_{4}
2	1.324	2.752	2.752		10.927	7.285		22.719		22.719	22.719
3	1.324	1.976	2.752	2.732	5.464	7.285		11.360	6.263	13.181	22.719
4	1.324	1.850	2.752	2.595	4.047	7.285	3.783	9.256	5.738	11.234	22.719
5	1.324	1.807	2.752	2.493	3.870	7.285	3.485	8.520	5.404	10.542	22.719
6	1.324	1.787	2.752	2.437	3.788	7.285	3.350	8.179	5.218	10.219	22.719
7	1.324	1.776	2.752	2.403	3.744	7.285	3.273	7.994	5.108	10.043	22.719
8	1.324	1.770	2.752	2.382	3.717	7.285	3.224	7.882	5.037	9.937	22.719
9	1.324	1.766	2.752	2.368	3.699	7.285	3.190	7.810	4.990	9.868	22.719
10	1.324	1.763	2.752	2.358	3.687	7.285	3.167	7.760	4.956	9.821	22.719
11	1.324	1.761	2.752	2.351	3.679	7.285	3.150	7.725	4.932	9.787	22.719
12	1.324	1.759	2.752	2.346	3.673	7.285	3.136	7.698	4.913	9.762	22.719
13	1.324	1.758	2.752	2.342	3.668	7.285	3.126	7.678	4.899	9.743	22.719
14	1.324	1.757	2.752	2.338	3.664	7.285	3.118	7.663	4.888	9.728	22.719
15	1.324	1.756	2.752	2.336	3.661	7.285	3.112	7.650	4.879	9.716	22.719
16	1.324	1.756	2.752	2.334	3.659	7.285	3.107	7.640	4.872	9.707	22.719
17	1.324	1.755	2.752	2.332	3.657	7.285	3.102	7.632	4.866	9.699	22.719
18	1.324	1.755	2.752	2.330	3.655	7.285	3.099	7.626	4.861	9.692	22.719
19	1.324	1.755	2.752	2.329	3.654	7.285	3.096	7.620	4.857	9.687	22.719
20	1.324	1.754	2.752	2.328	3.653	7.285	3.093	7.615	4.853	9.683	22.719
21	1.324	1.754	2.752	2.327	3.652	7.285	3.091	7.611	4.850	9.679	22.719
22	1.324	1.754	2.752	2.327	3.651	7.285	3.089	7.608	4.847	9.675	22.719
23	1.324	1.754	2.752	2.326	3.650	7.285	3.087	7.604	4.845	9.672	22.719
24	1.324	1.754	2.752	2.325	3.649	7.285	3.086	7.602	4.843	9.670	22.719
25	1.324	1.753	2.752	2.325	3.649	7.285	3.084	7.599	4.841	9.668	22.719

n	λ_{11}	λ_{2}	λ_{111}	λ_{21}	λ_{3}	λ_{1111}	λ_{22}	λ_{211}	λ_{31}	λ_{4}
2	1.000000	1		0.995272	0.995		3.869177		3.869177	3.869
3	0.223941	1	-0.476245	0.317801	0.995		1.534680	- 0.100491	1.099638	3.869
4	0.098263	1	-0.114548	0.144598	0.995	0.287080	1.228759	- 0.023874	0.501400	3.869
5	0.055039	1	-0.044536	0.081956	0.995	0.072565	1.127037	0.001521	0.284535	3.869
6	0.035157	1	-0.021797	0.052631	0.995	0.026558	1.080828	0.007403	0.182842	3.869
7	0.024389	1	-0.012256	0.036617	0.995	0.011953	1.055954	0.008238	0.127253	3.869
8	0.017908	1	-0.007565	0.026931	0.995	0.006156	1.041031	0.007708	0.093616	3.869
9	0.013705	1	-0.004994	0.020634	0.995	0.003488	1.031375	0.006861	0.071736	3.869
10	0.010826	1	-0.003468	0.016311	0.995	0.002122	1.024770	0.006015	0.056713	3.869
11	0.008767	1	-0.002506	0.013217	0.995	0.001364	1.020051	0.005258	0.045957	3.869
12	0.007244	1	-0.001869	0.010926	0.995	0.000917	1.016564	0.004607	0.037993	3.869
13	0.006087	1	-0.001431	0.009182	0.995	0.000639	1.013914	0.004055	0.031932	3.869
14	0.005186	1	-0.001120	0.007825	0.995	0.000458	1.011852	0.003588	0.027213	3.869
15	0.004471	1	-0.000893	0.006748	0.995	0.000338	1.010217	0.003192	0.023468	3.869
16	0.003895	1	-0.000723	0.005879	0.995	0.000254	1.008899	0.002854	0.020446	3.869

17	0.003423	1	-0.000594	0.005167	0.995	0.000195	1.007820	0.002566	0.017972	3.869
18	0.003032	1	-0.000494	0.004578	0.995	0.000152	1.006927	0.002317	0.015921	3.869
19	0.002704	1	-0.000415	0.004083	0.995	0.000120	1.006178	0.002102	0.014202	3.869
20	0.002427	1	-0.000352	0.003665	0.995	0.000096	1.005544	0.001915	0.012747	3.869
21	0.002190	1	-0.000301	0.003308	0.995	0.000078	1.005003	0.001752	0.011505	3.869
22	0.001987	1	-0.000260	0.003000	0.995	0.000064	1.004538	0.001608	0.010436	3.869
23	0.001810	1	-0.000225	0.002734	0.995	0.000053	1.004134	0.001480	0.009509	3.869
24	0.001656	1	-0.000197	0.002501	0.995	0.000044	1.003783	0.001368	0.008700	3.869
25	0.001521	1	-0.000173	0.002297	0.995	0.000037	1.003474	0.001267	0.007990	3.869

Table A3.2 Moments (non-centred, left; centred, right) of the adjusted mean deviations, $\bar{u}_{i.}$

n	μ'_1	μ'_2	μ'_3	μ'_4	μ_2	μ_3	μ_4	$\sqrt{\mu_2}$
2	1.3236	2.7519	7.2850	22.7195	1	0.9953	3.8692	1
3	1.3236	2.2346	4.4519	10.0549	0.4826	0.2166	0.7655	0.6947
4	1.3236	2.0756	3.7048	7.3452	0.3237	0.1006	0.3407	0.5689
5	1.3236	1.9960	3.3457	6.1307	0.2440	0.0578	0.1904	0.4940
6	1.3236	1.9479	3.1345	5.4489	0.1960	0.0375	0.1214	0.4427
7	1.3236	1.9157	2.9954	5.0138	0.1638	0.0263	0.0841	0.4047
8	1.3236	1.8926	2.8969	4.7125	0.1407	0.0194	0.0617	0.3751
9	1.3236	1.8752	2.8234	4.4916	0.1233	0.0149	0.0472	0.3511
10	1.3236	1.8617	2.7665	4.3229	0.1097	0.0119	0.0372	0.3313
11	1.3236	1.8508	2.7211	4.1899	0.0989	0.0096	0.0302	0.3145
12	1.3236	1.8419	2.6841	4.0823	0.0900	0.0080	0.0249	0.3000
13	1.3236	1.8345	2.6534	3.9935	0.0825	0.0067	0.0209	0.2873
14	1.3236	1.8282	2.6274	3.9190	0.0762	0.0057	0.0178	0.2761
15	1.3236	1.8228	2.6051	3.8556	0.0708	0.0050	0.0154	0.2662
16	1.3236	1.8181	2.5859	3.8009	0.0662	0.0043	0.0134	0.2572
17	1.3236	1.8140	2.5691	3.7534	0.0620	0.0038	0.0118	0.2491
18	1.3236	1.8104	2.5542	3.7117	0.0584	0.0034	0.0104	0.2417
19	1.3236	1.8071	2.5411	3.6747	0.0552	0.0030	0.0093	0.2349
20	1.3236	1.8042	2.5293	3.6418	0.0523	0.0027	0.0083	0.2287
21	1.3236	1.8016	2.5187	3.6122	0.0497	0.0024	0.0075	0.2229
22	1.3236	1.7993	2.5091	3.5856	0.0474	0.0022	0.0068	0.2176
23	1.3236	1.7971	2.5004	3.5615	0.0452	0.0020	0.0062	0.2126
24	1.3236	1.7952	2.4925	3.5395	0.0433	0.0019	0.0057	0.2080
25	1.3236	1.7934	2.4852	3.5193	0.0415	0.0017	0.0052	0.2036

Appendix 4

Numerical approximation for I_n

$$I_n = \int_0^{\sin^{-1}(1/(n-3))} \tan^{-1}\sqrt{\frac{2-(n-1)(n-4)\tan^2(\varphi)}{n(n-3)}}\,d\varphi$$

Firstly, note that $I_4 = \frac{\pi}{2}\tan^{-1}\sqrt{\frac{1}{2}} = 0.967$. For n>4, the integral can only be evaluated numerically.

Let

$\psi = \sqrt{\frac{(n-1)(n-4)}{2}}\tan(\varphi)$. Then, for n>4,

$$I_n = \frac{\sqrt{2(n-1)(n-4)}}{(n-2)(n-3)}\int_0^{\sqrt{(n-1)/2(n-2)}}\left[1-\frac{2}{(n-2)(n-3)}\left(1-\psi^2\right)\right]^{-1}\tan^{-1}\sqrt{\frac{2\left(1-\psi^2\right)}{n(n-3)}}d\psi .$$

We now expand the two expressions inside the integral. The integrand simplifies to

$$\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{1}{2}}+\frac{2(n+1)}{3(n-2)}\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{3}{2}}+\frac{13n^2-2n+12}{15(n-2)^2}\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{5}{2}}+\ldots$$

Each of these components can be integrated explicitly. Thus, I_n (n>4) is, to O(n^{-6}), the sum of the following three solutions:

i) $$\frac{\sqrt{2(n-1)(n-4)}}{(n-2)(n-3)}\int_0^{\sqrt{(n-1)/2(n-2)}}\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{1}{2}}d\psi$$

$$=2\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\frac{1}{(n-2)(n-3)}\frac{1}{2}\left\{\left[\frac{(n-1)}{2(n-2)}\left(1-\frac{(n-1)}{2(n-2)}\right)\right]^{\frac{1}{2}}+\sin^{-1}\sqrt{\frac{(n-1)}{2(n-2)}}\right\}$$

$$=\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\frac{1}{(n-2)(n-3)}\left\{\frac{\sqrt{(n-1)(n-3)}}{2(n-2)}+\tan^{-1}\sqrt{\frac{n-1}{n-3}}\right\}$$

ii) $$\frac{\sqrt{2(n-1)(n-4)}}{(n-2)(n-3)}\frac{2(n+1)}{3(n-2)}\int_0^{\sqrt{(n-1)/2(n-2)}}\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{3}{2}}d\psi$$

=

$$\frac{8}{3}\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\frac{n+1}{n(n-2)^2(n-3)^2}\left\{\frac{1}{8}\left(5-2\frac{(n-1)}{2(n-2)}\right)\left[\frac{(n-1)}{2(n-2)}\left(1-\frac{(n-1)}{2(n-2)}\right)\right]^{\frac{1}{2}}+\frac{3}{8}\tan^{-1}\sqrt{\frac{\frac{(n-1)}{2(n-2)}}{1-\frac{(n-1)}{2(n-2)}}}\right\}$$

$$=\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\,\frac{n+1}{n(n-2)^2(n-3)^2}\left\{\frac{(4n-9)\sqrt{(n-1)(n-3)}}{6(n-2)^2}+\tan^{-1}\sqrt{\frac{n-1}{n-3}}\right\}$$

iii) $$\frac{\sqrt{2(n-1)(n-4)}}{(n-2)(n-3)}\,\frac{13n^2-2n+12}{15(n-2)^2}\int_0^{\sqrt{(n-1)/2(n-2)}}\left[\frac{2}{n(n-3)}\left(1-\psi^2\right)\right]^{\frac{5}{2}}d\psi$$

$$=\frac{8}{15}\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\,\frac{13n^2-2n+12}{n^2(n-2)^3(n-3)^3}\quad x$$

$$\left\{\frac{1}{48}\left[15+10\left(1-\frac{(n-1)}{2(n-2)}\right)+8\left(1-\frac{(n-1)}{2(n-2)}\right)^2\right]\left[\frac{(n-1)}{2(n-2)}\left(1-\frac{(n-1)}{2(n-2)}\right)\right]^{\frac{1}{2}}+\frac{5}{16}\tan^{-1}\sqrt{\frac{\frac{(n-1)}{2(n-2)}}{1-\frac{(n-1)}{2(n-2)}}}\right\}$$

$$=\sqrt{\frac{(n-1)(n-4)}{n(n-3)}}\,\frac{13n^2-2n+12}{6n^2(n-2)^3(n-3)^3}\left\{\frac{\left(22n^2-97n+108\right)\sqrt{(n-1)(n-3)}}{30(n-2)^3}+\tan^{-1}\sqrt{\frac{n-1}{n-3}}\right\}$$

Appendix 5

Derivation of the product moments of the absolute residuals

In solving these integrals we tried several methods, including completing the squares to obtain Gaussian-type integrals. However, because of the range of integration, we were confronted by difficult forms of the so-called probability integral, and we could make no headway in many instances. We therefore used polar and spherical transformations, and needed results pertaining to the integral of certain rational functions.

To minimise notation, we will find the moments of the D_i and scale as in Eq (A1.8). Furthermore, without loss of generality in this section we will use $\sigma = 1$. To avoid the constant use of subscripts, we will use U, V, ... in place of D_1, D_2, ... inside the integrals.

1. **Second-order moments**

The joint pdf of (D_1, D_2) is, from Eq (A1.4),

$$f(v,w) = \frac{1}{\pi}\sqrt{\frac{n}{n-2}}\left\{ e^{-\frac{1}{2(n-2)}\left[(n-1)v^2+(n-1)w^2+2vw\right]} + e^{-\frac{1}{2(n-2)}\left[(n-1)v^2+(n-1)w^2-2vw\right]} \right\},\ v\geq 0,\ w\geq 0. \qquad \text{(A5.1)}$$

Hence, taking a ≥ b without loss of generality,

$$E\left(D_1^a D_2^b\right) = \frac{1}{\pi}\sqrt{\frac{n}{n-2}} \int_0^\infty v^a \, dv \int_0^\infty w^b \left\{ e^{-\frac{1}{2(n-2)}\left[(n-1)v^2+(n-1)w^2+2vw\right]} + e^{-\frac{1}{2(n-2)}\left[(n-1)v^2+(n-1)w^2-2vw\right]} \right\} dw$$

(A5.2)

With the use of the transformation

$$v = r\cos\theta$$

$$w = r\sin\theta$$

this integral becomes

$$E\left(D_1^a D_2^b\right) = \frac{1}{\pi}\sqrt{\frac{n}{n-2}} \int_0^{\pi/2} \cos^a\theta \sin^b \theta \, d\theta \int_0^{\infty} r^{a+b+1} \left\{ e^{-\frac{1}{2(n-2)}r^2[n-1+\sin 2\theta]} + e^{-\frac{1}{2(n-2)}r^2[n-1-\sin 2\theta]} \right\} dr$$

(A5.3)

Integrating out r requires two cases depending on whether (a + b) is even or odd.

Case a + b = 2 k

We firstly use 3.461(3) of Gradshteyn and Ryzhik (1994).

$$E\left(D_1^a D_2^b\right) = \frac{1}{\pi}\sqrt{\frac{n}{n-2}} 2^k k!(n-2)^{k+1} \int_0^{\pi/2} \left[\frac{1}{(n-1+\sin 2\theta)^{k+1}} + \frac{1}{(n-1-\sin 2\theta)^{k+1}}\right] \cos^a\theta \sin^{2k-a} \theta \, d\theta$$

.

(A5.4)

Particular cases have straightforward solutions. One method is described as follows. An alternative method is given in the next section. In the case a = 3, b = 1, we re-write $\cos^3 \theta \sin \theta$ as ¼(1 + cos x) sin x, where x = 2 θ. The integral becomes

$$E\left(D_1^3 D_2\right) = \frac{1}{\pi}\sqrt{\frac{n}{n-2}} (n-2)^3 \int_0^{\pi} \left[\frac{1}{(n-1+\sin x)^3} + \frac{1}{(n-1-\sin x)^3}\right] (1+\cos x) \sin x \, dx \, . \qquad \text{(A5.5)}$$

Note that the integral $\int_0^{\pi} \cos x \sin x \left[\frac{1}{(a+b\sin x)^3} + \frac{1}{(a-b\sin x)^3}\right] dx$ must be zero, being an odd function about the point $x = \frac{\pi}{4}$. To evaluate the remaining integral we need the following result, inferred from 2.551 of Gradshteyn and Ryzhik (1994) or obtained directly through a mathematical package such as Mathematica:

$$\int_0^{\pi} \frac{\sin x}{[a+b\sin x]^3} dx = \frac{1}{(a^2-b^2)^2} \left\{ \frac{2a^2+b^2}{a} + \frac{3ab}{\sqrt{a^2-b^2}} \left(\tan^{-1} \frac{b}{\sqrt{a^2-b^2}} - \frac{\pi}{2} \right) \right\} . \qquad \text{(A5.6)}$$

Thus

$$\int_0^{\pi} \sin x \left[\frac{1}{(\mathrm{n-1}+\sin x)^3} + \frac{1}{(\mathrm{n-1}-\sin x)^3} \right] \mathrm{d}x = \frac{2}{[\mathrm{n}(\mathrm{n-2})]^2} \left[\frac{2\mathrm{n}^2 - 4\mathrm{n} + 3}{\mathrm{n-1}} + \frac{3(\mathrm{n-1})}{\sqrt{\mathrm{n}(\mathrm{n-2})}} \tan^{-1} \frac{1}{\sqrt{\mathrm{n}(\mathrm{n-2})}} \right]. \qquad \text{(A5.7)}$$

Note that $\tan^{-1} \frac{1}{\sqrt{\mathrm{n}(\mathrm{n-2})}} = \sin^{-1} \frac{1}{\mathrm{n-1}}$, and so the final expectation can be expressed more simply as

$$E\left(D_1^3 D_2\right) = \frac{2}{\pi} \frac{1}{\mathrm{n}^2} \left[\frac{2\mathrm{n}^2 - 4\mathrm{n} + 3}{\mathrm{n}-1} \sqrt{\mathrm{n}(\mathrm{n}-2)} + 3(\mathrm{n}-1) \sin^{-1} \frac{1}{\mathrm{n}-1} \right] \qquad \text{(A5.8)}$$

Case a + b = 2 k + 1

Next we use 3.461(2) of Gradshteyn and Ryzhik (1994).

$$E\left(D_1^{\mathrm{a}} D_2^{\mathrm{b}}\right) = \sqrt{\frac{\mathrm{n}}{2\pi}} (2\mathrm{k}+1)!!(\mathrm{n}-2)^{\mathrm{k}+1} \int_0^{\pi/2} \left[\frac{1}{(\mathrm{n}-1+\sin 2\theta)^{\mathrm{k}+3/2}} + \frac{1}{(\mathrm{n}-1-\sin 2\theta)^{\mathrm{k}+3/2}} \right] \cos^{\mathrm{a}}\theta \sin^{2\mathrm{k}+1-\mathrm{a}} \theta \, \mathrm{d}\theta$$

(A5.9)

where (2k+1)!! = 1.3.5. ... .(2k+1).

Write (n - 1 + sin 2θ) as [(n - 1) ($\sin^2\theta + \cos^2\theta$) + 2 sin θ cos θ] and divide top and bottom inside the integral by $\cos^{2k+3}\theta$. Then change the variable of integration by $t = \tan\theta$ to obtain

$$E\left(D_1^{\mathrm{a}} D_2^{\mathrm{b}}\right) = \sqrt{\frac{\mathrm{n}}{2\pi}} (2\mathrm{k}+1)!!(\mathrm{n}-2)^{\mathrm{k}+1} \int_0^{\infty} \left[\frac{1}{\left((\mathrm{n}-1)t^2 + 2t + (n-1)\right)^{\mathrm{k}+3/2}} + \frac{1}{\left((\mathrm{n}-1)t^2 - 2t + (n-1)\right)^{\mathrm{k}+3/2}} \right] t^{2\mathrm{k}+1-\mathrm{a}} \, \mathrm{d}t$$

.

(A5.10)

Particular cases can be solved using integrals of rational functions. For example, using 3.252(7) of Gradshteyn and Ryzhik (1994) gives

$$E\left(D_1^2 D_2\right) = 3\sqrt{\frac{\mathrm{n}}{2\pi}} (\mathrm{n}-2)^2 \int_0^{\infty} \left[\frac{1}{\left((\mathrm{n}-1)t^2 + 2t + (n-1)\right)^{5/2}} + \frac{1}{\left((\mathrm{n}-1)t^2 - 2t + (n-1)\right)^{5/2}} \right] t \, \mathrm{d}t$$

$$= \sqrt{\frac{n}{2\pi}}(n-2)^2 \frac{1}{\sqrt{n-1}}\left[\frac{1}{\left[\left(\sqrt{n-1}\right)^2+1\right]^2}+\frac{1}{\left[\left(\sqrt{n-1}\right)^2-1\right]^2}\right]$$

$$= \sqrt{\frac{n}{2\pi}}(n-2)^2 \frac{1}{\sqrt{n-1}}\left[\frac{1}{n^2}+\frac{1}{[n-2]^2}\right]$$

$$= \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n(n-1)}}\frac{n^2-2n+2}{n} \tag{A5.11}$$

1. **Third-order moments**

The joint pdf of (D_1, D_2, D_3) is, from Eq (A1.4),

$$f(v,w,x) = \frac{2}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\left\{e^{-\frac{1}{2(n-3)}\left[(n-2)\left(v^2+w^2+x^2\right)+2vw+2vx+2wx\right]}+\ldots\right\},\ v\geq 0,\ w\geq 0,\ x\geq 0 \tag{A5.12}$$

where ... represents 3 similar exponential functions with each of v, w and x being negative in turn. We will evaluate $E(D_1 D_2 D_3)$ and $E(D_1^2 D_2 D_3)$.

$$E\left(D_1^a D_2 D_3\right) = \frac{2}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\int_0^\infty v^a\,dv\int_0^\infty w\,dw\int_0^\infty x\left\{e^{-\frac{1}{2(n-3)}\left[(n-2)\left(v^2+w^2+x^2\right)+2vw+2vx+2wx\right]}+\ldots\right\}dx\,. \tag{A5.13}$$

We change the variables to

$$v = r\cos\theta$$

$$w = r\sin\theta\cos\varphi$$

$$x = r\sin\theta\sin\varphi$$

where the Jacobian is now $r^2 \sin\theta$. The integral becomes

$$E\left(D_1^a D_2 D_3\right) = \frac{1}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\ \textit{multiplied by}$$

$$\int_0^{\infty} r^{a+4}\, dr \int_0^{\pi/2} \cos^a\theta \sin^3\theta\, d\theta \int_0^{\pi/2} \cos\varphi \sin\varphi \left\{ e^{-\frac{1}{2(n-3)} r^2 \left[(n-2)+2\sin\theta\cos\theta\left(\cos\varphi+\sin\varphi\right)+2\sin^2\theta\sin\varphi\cos\varphi\right]} + \ldots \right\} d\varphi$$

(A5.14)

Case 1 $E(D_1 D_2 D_3)$

Because of the symmetry of the original integral, the integral of the three functions not shown in the above equation are all identical, and so, after integrating out *r* using 3.461(3) of Gradshteyn and Ryzhik (1994), we obtain

$$E\left(D_1 D_2 D_3\right) =$$

$$\frac{2}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\int_0^{\pi/2} \cos\theta \sin^3\theta \, d\theta \int_0^{\pi/2} \cos\varphi \sin\varphi \left[\frac{2(n-3)}{(n-2)+2\sin\theta\cos\theta\left(\cos\varphi+\sin\varphi\right)+2\sin^2\theta\sin\varphi\cos\varphi}\right]^3 d\varphi$$

$$+ \frac{6}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\int_0^{\pi/2} \cos\theta \sin^3\theta \, dw \int_0^{\pi/2} \cos\varphi \sin\varphi \left[\frac{2(n-3)}{(n-2)-2\sin\theta\cos\theta\left(\cos\varphi+\sin\varphi\right)+2\sin^2\theta\sin\varphi\cos\varphi}\right]^3 d\varphi$$

(A5.15)

Now using 3.252(7) of Gradshteyn and Ryzhik (1994), we can show that

$$\int_0^{\infty} \frac{t^3}{\left(A + B t + C t^2\right)^3} dt = \frac{1}{C\Delta} + \frac{3B^2}{2C\Delta^2} - \frac{6AB}{\Delta^{5/2}}\left(\frac{\pi}{2} - \tan^{-1}\frac{B}{\sqrt{\Delta}}\right) \quad \text{(A5.16)}$$

where $\Delta = 4AC-B^2$.

Let A = (n - 2), B = 2(sin φ + cos φ) and C = (n - 2) + 2 sin φ cos φ, so that Δ = 4 (n - 3) (n - 1 + 2 sin φ cos φ) =

4 (n - 3) (C + 1). Then, the denominator inside each integral can be expressed as $\cos^6\theta \, [A \pm B\, t + C\, t^2]^3$ (depending on the integral), where $t = \tan\theta$. We therefore change the first variable of integration from θ to *t*. Furthermore, the two integrals differ only in the sign of B, and hence we can evaluate the expectation as

$$E\left(D_1 D_2 D_3\right) = \left(\frac{2}{\pi}\right)^{3/2} 2\left(n-3\right)^3\sqrt{\frac{n}{n-3}}\int_0^{\pi/2} \cos\varphi \sin\varphi \left\{\frac{4}{C\Delta} + \frac{6B^2}{C\Delta^2} + \frac{6AB}{\Delta^{5/2}}\left(\pi + 4\tan^{-1}\frac{B}{\sqrt{\Delta}}\right)\right\} d\varphi \ . \quad \text{(A5.17)}$$

There are now three different types of integral to evaluate.

Type 1

Firstly, note that $C = (n - 2) + \sin 2\varphi$ and $B^2 = 4(1 + \sin 2\varphi) = 4(C - (n - 3))$. Hence we can express the first two terms in braces as

$$\frac{4}{C\Delta} + \frac{6B^2}{C\Delta^2} \quad = \quad \frac{1}{2(n-3)}\left[-\frac{1}{C}+\frac{1}{C+1}+\frac{3(n-2)}{n-3}\frac{1}{(C+1)^2}\right].$$

Thus, the first type of integral reduces to

$$\left(\frac{2}{\pi}\right)^{3/2}(n-3)^2\sqrt{\frac{n}{n-3}}\int_0^{\pi/2}\sin 2\varphi\left\{-\frac{1}{C}+\frac{1}{C+1}+\frac{3(n-2)}{n-3}\frac{1}{(C+1)^2}\right\}d\varphi \ .$$

We change the variable of integration from φ to $y = 2\varphi$, and use the fact that the integrand is an even function about φ, obtaining

$$\left(\frac{2}{\pi}\right)^{3/2}(n-3)^2\frac{1}{2}\sqrt{\frac{n}{n-3}}\int_0^{\pi/2}\left[-\frac{\sin\varphi}{n-2+\sin\varphi}+\frac{\sin\varphi}{n-1+\sin\varphi}+\frac{3(n-2)}{n-3}\frac{\sin\varphi}{(n-1+\sin\varphi)^2}\right]d\varphi \ .$$

Using 2.551 and 2.552 of Gradshteyn and Ryzhik (1994) gives this integral, after some simplification, as

$$\left(\frac{2}{\pi}\right)^{3/2}(n-3)^2\sqrt{\frac{n}{n-3}}\left\{\frac{n-2}{\sqrt{(n-1)(n-3)}}\left[\tan^{-1}\sqrt{\frac{n-1}{n-3}}-\tan^{-1}\frac{1}{\sqrt{(n-1)(n-3)}}\right]+\frac{3}{2n(n-3)}\right.$$
$$\left.-\frac{n^3-4n^2+3n+3}{n(n-3)\sqrt{n(n-2)}}\left[\tan^{-1}\sqrt{\frac{n}{n-2}}-\tan^{-1}\frac{1}{\sqrt{n(n-2)}}\right]\right\} \qquad \text{(A5.17)}$$

Further simplification is possible on combining the pairs of $\tan^{-1}$ terms, using 1.625 of Gradshteyn and Ryzhik (1994). The final answer for this first type of integral is

$$\left(\frac{2}{\pi}\right)^{3/2}(n-3)^2\sqrt{\frac{n}{n-3}}\left[\frac{n-2}{\sqrt{(n-1)(n-3)}}\tan^{-1}\sqrt{\frac{n-3}{n-1}}+\frac{3}{2n(n-3)}-\frac{n^3-4n^2+3n+3}{n(n-3)\sqrt{n(n-2)}}\tan^{-1}\sqrt{\frac{n-2}{n}}\right]$$

Type 2

The second type of integral is solved by writing numerator and denominator in terms of tan2 and sec2, and then changing the variable of integration from φ to $t = \tan\varphi$. In the process we have to note that the integrand is an even function and separate its two equal parts. Thus

$$\left(\frac{2}{\pi}\right)^{3/2} 2(n-3)^3 \sqrt{\frac{n}{n-3}} \int_0^{\pi/2} \cos\varphi \sin\varphi \frac{6\pi AB}{\Delta^{5/2}} d\varphi = 3\sqrt{\frac{n}{2\pi}}(n-2) \int_0^{\infty} \frac{t(t+1)}{\left[(n-1)t^2+2t+(n-1)\right]^{5/2}} dt. \quad \text{(A5.18)}$$

The solution of this is obtained by immediate application of 3.252 (7 and 8) of Gradshteyn and Ryzhik (1994). The solution is $\sqrt{\frac{2n}{\pi(n-1)}}\frac{n-2}{n^2}$.

Type 3

The last integral is quite complex, and we proceed as before to reach

$$3\left(\frac{2}{\pi}\right)^{3/2} \sqrt{n}(n-2) \int_0^{\infty} \frac{t(t+1)}{\left[(n-1)t^2+2t+(n-1)\right]^{5/2}} \tan^{-1} \frac{t+1}{\sqrt{(n-3)\left[(n-1)t^2+2t+(n-1)\right]}} dt. \quad \text{(A5.19)}$$

Now $\frac{t+1}{\sqrt{(n-3)\left[(n-1)t^2+2t+(n-1)\right]}}$ is a symmetric, even function about the point t = 1, and so we can change the variable of integration to $u = \frac{t+1}{\sqrt{(n-3)\left[(n-1)t^2+2t+(n-1)\right]}}$ with the range of integration being from

$a = \sqrt{\frac{1}{(n-1)(n-3)}}$ to $b = \sqrt{\frac{2}{n(n-2)}}$. After simplification, the integral reduces to

$$\text{Type 3 integral} = 3\left(\frac{2}{\pi}\right)^{3/2} \sqrt{\frac{n-3}{n-2}} (n-1)(n-3) \int_a^b \frac{u\left(u^2-a^2\right)}{\sqrt{b^2-u^2}} \tan^{-1} u \, du. \quad \text{(A5.20)}$$

The indefinite integrals we need are, using Mathematica,

$$\int \frac{u}{\sqrt{b^2-u^2}} \tan^{-1} u \, du = -\sqrt{b^2-u^2} \tan^{-1} u - \tan^{-1} \frac{u}{\sqrt{b^2-u^2}} + \sqrt{1+b^2} \tan^{-1} \sqrt{\frac{1+b^2}{b^2-u^2}} u$$

and

$$\int \frac{u^3}{\sqrt{b^2 - u^2}} \tan^{-1} u \, du = \frac{u}{6}\sqrt{b^2 - u^2} - \frac{2b^2 + u^2}{3}\sqrt{b^2 - u^2} \tan^{-1} u + \frac{2 - 3b^2}{6} \tan^{-1} \frac{u}{\sqrt{b^2 - u^2}} + \frac{2b^4 + b^2 - 1}{3\sqrt{1 + b^2}} \tan^{-1} \sqrt{\frac{1 + b^2}{b^2 - u^2}} u$$

Applying these to the Type 3 integral produces, after simplification, Type 3 integral =

$$\left(\frac{2}{\pi}\right)^{3/2}\sqrt{\frac{n-3}{n-2}}\,\frac{1}{n}\left\{\left(n^3-4n^2+3n+3\right)\tan^{-1}\sqrt{\frac{n-2}{n}}-\left(n^3-4n^2+2n+4\right)\sqrt{\frac{(n-1)(n-2)}{n(n-3)}}\tan^{-1}\sqrt{\frac{n-3}{n-1}}\right..$$

$$\left.-\frac{\sqrt{n(n-2)}}{2}+2\sqrt{\frac{(n-2)^3}{n(n-1)(n-3)}}\tan^{-1}\sqrt{\frac{1}{(n-1)(n-3)}}\right\} \qquad \text{(A5.21)}$$

The final expectation is the sum of these three solutions. We note the following:

- The terms involving $\tan^{-1}\sqrt{\frac{n-2}{n}}$ cancel out.

- The terms involving $\tan^{-1}\sqrt{\frac{n-3}{n-1}}$ add to

$$(n-2)(n-3)\sqrt{\frac{n}{n-1}}-\frac{n^3-4n^2+2n+4}{n}\sqrt{\frac{n-1}{n}}=-\frac{2(n-2)}{n\sqrt{n(n-1)}} \text{ on simplification}$$

- The Type 2 solution can therefore be incorporated into the above term. Note that

$$\frac{\pi}{2}\frac{n-2}{n\sqrt{n(n-1)}}-\frac{2(n-2)}{n\sqrt{n(n-1)}}\tan^{-1}\sqrt{\frac{n-3}{n-1}} = \frac{n-2}{n\sqrt{n(n-1)}}\left[\frac{\pi}{2}-2\tan^{-1}\sqrt{\frac{n-3}{n-1}}\right]$$

$$= \frac{n-2}{n\sqrt{n(n-1)}}\left[\frac{\pi}{2}-\tan^{-1}\sqrt{(n-1)(n-3)}\right]$$ using 1.662 of Gradshteyn and

Ryzhik (1994)

$$= \frac{n-2}{n\sqrt{n(n-1)}}\tan^{-1}\frac{1}{\sqrt{(n-1)(n-3)}}.$$

- the terms involving $\tan^{-1}\sqrt{\frac{1}{(n-1)(n-3)}}$ now come together as

$$\frac{3(n-2)}{n\sqrt{n(n-1)}}\tan^{-1}\frac{1}{\sqrt{(n-1)(n-3)}}=\frac{3(n-2)}{n\sqrt{n(n-1)}}\sin^{-1}\frac{1}{n-2}.$$

Thus, the final expectation is

$$E(D_1D_2D_3)=\left(\frac{2}{\pi}\right)^{3/2}\left[\frac{n-2}{n}-\frac{3(n-2)}{n\sqrt{n(n-1)}}\sin^{-1}\frac{1}{n-2}\right]. \qquad \text{(A5.22)}$$

Case 2 $E\left(D_1^2\, D_2\, D_3\right)$

Here we will only indicate the steps we found necessary to evaluate this expectation.

$$E\left(D_1^2 D_2 D_3\right) =$$

$$\frac{1}{(2\pi)^{3/2}}\sqrt{\frac{n}{n-3}}\int_0^\infty r^6 dr \int_0^{\pi/2} \cos^2\theta\, \sin^3\theta\, dw \int_0^{\pi/2} \cos\varphi\, \sin\varphi \left\{ e^{-\frac{1}{2(n-3)}r^2\left[(n-2)+2\sin\theta\cos\theta\left(\cos\varphi+\sin\varphi\right)+2\sin^2\theta\,\sin\varphi\cos\varphi\right]} + \ldots \right\} d\varphi$$

(A5.23)

Integrating out *r* using 3.461(2) of Gradshteyn and Ryzhik (1994) leads to

$$E\left(D_1^2 D_2 D_3\right) = \frac{15\sqrt{n}(n-3)^3}{2\pi}\int_0^{\pi/2}\int_0^{\pi/2} \frac{\sin^3\theta\, \cos^2\theta\, \cos\varphi\, \sin\varphi}{\left[(n-2)+2\sin\theta\cos\theta\left(\sin\varphi+\cos\varphi\right)+2\sin^2\theta\,\sin\varphi\,\cos\varphi\right]^{7/2}}\, d\varphi\, d\theta$$

+ three similar integrals that differ only in the signs in the denominator.

(A5.24)

Next, we change the first variable of integration to $t = \tan(\theta)$ and use 3.252(7) of Gradshteyn and Ryzhik (1994) to obtain

$$E\left(D_1^2 D_2 D_3\right) = \frac{\sqrt{n}(n-3)^3}{\pi}\int_0^{\pi/2} \frac{\cos\varphi\, \sin\varphi}{\sqrt{(n-2)+2\sin\varphi\cos\varphi}\left[\sqrt{n-2}\sqrt{(n-2)+2\sin\varphi\cos\varphi}+\sin\varphi+\cos\varphi\right]^3}\, d\varphi$$

+ three similar integrals that differ only in the signs in the two denominators.

(A5.25)

We need to remove the square root inside the denominator, and this is achieved for the first of the four integrals by a change of variable as follows:

$$(n-2)\tan\varphi + 1 \quad = \quad \sqrt{(n-2)(n-2)}\, \tan y\,. \qquad \text{(A5.27)}$$

The remaining integrals work the same way, but care must be taken with the signs. The range of integration changes from $\left(0, \frac{\pi}{2}\right)$ to $\left(\pm\tan^{-1}\frac{1}{\sqrt{(n-1)(n-3)}}, \frac{\pi}{2}\right)$, depending on the signs in each integral. We will let $p = \frac{1}{n-2}$ and denote $q = \frac{p}{\sqrt{1-p^2}} = \frac{1}{\sqrt{(n-1)(n-3)}}$ so that the lower range of

integration is $\pm \tan^{-1} q$. We have therefore re-arranged the first of the four integrals as

$$E\left(D_1^2 D_2 D_3\right) = \frac{\sqrt{n}(n-3)^2}{2\pi(n-1)(n-2)^{3/2}} \int_{\tan^{-1} q}^{\pi/2} \frac{(\sin y - q \cos y) \cos y}{\left[1 + p \sin y + (1-p) q \cos y\right]^3} dy + \ldots \quad \text{(A5.28)}$$

We now combine the four integrals and let $x = \tan y$ and $q_1 = \sqrt{\frac{1-p}{1+p}} = \sqrt{\frac{n-3}{n-1}}$. We obtain

$$E\left(D_1^2 D_2 D_3\right) = \frac{2\sqrt{n}(n-3)^2}{\pi(n-1)(n-2)^{3/2}} \left\{ \int_{-q}^{\infty} \frac{\left[\left(1+3p^2\right)x^2 - 6p^2 q_1 x + \left(1+3p^2 q_1^2\right)\right](x+q)}{\left[\left(1-p^2\right)x^2 + 2p^2 q_1 x + \left(1-p^2 q_1^2\right)\right]^3} dx \right.$$

$$\left. + \int_{q}^{\infty} \frac{\left[\left(1+3p^2\right)x^2 + 6p^2 q_1 x + \left(1+3p^2 q_1^2\right)\right](x-q)}{\left[\left(1-p^2\right)x^2 - 2p^2 q_1 x + \left(1-p^2 q_1^2\right)\right]^3} dx \right\} \quad \text{(A5.29)}$$

Each of these integrals is a cubic polynomial in x, divided by the cube of a quadratic trinomial in x. Indefinite integrals for each element exist - see 2.175 of Gradshteyn and Ryzhik (1994). After tedious manipulation, we reach the final form for the expectation,

$$E\left(D_1^2 D_2 D_3\right) = \frac{2}{\pi n^2} \left[\frac{n^2 - 2n + 3}{n-1} \sqrt{n(n-2)} + (n-3) \sin^{-1} \frac{1}{n-1} \right] . \quad \text{(A5.30)}$$