

The Effects of Correlated Data and Correction Procedures for F-Test in Unbalanced Two Way Model

Ivan S. Kababchi
College of Education
University of Salahaddin

Received on: 20/11/2005

Accepted on: 24/7/2005

الملخص

أن استقلالية المشاهدات تعتبر أحد الفروض الأساسية في جدول تحليل التباين. حيث تتضمن كون حدود الأخطاء في النموذج متغيرات مستقلة ويكون توزيعها متماثل وتتبع التوزيع الطبيعي بمعدلات صفرية وتباينات متجانسة. في هذه الدراسة تم توضيح تأثير عدم استقلالية البيانات في تحليل التباين لنموذج ثابت ذي اتجاهين غير متوازن متشعب كما تم تطوير طريقة لتعديل التباين لهذه المشاهدات المرتبطة. هذا يعني أن حدود الأخطاء تكون مترابطة وسنسلط الضوء على تأثير إزالة شرط الاستقلالية على اختبار الفرضيات وذلك بواسطة حساب توقع معدل مجموع المربعات للأخطاء، وكذلك للمعالجات في هذا النموذج وتصحيح الإحصاء F لاختبار تأثيرات العامل. لهذا نفترض نموذجاً جميع القياسات فيه تمتلك نفس التباين σ^2 ومصفوفة التباين المشترك تتمتع ببنية معرفة كما يلي: كل زوج من القياسات التي تأتي من:

(1) تكرارات مختلفة ولكن من نفس الوحدة. (2) وحدات جزئية مختلفة بنفس الوحدة. (3) وحدات مختلفة لها تباين $s^2 r_1, s^2 r_2, s^2 r_3$ على التوالي.

ABSTRACT

Independence of observations is one of the standard assumption in analysis of Variance (ANOVA) table. Where the error terms in the model are independent, identically distributed normal variables with null means and homogeneous variances. In this paper investigate the effect of dependence of observations in ANOVA for unbalanced 2-way nested fixed model and developing a method for adjusting it. When the error terms are correlated and focus on the effects of departures from independence assumptions on hypothesis testing by determining the expect mean squares for errors as well as treatments for this model and correcting the F statistics for testing the factor effect. The model considered is one in which all measurements have same variance σ^2 , and the covariance matrix enjoy a structure defined as follows: every pair of measurements comes from:

- i) The same experimental observation and the same experimental unit;
 - ii) Different experimental observation, but in the same experimental unit;
 - ii) Different experimental unit;
- has covariance $s^2 r_1, s^2 r_2$ and $s^2 r_3$ respectively.

Introduction

The assumption of independence of observation in ANOVA table may seem like a reasonable assumption in examining data using experimental designs. Rarely is the

independent assumption verified in an ANOVA consequently, the analysis of data from experimental design is often hampered by lack of technique to correct the usual F-test for the effect of correlation. Some researches have been devoted to show that slight dependence in the form of serial correlation and interclass correlation can quickly inflate the Type I error rate. See [4], [8], [1] and [3] have shown that in an experimental design, certain forms of dependence can quickly invalidate the result of ANOVA. How small experiment is shown see [6] and see [5] for 2-way crossed and balanced nested classifications respectively.

The aim of this work explains a method for adjusting ANOVA table when observation are correlated, that is, when the error terms are correlated and focus on the effects of departures from independence assumption on hypothesis by determining the expected mean squares for errors as well as treatments for un balanced two way nested fixed effects model (balanced model is special case of un balanced model) and correcting the F statistics for testing the factor effect.

1- Defining the model

The model of study occurs in ANOVA, when we have a number of independent experimental units and each experimental unit has the same number of experimental observation and each experimental observation receives different number of treatment. See [2], [10], [11].

Let Y_{ijk} be the observation of the k th treatment of the j th experimental observation from the i th experimental unit with $i = 1, 2, \dots, m$, $j = 1, 2, \dots, p$, $k = 1, 2, \dots, n_{ij}$

Let

$$m_{ijk} = E(Y_{ijk}), \quad e_{ijk} = Y_{ijk} - m_{ijk} \quad \dots(1)$$

Where e_{ijk} is the error term of the measure Y_{ijk} . We assume that the Y_{ijk} distributed normally with mean m_{ijk} and that the measurements have the same variance s^2 , and every pair of measurements comes from

- i) The same experimental observation and the same experimental unit;
- ii) Different experimental observation, but in the same experimental unit;
- ii) Different experimental unit;

has covariance $s^2 r_1, s^2 r_2$ and $s^2 r_3$ respectively. Symbolically

$$\text{cov}(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} s^2 & \text{if } i=i', j=j', k=k' \\ s^2 r_1 & \text{if } i=i', j=j', k \neq k' \\ s^2 r_2 & \text{if } i=i', j \neq j' \\ s^2 r_3 & \text{if } i \neq i' \end{cases} \quad \dots(2)$$

In this study, we assume that

$$m_{ijk} = q + \alpha_i + g_{ij} \quad \dots(3)$$

Where $q \in R$ is the overall mean, α_i is the effect of the i th experimental unit and g_{ij} is the effect of the i th nested experimental observation in the j th experimental unit; such that

$$\sum_i \alpha_i = 0, \quad \sum_j g_{ij} = 0 \quad \forall i \quad \dots(4)$$

Thus from equations (1) and (3) we have

$$Y_{ijk} = q + a_i + g_{ij} + e_{ijk} \quad \dots(5)$$

The model we consider in this part, which is defined by equations (2), (5), is unbalanced 2-way nested effects model. Then from equation (1) the e_{ijk} are distributed normally with mean zero and

$$COV(e_{ijk}, e_{i'j'k'}) = COV(Y_{ijk}, Y_{i'j'k'}) \quad \dots(6)$$

First we discuss ANOVA, concerns with the F statistic for the equality of factor level means, and we will discuss correction for correlation.

2- Analysis of Variance

In this section see [2], [9] and [10] we will partition the total sum of square and degrees of freedom from the model given in equation (5), in the same way that we partition the total sum of square and degrees of freedom in the ANOVA for unbalanced 2-way nested effects model when the observation are independent .Where we partition the total sum of square in the ANOVA table for unbalanced 2-way nested effects model to sum of square of factor A, sum of square of factor B and sum of square of error when the observation are dependent as follows

$$SSTO = SSA + SSB + SSe \quad \dots(7)$$

$$SSTO = \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{...})^2 \quad \dots(8a)$$

$$SSA = \sum_{i=1}^m \sum_{j=1}^p n_{ij} (\bar{Y}_{i..} - \bar{Y}_{...})^2 \quad \dots(8b)$$

$$SSB = \sum_{i=1}^m \sum_{j=1}^p n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \quad \dots(8c)$$

$$SSe = \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (\bar{Y}_{ijk} - \bar{Y}_{i..})^2 \quad \dots(8d)$$

The degrees of freedom that corresponds to each of the sums SSA , SSB , SSe , and $SSTo$ are respectively $m-1$, $m(p-1)$, $mp(n_{ij}-1)$, $mpN-1$ where $N = n_{11} + n_{12} + n_{13} + \dots + n_{mp}$. Thus, we will determine the expected mean squares $E(MS)$. When the observation are not independent by using equation (2) each of the $MSSA$, $MSSB$ and $MSSe$.

2.1 Determining expected mean squares sum of square of error $E(MSe)$

Here, we use equation (8d) and the corresponding degrees of freedom for SSe

$$MSe = \frac{SSe}{mp(N-1)} \quad \dots(9)$$

Now using equation (5) we get

$$\bar{Y}_{ij.} = \frac{1}{N} \sum_{j=1}^{n_{ij}} (q + a_i + g_{ij} + e_{ijk}) = q + a_i + g_{ij} + \bar{e}_{ij.} \quad \dots(10)$$

Where $\bar{Y}_{ij.}$ represent the mean of the j th experimental observation; from the i th experimental unit; subtracting equation (10) from equation (5) we get

$$Y_{ijk} - \bar{Y}_{ij.} = e_{ijk} - \bar{e}_{ij.} \quad \dots(11)$$

By using equation (8d), (9) and (11) we get

$$\begin{aligned}
 MSe &= \frac{1}{mp(N-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 \\
 &= \frac{1}{mp(N-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (e_{ijk} - \bar{e}_{ij.})^2 \quad \dots(12) \\
 &= \frac{1}{mp(N-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (e_{ijk}^2 + \bar{e}_{ij.}^2 - 2e_{ijk}\bar{e}_{ij.})
 \end{aligned}$$

Now by taking mathematical expectation for (12) we get

$$E(MSe) = \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (A_1 + A_2 - 2A_3) \quad \dots(13)$$

Where

$$A_1 = E(e_{ijk}^2), \quad A_2 = E(\bar{e}_{ij.}^2), \quad A_3 = E(e_{ijk}\bar{e}_{ij.}) \quad \dots(14)$$

Now using the two equations (1), (2) and (6), we can determine A_h for $h=1,2,3$ as follows

$$A_1 = E(e_{ij}^2) = s^2 \quad \dots(15a)$$

And

$$\begin{aligned}
 A_3 &= E(e_{ij}\bar{e}_{i.}) = \frac{1}{n_{ij}} E(e_{ij}e_{i.}) \\
 &= \frac{1}{n_{ij}} E(e_{ij}(e_{i1} + e_{i2} + \dots + e_{in_i})) \\
 &= \frac{1}{n_{ij}} E(e_{ij}^2) + \sum_{\substack{j=1 \\ j \neq j'}}^{n_i} E(e_{ij}e_{ij'}) \\
 &= \frac{1}{n_{ij}} s^2 + (n_{ij} - 1)s^2 r_1 \\
 &= \frac{s^2}{n_{ij}} (1 + n_{ij}r_1 - r_1) \\
 &= \frac{s^2}{n_{ij}} (1 - r_1 + n_{ij}r_1) \quad \dots(15b)
 \end{aligned}$$

And

$$\begin{aligned}
 A_2 &= E(\bar{e}_{i.}^2) \\
 &= \frac{1}{n_i^2} E(e_{i.}^2) \\
 &= \frac{1}{n_i^2} E(e_{i1} + e_{i2} + \dots + e_{in_i})^2
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n_{ij}^2} \left[E\left(\sum_{j=1}^{n_i} e_{ij}^2\right) + E\left(\sum_{j=1}^{n_i} \sum_{\substack{j'=1 \\ j \neq j'}}^{n_i} e_{ij} e_{ij'}\right) \right] \\
 &= \frac{1}{n_{ij}^2} \left[\sum_{j=1}^{n_i} E(e_{ij}^2) + \sum_{j=1}^{n_i} \sum_{\substack{j'=1 \\ j \neq j'}}^{n_i} E(e_{ij} e_{ij'}) \right] \\
 &= \frac{1}{n_{ij}^2} [n_i s^2 + n_{ij}(n_{ij}-1)s^2 r_1] \\
 &= \frac{s^2}{n_{ij}} (1 - r_1 + n_{ij} r_1) \quad \dots(15c)
 \end{aligned}$$

By substituting equation (15a), (15b) and (15c) in equation (13) we get

$$\begin{aligned}
 E(MSe) &= \frac{1}{p(N-1)} \sum_{i=1}^p \sum_{j=1}^{n_{ij}} (A_1 - 2A_2 + A_3) \\
 &= \frac{1}{p(N-1)} \sum_{i=1}^p \sum_{j=1}^{n_{ij}} \left[s^2 - 2 \left(\frac{s^2}{n_{ij}} (1 - r_1 + n_{ij} r_1) \right) + \frac{s^2}{n_{ij}} (1 - r_1 + n_{ij} r_1) \right] \\
 &= \frac{1}{p(N-1)} pN \left(s^2 - \frac{s^2}{N} (1 - r_1 + N r_1) \right) \\
 &= \frac{N}{(N-1)} s^2 \left(1 - \frac{1}{N} (1 - r_1 + N r_1) \right) \\
 &= \frac{s^2}{(N-1)} (N-1)(1 - r_1) \quad \dots(16) \\
 &= (1 - r_1) s^2
 \end{aligned}$$

2.2 Determining expected mean squares sum of square of factor B E(MSB)

Using equation (8c) and the corresponding degrees of freedom for SSB

$$MSB = \frac{SSB}{m(p-1)} \quad \dots(17)$$

Now using equation (5) we get

$$\begin{aligned}
 \bar{Y}_{i..} &= \frac{1}{pN} \sum_{j=1}^p \sum_{k=1}^{n_{ij}} Y_{ijk} \\
 &= \frac{1}{pN} \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (q + a_i + g_{ij} + e_{ijk}) \quad \dots(18) \\
 &= q + a_i + \bar{e}_{i..}
 \end{aligned}$$

Subtracting equation (18) from (10) we get

$$\begin{aligned}
 \bar{Y}_{ij.} - \bar{Y}_{i..} &= q + a_i + g_{ij} + \bar{e}_{ij.} - q - a_i - \bar{e}_{i..} \\
 &= g_{ij} + \bar{e}_{ij.} - \bar{e}_{i..} \quad \dots(19)
 \end{aligned}$$

By using equation (8c), (17) and (19) then

$$\begin{aligned}
 MSB &= \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \\
 &= \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (g_{ij} + \bar{e}_{ij.} - \bar{e}_{i..})^2 \\
 &= \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (g_{ij}^2 + \bar{e}_{ij.}^2 + \bar{e}_{i..}^2 - 2(\bar{e}_{ij.} \bar{e}_{i..}) - 2(\bar{e}_{ij.} - \bar{e}_{i..})) \quad \dots(20)
 \end{aligned}$$

Now by taking the mathematical expectation for equation (20) we get

$$\begin{aligned}
 E(MSB) &= \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p E(n_{ij} (g_{ij}^2 + \bar{e}_{ij.}^2 + \bar{e}_{i..}^2 - 2(\bar{e}_{ij.} \bar{e}_{i..}) - 2g_{ij}(\bar{e}_{ij.} - \bar{e}_{i..}))) \\
 &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} g_{ij}}{m(p-1)} + \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (B_1 + B_2 - 2B_3 - 2B_4) \quad \dots(21)
 \end{aligned}$$

Where

$$B_1 = E(\bar{e}_{ij.}^2), B_2 = E(\bar{e}_{i..}^2), B_3 = E(\bar{e}_{ij.} \bar{e}_{i..}), B_4 = E(\bar{e}_{ij.} - \bar{e}_{i..}) \quad \dots(22)$$

Using equation (2) and (6) we can determine B_h for $h=1,2,3,4$ as follows

$$B_4 = E(\bar{e}_{ij.} - \bar{e}_{i..}) = E(\bar{e}_{ij.}) - E(\bar{e}_{i..}) = 0 \quad \dots(23a)$$

From equation (15c) we note that $B_1 = A_2$

$$B_1 = \frac{s^2}{n_{ij}} (1 - r_1 + n_{ji} r_1) \quad \dots(23b)$$

And

$$\begin{aligned}
 B_2 &= E(\bar{e}_{i..}^2) = \frac{1}{p^2 n_{ij}^2} E(e_{i..}^2) \\
 &= \frac{1}{p^2 n_{ij}^2} E(e_{i11} + e_{i12} + \dots + e_{i1n_{i1}} + e_{i11} + e_{i21} + e_{i22} + \dots \\
 &\quad + e_{i2n_{i2}} + e_{ij1} + e_{ij2} + \dots + e_{ijn_{ij}} + \dots + e_{ip1} + e_{ip2} + \dots + e_{ipn_{ip}})^2 \\
 &= \frac{1}{p^2 n_{ij}^2} E \left(\sum_{j=1}^p \sum_{k=1}^{n_{ij}} e_{ijk}^2 + \sum_{k=1}^p \sum_{\substack{k,k'=1 \\ k \neq k'}}^{n_{ij}} e_{ijk} e_{ij'k'} + \sum_{\substack{j \neq j' \\ j=1}}^p \sum_{k,k'}^{n_{ij}} e_{ijk} e_{ij'k'} \right) \\
 &= \frac{1}{p^2 n_{ij}^2} \left(\sum_{j=1}^p \sum_{k=1}^{n_{ij}} E(e_{ijk}^2) + \sum_{k=1}^p \sum_{\substack{k,k'=1 \\ k \neq k'}}^{n_{ij}} E(e_{ijk} e_{ij'k'}) + \sum_{\substack{j \neq j' \\ j=1}}^p \sum_{k,k'}^{n_{ij}} E(e_{ijk} e_{ij'k'}) \right) \\
 &= \frac{1}{p^2 n_{ij}^2} (pn_{ij} \sigma^2 + pn_{ij}(n_{ij} - 1) \sigma^2 \rho_1 + n_{ij}^2 (p-1) \rho_2) \\
 &= \frac{\sigma^2}{pn_{ij}} (1 + (n_i - 1) \rho_1 + (p-1) n_{ij} \rho_2)
 \end{aligned}$$

$$= \frac{\sigma^2}{pn_{ij}} (1 - \rho_1 + n_{ij}(\rho_1 - \rho_2) + n_{ij}p\rho_2) \quad \dots(23c)$$

And

$$\begin{aligned} B_3 &= E(\bar{e}_{ij} \bar{e}_{i...}) = \frac{1}{pn_{ij}^2} E(e_{ij} e_{i...}) \\ &= \frac{1}{pn_{ij}^2} E((e_{ij1} + e_{ij2} + \dots + e_{ijn_{ij}})(e_{i11} + e_{i12} + \dots + e_{i1n_{i1}} e_{i21} + e_{i22} + \\ &\quad \dots + e_{i2n_{i2}} + e_{i1j1} + e_{ij2} + \dots + e_{ijn_{ij}} + e_{ip1} + e_{ip2} + \dots + e_{ipn_{ip}})) \\ &= \frac{1}{pn_{ij}^2} E \left[\sum_{k=1}^{n_{ij}} e_{ijk}^2 + \sum_{\substack{k,k'=1 \\ k \neq k'}}^{n_{ij}} e_{ijk} e_{ijk'} + \sum_{\substack{j=1 \\ j \neq j'}}^p \sum_{ik=1}^{n_{ij}} \sum_{k'=1}^{n_{ij}} e_{ijk} e_{ij'k'} \right] \\ &= \frac{1}{pn_{ij}^2} \left[\sum_{k=1}^{n_{ij}} E(e_{ijk}^2) + \sum_{\substack{k,k'=1 \\ k \neq k'}}^{n_{ij}} (e_{ijk} e_{ijk'}) + \sum_{\substack{j=1 \\ j \neq j'}}^p \sum_{ik=1}^{n_{ij}} \sum_{k'=1}^{n_{ij}} (e_{ijk} e_{ij'k'}) \right] \\ &= \frac{1}{pn_{ij}^2} [n_{ij}\sigma^2 + n_{ij}(n_{ij} - 1)\rho_1\sigma^2 + n_{ij}^2(p - 1)\rho_2\sigma^2] \\ &= \frac{\sigma^2}{pn_{ij}} [1 + (n_{ij} - 1)\rho_1 + n_{ij}(p - 1)\rho_2] \\ &= \frac{\sigma^2}{pn_{ij}} [1 - \rho_1 + n_{ij}(\rho_1 - \rho_2) + pn_{ij}\rho_2] \quad \dots(23d) \end{aligned}$$

Substituting equation (23a), (23b), (23c) And (23d) in equation (21) we get

$$\begin{aligned} E(MSB) &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (B_1 + B_2 - 2B_3 - 2B_4) \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{\sigma^2}{n_{ij}} (1 - \rho_1 + n_{ij}\rho_1) + \frac{\sigma^2}{pn_{ij}} (1 - \rho_1) + n_{ij}(\rho_1 - \rho_2) + pn_{ij}\rho_2 + \frac{\sigma^2}{pn_{ij}} (1 - \rho_1 + n_{ij}(\rho_1 - \rho_2) + pn_{ij}\rho_2) \right] \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \frac{\sigma^2}{m(p-1)} \left[mp(1 - \rho_1) + \sum_{i=1}^m \sum_{j=1}^p n_{ij}\rho_1 - m(1 - \rho_1) - \sum_{i=1}^m \sum_{j=1}^p \frac{n_{ij}}{p} (\rho_1 - \rho_2) - \sum_{i=1}^m \sum_{j=1}^p n_{ij}\rho_2 \right] \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \frac{\sigma^2}{m(p-1)} \left[mp(p-1)(1 - \rho_1) + \sum_{i=1}^m \sum_{j=1}^p n_{ij}(\rho_1 - \rho_2) - \sum_{i=1}^m \sum_{j=1}^p \frac{n_{ij}}{p} (\rho_1 - \rho_2) \right] \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \frac{\sigma^2}{m(p-1)} \left[mp(p-1)(1 - \rho_1) + (1 - \frac{1}{p}) \sum_{i=1}^m \sum_{j=1}^p n_{ij}(\rho_1 - \rho_2) \right] \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} \gamma_{ij}}{m(p-1)} + \sigma^2 \left[(1 - \rho_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij}(\rho_1 - \rho_2) \right] \quad (24) \end{aligned}$$

2.2 Determining expected mean squares sum of square of factor A E(MSA)

Using equation (8d) and the corresponding degrees of freedom for SSA

$$MSA = \frac{SSA}{(m-1)} \quad \dots(25)$$

Now using equation (5) we get

$$\begin{aligned} \bar{Y}_{...} &= \frac{1}{mpN} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} Y_{ijk} \\ &= \frac{1}{pN} \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (\theta + \alpha_i + \gamma_{ij} + e_{ijk}) \\ &= \theta + \bar{e}_{...} \end{aligned} \quad \dots(26)$$

Subtracting equation (26) from (10) we get

$$\begin{aligned} \bar{Y}_{i..} - \bar{Y}_{...} &= q + a_i + \bar{e}_{i..} - q - \bar{e}_{...} \\ &= a_i + \bar{e}_{i..} - \bar{e}_{...} \end{aligned} \quad \dots(27)$$

By using equation (8),(25)and (27) then

$$\begin{aligned} MSA &= \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (\bar{Y}_{i..} - \bar{Y}_{...})^2 \\ &= \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (a_i + \bar{e}_{ij.} - \bar{e}_{i..})^2 \\ &= \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (a_i^2 + \bar{e}_{i..}^2 + \bar{e}_{...}^2 - 2(\bar{e}_{i..} \bar{e}_{...}) - 2a_i(\bar{e}_{i..} - \bar{e}_{...})) \end{aligned} \quad \dots(28)$$

Now by taking the mathematical expectation for equation (20) we get

$$\begin{aligned} E(MSA) &= \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} E(a_i^2 + \bar{e}_{i..}^2 + \bar{e}_{...}^2 - 2(\bar{e}_{i..} \bar{e}_{...}) - 2a_i(\bar{e}_{i..} - \bar{e}_{...})) \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} E(\bar{e}_{i..}^2) + E(\bar{e}_{...}^2) - 2E(\bar{e}_{i..} \bar{e}_{...}) - 2a_i E(\bar{e}_{i..} - \bar{e}_{...}) \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p \sum_{k=1}^{n_{ij}} (C_1 + C_2 - 2C_3 - 2a_i C_4) \end{aligned} \quad \dots(29)$$

Where

$$C_1 = E(\bar{e}_{i..}^2), C_2 = E(\bar{e}_{...}^2), C_3 = E(\bar{e}_{i..} \bar{e}_{...}), C_4 = E(\bar{e}_{i..} - \bar{e}_{...}) \quad \dots(30)$$

Using equation (2) and (6)we can determine B_h for $h=1,2,3,4$ as follows

$$C_4 = E(\bar{e}_{i..} - \bar{e}_{...}) = E(\bar{e}_{i..}) - E(\bar{e}_{...}) = 0$$

...(31a)

From equation (15c) we note that $C_1 = B_2$

$$C_1 = \frac{s^2}{pn_{ij}} [(1 - r_1) + n_{ji}(r_1 - r_2) + pn_{ji}r_2] \quad \dots(31b)$$

And

$$\begin{aligned}
 C_2 &= E(\bar{e}_{...}^2) = \frac{1}{m^2 p^2 n_{ij}^2} E(e_{...}^2) \\
 &= \frac{1}{p^2 n_{ij}^2} E(e_{111} + e_{112} + \dots + e_{11n_{11}} + \dots + e_{ij1} + e_{ij2} + \dots + e_{ijn_{ij}} + \dots + e_{1p1} + e_{1p2} + \dots + e_{1pn_{1p}} \\
 &\quad + \dots + e_{ip1} + e_{ip2} + \dots + e_{ipn_{ip}} + \dots + e_{mp1} + e_{mp2} + \dots + e_{mpn_{mp}})^2 \\
 &= \frac{1}{m^2 p^2 n_{ij}^2} E\left[\sum_{i,j,k=1}^{m,p,n_{ij}} e_{i..}^2 + \sum_{i,j=1}^m \sum_{k=1}^p \sum_{k' \neq k}^{n_{ij}} e_{ijk} e_{ijk'} + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} e_{ijk} e_{ij'k'} + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} e_{ijk} e_{ij'k'}\right] \\
 &= \frac{1}{m^2 p^2 n_{ij}^2} \left[\sum_{i,j,k=1}^{m,p,n_{ij}} E(e_{i..}^2) + \sum_{i,j=1}^m \sum_{k=1}^p \sum_{k' \neq k}^{n_{ij}} E(e_{ijk} e_{ijk'}) + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} E(e_{ijk} e_{ij'k'}) + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} E(e_{ijk} e_{ij'k'}) \right] \\
 &= \frac{1}{m^2 p^2 n_{ij}^2} (mpn_{ij}s^2 + mpn_{ij}(n_{ij}-1)s^2 r_1 + mp(p-1)n_{ij}^2 r_2 + m(m-1)p^2 n_{ij}^2 r_3) \\
 &= \frac{s^2}{mpn_{ij}} (1 + n_{ij}(r_1 - r_2) + n_{ij}(r_2 - r_3) + mpn_{ij}r_3 - r_1) \quad \dots(31c)
 \end{aligned}$$

And

$$\begin{aligned}
 C_3 &= E(\bar{e}_{i..} \bar{e}_{...}) = \frac{1}{mp^2 n_{ij}} E(e_{i..} e_{...}) \\
 &= \frac{1}{mp^2 n_{ij}} E((e_{i11} + e_{i12} + \dots + e_{i1n_{11}} + \dots + e_{ij1} + e_{ij2} + \dots + e_{ijn_{ij}} + \dots + e_{ip1} + e_{ip2} + \dots + e_{ipn_{ip}}) \\
 &\quad (e_{111} + e_{112} + \dots + e_{11n_{11}} + \dots + e_{ij1} + e_{ij2} + \dots + e_{ijn_{ij}} + \dots + e_{1p1} + e_{1p2} + \dots + e_{1pn_{1p}} + \dots + e_{ij1} + \\
 &\quad e_{ij2} + \dots + e_{ijn_{ij}} + \dots + e_{mp1} + e_{mp2} + \dots + e_{mpn_{mp}})) \\
 C_3 &= \frac{1}{mp^2 n_{ij}^2} E\left[\sum_{j,k=1}^{p,n_{ij}} e_{ijk}^2 + \sum_{j=1}^p \sum_{k,k'=1}^{n_{ij}} e_{ijk} e_{ij'k'} + \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} e_{ijk} e_{ij'k'} + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} e_{ijk} e_{ij'k'}\right] \\
 &= \frac{1}{mp^2 n_{ij}^2} \left[\sum_{j,k=1}^{p,n_{ij}} E(e_{ijk}^2) + \sum_{j=1}^p \sum_{k,k'=1}^{n_{ij}} E(e_{ijk} e_{ij'k'}) + \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} E(e_{ijk} e_{ij'k'}) + \sum_{i=1}^m \sum_{j,j' \neq j}^p \sum_{k,k'=1}^{n_{ij}} E(e_{ijk} e_{ij'k'}) \right] \\
 &= \frac{1}{mp^2 n_{ij}^2} [pn_{ij}s^2 + pn_{ij}(n_{ij}-1)s^2 r_1 + n_{ij}^2(p-1)s^2 r_2 + (m-1)p^2 n_{ij}^2 s^2 r_3] \\
 &= \frac{s^2}{mpn_{ij}} (1 + (n_{ij}-1)r_1 + (p-1)n_{ij}r_2 + (m-1)pn_{ij}r_3) \\
 &= \frac{s^2}{mpn_{ij}} (1 + n_{ij}(r_1 - r_2) + pn_{ij}(r_2 - r_3) + mpn_{ij}r_3 - r_1) \quad \dots(31d)
 \end{aligned}$$

Substituting equation(31a), (31b), (31c) And (31d) in equation (21) we get

$$\begin{aligned}
E(MSA) &= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (C_1 + C_2 - 2C_3 - 2a_i C_1) \\
&= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{s^2}{pn_{ij}} (1-r_1) + n_{ij}(r_1-r_2) + pn_{ij}r_2 + \frac{s^2}{mpn_{ij}} [1+n_{ij}(r_1-r_2) + mpn_{ij}r_3-r_3] - \right. \\
&\quad \left. \frac{2s^2}{mpn_{ij}} [1+n_{ij}(r_1-r_2) + n_{ij}(r_2-r_3) + mpn_{ij}r_3-r_1] \right] \\
&= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{n(p-1)} + \frac{s^2}{(m-1)} \left(\frac{1}{p} (1-r_1) + n_{ij}(r_1-r_2) + pn_{ij}r_2 + \frac{1}{mp} [1+n_{ij}(r_1-r_2) + mpn_{ij}r_3-r_3] \right) \\
&= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{n(p-1)} + \frac{s^2}{(m-1)} \left(\frac{1}{p} (1-\frac{1}{m})(1-r_1) + \frac{n_{ij}}{p} (1-\frac{1}{m})(r_1-r_2) + n_{ij}(1-\frac{1}{m})(r_1-r_2) \right) \\
&= \frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + s^2 \left[(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{1}{p} (r_1-r_2) + (r_2-r_3) \right] \right] \quad \dots(32)
\end{aligned}$$

4. Analysis of variance table

Analysis of variance table for the model of study, which is given in equation (5), is given in table 1.

Table 1: Analysis of variance

Sources of variation S.O.V	Degree of freedom d.f	Sum of squares S.S	Mean square M.S	Expected Mean Square E.M.S	
A	m-1	SSA	$\frac{SSA}{(m-1)}$	For independent Data	For dependent data
				$s^2 + \frac{1}{(m-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2$	$\frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} a_i^2}{(m-1)} + s^2 [(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} [\frac{1}{p} (r_1-r_2) + (r_2-r_3)]]$
B	m(p-1)	SSB	$\frac{SSB}{m(p-1)}$	$s^2 + \frac{1}{m(p-1)} \sum_{i=1}^m \sum_{j=1}^p n_{ij} g_{ij}^2$	$\frac{\sum_{i=1}^m \sum_{j=1}^p n_{ij} g_{ij}^2}{m(p-1)} + s^2 [(1-r_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (r_1-r_2)]$
Error	mp(N-1)	SSe	$\frac{SSe}{mp(N-1)}$	σ^2	$(1-r_1)s^2$
Total	mpN-1	SST			

Where SSA, SSB, SSe and SST are given in equation (8)

5. The F test for equality of Factor Level Means

In the last section, we have prepared the ANOVA table, So in this section we will discuss two case for two null hypothesis to know whether the factor level means are equal for the first factor and for the second factor, which is the nested factor. Hence, the alternative conclusions, which we want to, consider for the model of study, are the following .for the first factor is

$$H_0 : a_i = 0 \quad \forall i \quad \dots(33)$$

$$H_1 : a_i \neq 0 \text{ for some } i$$

And for the nested factor is

$$H_0 : g_{ij} = 0 \quad \forall i \text{ and } j \quad \dots(34)$$

$$H_1 : g_{ij} \neq 0 \text{ for at least one } i \text{ or } j$$

Because of the additive property for both total sum of squares and degrees of freedom, therefore according to cochrans theorem (if all n observations Y_i come from the same normal distribution with mean m and variance s^2 and sum square total $SSTO$ is decomposed into k sum square SSr and every sum with degree freedom df_r then SSR/s^2 independent variable distribute c^2 with df_r degree of freedom if $\sum_{r=1}^k df_r = n-1$) and under the null hypothesis.

$$\begin{aligned} W_A &= \frac{SSA}{s^2} \sim \left[(1-r_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (r_1 - r_2) \right] c_{(m-1)}^2 \\ W_B &= \frac{SSB}{s^2} \sim \left[(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{1}{p} (r_1 - r_2) + (r_2 - r_3) \right] \right] c_{m(p-1)}^2 \\ W_e &= \frac{SSe}{s^2} \sim (1-r_1) c_{mp(N-1)}^2 \end{aligned} \quad \dots(35)$$

Where W_A, W_B and W_e are independent of each other, having chi-square distribution each with respective degrees of freedom $(m-1)$, $m(p-1)$ and $mp(N-1)$. Therefore we can form the F_1^* statistic for testing the equality of factor level means of the first factor as follows

$$\begin{aligned} F_1^* &= \frac{W_A / \left[(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{1}{p} (r_1 - r_2) + (r_2 - r_3) \right] (m-1) \right]}{W_e / (1-r_1) mp(N-1)} \\ &= C_1 F_1 \end{aligned} \quad \dots(36)$$

Where

$$\begin{aligned} C_1 &= \frac{(1-r_1)}{\left[(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{1}{p} (r_1 - r_2) + (r_2 - r_3) \right] \right]} \\ F_1 &= \frac{MSA}{MSe} \end{aligned} \quad \dots(37)$$

Also, we can form the F_2^* statistic for testing the equality of factor level mean of the nested factor as follows

$$F_2^* = \frac{W_B / [(1-r_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij}(r_1 - r_2)]m(p-1)}{W_E / (1-r_1)mp(N-1)} \quad \dots(38)$$

$$= C_2 F_2$$

Where

$$C_2 = \frac{(1-r_1)}{\left[(1-r_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij}(r_1 - r_2) \right]} \quad \dots(39)$$

$$F_2 = \frac{MSB}{MSe}$$

5. Correcting for correlation's

The correcting factor C_1 and C_2 which is given in equation (37), (39) could be in the following three cases:

Case1: the correcting factor =1

Case2: the correcting factor >1

Case3: the correcting factor <1

When we are in the first case, then no correction is needed to correct F test . otherwise the correction is essential and we cannot ignore it.

To find the effect of correlation's for testing the factors effects in model (5), the correcting factors C_1 and C_2 are calculated for several different values of correlation's r_1, r_2 and r_3 and also for several different values of m, p, N are shown in tables 2, 3, 4 and 5 respectively.

$$C_1 = \frac{(1-r_1)}{\left[(1-r_1) + \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p n_{ij} \left[\frac{1}{p}(r_1 - r_2) + (r_2 - r_3) \right] \right]}$$

Table 2: shows the different values of the correcting factor C_1

			$m=2, p=2, N=18$ $n_{11}=2, n_{12}=3, n_{13}=4$ $n_{21}=2, n_{22}=3, n_{23}=4$	$m=2, p=2, N=30$ $n_{11}=14, n_{12}=6$ $n_{21}=7, n_{22}=3$	$m=3, p=3, N=60$ $n_{11}=5, n_{12}=6, n_{13}=7$ $n_{21}=4, n_{22}=6, n_{23}=12$ $n_{31}=3, n_{32}=9, n_{33}=8$
r_1	r_2	r_3			
.005	.05	.02	0.789	0.678	0.613
.001	.005	.002	0.985	0.869	0.75
.005	.001	.001	0.976	0.957	0.949
.05	.05	.05	1	1	1
.01	.05	.055	1.2	1.467	1.588
.01	.02	.01	1.398	2	2.75

$$C_2 = \frac{(1 - r_1)}{\left[(1 - r_1) + \frac{1}{mp} \sum_{i=1}^m \sum_{j=1}^p n_{ij} (r_1 - r_2) \right]}$$

Table 3: shows the different values of the correcting factor C_2

r_1	r_2	$m=2, p=2, N=18$ $n_{11}=2, n_{12}=3, n_{13}=4$ $n_{21}=2, n_{22}=3, n_{23}=4$	$m=2, p=2, N=30$ $n_{11}=14, n_{12}=6$ $n_{21}=7, n_{22}=3$	$m=3, p=3, N=60$ $n_{11}=5, n_{12}=6, n_{13}=7$ $n_{21}=4, n_{22}=6, n_{23}=12$ $n_{31}=3, n_{32}=9, n_{33}=8$
.05	.01	0.888	0.76	0.78
.005	.003	0.994	0.985	0.987
.05	.05	1	1	1
.001	.005	1.004	1.031	1.027
.01	.05	1.138	1.435	1.369

Table 4: shows the true alpha level for some different values of C_1

α C_1	0.01	0.025	0.05	0.1	0.25
0.4	0.056	0.109	0.19	0.348	0.408
0.6	0.04	0.08	0.143	0.265	0.355
0.8	0.025	0.053	0.097	0.183	0.303
1	0.01	0.025	0.05	0.1	0.25
1.2	0	0	0.003	0.017	0.197
1.4	0	0	0	0	0.145

Table 5: shows the true alpha level for some different values of C_2

α C_2	0.01	0.025	0.05	0.1	0.25
0.4	0.044	0.087	0.155	0.294	0.361
0.6	0.033	0.067	0.12	0.229	0.324
0.8	0.021	0.046	0.085	0.165	0.287
1	0.01	0.025	0.05	0.1	0.25
1.2	0	0.002	0.011	0.029	0.209
1.4	0	0	0.004	0.008	0.173

We can note from table 4,5, that the true α level in flate /deflate when the correction factor greater/less than 1, and this leads to have a bigger /smaller rejection region for the complete null hypothesis on testing factors. Hence the uncorrected F test can be liberal for some tests and conservative for others in ANOVA when the correlation structure is ignored.

6. The Relationship between a and true α

The relationship between values of α when data are uncorrelated and values of the true a when data are correlated for different values of the correction factor C_1, C_2 is shown in figure 1, 2 when $m = 2$, $p = 2$, $N = 18$, $n_{11} = 2$, $n_{12} = 3$, $n_{13} = 4$, $n_{21} = 2$, $n_{22} = 3$, $n_{23} = 4$.

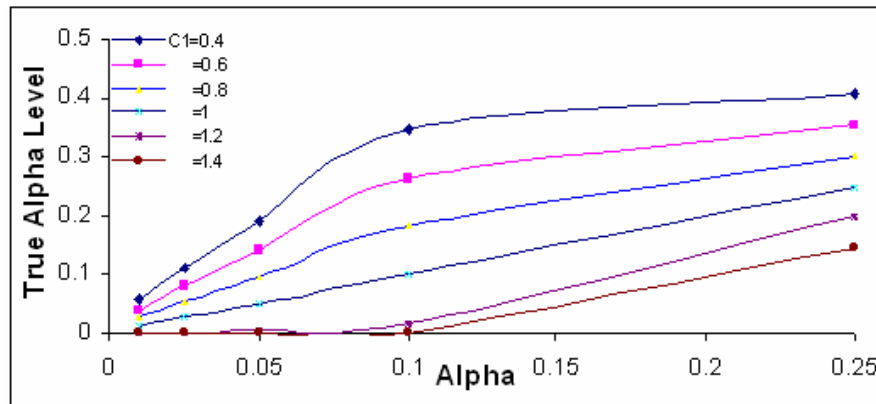


Figure 1: True Alpha versus Alpha for different values of C_1

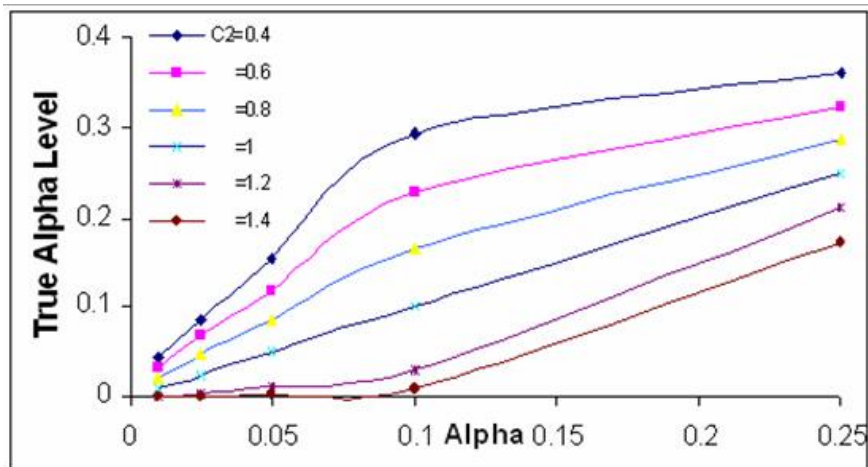


Figure 2: True Alpha versus Alpha for different values of C_2

7. Conclusion

Correlation's constant C_1, C_2 small correlation can be amplified by the number of treatment $n_{ij}, i = 1, 2, \dots, m$ $j = 1, 2, \dots, p$. And these correlations can thus easily inflate the type I error Rate. Table 1.4, 1.5 show that the true alpha inflate/deflate when the correction constant greater/less than 1, the uncorrected F test would be liberal test (high

type I error rate). In conclusion, The F statistic can be effected by very large correlations sometimes the same as very small correlations. That is, the F statistic is affected by the value of the correction constant C_1, C_2 and not the value of correlations.

REFERENCES

- [1] AL-Rabeaa, Z.A. 1995, "The Effect of Dependent Data on Type I Error Rates for Multiple Comparisons Procedures", M.Sc Thesis, College of Science, University of Basrah.
- [2] Arnold, S.F. 1981, "Theory of linear Models and Multivariate Analysis", John Wiley, New York.
- [3] Gabbara, S.D. and AL-Rabeaa, Z.A.1996, "The Effect of Dependent Data on Type I Error Rates for Multiple Comparisons Procedures", Journal of Iraqi Statistical Association Proceeding of the 8th conference.
- [4] Gastwirth, J.L. and Rubin, H. 1972, "Effect of Dependence of level of Some One-Sample Test", JASA, Vol. 66, No. 3, PP. 816-820.
- [5] Latif, I.S. 2000, "Determining The Expected Mean Squares For 2-way Nested Effects Model With Weak Correlated Data", M.Sc. Thesis, College of Science, University of Salahaddin.
- [6] Pavur, R.J. and Davenport, J.M. 1985, "The (Large) Effect of (small) Correlation In ANOVA and Correction Procedures", Amer. J. Math. & Manag. Sci., Vol. 15, No. 2, PP. 77-92.
- [7] Pavur, R.J. 1988, "Type I Error Rates for Multiple Comparisons procedures With Dependent Data", J. Amer. Statis. Assoc., Vol. 42, NO. 3, PP. 171-173.
- [8] Smith, J.H. and Lewis, T.O. 1980, "Determining the Effects of Interclass Correlation On Factorial Experiments", J. Amer. Statis. Assoc., Vol. 13, NO. 9, PP. 1553-1556.
- [9] Neter, J. and Wasserman, W. 1974, "Applied linear Statistical Models", IRWIN.
- [10] Ott, Lyman. 1984, "An introduction to method and data analysis", Duxbury Press. Wadsworth, Inc.
- [11] Rao, C.R. and Mitral, S.K. 1971, "Generalized Inverse of Matrices and Its Applications", New York: John Wiley and Sons, Inc. Canada.