Science Journal of Mathematics & Statistics **Published By Science Journal Publication**
ISSN: 2276-6324
http://www.sjpub.org/sjms.html

Research Article **Volume 2013 (2013), Article ID sjms-104, 6 Pages, doi: 10.7237/sjms/104**

Ties Adjusted Two Way Analysis Of Variance Tests with Unequal Observations Per Cell

Oyeka I.C.A

Department of Industrial Mathematics and Applied Statistics,
Ebonyi State University Abakaliki,Nigeria.

Accepted 2nd June, 2013

Abstract

This paper proposes a non-parametric statistical method for the analysis of factor effects in a two factor analysis of variance type model with unequal observations or replications per cell or treatment combination. It is here, for simplicity assumed that there are no interactions between the factors of interest or that such interactions have been removed by appropriate data transformation. The proposed test statistics are intrinsically and structurally adjusted for the possible presence of ties between observations in each cell or treatment combination, thereby obviating the need to require the sampled populations to be continuous or even numeric. The populations may be measurements on as low as the ordinal scale. The method enables the researcher test for the statistical significance of not only the effect of each factor level but also for the equally of several factor level effects.

The proposed method which is generally more robust than the corresponding classical 'F' test that may be used for the same purpose is illustrated with some sample data and shown to be as expected more powerful than an alternative ties-unadjusted method which is shown to have reduced degrees of freedom, for the same sample size.

Keywords: intrinsically and structurally adjusted, ties-unadjusted, level of factor A, level of factor A, replication

Introduction

The Friedmans Two Way Analysis Of Variance test by ranks is a non-parametric equivalence of the classical two way analysis of variance F test used when the data being analyzed conform with mixed effects model, with only one observation per cell that is without replication (Gibbons, 1973; Siegel,1956).Statistical methods also exist for use in data analysis when there are equal observations or replications per treatment combination. However, just as it may often be too difficult or too expensive to obtain more than one observation per treatment combination, it may also prove impossible to obtain equal number of observations per cell in a two factor analysis. For example even though an experiment may have been planned with equal number of observations per cell, some of the observations may end up missing for a variety of reasons. The required number of subjects in a given treatment combination may not be available to the experimenter at a specified time or location. Research subjects may for instance be lost to follow-up, withdrawal, death, migration, refusal, etc. With the sample size per treatment combination is not the same for all treatment combinations in a two factor analysis, the analysis of variance of factor effects, whether for a fixed-effects, random effects or mixed-effects model using the classical F-test becomes more complicated and the usual calculations are no longer directly applicable (Montgomery and Peck, 1982; Oyeka, 2013).

Approximate methods are now used for data analysis including the so called method of unweighted means provided all the assumptions for the use of 'F' test are satisfied. However the sums of squares are no longer additive under this approach (Montgomery and Peck,1982).In these situations a more general, easier and exact way to obtain the proper sums of squares for testing for factor effects and interactions is through the use of dummy variable regression techniques(Neter and Wasserman,1974).Oyeka and others (Oyeka et al,2012)have also developed a non-parametric statistical method for the analysis of effects in a two factor analysis of variance with unequal observations per cell. But the authors' method strictly speaking is only applicable when the sampled populations are continuous, that is when there are no ties between the observations themselves.

We here propose and present a non-parametric statistical method that is intrinsically and structurally adjusted for possible ties between observations in the cells. The factors or population of interest may now be measurements on as low as the ordinal scales that are not necessarily continuous or even numeric.

The Proposed Method

Suppose interest is studying the effects of two factors A and B on some Criterion or dependent variable. Let n_{lj} be the number of observations in the lth level of factor A and the jth level of factor B,that is the number of observations in the $(l, j)th$ cell of the A by B treatment or factor combinations for $l = 1, 2\ldots.a,\ j = 1, 2, .., b$ .Let x_{ilj} be the ith observation in the $(l, j)th$ cell,that is in the lth level of factor A and the jth level of factor B for $i = 1, 2\ldots, n_{lj}, l = 1, 2, \ldots a;\ j - 1, 2, \ldots, b$

Now instead of using the unweighted means approach based on the cell means, we here base our analysis on the median M of the combined sample observations. This approach is informed by the fact that even though the observations on the 'a' levels of factor A and the 'b' levels of factor B may separately have different medians, however because both Samples together comprise one data set, these observations when pooled together into one combined sample would have one common overall median, M say. Now let

$$
{}_{ilj} = \begin{cases} 1, if\ x_{ilj}\ is\ a\ higher(better, l\arg er) observation\ or\ score \\ than\ M; or\ x_{ilj} > M \\ 0, if\ x_{ilj}\ is\ the\ same\ score\ as(equal\ to)M; or\ x_{ilj} = M \\ -1, if\ x_{ilj}\ is\ a\ lower(worse, smaller) observation\ or\ score \\ than\ M; or\ x_{ilj} < M \end{cases} \quad (1)
$$

$$
for\ i = 1, 2, \ldots, n_{lj}; l = 1, 2, \ldots., a; j = 1, 2, \ldots., b
$$

Let

$$
\circledR_{lj}^{+} = P(u_{ilj} = 1) : \circledR_{lj}^{0} = P(u_{ilj} = 0) : \circledR_{lj}^{-} = P(u_{ilj} = -1) \quad (2)
$$

where

$$
\circledR_{lj}^{+} + \circledR_{lj}^{0} + \circledR_{lj}^{-} = 1 \quad (3)
$$

Also let

$$
W_{lj} = \sum_{i=1}^{n_{lj}} u_{ilj} \quad (4)
$$

Now

$$
E(u_{ilj}) = \circledR_{lj}^{+} - \circledR_{lj}^{-}; Var(u_{ilj}) = \circledR_{lj}^{+} + \circledR_{lj}^{-} - \left(\circledR_{lj}^{+} - \circledR_{lj}^{-}\right)^2 \quad (5)
$$

Also

$$
E(W_{lj}) = \sum_{i=1}^{n_{lj}} E(u_{ilj}) = n_{lj}\left(\circledR_{lj}^{+} - \circledR_{lj}^{-}\right) \quad (6)
$$

And

$$
Var(W_{lj}) = \sum_{i=1}^{n_{lj}} Var(u_{ilj}) = n_{lj}\left(\circledR_{lj}^{+} + \circledR_{lj}^{-} - (\circledR_{lj}^{+} - \circledR_{lj}^{-})^2\right) \quad (7)
$$

Now $®_{lj}^{+}, ®_{lj}^{-}$ *and* $®_{lj}^{0}$ are respectively the probabilities that the performance scores by subjects in the $(l, j)th$ cell or treatment combination are on the average higher (better, larger) than or lower (worse, smaller) than or the same as (equal to) the common overall median M. Their sample estimates are respectively

$$P_{lj}^{+} = ®_{lj}^{+} = \frac{f_{lj}^{+}}{n_{lj}}; P_{lj}^{-} = ®_{lj}^{-} = \frac{f_{lj}^{-}}{n_{lj}}; P_{lj}^{0} = ®_{lj}^{0} = \frac{f_{lj}^{0}}{n_{lj}} = \frac{n_{lj} - f_{lj}^{+} - f_{lj}^{-}}{n_{lj}} = 1 - f_{lj}^{+} - f_{lj}^{-} \quad (8)$$

where f_{lj}^{+}, f_{lj}^{-} *and* $f_{lj}^{0} = n_{lj} - f_{lj}^{+} - f_{lj}^{-}$ are respectively the total number of 1s,-1s and 0s in the $(l, j)th$ cell or treatment combination, that is the total number of 1s,-1s and 0s in the frequency distribution of the n_{lj} values of these numbers in $u_{ilj}, i = 1, 2, \ldots, n_{lj}; l = 1, 2, \ldots, a; j = 1, 2, \ldots, b$ .For subjects at the lth level of factors A, the total number of 1s,-1 and 0s,for all levels of factors B, that is the total number of times the performance or scores by subjects at the lth level of factors A are higher (better, larger) than, per lower (worse, smaller) than or the same as (equal to)the common median M

for the 'b' all levels of factor B are respectively

$$f_{l}^{+} = \sum_{j=1}^{b} f_{lj}^{+}; f_{l}^{-} = \sum_{j=1}^{b} f_{lj}^{-}; f_{l}^{0} = \sum_{j=1}^{b} f_{lj}^{0} = n_{l.} - f_{l.}^{+} - f_{l.}^{-} \quad (9)$$

Where

$n_{l.} = \sum_{j=1}^{b} n_{lj}$, is the total number of subjects at the lth level of factor A.

The corresponding sample proportions are respectively

$$P_{l.}^{+} = \frac{f_{l.}^{+}}{n_{l.}}; P_{l.}^{-} = \frac{f_{l.}^{-}}{n_{l.}}; P_{l.}^{0} = \frac{f_{l.}^{0}}{n_{l.}} = 1 - P_{l.}^{+} - P_{l.}^{-} \quad (10)$$

The corresponding frequencies and proportions for the jth factor of B for all levels of factor A are respectively

$$f_{.j}^{+} = \sum_{l=1}^{a} f_{lj}^{+}; f_{.j}^{-} = \sum_{l=1}^{a} f_{lj}^{-}; f_{.j}^{0} = n_{.j} - f_{.j}^{+} - f_{.j}^{-} \quad (11)$$

where

$$n_{.j} = \sum_{l=1}^{a} n_{lj}$$

And

$$P_{.j}^{+} = \frac{f_{.j}^{+}}{n.j}; P_{.j}^{-} = \frac{f_{.j}^{-}}{n.j}; P_{.j}^{0} = \frac{f_{.j}^{0}}{n.j}; = 1 - P_{.j}^{+} - P_{.j}^{-} \quad (12)$$

These results are summarized in Table 1

Table 1: Summary values of u_{ilj} (Equation 1) in a two way ANOVA with unequal observations per cell

Levels of factor A	Levels of factor B					
	Frequencies and proportions of 1s and 0	1	2……	….	j….	b	Total
1	f_{1j}^{+}	f_{11}^{+}	$f_{12}^{+} \cdots$	….	$f_{1j}^{+} \cdots$	f_{1b}^{+}	$f_{1.}^{+}$
	f_{1j}^{-}	f_{11}^{-}	$f_{12}^{-} \cdots$	…..	$f_{1j}^{-} \cdots$	f_{1b}^{-}	$f_{1.}^{-}$
	$f_{1j}^{0} = n_{11} - f_{11}^{+} - f_{11}^{-}$		f_{12}^{0}…..	…..	f_{1j}^{0}..	f_{1b}^{0}	$f_{1.}^{0}$
	Total n_{1j}	n_{11}	n_{12}……	…..	n_{1j}…	n_{1b}	$n_{1.}$
	P_{1j}^{+}	P_{11}^{+}	P_{12}^{+}……	……	P_{1j}^{+}	P_{1b}^{+}	$P_{1.}^{+}$
	P_{1j}^{-}	P_{11}^{-}	P_{12}^{-}	…..	P_{1j}^{-}	P_{1b}^{-}	$P_{1.}^{-}$
	$P_{1j}^{0} = 1 - P_{1j}^{+} - P_{1j}^{-}$	P_{11}^{0}	P_{12}^{0}	……	P_{1j}^{0}	P_{1b}^{0}	$P_{1.}^{0}$
	…..	…..	….	….	…..	…	….
	…..	….	…..	…	….	…	….
i	f_{lj}^{+}	f_{l1}^{+}	f_{l2}^{+}	…	f_{lj}^{+}	f_{lb}^{+}	$f_{l.}^{+}$
	f_{lj}^{-}	f_{l1}^{-}	f_{l2}^{-}	….	f_{lj}^{-}	f_{lb}^{-}	$f_{l.}^{-}$
	$f_{lj}^{0} = n_{lj} - f_{lj}^{+} - f_{lj}^{+}$	f_{l1}^{0}	f_{l2}^{0}	…..	f_{lj}^{0}	f_{lb}^{0}	$f_{l.}^{0}$
	Total n_{lj}	n_{l1}	n_{l2}	….	n_{lj}….	n_{lb}	$n_{l.}$
	P_{lj}^{+}	P_{l1}^{+}	P_{l2}^{+}	…..	P_{lj}^{+}	P_{lb}^{+}	$P_{l.}^{+}$
	P_{lj}^{-}	P_{l1}^{-}	P_{l2}^{-}	……	P_{lj}^{-}	P_{lb}^{-}	$P_{l.}^{-}$
	$P_{lj}^{0} = 1 - P_{lj}^{+} - P_{lj}^{-}$	P_{l1}^{0}	P_{l2}^{0}	……	P_{lj}^{0}	P_{lb}^{0}	$P_{l.}^{0}$
a	f_{aj}^{+}	f_{a1}^{+}	f_{a2}^{+}	……	f_{aj}^{+}	f_{ab}^{+}	$f_{a.}^{+}$
	f_{aj}^{-}	f_{a1}^{-}	f_{a2}^{-}	….	f_{aj}^{-}	f_{ab}^{-}	$f_{a.}^{-}$
	$f_{aj}^{0} = n_{aj} - f_{aj}^{+} - f_{aj}^{-}$	f_{a1}^{0}	f_{a2}^{0}	…	f_{aj}^{0}	f_{ab}^{0}	$f_{a.}^{0}$
	Total n_{aj}	n_{a1}	n_{a2}	…..	n_{aj}	n_{ab}	$n_{a.}$
	P_{aj}^{+}	P_{a1}^{+}	P_{a2}^{+}	….	P_{aj}^{+}	P_{ab}^{+}	$P_{a.}^{+}$
	P_{aj}^{-}	P_{a1}^{-}	P_{a2}^{-}	……	P_{aj}^{-}	P_{ab}^{-}	$P_{a.}^{-}$
	$P_{aj}^{0} = 1 - P_{aj}^{+} + P_{aj}^{-}$	P_{a1}^{0}	P_{a2}^{0}	……	P_{ab}^{0}	P_{aj}^{0}	$P_{a.}^{0}$

How to Cite this Article: Oyeka I.C.A, "Ties Adjusted Two Way Analysis Of Variance Tests with Unequal Observations Per Cell" Science Journal Of Mathematics and Statistics, Volume 2013 (2013), Article ID sjms-104, 6 Pages, doi: 10.7237/sjms/104

Total	$f_{.j}^{+}$	$f_{.1}^{+}$	$f_{.2}^{+}$	....	$f_{.j}^{+}$	$f_{.b}^{+}$	$f_{..}^{+} = f^{+}$
	$f_{.j}^{-}$	$f_{.1}^{-}$	$f_{.2}^{-}$	.....	$f_{.j}^{-}$	$f_{.b}^{-}$	$f_{..}^{-} = f^{-}$
	$f_{.j}^{0} = n_{.j} - f_{.j}^{+} - f_{.j}^{-}$	$f_{.1}^{0}$	$f_{.2}^{0}$	......	$f_{.j}^{0}$	$f_{.b}^{0}$	$f_{..}^{0} = f^{0}$
	Total $n_{.j}$	$n_{.1}$	$n_{.2}$	.....	$n_{.j}$	$n_{.b}$	$n_{..} = n$
	$P_{.j}^{+}$	$P_{.1}^{+}$	$P_{.2}^{+}$	....	$P_{.j}^{+}$	$P_{.b}^{+}$	$P_{..}^{+} = P^{+}$
	$P_{.j}^{-}$	$P_{.1}^{-}$	$P_{.2}^{-}$	.....	$P_{.j}^{-}$	$P_{.b}^{-}$	$P_{..}^{-} = P^{-}$
	$P_{.j}^{0} = 1 - P_{.j}^{+} - P_{.j}^{-}$	$P_{.1}^{0}$	$P_{.2}^{0}$	......	$P_{.j}^{0}$	$P_{.b}^{0}$	$P_{..}^{0} = P^{0}$

Now $®_{lj}^{+} - ®_{lj}^{-}$ is the probability that the performance or scores by subjects at the $(l, j)th$ cell or treatment combination are on the average higher (better, larger) less the probability that they are on the average lower (worse, smaller) than M. If $®_{lj}^{+} - ®_{lj}^{-} = 0$, then this would imply that the scores by subjects at the lth level of factor A and jth level of factor B are as likely to be higher as lower than M meaning that M is in fact equal or the same as the median score of the $(l, j)th$ treatment combination. Now if for a given level $'l'$ of factor A, $®_{lj}^{+} - ®_{lj}^{-} = 0$, for all the ‘b’ levels of factor B. This would mean that the median scores for each of these ‘b’ treatment combinations are equal and each is equal to M. Hence in general if $®_{lj}^{+} - ®_{lj}^{-} = 0$, for each and all of the ‘a’ levels of factor A for all the ‘b’ levels of factor B, then the effects of factor A on subjects for all levels of factor B would be equal, implying that the ‘b’ levels of factor B have equal medians for all levels of factor A.

Following a similar argument we would expect the ‘a’ levels of factor A to also have equal medians for all levels of factor B. Hence the null hypothesis that may need to be tested are that there is no significant difference between the effects of the (1) ‘b’ levels of factor B on subjects at all levels of factor A; and (2) ‘a’ levels of factor A for all the ‘b’ levels of factor B or in terms of the medians of these treatments, the null hypothesis that may be tested are respectively :

$$For\ \ factor\ B; H_0 : M_{1b} = M_{2b} = ... = M_{bb} = M_c, say \qquad 13$$

$$And$$

$$For\ \ factor\ A; H_0 : M_{1a} = M_{2a} = ... = M_{aa} = M_c, say \qquad .14$$

Versus any appropriate alternative hypothesis where M_{la}, M_{lb} are respectively the medians of the lth level of factor A and jth level of factor B and Mc is the common overall population median.

Appropriate test statistic for testing these null hypothesis can be developed by complete enumeration and determination of the distribution of W_{lj} of Equ 4 based on Equations 1-3 and 5-7.This approach would however be cumbersome and tedious. Instead we here propose an alternative method based on the Chi-square test of independence.

Now the observed frequencies of 1s,-1s and 0s, that is the total number of times the performance or scores by subjects at the $(l, j)th$ cell or treatment combination are higher (better, larger) than, or lower (worse, smaller) than or the same as (equal to) the common median M are respectively

$$O_{1lj} = f_{lj}^{+}; O_{2lj} = f_{lj}^{-}; O_{3lj} = f_{lj}^{0} = n_{lj} - f_{lj}^{+} - f_{lj}^{-} \qquad 15$$

We here assume for simplicity that there are no factors A by factor B interactions or that these interactions have all been removed by appropriate data transformation. Under this assumption and under the null hypothesis that subjects at the lth level of factor A do not experience differential treatment effects at all the ‘b’ levels of factor B, that is factor B has equal medians at the lth level of factor A, then the expected frequencies of 1s,-1s and 0s at the lth level of factor A and jth level of factor B are respectively

$$E_{1lj} = \frac{n_{lj} f_{l.}^{+}}{n_{l.}}; E_{2lj} = \frac{n_{lj} f_{l.}^{-}}{n_{l.}}; E_{3lj} = \frac{n_{lj} f_{l.}^{0}}{n_{l.}} = \frac{n_{lj}\left(n_{l.} - f_{l.}^{+} - f_{l.}^{-}\right)}{n_{l.}} \qquad 16$$

Note that Equation 16 yields the values of the expected frequencies of 1s,-1s and 0s respectively for the lth level of factor A at all levels ‘J’ of factor B, that is the expected number of times observations or scores at the lth level of factor A are higher (better, larger) than lower (worse, smaller) than, or the same as (equal to) the combined or common sample median M, for all the ‘b’ levels of factor B; in other words the expected number of observations or scores in the $(l, j)th$ cell or treatment combination that are higher (better, larger)than or lower (worse, smaller) than or the same as (equal to)M for some fixed level $'l'$ of factor A and all levels $' j' = 1, 2, \ldots, b$ of factor B.

Now under the null hypothesis of no differential treatment effects between the ‘b’ levels of factor B for subjects at the lth level of factor A, the test statistic

$$Q_{l.} = \sum_{i=1}^{3}\sum_{j=1}^{b} \frac{\left(O_{ilj} - E_{ilj}\right)^2}{E_{ilj}} \qquad 17$$

Has approximately the chi-square distribution with (3-1)(b-1)=2(b-1) degrees of freedom for sufficiently large $n_{l.}; l = 1, 2, \ldots, a$.Using Equations 15 and 16 in Equation 17 we have

$$Q_{l.} = \frac{\sum_{j=1}^{b} \frac{\left(f_{lj}^{+} - \frac{n_{lj} f_{l.}^{+}}{n_{l.}}\right)^2}{\frac{n_{lj} f_{l.}^{+}}{n_{l.}}} + \sum_{j=1}^{b} \frac{\left(f_{lj}^{-} - \frac{n_{lj} f_{l.}^{-}}{n_{l.}}\right)^2}{\frac{n_{lj} f_{l.}^{-}}{n_{l.}}} + \sum_{j=1}^{b}\left((n_{lj} - f_{lj}^{+} - f_{lj}^{-}) - \frac{n_{lj}(n_{l.} - f_{l.}^{+} - f_{l.}^{-})}{n_{l.}}\right)^2}{\frac{n_{lj}\left(n_{l.} - f_{l.}^{+} - f_{l.}^{-}\right)}{n_{l.}}}$$

Which when further simplified and evaluated yields the test statistic

$$Q_{l.} = \frac{n_{l.}}{f_{l.}^{+} f_{l.}^{-}(n_{l.} - f_{l.}^{+} - f_{l.}^{-})} \cdot \frac{\left(f_{l.}^{-}(n_{l.} - f_{l.}^{-})\right)\frac{\sum_{j=1}^{b}\left(f_{lj}^{+} - n_{lj} f_{l.}^{+}\right)^2}{n_{l.}} + f_{l.}^{+}(n_{l.} - f_{l.}^{+})\sum_{j=1}^{b}\frac{\left(f_{lj}^{-} - n_{lj} f_{lj}^{-}\right)^2}{n_{l.}} + 2 f_{l.}^{+} f_{l.}^{-} \sum_{j=1}^{b}\left(\frac{f_{lj}^{+} - n_{lj} f_{l.}^{+}}{n_{l.}}\right)\left(\frac{f_{lj}^{-} - n_{lj} f_{lj}^{-}}{n_{l.}}\right)}{n_{lj}}$$

Or Equivalently in terms of the sample proportions of Equations 8 and 10 becomes

$$Q_{l.} = \frac{1}{P_{l.}^{+} P_{l.}^{-}} \frac{\left(P_{l.}^{-}(1 - P_{l.}^{-})\right)\sum_{j=1}^{b} n_{lj}\left(P_{lj}^{+} - P_{l.}^{+}\right)^2 + P_{l.}^{+}(1 - P_{l.}^{+})\sum_{j=1}^{b} n_{lj}(P_{lj}^{-} - P_{l.}^{-})^2 + 2 P_{l.}^{+} P_{l.}^{-} \sum_{j=1}^{b} n_{lj}(P_{lj}^{+} - P_{l.}^{+})(P_{lj}^{-} - P_{l.}^{-})}{\left(1 - P_{l.}^{+} - P_{l.}^{-}\right)} \qquad 19$$

Which under the null hypothesis has approximately the chi-square distribution with 2(b-1) degrees of freedom for sufficiently large $n_{l.}, l = 1, 2, \ldots, a$.

An equivalent and easier to use computational formulae for Equation 19 is

$$Q_{l.} = \frac{1}{P_{l.}^{+} P_{l.}^{-}(1 - P_{l.}^{+} - P_{l.}^{-})}\left(P_{l.}^{-}(1 - P_{l.}^{-})\left(\sum_{j=1}^{b} n_{lj} P_{lj}^{+2} - n_{l.} P_{lj}^{+2}\right) + P_{l.}^{+}(1 - P_{l.}^{+})\left(\sum_{j=1}^{b} n_{lj} P_{lj}^{-2} - n_{l.} P_{lj}^{-2}\right) + 2 P_{l.}^{+} P_{l.}^{-}\left(\sum_{j=1}^{b} n_{lj} P_{lj}^{+} P_{lj}^{-}\right)\right) \qquad 20$$

How to Cite this Article: Oyeka I.C.A, "Ties Adjusted Two Way Analysis Of Variance Tests with Unequal Observations Per Cell" Science Journal Of Mathematics and Statistics, Volume 2013 (2013), Article ID sjms-104, 6 Pages, doi: 10.7237/sjms/104

Hence the null hypothesis of no difference effects between the ‘b’ levels of factor B on subjects at all the ‘a’ levels factor A (Equation 13) is tested using the test statistic

$$\chi^2 = \sum_{l=1}^{a} Q_{l.} = \sum_{l=1}^{a} \frac{1}{P_{l.}^{+} P_{l.}^{-}(1-P_{l.}^{+}-P_{l.}^{-})}\left(P_{l.}^{-}(1-P_{l.}^{-})\right)\sum_{j=1}^{b} n_{lj}\left(P_{lj}^{+}-P_{l.}^{+}\right)^2 + P_{l.}^{+}(1-P_{l.}^{+})\sum_{j=1}^{b} n_{lj}(P_{lj}^{-}-P_{l.}^{-})^2 + 2P_{l.}^{+}P_{l.}^{-}\sum_{j=1}^{b} n_{lj}(P_{lj}^{+}-P_{l.}^{+})(P_{lj}^{-}-P_{l.}^{-}) \quad 21$$

Which under the null hypothesis H_0 of equation 13 has approximately the chi-square distribution with 2a (b-1) degree of freedom for such sufficiently large $'n' = n_{..} = \sum_{l=1}^{a} n_{l.}$ the null hypothesis of Equation 13 is rejected at the α level of significance if

$$\chi^2 \geq \chi^2_{1-\alpha;2a(b-1)} \quad 22$$

Otherwise H_0 is accepted.

Similarly the null hypothesis of Equation 14 that factor A has no differential effects on subjects at all levels of factor B is tested using the test statistic

$$\chi^2 = \sum_{j=1}^{b} Q_{.j} = \sum_{j=1}^{b}\left(\frac{1}{P_{.j}^{+}P_{.j}^{-}(1-P_{.j}^{+}-P_{.j}^{-})}(P_{.j}^{-}(1-P_{.j}^{-})\sum_{l=1}^{a} n_{lj}(P_{lj}^{+}-P_{.j}^{+})^2 + P_{.j}^{+}(1-P_{.j}^{+})\sum_{l=1}^{a} n_{lj}(P_{lj}^{-}-P_{.j}^{-})^2 + 2P_{.j}^{+}P_{.j}^{-}\sum_{l=1}^{a} n_{lj}\left(P_{lj}^{+}-P_{.j}^{+}\right)\left(P_{lj}^{-}-P_{.j}^{-}\right)\right) \quad 23$$

Which under the null hypothesis of Equation 14 has approximately the chi-square distribution with 2b (a-1) degrees of freedom for sufficiently large n=n…The null hypothesis of Equation 14 is rejected at the A level

of significance if

$$\chi^2 \geq \chi^2_{1-\alpha;2b(a-1)} \quad 24$$

Otherwise H_0 is accepted

Finally the proposed method easily applicable when the cells or treatment combinations have equal observation; that is when they are equal replications per treatment combination. Thus if $n_{lj} = n$ for all $l = 1,2,...,a, j = 1,2,...,b$ then $n_{l.} = \sum_{j=1}^{b} n_{lj} = b_{.n}; n_{.j} = \sum_{l=1}^{a} n_{lj} = a_{.n}$ the sample proportions $P_{l.}^{+}, P_{l.}^{0}$ *and* $P_{l.}^{-}$ now becomes simply the unweighted sums of P_{lj}^{+}, P_{lj}^{0} *and* P_{lj}^{-} each computed using equal cell sizes of $n_{lj} = n$ respectively summed over the ‘b’ levels of factor B. Similarly the sample proportions $P_{.j}^{+}, P_{.j}^{0}$ *and* $P_{.j}^{-}$ are also now simply the unweighted sums of the corresponding cell proportions summed over the ‘a’ levels of factor A. The overall sample size now becomes n..=abn, thereby making calculations relatively much easier. Note that when not intrinsically adjusted for possible presence of ties, the test statistic for the null hypothesis of Equations of 13 and 14, namely Equations 21 and 23, reduce respectively to (Oyeka et al, 2012)

$$\chi^2 = \sum_{j=1}^{b} Q_{l.} = \sum_{l=1}^{a} \frac{1}{P_{l.}^{+}(1-P_{l.}^{+})}\left(\sum_{l=1}^{b} n_{lj}P_{lj}^{+2} - n_{l.}P_{l.}^{+2}\right) \quad 25$$

With a (b-1) degrees of freedom,
And

$$\chi^2 = \sum_{j=1}^{b} Q_{j.} = \sum_{j=1}^{b} \frac{1}{P_{.j}^{+}(1-P_{.j}^{+})}\left(\sum_{l=1}^{a} n_{lj}P_{lj}^{+2} - n_{.j}P_{.j}^{+2}\right) \quad 26$$

With b (a-1) degrees of freedom.

Illustrative Example.

Table 2 shows the letter grades earned by random samples of students who took a course offered during three semesters of the academic year in a University

Table 2: Letter Grades earned by Random Samples of students in a course

Semester	Academic Year			
1,Fall	1	2	3	4
	D,A+,A,F	E,E,C-,B-	B-,A+,B+,C+	A,A+,A,B+
	C,A+,D,C	B,A,A-,E	F,C+,A+,E	A+,A+,F,B+
	B-,C+,B	B,F,C-,C-,	E,C+	B-,A-A,A,
		F		B-A+
2,Winter				
	B-,D,A-	B+,C+,C-,B	B+,E,D,	B-,C-,D,C+
	C,B,D,	D,C+F,D	B-,F,B,	B,B+,E,B-
	A+,B,C+	C,C,A+	B+,C+,	A+C,B-,F,
3,Summer				
	C+,B+,F	D,A-,C-,C	F,C-,E,	B-,C-,C,C-
	E,D,E	A-,A+,A+,E,	C-,E,B-	B-,A-,F,C
	B	A,C+	B,A,D	C+,C-,A

Research interest is to determine whether students perform equally well in the course in all semesters and also in all years. To test the null hypothesis that students performance or scores in the course do not differ by year for all semesters and by semester irrespective of academic year, we here first pool the sample observations that is the letter grades in Table 2 into one combined sample of ‘n’=n..=125, and determine the overall common sample median which is found to be M=a C+ grade.

Now applying Equation 1 with M=C+ to the letter grades of Table 2 we obtain values of u_{ilj} which are summarized in Table 3.

Table 3: Summary values of u_{ilj} (Equation 1) and other statistics for the data of Table 1

Levels of factor A(semester) Fall	Frequencies and proportion	Levels of factor B(Academic year) $Q_{l.}$				
		1	2	3	4	Total
	f_{1j}^{+}	5	7	4	13	29($f_{1.}^{+}$)
	f_{1j}^{-}	5	6	3	1	15($f_{1.}^{-}$)
	f_{1j}^{0}	1	0	3	0	4 $f_{1.}^{0}$
	Total(n_{1j})	11	13	10	14	48($n_{1.}$)
	P_{1j}^{+}	0.455	0.538	0.400	0.929	0.604($P_{1.}^{+}$)
	P_{1j}^{-}	0.455	0.462	0.300	0.071	0.313($P_{1.}^{-}$)
	P_{1j}^{0}	0.090	0.000	0.300	0.000	0.083($P_{1.}^{0}$)
	$Q_{.j} = 1.955$					15.250($Q_{1.}$)
Winter	f_{2j}^{+}	5	3	4	6	18($f_{2.}^{+}$)
	f_{2j}^{-}	3	6	3	5	17($f_{2.}^{-}$)
	f_{2j}^{0}	1	2	1	1	5($f_{2.}^{0}$)
	Total (n_{2j})	9	11	8	12	40($n_{2.}$)
	P_{2j}^{+}	0.556	0.273	0.500	0.500	0.450($P_{2.}^{+}$)
	P_{2j}^{-}	0.333	0.545	0.375	0.417	0.425($P_{2.}^{-}$)
	P_{2j}^{0}	0.111	0.182	0.125	0.083	0.125($P_{2.}^{0}$)
	$Q_{.j} = 19.696$					$Q_{2.} = 12.458$
Summer	f_{3j}^{+}	2	5	3	4	14($f_{3.}^{+}$)
	f_{3j}^{-}	4	4	6	6	20($f_{3.}^{-}$)
	f_{3j}^{0}	1	1	0	1	3($f_{3.}^{0}$)
	Total(n_{3j})	7	10	9	11	37($n_{3.}$)
	P_{3j}^{+}	0.286	0.500	0.333	0.364	0.378($P_{3.}^{+}$)
	P_{3j}^{-}	0.571	0.400	0.667	0.545	0.541($P_{3.}^{-}$)
	P_{3j}^{0}	0.143	0.100	0.000	0.091	0.081($P_{3.}^{0}$)
	$Q_{.j} = 4.222$					$Q_{3.} = 6.588$
Total	$f_{.j}^{+}$	12	15	11	23	61 ($f_{..}^{+} = f^{+}$)
	$f_{.j}^{-}$	12	16	12	12	52 ($f_{..}^{-} = f^{-}$)
	$f_{.j}^{0}$	3	3	4	2	12 ($f_{..}^{0} = f^{0}$)
	Total $n_{.j}$	27	34	27	37	125(n..=n)
	$P_{.j}^{+}$	0.444($P_{.1}^{+}$)	0.411($P_{.2}^{+}$)	0.407($P_{.3}^{+}$)	0.622($P_{.4}^{+}$)	0.488 ($P_{..}^{+} = P^{+}$)
	$P_{.j}^{-}$	0.444($P_{.1}^{-}$)	0.471($P_{.2}^{-}$)	0.444($P_{.3}^{-}$)	0.324($P_{.4}^{-}$)	0.416 ($P_{..}^{-} = P^{-}$)
	$P_{.j}^{0}$	0.112($P_{.1}^{0}$)	0.088($P_{.2}^{0}$)	0.148($P_{.3}^{0}$)	0.054($P_{.4}^{0}$)	0.096 ($P_{..}^{0} = P^{0}$)
	$Q_{.j} =$					$Q_{.j} = 8.545$

Table 3 shows the values of 1s,-1s,0s and the corresponding sample proportions by semester and academic year the students took the course. Assuming that there are no interactions between the semester and year the course is taken, we may apply Equation 16 to obtain the corresponding expected frequencies under the null hypothesis of Equations 13 and 14.These observed and expected frequencies are used to calculate chi-square values to test the null hypothesis of no differential effects as hypothesized. For example the null hypothesis of no differential effects between academic year and scores by students during the 'Fall' semester is tested using the chi-square test statistic of Equation 20 and the proportions in Table 3 corresponding to the 'Fall' semester as

$$Q_{1.} = \frac{\left(\left(P_{1.}^{-}(1-P_{1.}^{-})\right)\left(\sum_{j=1}^{4} n_{1j}P_{1j}^{+2} - n_{1.}P_{1.}^{+2}\right) + P_{1.}^{+}(1-P_{1.}^{+})\left(\sum_{j=1}^{4} n_{1j}P_{1j}^{-2} - n_{1.}P_{1.}^{-2}\right) + 2P_{1.}^{+}P_{1.}^{-}\left(\sum_{j=1}^{4} n_{1j}P_{1j}^{+}P_{1j}^{-} - n_{1.}P_{1.}^{+}P_{1.}^{-}\right)\right)}{P_{1.}^{+}P_{1.}^{-}\left(1-P_{1.}^{+}-P_{1.}^{-}\right)}$$

$$= \frac{((0.315)(0.687)(19.716-17.520)+(0.604)(0.396)(6.016-4.704)+2(0.604(0.313)(7.638-9.072))}{(0.604)(0.313(0.033)}$$

$$= \frac{0.472+0.314-0.542}{0.016} = \frac{0.244}{0.016} = 15.250$$

The values of $Q_{2.}$ *and* $Q_{3.}$ for winter and summer are similarly calculated. The values of $Q_{.j}$, $j = 1, 2, 3, 4,$

For each of the four academic years are similarly calculated using Equation 23.The results are shown in Table 3.Hence from Equation 21,we have that the test statistic for the null hypothesis of Equation 13,namely that the median scores by students are the same for all academic years, is

$$\chi^2 = \sum_{l=1}^{3} Q_{l.} = 15.250 + 12.458 + 6.588 = 34.296(P - value = 0.0000)$$

Which with 2(3)(4-1)=18 degrees of freedom is highly statistically significant showing that year course is taken has significant effect on student performance or scores.

Similarly from Equation 23 and Table 3 we have that the test statistic for the null hypothesis of Equation 14 is

$$\chi^2 = \sum_{j=1}^{4} Q_{.j} = 1.955 + 19.696 + 4.222 + 8.545 = 34.418(P - value = 0.0000)$$

Which with 2(4)(3-1)=16 degrees of freedom is highly statistically significant, leading to a rejection of the null hypothesis of equation of Equation 14. If no adjustments for ties are made then Equation 20 would revert to Equation 25 so that using the data of Table 3 the test statistic for the null hypothesis of Equation 13 now becomes

$$\chi^2 = \sum_{l=1}^{3} Q_{l.} = \frac{(19.716 - 17.520)}{(0.604)(0.396)} + \frac{(9.606 - 8.120)}{(0.450)(0.550)} + \frac{(5.525 - 5.291)}{(0.378)(0.622)}$$
$$= 9.188 + 5.992 + 0.996 = 16.176,$$

Which with 9 degrees of freedom is not statistically significant ($\chi^2_{0.95;9} = 16.919$), with the result that the null hypothesis of Equation 13 is now not rejected. Similarly from Equation 26 and Table 3, we have that the corresponding test statistic for the null hypothesis of Equation 14 is

$$\chi^2 = \sum_{j=1}^{4} Q_{.j} = \frac{(5.632 - 5.319)}{(0.444)(0.556)} + \frac{(7.082 - 5.746)}{(0.411)(0.589)} + \frac{(4.519 - 4.482)}{(0.407)(0.593)} + \frac{(16.534 - 14.319)}{(0.622)(0.378)}$$
$$= 1.267 + 5.521 + 0.154 + 9.426 = 16.368\ (P - value = 0.0393)$$

Which with 8 degrees of freedom is statistically significant leading to a rejection of the null hypothesis of Equation 14.These results however show, as expected that intrinsically and structurally ties adjusted test statistics are less likely to lead to an acceptance of a false null hypothesis (Type II Error) and hence are likely to be more powerful than ties unadjusted test statistic.

Summary and Conclusion

We have in this paper proposed and developed a non-parametric statistical method that may be used for determining the significance of factor effects in a two factor analysis of variance with unequal cell frequencies, that is with unequal replications per treatment combination, assuming that there are no interactions between the factors or that interactions have been removed by appropriate data transformation. The proposed test statistic has been intrinsically and structurally adjusted to provide for the possible presence of ties between observations in each cell or treatment combination. The sample populations may therefore be measurements on as low as the ordinal scale and are not required to be continuous or numeric. The proposed enables one test for the significance of not only each factor level effect but also for the equality of the several factor level effects.

The method is generally more robust than the corresponding classical 'F' test that may also be used for the same purpose, because the resulting sums of squares in the 'F' test are no longer additive and the error sums of squares is calculated as the weighted sum of the cell error sums of squares, an approach that is often tedious in practical applications. The proposed method has been illustrated with some sample data and shown to be as expected more powerful than the corresponding ties-unadjusted method.

Acknowledgement

I sincerely appreciate the departmental typist who assisted in typesetting this work.

Reference

1. Gibbons JD (1973): Non- Parametric Statistical. An Introduction; Newbury Park: Sage Publication
2. Montgomery and Pack, 1982
3. Neter and Wasserman, 1974
4. J. NETER, W. WASSERMAN, M. H. KUTNER (1983), applied linear regression models (Richard D. Irwin Inc, Illinois).
5. Oyeka I.C.A, Afuecheta E.O, Ebuh G.U. and Nnanatu C.C. (2012): Partitioning the Total Chi-square for Matched Dichotomous Data. International Journal of Mathematics and Computations (IJMC), ISBN0974-570X (online), ISSN 0974-5718 (Print), Vol.16; Issue No 3, pp 41-50.
6. Siegel S.(1956).Non-parametric Statistics for the Behavioural Sciences McGraw Hill Kogakusha Ltd,Tokoyo.