stats

MDPI

Article

Component Analysis When Testing for Fixed Effects in Unbalanced ANOVAs

J. C. W. Rayner [1,2,*] and G. C. Livingston, Jr. [2]

[1] National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW 2522, Australia
[2] School of Information and Physical Sciences, University of Newcastle, Newcastle, NSW 2308, Australia; glen.livingstonjr@newcastle.edu.au
* Correspondence: john.rayner@newcastle.edu.au

Abstract: In possibly unbalanced fixed effects in ANOVAs, we examine both parametric and nonparametric tests for main and two-way interaction effects when the levels of each factor may be ordered or unordered. For main effects, we decompose the factor sum of squares into one degree of freedom components involving contrasts, albeit not necessarily orthogonal contrasts. For interactions, we develop what we call coefficients. These are an extension of part of the interaction sum of squares in potentially unbalanced designs. They may be used to test nonparametrically for focused interaction effects. The tests developed here provide focused and objective assessments of main and interaction effects and augment traditional methods.

Keywords: degree-degree coefficients; exploratory data analysis; orthogonal polynomials; level-degree coefficients; level-level coefficients; pairwise comparisons; unbalanced designs

check for updates

Academic Editor: Wei Zhu

Received: 11 April 2025
Revised: 12 June 2025
Accepted: 13 June 2025
Published: 16 June 2025

Citation: Rayner, J.C.W.; Livingston, G.C., Jr. Component Analysis When Testing for Fixed Effects in Unbalanced ANOVAs. *Stats* **2025**, *8*, 48. https://doi.org/10.3390/stats8020048

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The tests developed here build on and extend the notion of contrasts. In undergraduate courses and texts for users of statistics, contrasts are often met as a means to better understand a significant main effect by making comparisons between the factor level means. Such texts include [1,2] (see, for example, p. 477), and [3] (see Section 12.8) a more extensive account is given in [4], Section 3.2. Although occasionally ordered levels are mentioned, most accounts focus on when the levels of the factor are unordered. See, for example [5].

The idea of orthogonal contrasts is to decompose a statistic, such as an ANOVA sum of squares, into components that may be used to detect alternatives in important and distinct subsets of the parameter space. The orthogonality ensures that these components are independent. However, we have found that most accounts do not consider unbalanced designs where there are different numbers of observations in the levels of the factors; they rarely give any detail of when the factor levels are ordered, nor do they discuss contrasts for interactions.

The tests proposed here seek to extend the benefits of contrasts by aiming to detect alternatives in different, important subsets of the parameter space. While we focus on two-factor models, extensions to multifactor ANOVA models are certainly possible. However, they are not considered here. We view the processes we propose as exploratory data analysis, with each component providing an input. Too many inputs would typically be too difficult to incorporate into an overall picture. We return to this issue in Section 7.

In the next section we derive orthonormal contrasts for main effect tests. Importantly, our definition of contrasts allows for unbalanced designs. Although almost part of statistical

folklore, these parametric tests are, we believe, unfamiliar to many users of statistics. When the levels of a factor are ordered, our contrast-based tests allow testing for polynomial effects in the factor levels, such as increasing means and umbrella effects. When the levels of a factor are not ordered, we recommend tests based on pairwise comparisons. These involve non-orthogonal comparisons and, we believe, give useful objective insights for data analysts.

When considering tests for interaction effects, we note that in unbalanced ANOVAs, tests for main and interaction effects use regression methods that will be unfamiliar to some data analysts. Instead, we observe that for balanced designs, an important part of the definition of the interaction sum of squares involves a quantity that we call a coefficient. The coefficient definition can readily be extended to unbalanced designs and used in focused nonparametric tests for aspects of the interaction effect. We give examples where the levels of neither, one, or two factors are ordered.

Most importantly, the nonparametric interaction tests developed here allow objective assessment of effects usually gleaned subjectively from interaction plots. We contend that both are important and provide useful insights in data analysis.

2. The Main Effects Test Statistics

The parametric model of interest here is a multifactor fixed-effects ANOVA, but as previously indicated, we focus only on two factors, A and B, say. In Section 3 the model assumes there is an AB interaction. Suppose Y_{ijk} is the kth of $n_{ij} > 1$ observations of the ith of r levels of factor A and the jth of c levels of factor B. The design is balanced if all n_{ij} are equal. All observations are mutually independent and normally distributed with constant variance σ^2.

We now focus on parametric tests for main effects. Although the design may have multiple factors, we focus on one only, say A. Suppose Y_{ij} is the jth of n_i observations of the ith of r levels of a factor A. There are $n = \sum_i n_i$ observations in all. A design is balanced if all n_i are equal. The completely randomized design is an example of a design that is in general unbalanced. The randomized block, balanced incomplete block, and Latin square designs are all balanced.

At this point the levels of A may or may not be ordered. Write Y_i for the sum of the observations for level i and $\overline{Y}_{i.}$ for the mean of the same.

Under the null hypothesis of no factor A effect, the unconditional factor A sum of squares, $SS_F = \sum_i n_i \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 = \sum_i Y_{i.}^2/n_i - Y_{..}^2/n$, has the $\sigma^2\chi^2_{r-1}$ distribution. As usual, the dot notation denotes summation. Note that unbalanced multifactor designs are not in general orthogonal. Thus, in R, the first factor called is unconditional, and the second is conditional on the first factor called. Subsequently, here we only work with the unconditional factor A sum of squares.

Put, for $i = 1, \ldots, t$, $p_i = n_i/n$ and $Z_i = (Y_{i.} - p_i Y_{..})$, and $Z = (Z_i)$ and $D = \mathrm{diag}(\sqrt{p_i})$. Suppose H is a $t \times t$ orthogonal matrix with columns $h_1, \ldots, h_t$ with $h_t = 1_t/\sqrt{t}$. Put $H = (h_1 \mid \ldots \mid h_t)$ so that $H\,H^T = h_1 h_1^T + \ldots + h_t h_t^T = I_t$. Then

$$n\,SS_F = Z^T\,D^{-2}\,Z = Z^T\,D^{-1}\,H\,H^T\,D^{-1}\,Z =$$

$$= Z^T D^{-1}(h_1 h_1^T + \ldots + h_t h_t^T) D^{-1} Z = \sum_{i=1}^{t} \left(h_i^T D^{-1} Z\right)^2.$$

We now give two constructions: first for the levels of factor A being ordered, and second, those levels being unordered.

If the levels of the factor are ordered, one choice for H is to first construct G with columns $g_1, \ldots, g_t$, with g_r the orthonormal polynomial of degree r on the weight function (p_i). The degree zero polynomial is taken to be identically one and is specified as the

tth: $g_t = 1_t$. Now as $\sum_i g_{ri}g_{si}p_i = \delta_{rs}$ (the Kronecker delta) if as above, D = diag($\sqrt{p_i}$), then $G^TD^2G = I_t$ and H = DG is an orthogonal matrix. If G has columns $g_1, \ldots, g_t$, and H has columns $h_1, \ldots, h_t$, then H = $(h_1 \mid \ldots \mid h_t)$ = D$(g_1 \mid \ldots \mid g_t)$ = (Dg_1 $\mid \ldots \mid$ Dg_t) and h_i = Dg_i.

Now, as above, n SS$_F$ = $\sum_{i=1}^{t} (h_i^T D^{-1} Z)^2$. However, for $i = 1, \ldots, t$ $h_i^T D^{-1} = g_i^T$ and $g_t^T Z = 1_t Z = \sum_i Z_i = 0$ so that n SS$_F$ = $\sum_{i=1}^{t-1} (g_i^T Z)^2$. (See [6] for a definition of contrast in unbalanced designs.) The $g_i^T Z$ are contrasts because, as just shown, $\sum_i Z_i = 0$ (the Z_i are centred) and $\sum_i g_{ri}p_i = E[g_r] = E[g_r g_t] = \delta_{rt} = 0$ for $r = 1, \ldots, t$: the contrast coefficients sum to zero. The $g_i^T Z$ decomposes SS$_F$ into $t - 1$ orthonormal contrasts that have the same interpretation as in the balanced case, for which again see [6].

Now suppose the levels of the factor are not ordered. We take $p_i = n_i/n$ for $i = 1, \ldots, t$, $h_t = (\sqrt{p_i})$, and D = diag($\sqrt{p_i}$). Then $h_t^T D^{-1} Z = \sum_{i=1}^{t} \sqrt{p_i}(Y_{i.} - p_i Y_{..})/\sqrt{p_i} = 0$ and $SS_F = n\sum_{i=1}^{t-1} \left(h_i^T D^{-1} Z\right)^2$, since, as before, the Z_i are centered random variables. Thus, the $h_i^T D^{-1} Z/\sqrt{n}$ decompose SS$_F$ into $t - 1$ orthonormal components.

One choice for H is to take

$$h_{11} = \sqrt{\frac{p_2}{p_1 + p_2}} \text{ and } h_{21} = -\sqrt{\frac{p_1}{p_1 + p_2}} \text{ and } h_{i1} = 0 \text{ for } i = 2, \ldots, t - 1.$$

Clearly $h_1^T D^{-1} Z$ gives a comparison of the first two levels of the factor. For it to be a contrast requires that $h_1^T D^{-1} 1_t = 0$: the potential contrast coefficients come to zero. Here $h_1^T D^{-1} 1_t = \frac{1}{\sqrt{p_1}}\sqrt{\frac{p_2}{p_1+p_2}} - \frac{1}{\sqrt{p_2}}\sqrt{\frac{p_1}{p_1+p_2}} \neq 0$ unless $p_1 = p_2$. Thus $h_1^T D^{-1} Z$ is not in general a contrast.

For $r = 2, \ldots, t - 1$, the rth column of H can be taken to have r ones and zeros thereafter, and the Gram–Schmidt orthogonalization process applied. This gives a decomposition of SS$_F$ into $t - 1$ orthonormal components, the first of which is $h_1^T D^{-1} Z/\sqrt{n}$.

However, it is more convenient to proceed as in the balanced case, discussed in [7]. We construct a matrix M with the last row $(\sqrt{p_j})$ and, for all $r < s$, rows with $\sqrt{(p_s/(p_r + p_s))}$ in the rth position, $-\sqrt{(p_r/(p_r + p_s))}$ in the sth and zeros elsewhere. As above, the rows of $MD^{-1}Z/\sqrt{n}$ are not, in general, contrasts. Moreover, such a matrix will not be orthogonal, and so the sum of the squared contrasts is not the treatment sum of squares. However, the elements of $MD^{-1}Z/\sqrt{n}$ give all the pairwise comparisons, and although these are not mutually orthonormal comparisons, their squares give the basis for an objective comparison between all pairs of the levels of the factor.

For parametric testing, whether the levels are ordered or not, we seek to test the null hypothesis that the expectation of the contrast/component is zero, usually against two-sided alternatives. This may be done parametrically by referring $(h_i^T D^{-1} Z)^2/(n\ ems)$ to the $F_{1,edf}$ distribution, where *ems* is the error mean square and *edf* are the error degrees of freedom. See [7], Section 2 for a proof. When the levels of the factor of interest are ordered and the orthonormal polynomials are used, a significant $(g_i^T Z)^2/(n\ ems)$ result suggests a degree i effect in the levels of the factor. In most applications it would be expected that linear and quadratic effects are of most interest. In nonparametric testing, we use permutation testing p-values based on the $(g_i^T Z)^2/(n\ ems)$ often comparing them with the corresponding parametric p-values based on the F distribution.

Steroid production example. The data in Table 1 are given in [4], p. 224, exercise 10. The response is steroid production per 100 mg of gland per hour for each of two treatments with the glands taken from rats at four different stages of growth. We assume the stages are ordered, but the treatments are not.

Table 1. Steroid production data with cell means in parentheses.

	Treatment		
Stage	**1**	**2**	**Means**
1	6.98, 6.58 (6.78)	8.62, 9.40, 9.20 (9.073)	8.156
2	6.07, 7.16, 6.34 (6.523)	9.42, 6.67, 8.64 (8.243)	7.383
3	5.38, 7.31, 6.65, 7.44 (6.695)	4.96, 6.8, 7.61 (6.457)	6.593
4	7.02, 9.23, 7.32 (7.857)	7.17, 7.65, 6.52, 6.86 (7.05)	7.398
Means	6.957	7.655	7.32

The design is not equally replicated and therefore not orthogonal. So different calls in R give different output. In R we are interested in the call that gives the stages sum of squares in the form assumed in this section:

anova(lm(Sterpro ~ Stage + Treatment + Stage: Treatment))

The ANOVA is given in Table 2. We note that the Shapiro–Wilk test for normality of the residuals had a p-value of 0.6287. There is no reason to doubt the parametric model.

Table 2. Analysis of variance table for the steroid production data.

	Degrees of Freedom	**Sums of Squares**	**Mean Squares**	**F Values**	***p*-Values**
Stage	3	7.2598	2.4199	2.7448	0.0751
Treatment	1	2.0454	2.0454	2.3200	0.1461
Stage: Treatment	3	9.9163	3.3054	3.7491	0.0311
Residuals	17	14.9880	0.8816		

The parametric p-values are 0.0751 for stages with orthonormal contrasts of degree one, two, and three, and 0.1274, 0.0442, and 0.3456, respectively.

We also calculated permutation test p-values: are they similar to the parametric p-values? Observed values of the F statistics for stages and their contrasts were calculated. Then residuals were calculated by removing treatment and interaction effects but not stage effects, since these are zero under the null hypothesis. See Appendix A to see how the residuals are defined. The residuals were then permuted, and for each permutation, the F statistics were calculated. The proportion that exceeded the observed values gave the required p-values.

With 1,000,000 permutations, we found p-values of 0.0755 for stages and 0.1275, 0.0445, and 0.3440 for the degree one, two, and three contrasts, respectively. These agree well with the parametric p-values.

Although the factor stage is not significant at the 0.05 level, its degree two contrast is. This is reflected in the cell mean plot in Figure 1. The treatment mean in the plot has a clear quadratic shape. We also note that initially treatment 1 is roughly constant, then increases at stage 4, while initially treatment 2 decreases, then increases at stage 4. The significant interaction manifests in that the two treatments are behaving differently; their plots are not parallel. This, however, is a subjective conclusion.

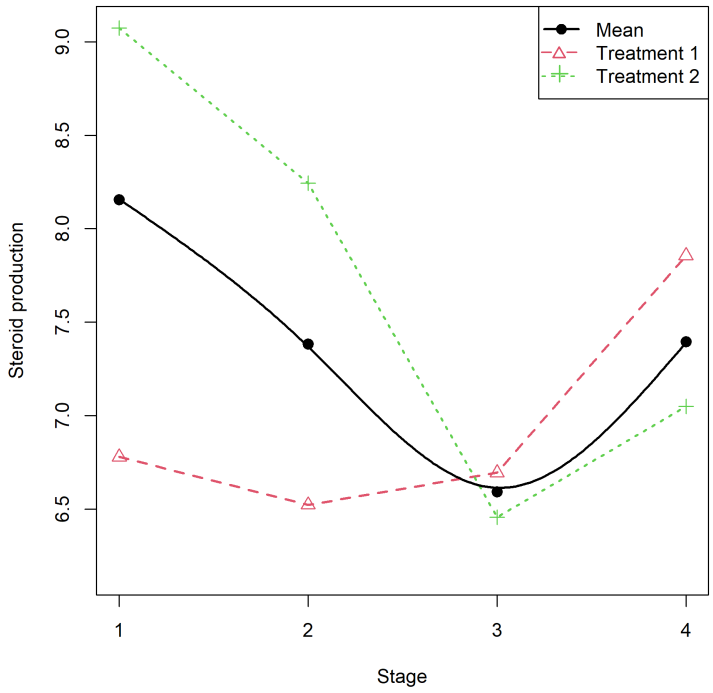


Figure 1. Cell mean plot for two treatments and their mean.

Even when the parametric model is in serious doubt, the parametric *p*-values may still be substantially correct, as the following example shows.

Biomass example. The data in Table 3 are modified from data that at the time of writing could be found at [8]. The purpose of the modification was to achieve an unbalanced two-way fixed effects ANOVA, and to that end six randomly selected responses were removed from the original data set.

Table 3. Biomass yield for four levels of fertilizer and four levels of irrigation.

	Fertilizer			
Irrigation	**Control**	**100 lb**	**150 lb**	**200 lb**
A	2700, 2801, 2720, 2390, 2890	3250, 3151, 3170, 3290	3235, 3025, 3165, 3120	3500, 3455, 3100, 3600, 3250
B	3101, 3035, 3205, 3007	2700, 2935, 2495, 2850	3050, 3110, 3033, 3195, 4250	3100, 3235, 3005, 3095, 3050
C	101, 97, 106, 142, 99	400, 302, 296, 315, 390	630, 624, 595, 595	400, 325, 200, 375, 390
D	121, 174, 88, 100, 76	100, 125, 91, 222, 219	60, 28, 112, 89, 67	201, 223, 120, 180

Output for the parametric ANOVA, both with fertilizer preceding irrigation and with irrigation preceding fertilizer, is given in Tables 4 and 5. The conclusions for both analyses are similar: that both main effects and interaction are all significant at levels less than 0.001. However, for both analyses, the Shapiro–Wilk test for normality of the ANOVA residuals returned a *p*-value of less than 2.618×10^{-10}. Thus the conclusions of the parametric analyses are not valid. For comparison purposes, we will proceed to analyze these data both parametrically and nonparametrically.

Table 4. Analysis of variance table for the biomass yield data, fertilizer preceding irrigation.

	Degrees of Freedom	Sums of Squares	Mean Squares	F Values	*p*-Values
Fertilizer	3	2,519,574	839,858	29.3628	1.121×10^{-11}
Irrigation	3	150,067,311	50,022,437	1748.8681	$<2.2 \times 10^{-16}$
Fertilizer: irrigation	9	1,609,794	178,866	6.2535	3.985×10^{-6}
Residuals	58	1,658,960	28,603		

Table 5. Analysis of variance table for the biomass yield data, fertilizer succeeding irrigation.

	Degrees of Freedom	Sums of Squares	Mean Squares	F Values	*p*-Values
Irrigation	3	151,595,613	50,531,871	1766.6788	$<2.2 \times 10^{-16}$
Fertilizer	3	991,272	330,424	11.5522	4.868×10^{-6}
Irrigation: fertilizer	9	1,609,794	178,866	6.2535	3.985×10^{-6}
Residuals	58	1,658,960	28,603		

Subjective conclusions can be drawn from the cell mean plots in Figure 2. The left hand panel plots yield against fertilizer, while the right hand panel plots yield against irrigation.

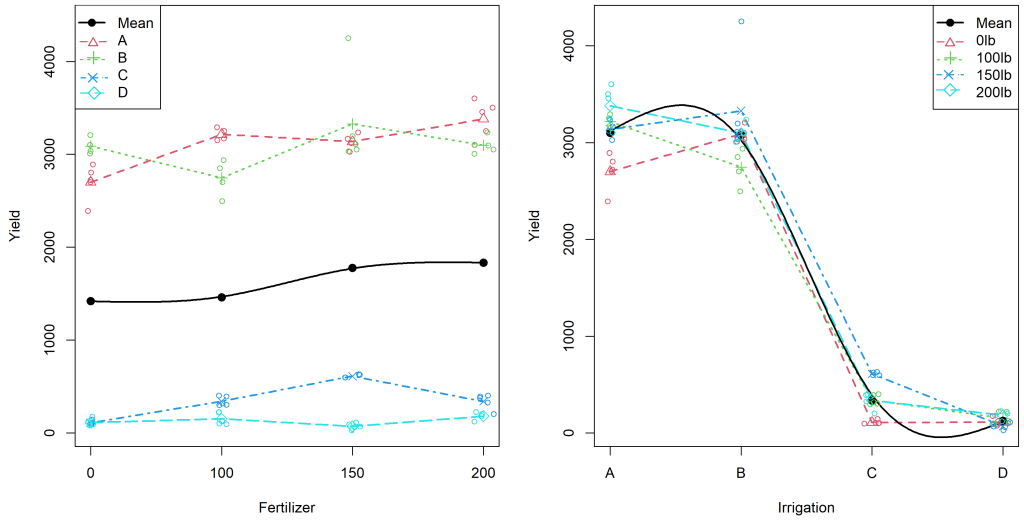


Figure 2. Cell mean plots of yield against irrigation in the left panel and against fertilizer in the right panel.

The levels of the fertilizer factor are clearly ordered. We find a strong parametric fertilizer main effect with contrast *p*-values of less than 0.0001, 0.8756, and 0.0040 for linear, quadratic, and cubic effects, respectively. There is strong evidence of both linear and cubic effects, but not of quadratic effects.

We assume the irrigation effects are ordered, from A to B to C and then to D, perhaps by quantity of water released. There is evidence of a strong parametric irrigation effect, with a *p*-value less than 0.0001. The parametric irrigation contrast *p*-values are less than 0.0001, 0.0965, and less than 0.0001 for linear, quadratic, and cubic effects, respectively. There is strong evidence of both linear and cubic effects, with only weak evidence, at the 0.1 level but not at the 0.05 level of a quadratic effect.

If we take the irrigation levels to not be ordered, pairwise comparisons are appropriate. The irrigation means are 3100.6667, 3080.6111, 335.8947, and 126.1053 for A, B, C, and D,

respectively. All pairwise parametric p-values are less than 0.0001 except for A and B, with a p-value of 0.7233, and C and D with a p-value of 0.0003. These are consistent with the left panel cell mean plot of yield against fertilizer in Figure 2.

Using 1,000,000 permutations, p-values were calculated for fertilizer main effects. Residuals were calculated by removing the irrigation and interaction effects but not the fertilizer effects, since these are zero under the null hypothesis. We find linear, quadratic, and cubic effect p-values for the contrasts of less than 0.0001, 0.8786, and 0.0044, respectively. The agreement with the parametric p-values is, given the rejection of the parametric model, surprisingly good.

For irrigation, we considered levels both ordered and not. For ordered levels and 1,000,000 permutations, the overall p-value was less than 0.0001, while for contrasts, they were less than 0.0001 for linear effects, 0.0960 for quadaratic effects, and less than 0.0001 for cubic effects. For levels not ordered and 1,000,000 permutations, the overall p-value was less than 0.0001, while all pairwise comparisons returned p-values less than 0.0001 except for comparing A and B, with a p-value of 0.7226, and comparing C and D, with a p-value of 0.0004.

For clearer comparison of the parametric and nonparametric p-values, these are given in Tables 6 and 7.

Table 6. p-values for testing for fertilizer and irrigation main effects and their contrasts, both parametrically and nonparametrically. Both fertilizer and irrigation have ordered levels.

		Overall	**Linear**	**Quadratic**	**Cubic**
Fertilizer	Parametric	<0.0001	<0.0001	0.8756	0.0040
	Nonparametric	<0.0001	<0.0001	0.8786	0.0044
Irrigation	Parametric	<0.0001	<0.0001	0.0965	<0.0001
	Nonparametric	<0.0001	<0.0001	0.0960	<0.0001

Table 7. p-values for testing for fertilizer and irrigation main effects and their contrasts, both parametrically and nonparametrically. The levels of irrigation are not ordered.

	Overall	**A, B**	**C, D**	**All Others**
Parametric	<0.0001	0.7233	0.0003	<0.0001
Nonparametric	<0.0001	0.7226	0.0004	<0.0001

Even if the parametric model is seriously compromised, the parametric p-values may be indicative. However, to avoid invalid inference, the nonparametric p-values should always be calculated.

3. The Interaction Effect Test Statistics

The parametric model of interest here is a multifactor fixed effects ANOVA, but as previously indicated, we focus only on two factors, A and B, say. The model assumes there is an AB interaction. Suppose Y_{ijk} is the kth of $n_{ij} > 1$ observations of the ith of r levels of factor A and the jth of c levels of factor B. The design is balanced if all n_{ij} are equal. All observations are mutually independent and normally distributed with constant variance σ^2. Write $n_i = \sum_j n_{ij}$, $n = \sum_i n_i$, $\overline{Y}_{ij.}$ for the mean of the observations for level i of factor A and level j of factor B, $\overline{Y}_{i..}$ for the mean of the observations for level i of factor A, $\overline{Y}_{.j.}$ for the mean of the observations for level j of factor B, and $\overline{Y}_{...}$ for the mean of all of the observations.

We now focus on the balanced parametric model with $n_{ij} = s$ for all i and j. The interaction sum of squares is

$$SS_{\text{AB}} = s\sum_i \sum_j \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2.$$

Note that $\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$ is the (i, j)th aligned cell mean. Alignment is a tool sometimes used to strip main effects from the observations. See, for example [9], Section 9.4. Put

$$Z_{\text{I}} = (Z_{11}, \ldots, Z_{1c}, Z_{21}, \ldots, Z_{2c}, \ldots, Z_{r1}, \ldots, Z_{rc})^{\text{T}}$$

in which $Z_{ij} = \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)\sqrt{s}$.

It was shown in [6] that

$$SS_{\text{AB}} = \sum_{t=1}^{(r-1)(c-1)} \left(m_t^{\text{T}} Z_{\text{I}}\right)^2$$

in which $m_1, \ldots, m_{(r-1)(c-1)}$ are the eigenvectors corresponding to the eigenvalues 1 of an idempotent matrix with rank $(r-1)(c-1)$ and t indexes the cells. The m_{ts} are not uniquely defined; it is only required that they are mutually orthonormal and orthogonal to 1_{rc}. A useful approach is to first suppose A is any $r \times r$ idempotent matrix of rank $r-1$ and B is any $c \times c$ idempotent matrix of rank $c-1$. Then A has $r-1$ eigenvalues of one and one eigenvalue of zero, and B has $c-1$ eigenvalues one and one eigenvalue zero. It follows that if $\otimes$ represents the Kronecker product, then A $\otimes$ B is idempotent and has $(r-1)(c-1)$ eigenvalues of one and $r+c-1$ eignvalues of zero. In particular suppose the eigenvectors of A corresponding to the eigenvalue one are $a_1, \ldots, a_{r-1}$ and the eigenvectors of B corresponding to the eigenvalue one are $b_1, \ldots, b_{c-1}$. The $a_u \otimes b_v$, $u = 1, \ldots, r-1$ and $v = 1, \ldots, c-1$, are an appropriate choice for the eigenvectors of A $\otimes$ B corresponding to the eigenvalue 1 and are mutually orthonormal.

Take $m_t = a_u \otimes b_v$, where t indexes (u, v), $u = 1, \ldots, r-1$, and $v = 1, \ldots, c-1$. Now define Z to be the $r \times c$ matrix (Z_{ij}), quite distinct from but obviously related to Z_{I}. Then

$$m_t^{\text{T}} Z_{\text{I}} = (a_u \otimes b_v)^{\text{T}} Z_{\text{I}} = \sum_e \sum_f a_{ue} b_{vf} Z_{ef} = a_u^{\text{T}} Z b_v.$$

Parametric tests for testing $\text{E}\left[a_u^{\text{T}} Z b_v\right] = 0$ against its two-sided negation can be based on $F_t = (a_u^{\text{T}} Z b_v)^2 / \{SS_{\text{E}}/edf\}$, which has the $\text{F}_{1,edf}$ distribution. The expression here for F_t seems to be the most intuitively appealing form of the contrast test statistic.

A coherent interpretation of $a_u^{\text{T}} Z b_v$ is that it is the projection of the aligned cell means into the parameter space spanned by $a_u \otimes b_v$, or, alternatively, it gives the degree (u, v) interaction effect. We call them $a_u^{\text{T}} Z b_v$ *coefficients* and distinguish the degree-degree, level-degree, and level-level coefficients depending on whether the levels of the factors are ordered or not.

Whatever is assumed about the ordering of the levels of the factors, a test for an overall interaction effect can be based on the sum of the F_t.

For unbalanced designs with $n_{ij} \neq s$ some i and j, we instead define $Z_{ij} = \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)\sqrt{n_{ij}}$ and proceed as in the balanced case, except that now the distribution of F_t is not clear, so instead we use permutation testing to obtain p-values.

4. Crop Yield Example

The data set in Table 8 was available at the time of writing at [10], where it is analyzed using regression techniques. No scenario is given, but it is apparent that three crops are subjected to two fertilizers and the yields noted. All other influences are assumed to have been randomized. Clearly the levels of neither factor are ordered.

Table 8. Crop yield data.

Fertilizer	**Crop**		
	Corn	Soy	Rice
Blend X	128	166	151
	150	178	125
	174	187	117
	116	153	155
	109		158
Blend Y	175	140	167
	132	145	183
	120	159	142
	187	131	167
	184		

The data are analyzed in R that uses type 1 sums of squares in which the order of the factors in the call is important. Interestingly, the *p*-values resulting from calling both crop before fertilizer and fertilizer before crop are the same, although there are slight differences in some of the test statistics. The crop called before fertilizer ANOVA is given in Table 9.

Table 9. Analysis of variance table for the crop yield data, fertilizer succeeding crop.

	Degrees of Freedom	**Sums of Squares**	**Mean Squares**	**F Values**	***p*-Values**
Crop	2	433.7	216.85	0.4515	0.6427
Fertilizer	1	506.2	506.20	1.0539	0.3163
Crop: fertilizer	2	3675.5	1837.74	3.8261	0.0383
Residuals	21	10,086.7	480.32		

From the output, neither main effect is significant at even the 0.1 level, with a crop *p*-value of 0.6427 and a fertilizer *p*-value of 0.3163. The interaction *p*-value of 0.0383, significant at the 0.05 level, while the Shapiro–Wilk normality test returns a *p*-value of 0.6226. There is no apparent reason to doubt the parametric *p*-values.

Based on 1,000,000 permutations and permuting the original data, the fertilizer *p*-value is 0.3128, while for the crop main effect the *p*-value is 0.6370 with a component *p*-value of 0.3493 comparing corn and soy, 0.6800 comparing corn and rice, and 0.5947 comparing soy and rice. These compare with the corresponding parametric component *p*-values of 0.3530, 0.6832, and 0.5976. Thus there is good agreement between the parametric and nonparametric *p*-values. No significant pairwise differences are found, as might be expected with such a large fertilizer *p*-value. Note that the crop *p*-value is calculated based on the crop F statistic in the parametric model.

Figure 3 gives cell mean plots of these data. In the left panel the fertilizer lines are clearly not parallel, but that is a subjective judgement. The right panel gives the plot without the interference of the main effects. The parametric analysis suggests there is weak evidence of an interaction, but it would be useful to have focused component tests to see whether there is an effect that is masked by non-significant effects.

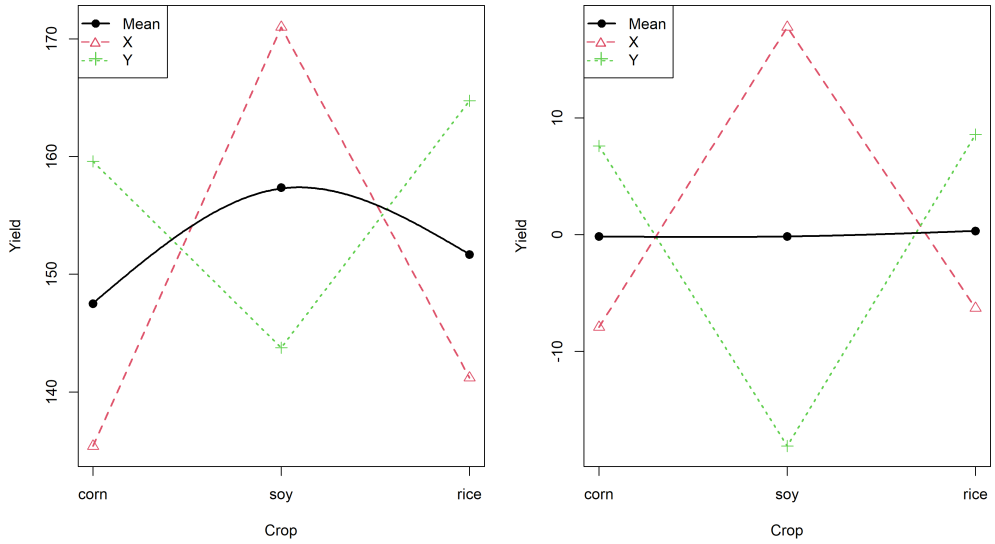


Figure 3. Cell mean plots for the crop yield data. Raw data in the left panel, aligned in the right panel.

If we align by removing the main effects, permutation testing for level-level effects based on 1,000,000 permutations gave p-values of 0.0276, 0.9695, and 0.0254 for pairwise effects reflecting effects due to soy and rice, corn and rice, and corn and soy, respectively. The corresponding p-values without aligning, just permuting the raw data, were 0.0281, 0.9697, and 0.0252 for the pairwise effects: very similar. This is not surprising, since the main effects were relatively small.

Apparently corn and rice do not contribute to the interaction effect, but both soy and rice, and finally corn and soy, do. This is reflected in the Figure 3 plots if the crops are visualized two at a time. Subjectively it is clear that soy is responding to the fertilizer blends in the opposite manner to corn and rice.

5. Steroid Production Example Continued

In Section 2 we introduced the steroid production data. The factor stages had a parametric p-value of 0.0751 with p-values of 0.1274, 0.0442, and 0.3456 for contrasts of degree one, two, and three, respectively. The interaction had p-value of 0.0311. It is obviously of interest to examine the interaction to see what is causing the significance. Given the Shapiro–Wilk test p-value of 0.6287, there is no reason to doubt the parametric model.

Here we calculate permutation test p-values of the degree-level coefficients.

Based on 1,000,000 permutations of the aligned (by removing the main effects) data, the resulting p-values for the coefficient tests are 0.0049, 0.9607, and 0.4033 for comparisons between treatment 1 and treatment 2 and of degrees one, two, and three, respectively.

Based on 1,000,000 permutations of the raw data the resulting p-values for the coefficient tests are 0.0048, 0.9604 and 0.4033 for comparisons between treatment 1 and treatment 2 and of degree one, two and three respectively. These p-values are very similar to those in the preceding paragraph.

Of the contrasts for the main effect stages, only that of degree two is significant at the 0.05 level. Coefficient testing suggests that the contrast between the two treatments has a strong linear effect. Subjectively, the latter is apparent in Figure 1 but more apparent in Figure 4.

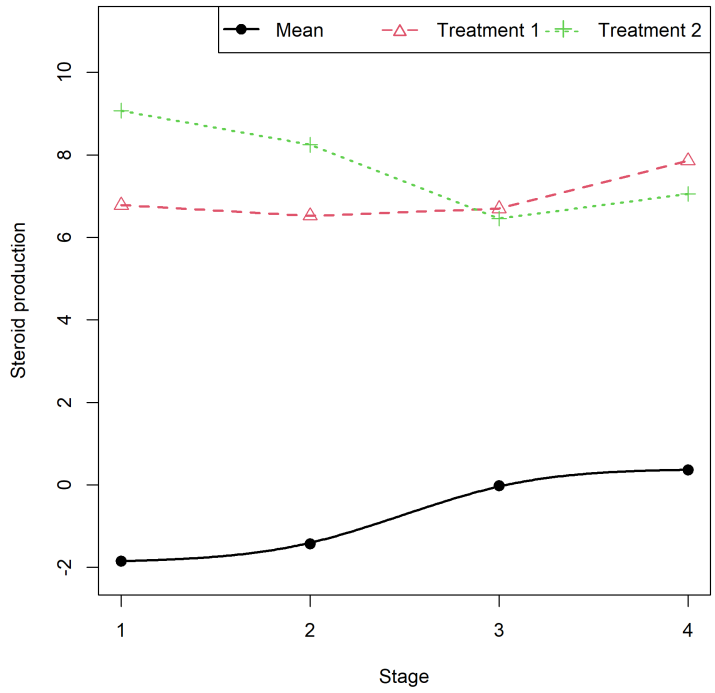


Figure 4. Plot for the steroid production data showing steroid production for the two treatments and the contrast between them.

6. Biomass Example Continued

We return now to the biomass example introduced previously at the end of Section 2, where the main effects were discussed.

Figure 2 contains cell mean plots, which also show jittered values of the yield. All lines appear to be roughly parallel, subjectively suggesting there are minimal interaction effects. Of course the parametric analysis indicated a strong interaction effect.

As before, we assume the irrigation levels are ordered from A to D, possibly by decreasing water volume. The levels of fertilizer are clearly ordered.

Based on 1,000,000 permutations of the data aligned by removing both fertilizer and irrigation effects, we find

- 0.0011, 0.4862, and 0.8814 for irrigation effects of degree one and fertilizer effects of degrees one, two, and three, respectively,
- 0.3813, 0.3951, less than 0.0001 for irrigation effects of degree two and fertilizer effects of degrees one, two, and three, respectively,
- 0.0727, 0.0037, and 0.0130 for irrigation effects of degree three and fertilizer effects of degrees one, two, and three, respectively.

Although it is not easy to interpret some of these coefficient tests, there is evidence of a linear-linear effect: with increasing fertilizer and increasing water (irrigation level from D to A), the biomass yield increases. There are other higher-order degree-degree effects present, which at first might appear at odds with the observation that the lines in Figure 2 are all roughly parallel. With the jittered yield values shown in Figure 2, we can observe that the within-cell variation is relatively small, which results in a small error mean square. As a result, we are able to detect very small interaction effects.

7. Conclusions

We noted in the introduction that contrasts "detect alternatives in different, important subsets of the parameter space". We developed contrasts for main effects in unbalanced

ANOVA designs and showed how to test for polynomial effects when the factor levels are ordered and for pairwise effects when they are not. Although here we focus on these choices, we note that other choices are possible. For interaction effects, a quantity in the interaction sum of squares in balanced designs was generalized to unbalanced designs and used to test for 'important subsets of the parameter space'. Our preferred test statistics are based on contrasts and components with F distributions when the design is balanced, but for interaction effects in unbalanced designs, we use permutation testing.

The main advantages of our approach are that we are able to analyze

- unbalanced designs,
- interactions and
- nonparametrically when the parametric assumptions are dubious.

In particular, Ref. [11] noted that "Since the data obtained in many areas of psychological inquiry are . . . frequently unbalanced, researchers using the conventional procedure will erroneously claim treatment effects when none are present . . .". We hope our treatment of unbalanced designs will appeal to these researchers.

In two-factor designs in which the levels of both factors are ordered, interpreting coefficient effects can be problematic. While what we might call a degree (1, 2) effect is an umbrella effect, it is not clear how to interpret, for example, a degree (2, 3) effect, beyond indicating a complex interaction effect.

As noted in the introduction we have not developed coefficient tests for multiway interactions. Many, as just noted, would not be interpretable. Moreover, when testing at the 0.05 level, 5% of the effects would be significant even when there was no interaction effect. Even when a significant effect is interpretable, it may be spurious, one of the 5%. In two-way ANOVAs the interaction degrees of freedom are usually much fewer, and a single significant interaction effect may be evident from the cell mean plot or discounted from the same and thus identified as a complex and unimportant effect.

In the examples we have examined in which both the parametric and nonparametric analyses are available, we have found that the parametric and nonparametric p-values for a particular effect are almost always similar. This suggests that the nonparametric tests are not noticeably inferior to their parametric competitors. However, even an indicative power study would be a sizable undertaking, and we leave that for another time.

In unbalanced ANOVAs, tests for main and interaction effects use regression methods. The interaction tests that we propose are routine extensions of tests for balanced ANOVAs, and we believe they will be especially appealing to data analysts familiar only with balanced ANOVA testing.

Author Contributions: Conceptualization, J.C.W.R.; methodology, J.C.W.R. and G.C.L.J.; software, G.C.L.J.; validation, J.C.W.R. and G.C.L.J.; formal analysis, J.C.W.R. and G.C.L.J.; investigation, J.C.W.R. and G.C.L.J.; resources, J.C.W.R. and G.C.L.J.; data curation, G.C.L.J.; writing—original draft preparation, J.C.W.R.; writing—review and editing, J.C.W.R. and G.C.L.J.; visualization, J.C.W.R. and G.C.L.J.; supervision, J.C.W.R.; project administration, J.C.W.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Estimating the Parameters in the Unbalanced Two-Factor ANOVA with Replications

Assume the model specifies

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}$$

with $i = 1, \ldots, r$, $j = 1, \ldots, c$, and $k = 1, \ldots, n_{ij,}$ and constraints forthcoming. Summing over cells, columns, rows, and then both rows and columns gives

$$Y_{ij.} = \mu n_{ij} + \alpha_i n_{ij} + \beta_j n_{ij} + (\alpha\beta)_{ij} n_{ij},$$

$$Y_{i..} = \mu n_{i.} + \alpha_i n_{i.} + \textstyle\sum_j \beta_j n_{ij} + \sum_j (\alpha\beta)_{ij} n_{ij},$$

$$Y_{.j.} = \mu n_{.j} + \textstyle\sum_i \alpha_i n_{ij} + \beta_j n_{.j} + \sum_i (\alpha\beta)_{ij} n_{ij} \text{ and}$$

$$Y_{...} = \mu n_{...} + \textstyle\sum_i \alpha_i n_{i.} + \sum_j \beta_j n_{.j} + \sum_{i,j} (\alpha\beta)_{ij} n_{ij}.$$

We assume the constraints

$$\textstyle\sum_i \alpha_i n_{ij} = \sum_j \beta_j n_{ij} = \sum_i (\alpha\beta)_{ij} n_{ij} = \sum_j (\alpha\beta)_{ij} n_{ij} = 0.$$

Then, with $n = n_{...}$ and $\overline{Y} = \sum_{i,j,k} Y_{ijk}/n$, we estimate

- μ by $\overline{Y}$,
- α_i by $\overline{Y}_{i..} - \overline{Y}$,
- β_j by $\overline{Y}_{.j.} - \overline{Y}$, and
- $(\alpha\beta)_{ij}$ by $\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}$.

Under the full model, the residual $E_{ijk} = Y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}$ is estimated by

$$Y_{ijk} - \overline{Y} - \left(\overline{Y}_{i..} - \overline{Y}\right) - \left(\overline{Y}_{.j.} - \overline{Y}\right) - \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}\right) = Y_{ijk} - \overline{Y}_{ij.}$$

If there are no α_i in the model, as, for example, under the null hypothesis, then the residual is $E_{ijk} = Y_{ijk} - \mu - \beta_j - (\alpha\beta)_{ij}$ which may be estimated by

$$Y_{ijk} - \overline{Y} - \left(\overline{Y}_{.j.} - \overline{Y}\right) - \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}\right) = Y_{ijk} - \overline{Y}_{ij.} + \overline{Y}_{i..} - \overline{Y}.$$

This is the residual used in the permutation tests for main effects.

We note that the effects are being estimated only. When we say an effect has been 'removed', it is more accurate to say 'reduced'. A reduced effect may still be significant.

References

1. Box, G.E.; Hunter, W.G.; Hunter, J.S. *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; Wiley: New York, NY, USA, 2005.
2. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 3rd ed.; Wiley: New York, NY, USA, 1998. [CrossRef]
3. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, 8th ed.; Iowa State University Press: Ames Iowa, IA, USA, 1989.
4. Kuehl, R.O. *Design of Experiments: Statistical Principles of Research Design and Analysis*; Duxbury Press: Pacific Grove, CA, USA, 2000.
5. Available online: https://en.wikipedia.org/wiki/Contrast_(statistics)#cite_note-9 (accessed on 15 May 2025).
6. Rayner, J.C.W.; Livingston, G.C., Jr. Orthogonal contrasts for both balanced and unbalanced designs and both ordered and unordered treatments. *Stat. Neerl.* **2024**, *78*, 68–78. [CrossRef]
7. Rayner, J.C.W.; Livingston, G.C., Jr. Orthonormal F contrasts for factors with ordered levels in two-factor fixed effects ANOVAs. *Stats* **2023**, *6*, 920–930. [CrossRef]
8. Available online: https://milnepublishing.geneseo.edu/natural-resources-biometrics/chapter/chapter-6-two-way-analysis-of-variance/ (accessed on 15 April 2025).

9. Rayner, J.C.W.; Livingston, G.C., Jr. *Introduction to Cochran–Mantel–Haenszel Testing and Nonparametric ANOVA*; Wiley: Chichester, UK, 2023; ISBN 978-1-119-83200-3.
10. Available online: https://real-statistics.com/multiple-regression/unbalanced-factorial-anova/ (accessed on 15 May 2025).
11. Keselman, H.J.; Algina, J.; Kowalchuk, R.K. Analysis of repeated measures designs: A review. *Br. J. Math. Stat. Psychol.* **2001**, *54*, 1–20. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.