I don’t like interacting, but it’s necessary: a brief overview of factorial experiments in animal science

Animal experiments, experiment planning, experimental statistic.
.

Edenio Detmann

Animal Scientist, D.Sc., Professor of Ruminant Nutrition, Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Uppsala, Sweden (edenio.detmann@slu.se)

Nutri·Time

Revista Eletrônica

Vol. 19, Nº 05, set/out de 2022
ISSN: 1983-9006
www.nutritime.com.br

A Nutritime Revista Eletrônica é uma publicação bimestral da Nutritime Ltda. Com o objetivo de divulgar revisões de literatura, artigos técnicos e científicos bem como resultados de pesquisa nas áreas de Ciência Animal, através do endereço eletrônico: http://www.nutritime.com.br.

ABSTRACT

The factorial experiments are useful in animal science, as they allow us to evaluate how different causative factors affect directly the animal responses and, mostly important, how the factors can interfere with each other on that responses. The animal production is naturally an interactive process. Therefore, interactions must be an important aspect to be considered when we look at a better understanding of nutritional and metabolic aspects of animal production. In this overview, I present the most basic aspects of factorial experiments applied do animal science and some ways to avoid the most common mistakes that can be made when carrying out this kind of experiment.

Keywords: animal experiments, experiment planning, experimental statistic.

EU NÃO GOSTO DE INTERAGIR, MAS É NECESSÁRIO: UM BREVE VISÃO SOBRE EXPERIMENTOS FATORIAIS EM ZOOTECNIA

RESUMO

Os experimentos fatoriais são úteis em zootecnia, pois permitem avaliar como diferentes fatores causadores afetam diretamente as respostas dos animais e, principalmente, como os fatores podem interferir entre si nessas respostas. A produção animal é naturalmente um processo interativo. Portanto, as interações devem ser um aspecto importante a ser considerado quando buscamos uma melhor compreensão dos aspectos nutricionais e metabólicos da produção animal. Neste artigo eu apresento os aspectos básicos dos experimentos fatoriais aplicados à zootecnia e também algumas maneiras de se evitar os erros mais comuns que podem ser cometidos ao se realizar esse tipo de experimento.

Palavras-chave: estatística experimental, experimentos com animais, planejamento experimental.

In my opinion, one of most underrated tales in the Bible is the story of Jonah. I use the adjective underrated because many people highlight this tale just to state how unlikely is to live inside a fish stomach. I am writing this article neither to discuss religion nor to debate about the reliability of the Bible. Nevertheless, I think there is something different and useful that we can learn from the Jonah's "adventure". I intend going beyond the so-called Jonah's "fish situation".

According to that story, God decided to send Jonah for preaching His word in Nineveh. However, Jonah did not like that order. He did not tell that to God, but we can easily understand it by reading between the lines. I do not know exactly why, but Jonah decided to go in the opposite direction. Maybe he did not like working or he just wanted to avoid getting in touch with the crowd. Anyway, Jonah took a ship to go to Tarshish. A huge storm came over them, and Jonah's shipmates decided to throw him overboard (he was the outsider and understood to be the most probable cause of the climate's fury). After that, the storm was no longer raged, but according to the story, a big fish swallowed Jonah and kept him inside its stomach for three days before dropped him out in a beach near to Nineveh. Hence, Jonah did his work. Wheter you believe the Bible or not, there is a take-home message here inside this story. Jonah would not like to interact with people, but it was necessary. He could not run away his responsibilities, as we must not ignore our own duties. I must confess that Jonah and I had similar stories (in my case, a history). I did not like interacting with people. That is why I decided to study animal science back in the 1990s. In my opinion, cows were much more interesting than people were. However, something analogous to a big fish swallowed me, threw me out next to a classroom, and that is how I became a professor. I did not like to interact, but it became into a necessary task for me. Life has strange paths for everyone and, believe me, your paths have many interactions.

As I wrote before, I am not writing this text to talk about living inside a fish or about your personal responsibilities. Indeed, I would like to talk about something less personal and associated with animal experiments: the factorial arrangements and the so-called "controversial" interactions. I say "controversial" because I have heard many bad things about interactions. Actually, it seems that interactions in animal experiments are a kind of curse for experimenters. A significant interaction seems to represent an unavoidable damnation and the final day of weeping and cries of sorrow. Is that overview of interactions correct? Is that what underlies the study of interaction effects in our experiments? Suffering and crying? I am pretty sure that is not correct. I refuse to think that Sir Ronald Fisher and other giants who established the groundwork for factorial arrangements had had an intention to make people suffer. It was completely the opposite of that. Interactions are extremely useful in animal science experiments, once the experimenter can understand their meaning and usefulness.

Nature is essentially interactive. This is factual. No living being would exist without interacting with the environment and other living beings. Nothing happens in Nature without an interaction. When we look inside a living being, we can see the same pattern, as metabolism is an interactive mechanism at all. However, what does an interaction mean in biology or animal science? I will give you the simplest answer I can. Every single event that occurs within an animal depends on a series of other events and factors. Nothing simply happens by itself. Let's move to some examples. I will not explore the mechanisms in their totality, as it would be boring, long, and out of our scope here. Then, I will try to highlight just a few pinpoints in a way that we can get it easily:

1. Pepsin is secreted in the stomach (or abomasum in ruminants) as pepsinogen, its zymogenic form. Pepsinogen itself is unable to digest protein (it must be like that, otherwise pepsin could digest the cells that produce it). Pepsinogen needs to be cleaved by hydrochloric acid to become an active enzyme. Thus, the amount of pepsin formed in the stomach depends on the presence and concentration of hydrochloric acid;

2. Supplements for grazing beef cattle are different during the dry and rainy seasons. The main reason for that is that the nutritional

characteristics of pasture vary according to annual climatic oscillation. Thus, the animal's response to the supplements depends on forage quality; and

3. Rumen microorganisms are able to synthetize protein from ammonia nitrogen. However, the effectivity of this process depends on the availability of adequate carbon skeletons and energy.

I gave you three quite simple examples. Maybe you wonder what interactions have to do with them, as I have not used the word "interaction" any moment. If you did that, you are right. I did not use that word. However, I did that purposely, as you could get it easily. Indeed, I used the same expression in every example above, which is **depends on**. That is the essence of an interactive event. How a given factor works **depends on** the presence (or concentration, etc.) of another factor. When our best explanation of how a biological event occurs starts with **"it depends on"**, we are facing an interaction. The positive response of grass growth to nitrogen fertilization do occur. We already knew that. However, the way in which this occurs **depends on** the species, soil type, phosphorus and potassium fertilization, etc. Using a correct protein content in cattle diets is essential, but it **depends on** animal category, genetic group, sex, physiological state, dietary energy, etc.

Interactions are the rules, rather than exceptions, for all biological systems, which also include animal production at any level (i.e., management, bioclimatology, nutrition, reproduction, metabolism, etc.). Technicians and producers work surrounded by interactions all the time. However, scientists in particular must be aware of interactions when they decide to study the effects of something on animal production. This is one of the pillars of the scientific method, which can be named as "control of variables". When a study on a specific factor is about to be performed, a scientist must assure that all other factors remain constant for each experimental unit. Otherwise, unwanted interactions would occur, which could compromise the experimental comparison and the reliability of the experimental responses. For example, if one wants to measure the effects of different dietary protein contents for steers, the only "thing" that must vary across experimental units is just the dietary content of protein[1]. If, for instance, the different protein contents were implemented using different protein sources, the variation among treatments would no longer represent exclusively the dietary protein content. There would be a confounding effect between protein content and source, two factors that may interact with each other. The experimental responses would not allow a valid conclusion on the effect of protein contents and the experiment would be invalid as well.

Experiments in which the effect of a single factor is evaluated are useful at some extent. However, as we saw earlier, a factor involved in a biological event does not act alone to produce a response and this is just the main limitation of that kind of experiment. A dietary protein content considered optimal for bulls cannot be the best for heifers. Then, an experiment carried out with bulls cannot be completely useful to understand how the animal's response to dietary protein would occur if the sex of the animals was different. A broader answer for that could be obtained by performing more than one experiment, one with each sex. However, there will always be a random and non-observable effect between experiments. It is unavoidable and makes our comparisons between different experiments more difficult to be built with confidence.

A valid alternative is to include one or more factors in the same experiment. In this case, if we can assure the adequate control of variables, our inferences can be valid and, more importantly, we may create a way of measuring how the factors interact with each other to define the experimental response. Hence, by adopting an adequate background, we define a way to study more than one factor in the same experiment using a **factorial arrangement**.

Unfortunately, I have constantly heard an incorrect and uncomfortable expression when some scientists perform experiments like that. Those people used

[1] Of course, there will be uncontrolled variations within the experiment, which are due to causes that cannot be anticipated by the experimenter. The variations that do not have a known cause are represented by residual variation, i.e. the "rest" of experimental variation that could not be controlled or assigned to a known cause.

say that they have a "factorial design". With respect, but that is a complete mistake. There is a huge difference between arrangement and design in experimental statistics. The term design defines the way we assign the treatments to the experimental units. Of course, the basis for that must be the randomization. However, randomization can be applied with or without any restriction and this is the main point used to differentiate among the several types of designs. When randomization is applied without any restriction, the experiment is carried out according to a completely randomized design (CRD). Any simple restriction in randomization changes the experimental design. Simply put, that is how randomized block designs or Latin squares become different from CRD. However, randomization is out of our scope here. If you want to read more about that, you may take a look at my previous published book (Detmann, 2018).

Now, we know what an experimental design is. What about the arrangement? In a simple sense, arrangement is a way of organizing the treatments that are about to be evaluated in the experiment, in order to study the effects of different factors and their interactions. You may easily see how different the two terms are and how they connect to each other. The arrangement helps us to define the treatments. Following that, the design will tell us how to assign those treatments to the experimental units. The two concepts define different stages of the experimental process. Thus, they are completely different and cannot be confounded. Our experiments mandatorily follow a design, but they can or cannot be performed following an arrangement.

Using a factorial arrangement means that we intend to study the effects of two or more factors in the same experiment. Thereafter, we must follow its theoretical background in order to define the treatments in the best way so we can understand how the factors affect the experimental response independently or, perhaps, interacting with each other. However, there are more than one type of factorial arrangement. Here, we will discuss only about the main and most common type, the complete factorial arrangement. Hereafter, when I say factorial, you must keep in mind that I am talking about the complete form of this arrangement. Two points must be also highlighted before we move on. First, many of the points I will discuss here could be also applied to other types of arrangements, such as the split-plot arrangements. However, split plots have a rather limited applicability for animal experiments, besides having many crucial differences when compared to factorials. Thus, what I present here is completely valid for factorials. Second, my discussion here would be completely valid if, and only if, all the factors encompassed by the factorial arrangement are fixed effects. I will not discuss factorials where one or more factors are random effects. In general, these types of arrangements are rare in animal production, yet common in feed analysis.

When we intend to study more than one factor in an experiment, a complete factorial arrangement teaches us that the combinations of all levels of the different factors must form the treatments. I know it can sound strange at first glance. However, believe me, this advice is completely worthy and help us to set up good experiments, and, most importantly, it avoids many mistakes and gives us straightforward tools to interpret experimental responses.

It seems better to go through a hypothetical example in order to assimilate the definition that I presented in the previous paragraph. Let us suppose that you intend to study the effects of the crude protein (CP) content in supplements offered to heifers under grazing. However, you have some previous and strong evidences that the protein source in the concentrate can influence how the variation in CP content affects animals' performance. Thus, this is an appropriate time to use a factorial arrangement. Previously, you had planned to study supplements with 20, 25, and 30% CP. Now, you want to check the effects of those levels, but by varying the protein source in the supplement: soybean meal or cottonseed meal.

The first step in establishing a factorial arrangement is to define which are the factors to be studied. In our example, it is clear that we have two factors: the protein content and the protein source. After that, the second step comes by defining how many and which are the levels of each factor. In the case of protein content, we have three levels, which correspond to the contents in the supplements: 20, 25, and 30%. On the other hand, the protein source factor has two

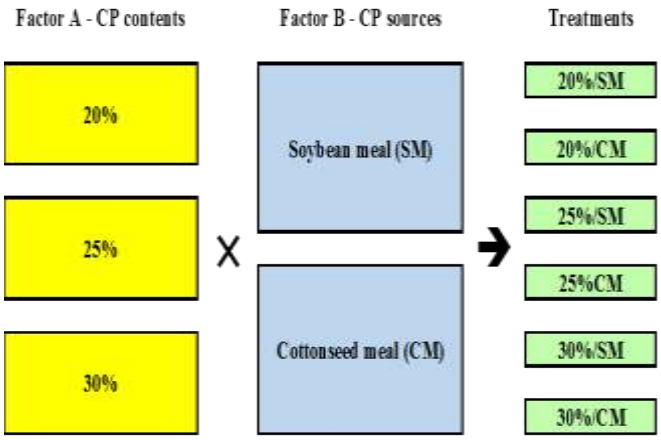
levels: soybean meal and cottonseed meal. From this information, we have the elements to name our arrangement, which is a 3 × 2 factorial arrangement. The first factor (CP content) has three levels and the second factor (CP source) has two levels. It does not matter which factor comes first. Thus, if you call this a 2 × 3 factorial arrangement, nothing will change (i.e., putting source before content). A complete factorial arrangement is always symmetric and the order of the factors does not change the product (i.e., data pattern and interpretation).

After defining the factorial arrangement, it is time to apply what it teaches us: how to organize the treatments that are about to be evaluated. This way reflects our planning and must be in accordance with what we defined in the previous paragraph. Let us assume that Factor A is indexed by "i" (i = 1, 2, …, I) and has I levels. In turn, factor B is indexed by "j" (j = 1, 2, …, J) and has J levels. Then, the number of treatments in a complete factorial arrangement will be I × J. According to the previous example, our experiment must encompass 3 × 2 or six treatments.

Now, we already knew the number of treatments. The final step of this process is to define what the six treatments are. In a complete factorial arrangement, treatments are produced by combining all the levels of each factor. Using our example, we must combine all CP contents and sources in all possible ways, which results in six combinations (i.e., 3 × 2). Every single combination corresponds to one treatment that must be applied to the experimental units (Figure 1). Now, you can proceed to the randomization and perform the experiment.

After the experiment is done, it's time to analyse the results. If you planned your experiment according to the factorial arrangement, it is expected that you will use the same reasoning when analysing your data. Leaving the factorial background behind when you are evaluating the results is not the correct thing to do. You applied a powerful tool to plan and perform your experiment aiming at concluding in a more functional way. The most rational decision here is to keep moving forward using the background that the factorial arrangement has given you.

FIGURE 1. Schematic representation of how to define treatments in a 3 × 2 factorial arrangement supplements for grazing heifers containing different crude protein contents and sources (see details in the text).



Actually, the analysis through a complete factorial arrangement represents an orthogonal decomposition of the treatment degrees of freedom and sum of squares towards the individual effects of factors and their interactions (Table 1). When we decide to do so, the overall hypotheses testing for the "treatment" effects is no longer necessary. Indeed, using it can lead us to some misinterpretations in data analysis and conclusions.

Following the orthogonal decomposition, we become able to obtain more functional information when compared to an overall test for treatment effects. Many important questions can be answered about the animal response pattern caused by each individual factor and, most importantly, if the different factors mutually influence their response pattern or not.

Using our example (Table 1), we can see that the variation caused by treatments is now orthogonally decomposed into three different sources of variation and, consequently, each one must be associated with a different hypotheses test. However, what are the hypotheses to be tested here? That is a very interesting question and a corrected answer for that can define how well your data interpretation will be.

Defining null and alternative hypotheses is always a concern for me, as different references may bring different approaches that reflect different interpretations by the authors. My intention here is not to cause conflict or overwhelm you with theoretical statistics. Therefore, I decided to go for a

TABELA 1. Partitioning the treatment degrees of freedom following an A × B factorial arrangement

General scheme		Example of grazing heifers	
Sources of variation	Degrees of freedom	Sources of variation	Degrees of freedom
[Treatments][a]	[(I×J – 1)]	[Treatments]	[2 × 3 -1 = 5]
Factor A	I – 1	CP contents	3 -1 = 2
Factor B	J – 1	CP sources	2 – 1 = 1
Interaction A × B	(I-1) × (J-1)	Contents × Sources	(3 -1) × (2 -1) = 2

[a] It easy to show that the partitioning of treatment degrees of freedom is orthogonal, as (I-1) + (J-1) + [(I -1) × (J – 1)] = (I × J – 1). In the example, 2 + 1 + 2 = 5. The same reasoning is applied to the decomposition of treatment sum of squares.

more didactic approach without using any mathematical notation. I will only refer to the null hypothesis and leave implicit the alternative hypothesis (i.e., there is no agreement with what is stated by the null hypothesis). Using a hypothetical A × B factorial arrangement, the null hypotheses tested in the analysis of variance (ANOVA, Table 1) should be:

Factor A – H_0: there is no difference in the responses caused by the different levels of A, if A is independent of B;
Factor B – H_0: there is no difference in responses caused by the different levels of B, if B is independent of A; and
Interaction A × B – H_0: responses to A are independent of B and responses to B are independent of A (there is a symmetry here, as I stated before).

I purposely used a textual form for the null hypotheses, as the meaning behind each one can be understood as clear as possible. The test for factors A and B will only propitiate a correct decision if, and only if, A and B are independent of each other. Who does allow us to check that condition? The answer for that is clear: the hypotheses test for interaction. Our first conclusion must be: the first thing to look out after running ANOVA is the test for interaction. It will guide us on how to proceed in data interpretation.

Let's analyse the first scenario where interaction effect is not significant (i.e., null hypothesis is accepted). Then, we would have empirical evidences that factors A and B do not interfere with each other in the experimental responses and discussions and conclusions can be built separately for each factor. Actually, the own conclusion drawn from the interaction test can be a worthy and useful conclusion. However, the most important issue here is that we are now allowed to study the factors separately. Then, we can look at the results of the other hypotheses tests, as the conditional established in each null hypotheses was assured by the interaction test.

The second scenario is completely different. If the null hypothesis for the interaction effect was rejected, the conditional of independency between factors was not met. Then, the individual tests cannot be performed (please, read the null hypotheses again). If the experimental effects of the factors are dependent on each other, how could we test hypotheses whose prerequisite relies on independency? The answer to that is clear and direct: we could not. Without fulfilling the conditional, the hypotheses for the factors are meaningless. Would you like to conclude something from your study based on a meaningless hypothesis? I am pretty sure you would not want to do that.

There are two main take-home messages here: interactions provide a very useful information on how to interpret our experiment, and show us how to analyse the experiment correctly. Regarding the first message, let's go back to our heifers' example.

A non-significant interaction would mean that we could verify the best CP content in the supplement without any connection with the protein source. In other words, the animal response pattern according to the CP contents in the supplements would not be affect if the animals are fed on a soybean- or cottonseed-based supplement. On the other hand, if the interaction would be found significant, our first conclusion would be the opposite of that. The animal response pattern to the supplement CP content would vary according to the protein source we feed to the animals. Think about it for a while. Is this kind of interpretation useful or not from a biological point of view? My answer for that is the information we can get from interaction is completely useful. Thus, we got a first and important reason to not hate interactions.

Understanding the second take-home message I highlighted in the previous paragraph takes a little more of effort on our part. When we analyse any hypothetical A × B factorial arrangement, there will be eight probable scenarios for the ANOVA results[2], which are exemplified in Table 2. Scenarios I to IV occur under a non-significant interaction. Therefore, it is allowed to use individual tests for each factor and each scenario will follow a different way of explaining the pattern of experimental responses. Conversely, the scenarios V to VIII will have strictly the same conclusion. Responses to A depends on B and vice-versa. As we saw earlier, the null hypotheses for individual factors require mutual independency, which is not supported by the empirical evidence. Hence, the results for individual factor tests mean nothing and the significance indicators for them are merely a standard way of showing results in a paper. The message here is that we must seek a deeper evaluation of results, considering the mutual influence of one factor on the other.

TABLE 2. Simulation of different scenarios for the results of the hypotheses test in an analysis of variance for a A × B factorial arrangement

	Scenarios[a]							
Sources of variation	I	II	III	IV	V	VI	VI	VIII
Factor A	ns	*	ns	*	ns	*	ns	*
Factor B	ns	ns	*	*	ns	ns	*	*
Interaction A × B	ns	ns	ns	ns	*	*	*	*

[a] ns, non-significant; *, significant.

Before proceeding with the discussion of interactions, I will present the two usual ways of showing the means obtained from a factorial arrangement (Table 3). The first way is to show the individual treatment means. I think I do not need to explain much about how to calculate a treatment mean. The second way is to show the so-called marginal means. Marginal means of one factor are calculated across all the levels of the other factor. Using our recurrent example, the marginal mean for the animal response to soybean meal would be calculated using the information of all CP contents in the supplements. The marginal mean for the response to cottonseed meal would be obtained in the same way. It is clear that the marginal means come from a larger number of observations and have smaller standard errors compared to the treatment means. However, these characteristics must not weigh in our decision about which sort of mean value we should show in our papers. There is more here and some details can lead to major misinterpretations.

TABLE 3. Schematic representation of different types of mean values in a theoretical 3 × 2 factorial arrangement (balanced design)

	Levels of Factor A[a]			
Levels of Factor B	1	2	3	Marginal means of B
1	$m_{11}(n)$	$m_{21}(n)$	$m_{31}(n)$	$M_{.1}(3n)$
2	$m_{12}(n)$	$m_{22}(n)$	$m_{32}(n)$	$M_{.2}(3n)$
Marginal means of A	$M_{1.}(2n)$	$M_{2.}(2n)$	$M_{3.}(2n)$	-

[a] Treatment means are indicated by "m" and marginal means by "M". "n" represents the number of replicates per treatment. Consequently, the information in parenthesis represents the number of experimental units used to calculate each mean value.

At first glance, marginal means are seductive, as they summarize what has happened across the levels of a same factor. You could take a look at them and quickly understand what happened when you fed cottonseed meal or soybean meal. However,

[2] I am following here what is defined by the Neyman-Pearson lemma. It means that a significance threshold is defined (e.g., P<0.05) and only two decisions are allowed: accept or reject the null hypothesis.

we must keep in mind that every time we compact data to improve the understandability of numbers, we also loose information. A treatment mean is already a compacted value, as we summarize information from all replicates in only one number. It is better to understand the average pattern of the response, but we must also understand that a mean value does not have the same level of information compared to the whole sample. However, it is necessary; otherwise, our tables would be like the old phone books and getting useful information from them would be very unlikely. On the other hand, a marginal mean is always obtained by a double data compaction. Besides the loss of information on sample variation, we also loose the information, for instance, on how the responses to soybean vary according to the CP content in the supplements.

To illustrate that, I will use as an example some results taken from the experiment of Souza et al. (2010), who studied the effects of nitrogen and starch supplementation on heifers fed a low-quality tropical forage. This is a typical 2 × 2 factorial arrangement, where the factors were nitrogen supplementation and starch supplementation and each one had only two levels (with or without). I express the results using both treatment means and marginal means (Table 4). It must be noticed that the experiment was planned to use 0.10 as the significance threshold.

First, let's evaluate the results of dry matter intake. Following the logical sequence, there was no interaction between nitrogen and starch on intake (P 0.29, = Table 4). Thus, we are allowed to check the individual hypotheses test. Starch supplementation did not affected intake (P=0.22), but nitrogen supplementation did (P<0.01). Looking at the treatment means, it is quite easy to see that intake increased when nitrogen was supplemented. You can get the same impression by looking at both treatment means and marginal means. One may wonder if the same insight can be gotten from both, why not use marginal means. It seems easier to explain things using them. However, things are not always like this. The microbial nitrogen production showed an interaction between nitrogen and starch supplementation (P=0.068). Thus, we are not allowed to use the individual hypotheses tests. We will go further into this example later. However, now, when we look at the treatment means, it is very easy to see why the interaction occurred. Giving only nitrogen or starch to the animals was not enough to alter the microbial synthesis in the rumen. However, when nitrogen was supplied along with starch, the microbial production increased more than 50%. The response to nitrogen supplementation depended on the presence of starch. It is a classic interaction. Now, try to get the same perception from marginal means. The results of hypotheses testing are the same Did you get it? I hope you did not, because

TABLE 4. Dry matter intake (DMI, kd/d) and microbial nitrogen production in the rumen (NMIC, g/d) in heifers fed a low-quality tropical forage and supplemented with nitrogen and, or starch (adapted from Souza et al., 2010)

Scenario I – Using treatment means							
	No nitrogen		Nitrogen		P values[a]		
Item	No starch	Starch	No starch	starch	N	S	N×S
DMI	3.46	3.49	4.48	4.94	0.001	0.22	0.29
NMIC	40.7	40.6	41.1	67.0	0.063	0.070	0.068
Scenario II – Using marginal means							
	Nitrogen		Starch		P values		
Item	without	with	without	with	N	S	N×S
DMI	3.48	4.71	3.97	4.22	0.001	0.22	0.29
NMIC	40.7	54.1	40.9	53.8	0.063	0.070	0.068

[a] N, S, and N × S, effects of supplemental nitrogen, starch, and their interaction, respectively. The experiment was planned to declare significance at P<0.10.

this information is not there. We are not able to perceive the interaction between nitrogen and starch because this information was lost in the double compaction behind the calculation of marginal means.

There is no rule that forbids showing marginal means. However, we are working with science and official regulations cannot be the best guidelines for us. We must show the results as clear as possible, as the reader should be able to understand the “message”, judge the results and, perhaps, apply the knowledge latter. In this sense, marginal means represent a lot of information loss. They are attractive, but less informative. If the experiment is balanced, you are able to calculate the marginal means from the treatment means. However, you cannot calculate treatment means from marginal means (unless the interaction effect is numerically equal to zero, which is almost impossible). Think about that: the greater the quality of information you show, the greater the quality of science you do. How would you choose to present your data? I already have my answer to this question. What about you?

After this brief pit stop, let’s get back to discussing interactions. We already know that a significant interaction denotes that individual hypotheses are not useful. Then, how to proceed with data analysis? First, I would like to make it clear that I will not discuss about methods to compare treatment means. It is a very extensive subject and each case is a different case. What I am about to discuss is a general guide, which should be followed by an adequate method of comparing treatments.

First, as the factors are dependent on each other, our study must take that into account. To do that, we must proceed to a nested evaluation of factors. It means that we must study the response pattern of one factor within each level of the other factor. Why? Because the responses of one factor will be dependent on which level of the other factor was applied to the experimental units. It seems complicated at first glance, but it is not. Data analysis software can easily do this today, as long as you know how to ask for that properly. Normally, these nested analyses are called “slicing” in the statistical packages. Studying one effect nested within another would be similar to when you slice a cake. Nice analogy.

Actually, there are two ways to perform such analysis (or two ways to slice the data). I will named these as one-handed and double-handed slicing procedure. In the one-handed slicing procedure, you choose which way would be more informative for you and the readers: studying A nested within B or B nested within A. Let’s go back to our heifers’ example. We can study either what happens to the response when we vary the CP content within each protein source or what happens when we change the protein source within each CP content. Which one is more informative? I cannot tell you that. It is up to you and how you want to explore your results in a more informative way. Particularly, if I was responsible for that experiment, I would probably choose to study the variation in CP contents within each protein source. Anyway, both ways are licit and, in some terms, have the same statistics background. What we do is reorganizing the degrees of freedom and sums of squares in order to perform the slicing procedure. You can see that in Table 5. Both forms of one-handed slicing procedure are orthogonal and use exactly the treatment degrees of freedom.

We can see an example of the one-handed slicing in the experiment performed by Palma et al. (2015), who studied different acid digestion methods for mineral analysis in different materials used in animal trials (Table 6). The experiment is bigger than what I show here. I just pinched a small piece, where they found a significant interaction ($P<0.01$) between the acid ratios used in the digestion procedure and the material that was digested on the calcium contents. The authors chose to study the variation in the acid ratios nested within each material, a typical one-handed slicing procedure. Why did they choose to study methods nested within materials and not materials within methods? I can tell you because I was one of the authors. Because we already knew that materials were different and the comparison among methods nested within each material would be much more informative according to our objective, which was to find the “best” method for acid digestion. As so, for most materials, there was no effect of different acid ratios. However, for bones, the lower the ratio, the higher the calcium content. The more rigid the matrix to be digested, the greater the amount of perchloric acid necessary to do the job. You can get two messages here. First, the slicing is not that difficult to perform and understand. Second, this is one more classic example of interaction. The effect of acid ratio depended on the material that has been digested.

TABLE 5. Schematic representation of different ways of partitioning the degrees of freedom when studying interaction effects

Source of variation	Degrees of freedom	Total degrees of freedom[b]
Factorial analysis		
Contents	2	
Sources	1	5
Contents × Sources	2	
Study of contents nested within sources[a]		
Sources	1	
Contents/Soybean meal	2	5
Contents/Cottonseed meal	2	
Study of sources nested within contents[a]		
Contents	2	
Sources/20%	1	5
Sources/25%	1	
Sources/30%	1	
Double-handed study of interaction[a]		
Contents/Soybean meal	2	
Contents/Cottonseed meal	2	
Sources/20%	1	7
Sources/25%	1	
Sources/30%	1	

[a] The symbol "/" must be read as "nested within" or simply "within". Thus, "contents/soybean meal" means studying the variation among CP contents when soybean meal is the supplemental protein source. Conversely, "sources/20%" means studying the difference between soybean meal and cottonseed meal when the supplement has 20% CP.
[b] Number of degrees on freedom used to compare treatments.

The double-handed slicing procedure is quite simple to understand. Unlike you choosing which way you will perform the one-handed slicing, you do both and use them to interpret the data. Again, there is no law against doing this. The only concern is that the rearrangement of degrees of freedom and sums of squares is no longer orthogonal. As you can see in Table 5, the double-handed slicing uses more degrees of freedom than the available number of degrees of freedom for the comparison among treatments. It can be a problem, as a non-orthogonal partition of treatment sum of squares can open some additional doors for the occurrence of type I error. As so, the probability of you wrongly point out a difference may increase.

TABLE 6. Study of interaction between digested material and nitric acid to perchloric acid ratio on the calcium content (g/kg dry matter) (adapted from Palma et al., 2015)

	Nitric to perchloric acid ratio			
Material	2:1	3:1	4:1	P value
Carcass	48.3	50.2	48.5	0.85
Bones	183.9a	172.8b	163.4b	<0.001
Excreta	10.5	9.82	9.85	0.97
Concentrates	1.40	1.32	1.37	>0.99
Grasses	6.47	5.81	6.34	0.98
Faeces	5.97	5.58	5.35	0.98

[a] Means in the same row followed by different letters differ at P<0.01.

I took an example of double-handed slicing from the study by Barbosa et al. (2017). Those authors evaluated methods to analyse the ether extract content in forages and cattle faeces (Table 7). Once more, it must be noted that the study is broader than what I have presented here. They found a significant interaction (P<0.01) between method and material. When they studied the materials nested within methods, they did not see any difference between materials regarding the ether extract content (P≥0.35). On the other hand, when comparing the methods within each material, they found that method B produced higher ether extract content for both forages and faeces (P<0.01); however, the difference between methods was greater when forages were evaluated and that was why interaction occurred.

TABLE 7. Study of interaction between method and material on the ether extract content (g/kg dry matter) (adapted from Barbosa et al., 2017)

	Material		
Method	Forages	Cattle faeces	P value
A	20.9	26.2	0.35
B	34.7	32.3	0.67
P value	<0.001	0.001	-

Some mistakes can be made when using factorial arrangements. I will try to discuss something about the four main ones, which I named the "ignorance mistake", the "donation mistake", the "neglect mistake", and the "discrimination mistake".

The ignorance mistake occurs when the researcher finds a significant interaction, but decide to ignore that and study the factors individually. I can tell you that this is quite dangerous for the conclusions. As we saw before, tests for individual factors require independency between them, but a significant interaction says this condition has not been met. I performed a simulation of the ignorance mistake based on the results of Souza et al. (2010). Previously, we found that there was an interaction between supplemental nitrogen and starch on rumen microbial production. The one-handed slicing of this effect based on treatment means shows clearly that improvements in microbial production were caused by nitrogen supplementation, but only when starch was concomitantly supplied (Table 8, at the centre). That was the authors' conclusion. However, let's suppose that the authors completely ignore the interaction and proceed to study nitrogen and starch separately using the marginal means. Surprisingly (or not), the conclusions will change. The study of marginal means (Table 8, at the borders) shows that both starch and nitrogen increased microbial production in the rumen regardless of whether they were fed together or not. This is obviously a complete mistake, and the conclusions drawn from this would be false as well.

TABLE 8. Study of the interaction between supplemental nitrogen and starch on microbial nitrogen production in the rumen (g/d) in heifers fed a low-quality tropical forage (adapted from Souza et al., 2010. See text for more details)

	Nitrogen[a, b]		
Starch	Without	With	Marginal starch
Without	40.7a	41.1a	40.9B
With	40.6b	67.0a	53.8A
Marginal nitrogen	40.7B	54.1A	-

[a] Treatments means in the same row followed by different lower case letters differ at P<0.10.
[b] Marginal means in the same column or row followed by different capital letters differ at P<0.10.

The donation mistake occurs when the experimenter finds a non-significant interaction, decide for omitting this of the model/analysis, and "donate" its degrees of freedom to the residual. The intention here would be to improve the precision of the experiment by increasing the residual degrees of freedom. I may surely state that this is a mistake. First, you may find interactions in many types of analysis, and each one has a different theoretical background. Thus, you must understand that I am referring here to interactions in complete factorial arrangements. For other types of interactions and analyses, what I present here cannot be completely valid. Considering this, the first point we must highlight is that the interaction is part of the treatment effects (Table 1). If you decide to "donate" the interaction to the residual, you are assuming that an important part of treatment effects has no known cause. Does it make sense? Interactions as a part of treatment effects can show many useful things. They can even help us to build more elaborate conclusions. The reader must know about the interactions and their "message". The second point is that the treatments are the basis for applying randomization. Remember that experimental design is the way how the treatments are designated to the experimental units, which is based on randomization; then, treatments are intimately connected to how randomization is applied and their degrees of freedom cannot be considered as residual degrees of freedom. Thus, the interaction, whether significant or not, must remain in the model and analysis. If you need mode residual degrees of freedom, do so during the experiment planning and increase the number of experimental units. Donating part of the treatment degrees of freedom must not be the way to do that.

The neglect mistake occur when you plan and perform the experiment according a factorial arrangement, but decide to neglect this when analysing the results. This is similar to buying a car, but going for a trip by walking because you refuse to put gasoline in your car. You have a powerful tool, but neglect to use that. Why? I have no simple explanation for that. I prefer showing you an example and leave the conclusion at your discretion. The data in Table 9 was obtained from an experiment performed by a dear friend of mine. To preserve the ethical aspects of this paper, I will not reveal the name or any details of the experiment. I just reproduced the numbers. My friend authorized me to use the data after we had a nice conversation about the way he/she analysed the results. Please, take a look at Table 9. Despite the factorial arrangement, the authors applied a multiple comparison procedure across all the treatments. Try to extract any useful information about factor A, factor B, or their interaction. Did you get it? Maybe it is possible, but quite hard. Probably, the comparisons bring much more confusion than information.

TABLE 9. Comparisons among treatment means in an experiment carried out following a 3 × 4 factorial arrangement (please, refer to the text for more details)

Levels of Factor A	Levels of Factor B	Treatment means
I	I	238f
	II	235f
	III	238f
	IV	238f
II	I	298d
	II	291e
	III	292e
	IV	295de
III	I	357a
	II	350bc
	III	347c
	IV	354ab

[a] Means in the column followed by different letters differ at P<0.05.

Last, but not least, the discrimination mistake occurs when an author decides to use different significance thresholds for the tests on individual factors and interaction. I do not know exactly why, but this mistake has become very common nowadays. Those who "opt" for that normally say something like this: "the significance for the factors were declared at P<0.05 and for interaction at P<0.15". Refuting this "option" is quite simple and I do not need to put much effort on it. The factors and their interaction are part of the same effect, which is the treatment effect (Table 1). If you adopt this "advice", it means that you are judging different parts of the same effect using different standards of thresholds. Does it make sense? This is a kind of discrimination regarding the different parts of the same effect. If the effect is unique, so must be the judgment on this.

Finally, I hope this paper had been helpful to you. Of course, it would be impossible to cover all aspects of factorial arrangements in a single article. What I have tried to do here is to present a more pragmatic overview of the main characteristics of complete factorial arrangements and to point out how we can get useful information from them. However, keep in mind that I did not say you must use factorial arrangements. What I have said is they are useful and should be used when they might help you to test your hypothesis. In this case, they will be a strong tool for you to try to achieve your objectives. You may not like to interact. However, interactions surround us and must be studied and understood. This is necessary for making good science.

ACKNOWLEDGMENTS

I would like to thank Dr. Marcia Franco (LUKE, Finland) and Luiz Carlos de Oliveira Sousa (Universidade Federal de Viçosa, Brazil) for the criticisms and suggestions during the manuscript preparation.

REFERÊNCIAS

Barbosa, M.M.; Detmann, E.; Valadares Filho, S.C.; Detmann, K.S.C.; Franco, M.O.; Batista, E.D.; Rocha, G.C. Evaluation of methods for the quantification of ether extract contents in forage and cattle feces. **Anais da Academia Brasileira de Ciências**, v.89, p.1295-1303, 2017.

Detmann, E. **Não seja como as vaquinhas! Uma abordagem informal sobre formalidades dos experimentos com animais de produção**. 2nd ed. Viçosa: Edenio Detmann, 2018. 373p.

Palma, M.N.N.; Rocha, G.C.; Valadares Filho, S.C.; Detmann, E. Evaluation of acid digestion procedures to estimate mineral contents in materials from animal trials. **Asian-Australasian Journal of Animal Science**, v.28, p.1624-1628, 2015.

Souza, M.A.; Detmann, E.; Paulino, M.F.; Sampaio, C.B.; Lazzarini, I.; Valadares Fillho, S.C. Intake, digestibility and rumen dynamics of neutral detergent fibre in cattle fed low-quality tropical forage and supplemented with nitrogen and/or starch. **Tropical Animal Health and Production**, v.42, p.1299-1310, 2010.