Turkish Journal of Veterinary and Animal Sciences

http://journals.tubitak.gov.tr/veterinary/

Research Article

Turk J Vet Anim Sci (2022) 46: 535-542
© TÜBİTAK
doi: 10.55730/1300-0128.4224

A biological perspective on interpreting interaction effect

Serdar GENÇ[1,*], **Mehmet MENDEŞ**[2]

[1]Department of Agricultural Biotechnology, Faculty of Agriculture, Kırşehir Ahi Evran University, Kırşehir, Turkey
[2]Faculty of Agriculture, Biometry and Genetics Unit, Çanakkale Onsekiz Mart University, Çanakkale, Turkey

Received: 19.10.2021 • **Accepted/Published Online:** 02.04.2022 • **Final Version:** 13.06.2022

Abstract: Factorial experiments are commonly employed in agricultural research as in other branches of applied sciences. In these experiments, inferences related to the interaction are essential. However, many researchers are still unable to analyze this type of experiment and interpret the results in the correct way. This is because researchers focus on interpreting the main effects although there is a significant interaction effect. Of course, meaningful main effects can exist even in the presence of an interaction, especially if interactions do not affect the main effects. Therefore, it is extremely important to understand thoroughly in which situations only the interaction effect(s), in which cases only the main effect(s), and in which cases the interpretation of the main effects will be meaningful although the interaction effect is significant. In this study, evaluating factorial experiments has been discussed in detail, especially in studies related to animal science. It has also been focused on the importance of considering both statistical and practical significance while interpreting the statistical analysis results.

Key words: Factorial experiments, interaction, main effect, effect size

1. Introduction

Due to some advantages of the factorial experiments they are commonly conducted by scientists and researchers wishing to investigate the effect of two or more independent variables on a single dependent variable almost in all branches of applied sciences [1–3]. In factorial experiments, although evaluating the interaction effect is essential, it is noticeable that many researchers only do consider main effects rather than interaction effects even if significant interaction effect. And, it has been also noticed that many researchers are still unable to analyze and interpret the factorial experiments in a proper way [1,3, 4–7]. However, ignoring the significant interaction effect(s) may cause crucial problems especially in the stage of interpreting the results and making inferences. It is because when there is a significant interaction effect, the factors are dependent, and thus, the combinations of the levels of the factors may affect the data in various ways [3–6, 8–11].

On the other hand, although trying to interpret the main effects are always not appropriate in the presence of interaction (Figure 1 and Figure 2), in some cases, interpretation of the main effects can be meaningful even if the interaction effect is significant. For example, consider a study that was carried out to investigate the effect of two different ration types on the live weight gain of the lambs in two different breeds and suppose there is a significant interaction between ration type and breed (ration type × breed interaction). For such cases, since the effect of ration type on live weight gain will be different for the lambs in breed 1 and breed 2, trying to make a general statement about the effect of ration type and breeds separately (main effects) will be misleading. Therefore, a significant main effect of the ration type does not necessarily indicate that the live weight gain of the lambs who fed with ration 1 is significantly higher than that of the lambs who fed with ration 2. In this case, there will be two simple effects of the ration types: the effect of ration types for breed 1 and the effect of ration types for breed 2. Since the presence of interaction means that the main effect is not representative of the simple effects, the effect of the ration type should be compared for each breed separately. As a result, when a researcher wants to know whether a factor has an effect at each level of a second factor, he/she should test the simple effects. It is possible to see these situations from the interaction plots (Figure 1 and Figure 2) easily.

The crossed lines on the figure suggest that there may be a ration × breed interaction effect. However, it should not be forgotten that the final result will be reached by hypothesis testing procedure (p-value) [3,7,12]. For example, Figure 1 shows that the live weight gains of the lambs in the breed 1 are higher when the first ration type is used. Conversely, the live weight gains of the lambs

* Correspondence: serdargenc1983@gmail.com

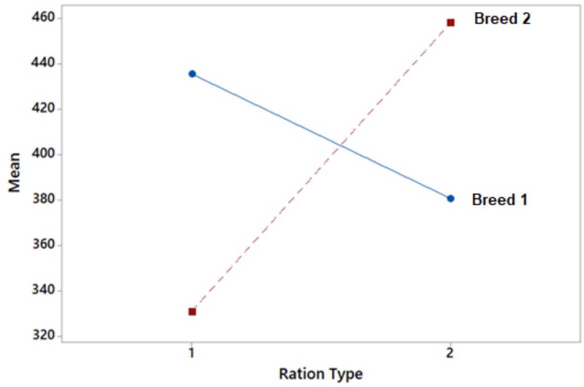

Figure 1. Interaction plot for live weight gain.

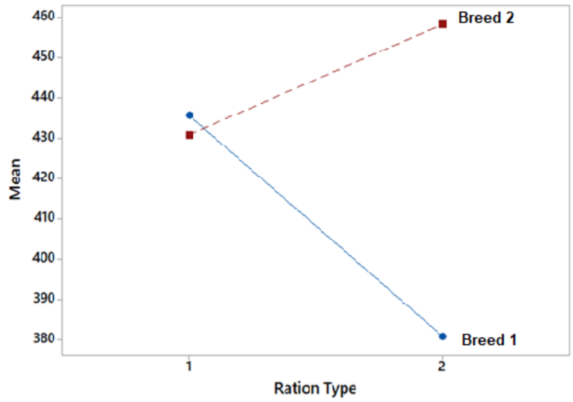

Figure 2. Interaction plot for live weight gain.

in breed 2 are higher when the second ration type is used. Therefore, the desired live weight gains will not be achieved for such cases if the lambs in breed 1 feed with the second ration and the lambs in breed 2 with the first ration type. Likewise, it can be understood from Figure 2, the live weight gains of the lambs in breed 1 and breed 2 are quite similar when the first ration type is used or significant differences are not observed between live weight gains of the lambs in both breeds when ration 1 is used in feeding the lambs. Live weight gains of the lambs in breed 2, however, are higher when the second ration type 2 is used. Therefore, the desired live weight gains will not be achieved for such cases if the lambs in breed 2 feed with the first ration and the lambs in breed 1 feed with the second ration. In this case, the following is asked: which ration type is better or which type of ration should be used in feeding lambs in breed 1 and breed 2? It depends on the breed. That is why an interaction effect is also known as **'it depends effect'**. Thus, since it can be able to cause getting unreliable results and limit the generalizability of the results, it will not be convenient trying to interpret the main effect(s) without considering the interactions if there is a statistically significant interaction effect(s). As it can be easily seen in the above examples, they cannot answer the question about which ration type is better without knowing the breed. Let us assume a researcher wants to determine which ration is the best for lambs in breed 1 and breed 2. However, suppose that he forgot to include the interaction effect and assessed only the main effect of type ration and breed. In this case, he will make his decision only based on the main effects plots (Figure 3 and Figure 4) below.

Based on main effect plots, the researcher would choose the second ration for the lambs in breed 1 because they each produce higher live weight gain (Figure 3). For the situation presented in Figure 4, the researcher would choose the first ration for the lambs in breed 2. However, since the interaction effects are significant for both cases, it will be quite misleading to reach such conclusions. A main effect here is the effect of a factor on an interested variable (dependent variable)–ignoring all other factors. Therefore, a main effect for breeds says that there is a difference between the breeds, regardless of ration types. Likewise, a main effect for ration types says that there is a difference between the ration types, regardless of the breeds. As it can be seen from Figure 1, there is clearly an interaction effect here ($p < 0.001$). As it is noticed from Figure 1, there is an obviously large change in the mean of the live weight gain for breed 2, but not for breed 1. Actually, this is a good example of a case where both main effects will be significant ($p < 0.001$ and $p = 0.002$) alongside the interaction effect, but main effects are not meaningful here [3,12].

Although there is a statistically significant interaction effect, in some cases, it is possible to interpret the main effects (for example, if there is no parallelism in the interaction chart, but one level of a factor is always higher in all levels of the other factor). However, in general, (especially there is a cross line on the figure) interpretation of the main effects will not be appropriate in case of the presence of an interaction effect. As a result, when you have statistically significant interactions, you cannot interpret the main effect without considering the interaction effects. Likewise, when Figure 2 is examined, it is also clear that there is a significant interaction effect. If one tries to interpret the main effect of the breed without considering the ration type × breed interaction he or she will conclude that there is a significant difference between the means of the live weight gains of the breeds, regardless of ration types. That may be technically true if averages of the breeds are compared regardless of the ration types. However, this conclusion or approach is only valid on average across the ration types because of the large difference in the second ration. Therefore, it is not true for each ration type. As a result, to conclude that the means generally differ across the breeds, regardless of ration types, is not really accurate.

However, there are some situations where the interaction does not affect the main effects. In such cases, interpretation of the main effects can be meaningful (Figure 5 and Figure 6). As it is well known that parallel lines in the interaction plots indicate that there is no interaction effect while different slopes suggest that one might be present. As it can be seen from Figure 5, the lines are not parallel. This shows that a significant interaction effect might have occurred. Although this figure is basically identical to Figure 2, what makes the main effect of breed meaningful here, despite the interaction, is that the first breed's mean is always higher than the second breed's for both ration types. Therefore, that is a meaningful main effect here and it says that the lambs in the first breed do generally have higher live weight gain means, regardless of the types of ration you feed the lambs. Therefore, it will be useful to consider the main effects as well in such cases although there is a significant interaction effect. It is because if the main effects are ignored in such cases, it might cause to ignore the fact that the lambs in the first breed not only changed more but started higher (Figure 5.).

Therefore, it will be beneficial to keep in our mind that in cases where the interaction effect is significant, it might be useful also to consider the main effect(s) alongside the interaction effect. As a result, if one has both a significant main effect and a significant interaction, it will be useful and it does not assume the main effect will be meaningless. It may be important.

One of the other important points that need to be considered is that the researchers should consider effect size when they interpret the results. That way, it will be possible to get information on both the statistical and practical significance of the observed differences. It is because, in practice, the researchers are commonly reported the p-value that shows only the statistical significance of the observed difference, finding a statistically significant difference among the group does not mean that this difference is also practically significant. Notwithstanding, most of the researchers believe that finding a smaller p-value shows that the observed difference among the group means it is very significant [13]. However, statistical significance is a function of sample size. Thus, very small differences may be found as statistically significant when studying with large samples while huge differences may not be found statistically significant in case of working with small sample sizes [3,14–16]. That is why only the reporting of p-value is not enough for both evaluating statistical and practical significances of observed differences among the treatment groups. Therefore, especially in studies related to applied science, it will be very beneficial to evaluate both

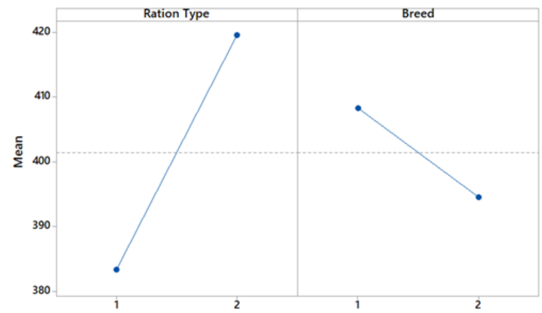

Figure 3. Main effects plot for live weight gain.

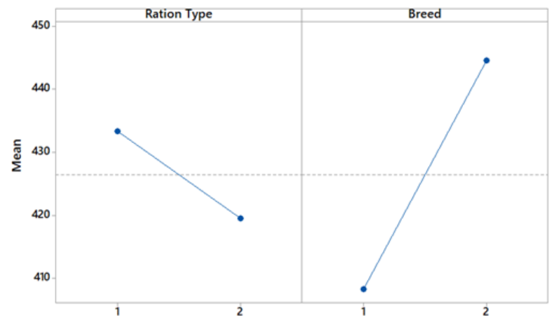

Figure 4. Main effects plot for live weight gain.

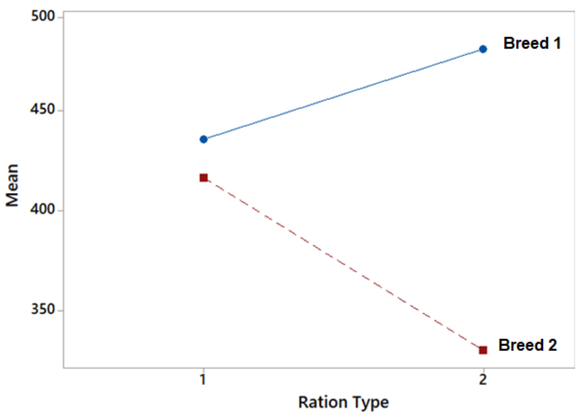

Figure 5. Interaction plot for live weight gain.

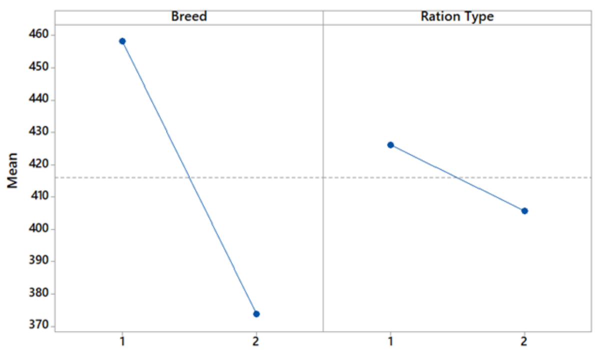

Figure 6. Main effects plot for live weight gain.

the statistical and the practical significance of the observed difference simultaneously. That way, it will be possible to get more detailed information about the effect of the factor(s) in the study. For this aim, different effect size measures have been proposed namely Eta-Squared, Partial Eta-Squared, Omega-Squared, and Epsilon-Squared [15–22]. That is why most reputable journals are looking for such authors for reporting some effect size measures that would provide information regarding practical significance along with p-value. Therefore, in this study, we will discuss the factorial experiments with a different perspective to show how we get more detailed and generalizable results. All discussions will be done based on three different scenarios generated from using the mean and standard deviation of a real data set which obtained an experiment conducted to investigate the effect of two factors (ration type and breed) on the live weight gain of lambs.

2. Materials and methods

The material of this study consisted of random numbers generated from three normal distributed populations with equal variances for two factors namely ration type and breed. Each subgroup contained 10 observations. Therefore, number of replications for this study was 10. Average and standard deviations of a real study which carried out to investigate the effect of different ration types on live weight gains of lambs in two different breeds were used in generating random numbers.

Since there are two factors namely ration type and breed following statistical models have been used in analyzing data sets.

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

Where

Y_{ijk}: is the live weight gain of the kth lamb

μ: General population mean

α_i : Effect of ith breed (i = 1, 2)

β_j: Effect of jth ration type (j = 1, 2)

$(\alpha\beta)_{ij}$: Effect of breed by ration type interaction

ε_{ijk}: Random error term

3. Results

Results of factorial ANOVA for three experimental cases and interaction plots for ration type by breed have been presented in Table 1, Table 2, Table 3, and Figure 7, Figure 8, and Figure 9, respectively. How the results of the factorial experiments should be interpreted and in which cases it may be meaningful to interpret the main effects together with the interaction effect are discussed in detail

Table 1. Anova results for the first scenario.

Source of variation	p-value	Effect size (contribution)	Total effect size
Breed	0.142	1.65	85.88
Ration type	0.001	11.48	
Breed × ration type int.	<0.001	72.75	

Note 1: Contribution stands for effect size values
Note 2: The effect size value of above 0.20 is generally evaluated as practical significant as well

Table 2. Anova results for the second scenario.

Source of variation	p-value	Effect size (contribution)	Total effect size
Breed	0.002	3.92	66.48
Ration type	0.142	27.26	
Breed × ration type int.	<0.001	35.30	

Table 3. Anova results for the third scenario.

Source of variation	p-value	Effect size (contribution)	Total effect size
Breed	<0.001	2.34	96.33
Ration type	0.002	86.29	
Breed × ration type int.	<0.001	7.60	

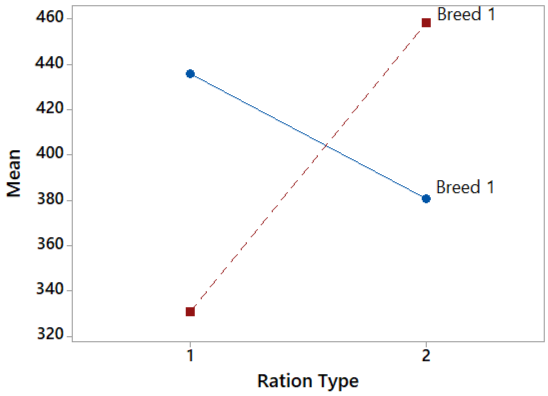

Figure 7. Interaction plot for live weight gain for case 1.

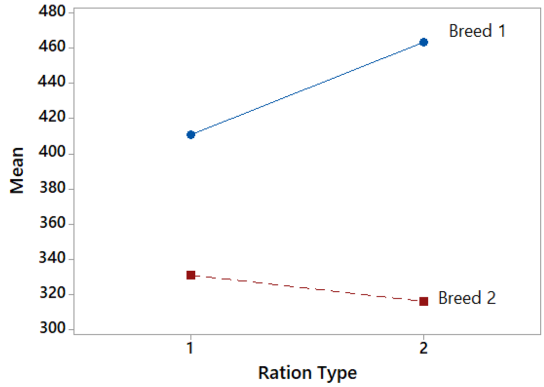

Figure 9. Interaction plot for live weight gain for case 3.

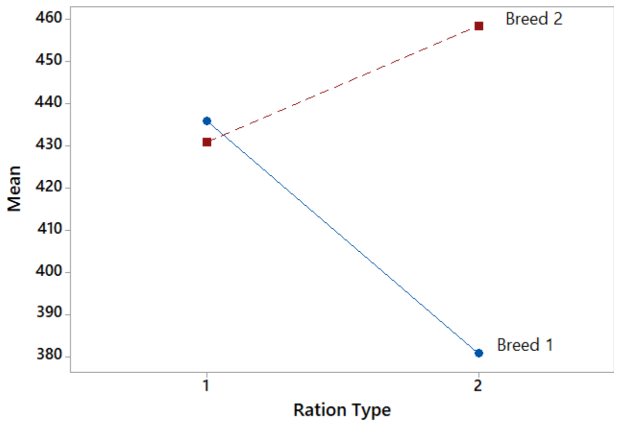

Figure 8. Interaction plot for live weight gain for case 2.

on these results. The question if interpreting main effects might be meaningful despite a significant interaction effect rises especially when both interaction and main effect(s) are important. In order to answer this question, we will focus on the results of three scenarios namely cases 1, 2, and 3, respectively. That way, it will be possible to give an answer to this question more easily and clearly. When interaction plots and ANOVA results related to case 1, case 2, and case 3 are evaluated together, it can be easily seen that interaction terms are significant for all three cases. Let us begin to evaluate the results of case 1.

3.1. Results of case 1

For the case 1, both interaction effect ($p < 0.001$) and main effect of ration type are significant ($p = 0.001$). That means the effect of ration type on weight gain of the lambs varies depending on ration type. In this case, it will not be appropriate to evaluate the effect of breed and ration type separately. The effect sizes related to the interaction and ration type are another indication that it is not appropriate to interpret the main effects separately. As can be seen from the ANOVA table, the effect size value of interaction (72.75%) is obviously higher than that of the ration type (11.48%). Therefore, 72.75% of the variation in the weight gains of the lambs can be explained by interaction while only a little part of variation can be explained by the ration types. This result is also one of the important indicators that the interaction effect is both statistically and practically significant and it shows that interpreting the main effects will not be meaningful.

3.2. Results of case 2

When ANOVA table related to case 2 is examined, it is seen that the interaction effect is significant as in case 1. However, interaction plots for case 1 and case 2 show a little bit different patterns. This is because, in case 1, while the effects of interaction and ration type are significant, in case 2, however, the effects of the breed and interaction are significant. However, since there is a cross line in both charts (Figure 7 and Figure 8), the mean of one level of one factor will always not be higher in both levels of the other factor. As can be seen from the ANOVA table, the effect size value of interaction (35.30%) is still higher than that of the breed (27.26%). This is also one of the indicators that the interaction effect is both statistically and practically significant.

3.3. Results of case 3

When ANOVA table and interaction plot related to case 3 are examined, it is seen that the interaction effect is significant as in case 1 and case 2. However, the interaction plot for case 3 obviously shows a different pattern. As it can be seen from the interaction plot, the mean of the first breed is always higher than the second breed for both ration types. Therefore, interpreting the main effect might be meaningful in such cases even the presence of a significant interaction effect. The obviously high effect size value of the breed is another indication that especially interpreting the main effect of the breed might be meaningful. It is because; the breed can explain 86.29% of the variation in the live weight gains of the lambs. However, a very small part of the variation (only 7.60%) in the weight gains of the lambs can be explained by the interaction term (despite presence of a significant interaction effect). This amount of variation is not evaluated as practically significant. Therefore, it will be beneficial to keep in our mind that in cases where the interaction effect is significant, it might be useful also to consider the main effect(s) alongside the interaction effect. As a result, if one has both a significant main effect and a significant interaction, it will be useful, and it does not assume the main effect will be meaningless. It may be important.

4. Discussion

Because of the different advantages of the factorial experiments, they are commonly used in agricultural research as in other branches of applied science. One of the biggest advantages of factorial designs is that they allow researchers to look for interactions between the factors [23–25]. Factorial experiments are also very efficient, can have high test power even if they have relatively few observations per experimental condition or subgroup, and provide extra information which cannot be obtained when using single factor designs. Despite important advantages of factorial experiments over single factor experiments, many researchers are still unable to analyze factorial experiments, interpret and present the results in the correct and efficient way.

In practice, especially in agriculture and biological science-based studies, one of the most frequent questions to statisticians is when the interaction effect is significant, whether the main effect(s) can be interpreted or not. From the statistician's point of view, if the interaction effect is significant, in this case, the interpretation of the main effects is meaningless. Since the presence of the interaction effect indicates that the effect of one factor depends on the other factor(s), and thus it will not be appropriate to evaluate the main effect of each factor separately. This is because it will be difficult to make a general statement about the effect of a factor when the size of the effect depends on the level of a second factor. When an interaction is large, the corresponding main effects have very little practical meaning. Consequently, when the interaction is present, the main effects of the factors involved in the interaction may not have much meaning. As it is stated by De Gonzales and Cox (2007) [26] interaction is one of the fundamental concepts of statistical analysis of factorial experiments. Establishing the presence or absence of interaction may be a key to the correct interpretation of data. Therefore, since the presence of interaction between the factors limits the generalizability of main effects, it will not be appropriate to focus on main effects when there is a significant interaction. If the interaction is present, there is an indication that the differences among the levels of a factor depend on the level of the other factor [27]. However, due to the difficulty and inability of the researchers to interpret the results of factorial experiments, erroneous inferences about the effect of treatments on response were observed [4]. Cardellino and Siewerdt (1992) [9] reported that a comparison of marginal averages without considering possible interactions is an example of such an error. Bertoldo et al. (2008) [5] noted that 72% of the published studies in the factorial experiments were incorrect when tests of average comparison were conducted. They reported that the reason for this problem was that the significant interaction effect was ignored and the factors were evaluated separately [2]. As Silva (1999) [28] reported that for experiments designed in a factorial scheme, the conclusion to be drawn will be changed depending directly on the presence or absence of interaction. If ANOVA results show nonsignificant interaction effect, in this case, complementary procedures of the main effects of factors are carried out, and the effect of the interaction is disregarded [29]. However, in the presence of significant interaction, an evaluation of the results requires a comparison of the levels of a factor inside the fixed levels of another factor [11]. In other words, the inferences about one of the factors depend directly on the level of the other factor [2,30].

As a result, the factorial experiments are the only way to discover interactions between variables. The presence of interaction shows how the factors or independent variables work together in terms of impacting the dependent variable. In another way, the presence of a significant interaction indicates that the effect of one factor depends on the level of the other factor. Therefore, including interaction terms in the model is extremely important since it provides the researcher with a better representation and understanding the relationship between the dependent and independent variables and helps explain more of the variability in the dependent variable. An omitted interaction effect from a model where a nonnegligible interaction does in fact exist may result in a misrepresentation of the relationship between the independents and dependent variables. It could also lead to a bias in estimating model parameters. However, it will not always be a correct approach to think that the main effects will not have any meaning in cases where the interaction is significant. As shown in case 3 above, in some cases the interpretation of the main effects may be meaningful, despite a significant interaction effect. Of course, when deciding on this, it should be remembered that ANOVA results, interaction plots, and effect size values should be evaluated together. Another important issue when interpreting and reporting statistical analysis results is the practical evaluation of the observed difference or effect. For this purpose, effect size measures are used. Effect size (ES) is a measure of the size or magnitude of a treatment effect. Therefore, effect size answers the question "How big is the difference between the group means?". The answer of this question is important for making decision. Because effect size measures help us to evaluate if the size of the effect in the population is large enough to be interest (or evaluate practical significance of the observed difference). That is why, especially recently the journals increasingly require the reporting of the effect size for publications.

References

1. Yassin N, Morais AR, Muniz JA. Analysis of variance of factorial experiments with two factors and additional treatments. Ciência e Agrotecnologia 2002; 26 (7): 1541-1547.
2. Rocha F, do Vale NM, Barili JLOC, Guidolin AF, Bertoldo JG. An approach to the decomposition of interaction in a factorial experiment with five factors. Acta Scientiarum Agronomy 2012; 34 (1). doi: 10.4025/actasciagron.v34i1.12104
3. Mendeş M. Statistical Methods and Experimental Design. Kriter Publication. İstanbul 2019 (in Turkish).
4. Bezerra NF, Nunes GHS, Negreiros MZ. Assessment of multiple-comparison procedures of published articles in the journal Horticultura Brasileira from 1.983 to 2.000. Horticultura Brasileira 2002; 20 (1): 5-9. doi: 10.1590/S0102-05362002000100001
5. Bertoldo JG, Coimbra JLM, Guidolin AF, Mantovani A, Vale NM. Difficulties related to the use of tests of comparison of averages in scientifi c articles. Biotemas 2008; 21 (2): 145-153. doi: 10.5007/2175-7925.2008v21n2p145
6. Bertoldo JG, Coimbra JLM, Guidolin AF, Miqueloto A, Toaldo D. Use or abuse in mean comparison tests: scientific or empiric knowledge?. Ciência Rural 2008; 38 (4): 1145-1148. doi: 10.1590/S0103-84782008000400039
7. Mendeş M, Yigit S. Type I error and test power of different tests for testing interaction effects in factorial experiments. Statistica Neelandica 2013; 67: 1-26. doi: 10.1111/j.1467-9574.2012.00528.x
8. Siemiatycki J, Thomas DC. Biological models and statistical interactions: An example from multistage carcinogenesis. International Journal of Epidemiology 1981; 10: 383-387. doi: 10.1093/ije/10.4.383
9. Cardellino RA, Siewerdt F. Use and misuse of statistical tests for comparison of means. Revista Brasileira de Zootecnia 1992; 21 (6): 985-995.
10. Wechsler FS. Unbalanced factorials: a misunderstood analysis. Pesquisa Agropecuária Brasileira 1998; 33(3): 231-26.
11. Nogueira MCS, Corrente JE. Triple interaction fractioning using PROC GLM CONTRAST/SAS applied to a three-way layout for balanced data. Bragantia 2000; 59 (1): 109-115. doi: 10.1590/S0006-87052000000100017
12. Starmer CF, Lastra AA, Nesterenko VV, Grant AO. A proarrhythmic response to sodium channel blockade: Theoretical model and numerical experiments. Circulation 1991; 84:1364-1377.
13. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. Psychol Methods 2000; 5: 241-301.
14. Fan X. Statistical significance and effect size in education research: two sides of a coin. The Journal of Educational Research 2001; 94: 275-282. doi: 10.1080/00220670109598763
15. Yiğit S, Mendeş M. Which effect size measure is appropriate for one-way and two-way anova models? A monte carlo simulation. Revstat Statistical Journal 2016; 16 (3): 295-313.
16. Yiğit S, Mendeş M. ANOM technique for evaluating practical significance of observed difference among treatment groups. International Journal of Agricultural Science Research 2017; 6 (1): 001-007.
17. Cohen J. Eta-squared and partial eta-squared in fixed factor anova designs. Educational and Psychological Measurement 1973; 33: 107-112. doi: 10.1177/001316447303300111
18. Carroll RM, Nordholm, LA. Sampling characteristics of Kelley's and Hays'. Educational and Psychological Measurement 1975; 35: 541-554. doi: 10.1177/001316447503500304
19. Keselman HJ. A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. Canadian Psychological Review 1975; 16: 44-48. doi: 10.1037/h0081789
20. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. Journal of Experimental Psychology: General 2012; 141 (1): 2-18. doi: 10.1037/a0024338
21. Okada K. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. Behaviormetrika 2013; 40: 129-147. doi: 10.2333/bhmk.40.129
22. Skidmore ST, Thompson B. Bias and precision of some classical ANOVA effect sizes when assumptions are violated. Behavior Research Methods 2012; 45 (2): 536-546. doi: 10.3758/s13428-012-0257-2
23. Wu C, Hamada M. Experiments: Planning, Analysis, and Parameter Design Optimization. New York: Wiley 2000.
24. Lewis SM, Dean AM. Detection of interactions in experiments on large number of factors. Journal of the Royal Statistical Society Series B 2001; 63 (4): 633-672. doi: 10.1111/1467-9868.00304
25. Collins LM, Dziak JJ, Li R. Design of experiments with multiple ındependent variables: a resource management perspective on complete and reduced factorial designs. Psychol Methods 2010; 14 (3): 202-224. doi: 10.1037/a0015826
26. De Gonzalez AB, Cox RD. Interpretation of interaction: a review. The Annals of Applied Statistics 2007; 1 (2): 371-385. doi: 10.1214/07-AOAS124

27. Chew V. Comparing treatment means: a compendium. Hort Science 1976; 11(4): 348-335.

28. Silva JGC. The consideration of the unit structure in ınferences from the experiment. Pesquisa Agropecuária Brasileira 1999; 34 (6): 911-925. doi: 10.1590/S0100-204X1999000600001

29. Perecin D, Cargnelutti Filho A. Comparisonwise and experimentwise effects in factorial experiments interactions. Ciência e Agrotecnologia 2008; 32 (1): 68-72. doi: 10.1590/S1413-70542008000100010

30. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. American Journal of Epidemiology 1980; 112: 467-470. doi: 10.1093/oxfordjournals.aje.a113015