

Comparisons of treatment means when factors do not interact in two-factorial studies

Jiawei Wei · Raymond J. Carroll ·
Kathryn K. Harden · Guoyao Wu

Received: 20 December 2010 / Accepted: 22 April 2011 / Published online: 6 May 2011
© Springer-Verlag 2011

Abstract Scientists in the fields of nutrition and other biological sciences often design factorial studies to test the hypotheses of interest and importance. In the case of two-factorial studies, it is widely recognized that the analysis of factor effects is generally based on treatment means when the interaction of the factors is statistically significant, and involves multiple comparisons of treatment means. However, when the two factors do not interact, a common understanding among biologists is that comparisons among treatment means cannot or should not be made. Here, we bring this misconception into the attention of researchers. Additionally, we indicate what kind of comparisons among the treatment means can be performed when there is a nonsignificant interaction among two factors. Such information should be useful in analyzing the experimental data and drawing meaningful conclusions.

Keywords Factor level means · Interaction · Main effects · Multiple comparison · Treatment means · Two-factor studies

Abbreviations

ANOVA	Analysis of variance
Arg	L-Arginine
DIO	Diet-induced obese
HF	High fat
LF	Low fat

Introduction

Two-factorial studies are common in nutrition and other biomedical researches, as evidenced in papers published in influential journals, including *Amino Acids* (e.g., Jobgen et al. 2009a; Willoughby et al. 2007), *Journal of Nutrition* (e.g., Eller and Reimer 2010; Vicario et al. 2007), and *Journal of Nutritional Biochemistry* (e.g., Liu et al. 2010; Mochizuki et al. 2001). This is primarily because scientists are interested not only in the main effects of two independent factors (e.g., amino acids and fats; protein and age; food intake and exercise), but also in their interaction (Fu et al. 2010). Take the study of amino acids in Jobgen et al. (2009a, b) as an example. These authors reported that dietary L-arginine (Arg) supplementation reduced white-fat gain in diet-induced obese (DIO) rats (Jobgen et al. 2009b) and modulated gene expression in adipose tissue (Jobgen et al. 2009a). In their study, both diet (high or low fat) and amino acids (with or without Arg) are the factors that affect white-fat gains in DIO rats, but whether these two factors exert effects separately or additively depends on whether an interaction effect exists.

Neter et al. (1996) have constructed a basic strategy for the analyses of two-factorial studies. Here, we create a flowchart (Fig. 1) of this method for a simple, parsimonious explanation. It is widely recognized that the analysis of factor effects is generally based on treatment means when the interaction of

J. Wei · R. J. Carroll
Department of Statistics, Texas A&M University,
3143 TAMU, College Station, TX 77843, USA

K. K. Harden
American Society of Nutrition, 9650 Rockville Pike,
Bethesda, MD 20814-3990, USA

G. Wu (✉)
Department of Animal Science, Faculty of Nutrition,
Texas A&M University, 2471 TAMU, College Station,
TX 77843, USA
e-mail: g-wu@tamu.edu

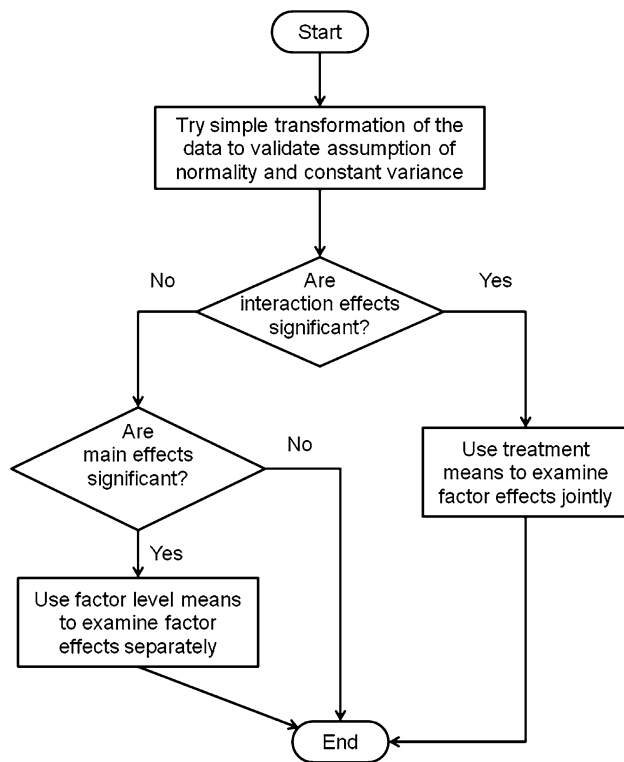


Fig. 1 Strategy for analysis of two-factorial studies. The analysis of factor effects is generally based on treatment means when the interaction of the factors is statistically significant, and involves multiple comparisons of treatment means. However, when the two factors do not interact in analysis of variance, multiple comparisons of treatment means are not generally indicated in the classic statistics textbooks

the factors is statistically significant, and involves multiple comparisons of treatment means (e.g., Neter et al. 1996; Steel et al. 1997). However, when the two factors do not interact in analysis of variance (ANOVA), multiple comparisons of treatment means are not generally indicated in the classic statistics textbooks (e.g., Steel et al. 1997). Therefore, a common understanding among biologists is that comparisons among treatment means cannot or should not be made when interaction of two factors is statistically nonsignificant (e.g., Eller and Reimer 2010; Liu et al. 2010; Mochizuki et al. 2001; Willoughby et al. 2007; Vicario et al. 2007). In the present article, we bring this misconception into the attention of biomedical and life science researchers. Additionally, we indicate what kind of comparisons among the treatment means can be performed when there is a nonsignificant interaction among two factors.

General considerations

When factors do not interact

Based on Fig. 1, the first step in two-way ANOVA is to validate the assumptions of normality and constant variance

for the experimental data using such a statistical method as the Levene's test (Levene 1960) or the Brown-Forsythe test (Brown and Forsythe 1974). If the data do not have homogenous variance, they should undergo transformation (e.g., logarithm transformation) to meet the necessary assumptions of ANOVA so that an appropriate statistical inference can be obtained (Fu et al. 2010). When the two factors do not interact (e.g., $P > 0.05$) in ANOVA, the main effects of these two factors should be determined, with the effect of one factor being averaged across the levels of the other factor. Similar strategies for analysis of two-factor studies have been proposed by some authors (e.g., Milliken and Johnson 1984). In all these methods, the analyses are focused on the main effects when the interaction is statistically nonsignificant. However, scientists may be interested in the comparison among treatment means even when the two factors do not interact. As in the nutritional study of Jobgen et al. (2009b), although there is no significant interaction between the two factors on white-fat mass, it is still interesting and important to determine whether the high-fat (HF) diet along with Arg supplementation has the same effect as the low-fat (LF) diet alone on adiposity.

Let the two factors be called A and B, and suppose each of them has two levels: A1, A2 and B1, B2. If there is no evidence of a statistically significant interaction between A and B, the analysis of factor effects usually involves only the factor level means. Inferences about the treatment means can be made by comparing the mean treatment responses across the levels of A averaged over the levels of B. This is possible because when there is no interaction between A and B, the difference in the response across the levels of A is the same at all levels of B.

Figure 2a plots the treatment means for different levels of A and B when there is no interaction between A and B. On the solid line are the mean treatment responses μ_{11} and μ_{21} for groups with (A1, B1), and (A2, B1), respectively. On the dashed line are the mean treatment responses μ_{12} and μ_{22} for groups with (A1, B2), and (A2, B2), respectively. The two lines are parallel when the two factors do not interact, thus $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$. Hence, we can examine the differences in the levels of A averaged across the levels of B, and vice versa.

When the interactions of factors are significant

If there is evidence of a significant interaction between A and B, inferences concerning the differences in the mean treatment responses for A must be conducted separately for each level of B, because the differences in the treatment mean responses across the levels of A may differ, depending on the level of B. Figure 2b plots the treatment means for different levels of A and B when there is an interaction. The two lines are not parallel when the interaction is

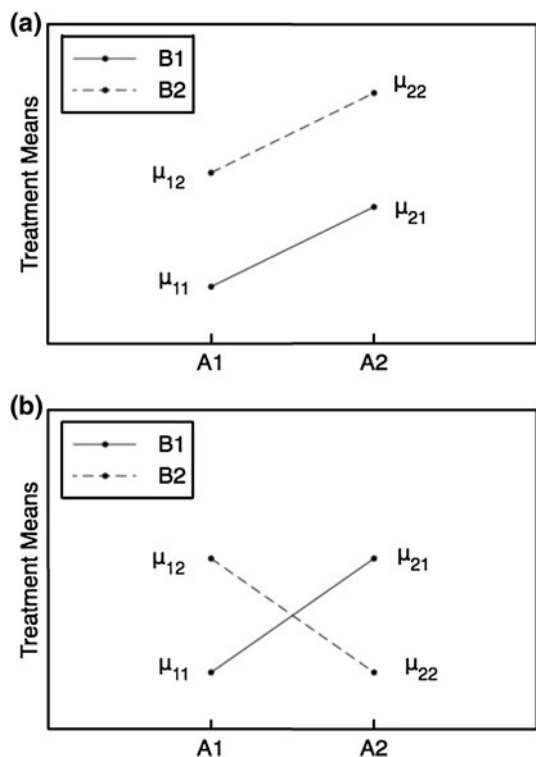


Fig. 2 Treatment means when the two factors do (a) or do not (b) interact significantly. On the *solid lines* are the mean treatment responses μ_{11} and μ_{21} for groups with (A1, B1), and (A2, B1), respectively. On the *dashed lines* are the mean treatment responses μ_{12} and μ_{22} for groups with (A1, B2), and (A2, B2), respectively

significant, thus $\mu_{11} - \mu_{12} \neq \mu_{21} - \mu_{22}$. Hence, comparisons among the treatment means can be constructed to answer particular questions posed by the researchers. Another salient point in Fig. 2b is that $\mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}$, and $\mu_{11} + \mu_{21} = \mu_{12} + \mu_{22}$, indicating that none of the main effects is significant. In such a case, it is justified to present statistically significant interaction, which may have a highly relevant biological interpretation. In statistical analysis, this case is also very interesting and important. Although $\mu_{11} - \mu_{12} \neq \mu_{21} - \mu_{22}$, $\mu_{11} - \mu_{12} = -(\mu_{21} - \mu_{22})$, which means that the differences in the treatment means for the level A1 is the opposite of that for the level A2, depending on the level of B.

Comparisons of treatment means when factors do not interact

Factor A has two levels, factor B has two levels

As noted previously, based on the general analysis of factor effects, no comparison among the treatment means is usually suggested when the two factors do not interact. However, this does not mean that comparisons among

treatment means cannot or should not be made. For simplicity, we first develop our strategy for the case when both A and B have two levels, and then extend to the general case in which A has *a* levels and B has *b* levels. From the preceding discussion, it is obvious that the detailed multiple comparisons of μ_{11} versus μ_{12} should not be done, as with the analysis of the main effects for B, namely μ_{11} versus μ_{21} , μ_{12} versus μ_{22} and μ_{21} versus μ_{22} . The only comparisons not covered by the main-effect analysis are μ_{12} versus μ_{21} and μ_{11} versus μ_{22} . In this section, we discuss whether the post hoc test to compare μ_{12} and μ_{21} , μ_{11} and μ_{22} can be performed when there is no significant interaction.

In the two-factor studies with two levels in each factor, there are four different situations that might occur when the two factors do not interact (Table 1). In the first case, there is no main effect in A and no main effect in B; in this case, no comparisons among μ_{12} and μ_{21} , μ_{11} and μ_{22} are needed because the four treatment means do not differ. In the second case, there is a main effect in A but not in B; thus, no comparisons between μ_{12} and μ_{21} and μ_{11} and μ_{22} are needed because logically $\mu_{11} \neq \mu_{22}$ and $\mu_{21} \neq \mu_{12}$. The third case is similar to the second case, when there is a main effect in B but not in A. The fourth case is the most complicated, when there are main effects in both A and B. Comparison among μ_{12} and μ_{21} or μ_{11} and μ_{22} is needed because we do not know whether $\mu_{11} = \mu_{22}$ or $\mu_{21} = \mu_{12}$. Specifically, if the treatments means are increasing from A1 to A2 at all levels of B (Fig. 2a), then $\mu_{11} < \mu_{22}$, and only the comparison between μ_{12} and μ_{21} is needed. Similarly, if the treatments means are decreasing from A1 to A2 at all levels of B, then $\mu_{12} > \mu_{21}$, and only the comparison between μ_{11} and μ_{22} is needed.

Factor A has *a* levels, factor B has *b* levels

In the two-factorial studies, there are sometimes more than two levels in each factor. Suppose factor A has *a* levels and factor B has *b* levels, and let μ_{ij} be the treatment means for

Table 1 Treatment means when the two factors do not interact

μ_{11}	μ_{21}	μ_{12}	μ_{22}	Main effects
<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	No main effect in A, no main effect in B; $\mu_{11} = \mu_{21} = \mu_{12} = \mu_{22}$
<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	Main effect in A, no main effect in B; $\mu_{11} \neq \mu_{22}, \mu_{21} \neq \mu_{12}$
<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	No main effect in A, main effect in B; $\mu_{11} \neq \mu_{22}, \mu_{21} \neq \mu_{12}$
<i>a</i>	<i>b</i>	<i>b</i> or <i>c</i>	<i>a</i> or <i>d</i>	Main effect in A, main effect in B; we do not know whether $\mu_{11} = \mu_{22}$ or $\mu_{21} = \mu_{12}$

i th level in A and j th level in B. Then, based on a procedure similar to that in the previous section, we have the following possible situations. If there is no main effect in either factor, no multiple comparisons among treatment means are needed. If there is a main effect in only one factor, we may only conclude that at least two of the treatment means in this factor are significantly different. Further, multiple comparisons among the treatment means are required to identify which specific means differed. If there are main effects in both A and B, and the treatments means are increasing from A1 to Aa at all levels of B, then multiple comparisons of $(\mu_{1,j+1}, \dots, \mu_{i-1,j+1}, \dots, \mu_{1,b}, \dots, \mu_{i-1,b})$ versus μ_{ij} for $i = 2, \dots, a; j = 1, \dots, b - 1$ are needed. Similarly, if the treatments means are decreasing from A1 to Aa at all levels of B, then multiple comparisons of $(\mu_{i+1,1}, \dots, \mu_{i+1,j-1}, \dots, \mu_{a,1}, \dots, \mu_{a,j-1})$ versus μ_{ij} for $i = 1, \dots, a - 1; j = 2, \dots, b$ are needed.

Example for multiple comparisons of treatment means when two factors do not interact

We use the experimental data of Jobgen et al. (2009b) to illustrate our strategy in performing multiple comparisons of treatment means when two factors do not interact. Rats were fed a LF or HF diet, and those which were fed the HF diet became obese. After a 15-week period of LF or HF feeding, the rats were unsupplemented or supplemented with Arg. Thus, there were four treatment groups: LF – Arg (LF without Arg), LF + Arg (LF with Arg), HF – Arg (HF without Arg), and HF + Arg (HF with Arg). Note that there were two levels of dietary fat (low vs. high) and two levels of Arg (+ vs. –). The relative weight of the white adipose tissue (% of body weight) is our outcome variable, and we are interested in whether the response is the same in the HF + Arg and LF – Arg groups, which, if true, would indicate that Arg supplementation prevents HF-induced obesity in adult rats.

Based on two-way ANOVA, there is no significant interaction ($P = 0.69$), and the main effects of dietary fat and Arg supplementation are both significant ($P < 0.01$). Figure 3 shows the mean relative weights of mesenteric adipose tissue in the four groups of rats. The treatment means clearly increase from LF to HF at both levels of Arg. The only comparison which is not clear is that between LF – Arg and HF + Arg, and differences between all other means are significant. Therefore, we use the t test, which allows us to determine if there is any difference between treatment means of the LF – Arg and HF + Arg groups. This is exactly what we are interested in, and from the post hoc t test, there is no statistically significant difference ($P = 0.997$) between these two treatment means. These results indicate that supplementing Arg to adult

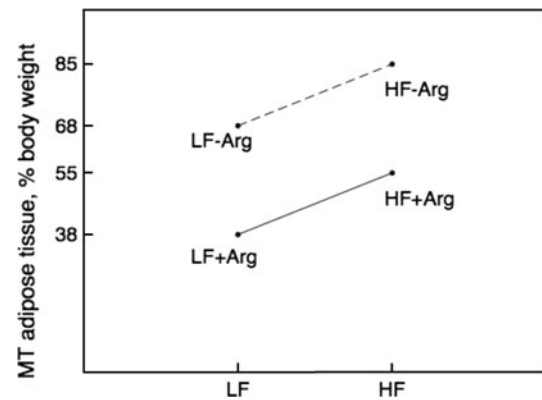


Fig. 3 Mean relative weights of mesenteric adipose tissue in rats fed high- or low-fat diets and either unsupplemented or supplemented with Arg. Data are adapted with permission from Jobgen et al. (2009b)

obese rats can effectively reduce the white-fat mass to the level observed in lean rats that are not supplemented with Arg (Jobgen et al. 2009b). This finding has important implications for preventing and treating obesity in both humans and animals (McKnight et al. 2010; Wu et al. 2009).

In the study of Jobgen et al. (2009b), the Tukey multiple comparison test was performed to identify which specific means differed. This is sound in the statistical principles and has been adopted in the subsequent studies (Lassala et al. 2010; Satterfield et al. 2010; Tan et al. 2010). Note that, in the work of Jobgen et al. (2009b), the only comparison needed is that between LF – Arg and HF + Arg. Therefore, a simple post hoc t test is sufficient to achieve the P value and also provides a greater power in statistical analysis. This new strategy can be used in the future studies involving multiple comparisons of treatment means when there is a nonsignificant interaction between two factors.

Conclusion

Comparison among treatment means when there is no interaction is meaningful for some specific situations. When we analyze the main effects of the two factors, no comparison among treatment means is needed if there is no main effect for either factor. If there is a main effect for only one factor, multiple comparisons among the treatment means for this factor are required to identify which specific means differ. If there are main effects for both factors and each factor has two levels, comparison among μ_{12} and μ_{21} , or μ_{11} and μ_{22} is needed. Similar conclusions are made for the general case in which A has a levels and B has b levels. In a two-factorial study, the basic principles of statistical analysis allows for comparison among treatment means when the two factors do not interact. This clarification will

help the biomedical and life science researchers to analyze their experimental data and answer specific scientific questions.

Acknowledgments This work is supported by grants from the National Cancer Institute (R25T-CA090301 and R37-CA057030), King Abdullah University of Science and Technology (KUS-CI-016-04; RJC), National Research Initiative Competitive Grants from the Animal Reproduction Program (2008-35203-19120) and Animal Growth and Nutrient Utilization Program (2008-35206-18764) of the USDA National Institute of Food and Agriculture, AHA (10GRNT4480020), and Texas AgriLife Research (H-8200).

References

- Brown MB, Forsythe AB (1974) Robust tests for equality of variances. *J Am Stat Assoc* 69:364–367
- Eller LK, Reimer RA (2010) A high calcium, skim milk powder diet results in a lower fat mass in male, energy-restricted, obese rats more than a low calcium, casein, or soy protein diet. *J Nutr* 140:1234–1241
- Fu WJ, Stromberg AJ, Viele K et al (2010) Statistics and bioinformatics in nutritional sciences: analysis of complex data in the era of systems biology. *J Nutr Biochem* 21:561–572
- Jobgen W, Fu WJ, Gao H et al (2009a) High fat feeding and dietary L-arginine supplementation differentially regulate gene expression in rat white adipose tissue. *Amino Acids* 37:187–198
- Jobgen W, Meininger CJ, Jobgen SC et al (2009b) Dietary L-Arginine supplementation reduces white fat gain and enhances skeletal muscle and brown fat masses in diet-induced obese rats. *J Nutr* 139:230–237
- Lassala A, Bazer FW, Cudd TA et al (2010) Parenteral administration of L-arginine prevents fetal growth restriction in undernourished ewes. *J Nutr* 140:1242–1248
- Levene H (1960) Robust tests for equality of variances. In: Olkin I, Hotelling H (eds) *Contributions to probability and statistics*. Stanford University Press, CA, pp 278–292
- Liu X, Ogawa H, Kishida T, Ebihara K (2010) The effect of high-amylose cornstarch on lipid metabolism in OVX rats is affected by fructose feeding. *J Nutr Biochem* 21:89–97
- McKnight JR, Satterfield MC, Jobgen WS et al (2010) Beneficial effects of L-arginine on reducing obesity: potential mechanisms and important implications for human health. *Amino Acids* 39:349–357
- Milliken GA, Johnson DE (1984) *Analysis of messy data, volume 1: designed experiments*. Chapman & Hall, New York
- Mochizuki H, Oda H, Yokogoshi H (2001) Dietary taurine potentiates polychlorinated biphenyl-induced hypercholesterolemia in rats. *J Nutr Biochem* 12:109–115
- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied linear statistical models*. McGraw-Hill/Irwin, New York
- Satterfield MC, Bazer FW, Spencer TE, Wu G (2010) Sildenafil citrate treatment enhances amino acid availability in the conceptus and fetal growth in an ovine model of intrauterine growth restriction. *J Nutr* 140:251–258
- Steel RGD, Torrie JH, Dickey DA (1997) *Principles and procedures of statistics*. McGraw-Hill, New York
- Tan BE, Yin YL, Kong XF et al (2010) L-Arginine stimulates proliferation and prevents endotoxin-induced death of intestinal cells. *Amino Acids* 38:1227–1235
- Vicario M, Amat C, Rivero M, Moreto M, Pelegrí C (2007) Dietary glutamine affects mucosal functions in rats with mild DSS-induced colitis. *J Nutr* 137:1931–1937
- Willoughby DS, Stout JR, Wilborn CD (2007) Effects of resistance training and protein plus amino acid supplementation on muscle anabolism mass, and strength. *Amino Acids* 32:467–477
- Wu G, Bazer FW, Davis TA et al (2009) Arginine metabolism and nutrition in growth, health and disease. *Amino Acids* 37:153–168