AAB BIOFLUX

**Advances in Agriculture & Botanics-
International Journal of the Bioflux Society**

Measuring magnitude of treatment effects in agricultural experiments

Dennis A. Apuan

Department of Agricultural Sciences, College of Agriculture, Xavier University, Cagayan de Oro City, Philippines. Corresponding author: D. A. Apuan, dennis_apuan@yahoo.com

Abstract. The estimation of treatment effect is an aspect of data analysis in agricultural experiments that is neglected. Such an effect size value cannot be determined through the usual Null Hypothesis Significant Testing (NHST) in agricultural research. Thus, the current study aimed to describe the foundation supporting the estimation and reporting of effect size. In the re-analysis of two unpublished data sets concerning the effects of four feeding treatments of Madre de Agua (*Trichanthera gigantea*) on the average final weight of pigs and on the feed cost per pig, effect sizes of treatments were measured following the omega squared algorithm procedure. This algorithm is based on the ratio of variances, which result is a unit less measure on the magnitude of treatment effect. Results of re-analysis successfully estimated the magnitude of effect equivalent to ω^2=.08 or 8 % for the average final weight of pigs and ω^2 = .9911 or 99.11 % for the feed cost per pig. These treatment effects are considered large based on the previously reported standards, and therefore substantial and appropriate recommendation may be given. The importance, implications and practical application of this effect size measure are discussed.

Key Words: Agriculture, data analysis, analysis of variance, effect size, null hypothesis testing.

Introduction. How large is the “Treatment Effect”? This is the question with direct field significance often asked by agriculturists from experiments for practical decision making. The vicious practice of Null Hypothesis Significance Testing (NHST) in agricultural research can never answer such question (Nakagawa & Cuthill 2007; Field 2005). NHST has limitation on its own – it tells only whether or not there is significant difference among treatment means (Rinella & James 2010). The statistical techniques that follow NHST, like the planned contrast and the Post hoc analysis (Field 2005) cannot measure the magnitude of treatment effect as well. Therefore, the substantive value and meaningfulness of the treatment effect cannot be assessed, and large information important in agricultural practice is lost when agricultural research stopped short at NHST and Post hoc analysis.

There are prevailing issues concerning the use of NHST (the F test, which is commonly used and referred to in agriculture as ANOVA, although t-test, chi square test and other statistical tests are also used). Foremost, is that the binary decision whether to reject or accept the null hypothesis in reference to an arbitrary cutoff point (known as α error) is simply based on probabilistic reasoning (Cohen 1994). This probability however is affected by the sample size (Aycaguer et al 2010). In fact, there were experiments in agriculture that failed to reject the null hypothesis because of sample size limitations despite the presence of visible and obvious treatment difference.

When result of experiment shows a mean difference, it is indicative of a fact that there is a treatment effect, and failure to quantify such an effect often leads to misinterpretation of results. Fidler et al (2006) reported that 63 % of the articles in conservation biology misinterpreted the non-significant results as evidence of no effect as they attributed this inability of NHST to detect significant difference among treatment means to small sample size.

While the practice of NHST is widely accepted in agriculture, Volker (2006) however explicitly pointed out that this probabilistic decision does not include actual size of treatment effect at the statistical level. Having statistically significant result is not enough – it is less informative and not meaningful for an agriculturist. Karl Pearson himself suggests that statistical significance must be supplemented because it provides only a partial explanation of the importance of the results (Aycaguer et al 2010; Kirk 1996).

One misconception among workers in agricultural research is that, significant results from ANOVA (NHST) and Post hoc analysis constitute the size of treatment effect; and that the treatment mean difference which is insignificant is construed as absence of effect while treatment difference with significant result has important effect. This measure of magnitude of treatment effect is not valid, since such a binary decision is simply based on arbitrary standards of probability and not based on actual size of effect.

Furthermore, significant results may occur in experiments even though the actual mean difference is small, so long as the sample size is large enough. Here obviously, the power of NHST to detect mean difference is high when sample size is large, but some agricultural researches often suffer from small sample size owing to constraints in land, labor, and resources. Hence, in this situation the test has low power that there may be treatment difference that really exists but is not detected. In either case, results from NHST may have low practical value and meaningless. Agricultural research needs some measure of magnitude of treatment effect that is free from the influence of sample size.

The basic conceptual problem and insufficiency of the NHST presented above, explicitly tell us the need to adopt a technique to measure the magnitude of treatment effect in agricultural experiments. Measuring the magnitude of treatment effect, otherwise known in other fields of science as “effect size” (Sun et al 2010; Field 2005; Volker 2006), and reporting of such in scientific journals are imperative in the field of Psychology. In fact, the 5th edition of its manual declares that publishing research result that does not report “effect size” is inferior (APA 2001).

In this paper I describe the foundation that supports the reporting of effect size in agricultural experiments using two data sets from experiments in animal science as examples to demonstrate the calculation and interpretation of effect sizes. A reporting technique is also demonstrated to effectively facilitate the understanding of readers.

Material and Method. One unpublished study was selected from the collections of student and faculty researches of Xavier University library. The selection criteria were:
1) the study must be an experimental type;
2) it comes from the agriculture discipline;
3) it is recently conducted experiment;
4) it includes raw data in the appendices.

The selected study was a dissertation on the “Utilization of Madre de Agua (*Trichanthera gigantean* var. guianensis) leaf meal as feed for growing-finishing pigs (*Sus domesticus*)” from the field of animal science.

With written permission and approval of the dissertation author, a close inspection and description of the statistical methods and result presentations of the selected study was done. Only two sets of data from among twenty-nine data sets in the selected study were chosen to build the case and subject to statistical analysis, *viz.* the average final weight of pigs and the total feed cost per pig. Both variables were responses of pigs to four different feeding treatments including control, in an experiment following completely randomized design, which were replicated three times.

Each data set in excel format was transported to the platform of a statistical software known as PAST (Paleontological Statistics) ver. 2.03 developed by Hammer et al (2001). The re-analysis was run in PAST ver. 2.03 to compute for the omega square (ω^2) as the overall measure of the magnitude of treatment effect.

Since PAST ver. 2.03 has limitations in terms of data summary visualization, graphical visualization of treatment differences was generated through SPSS ver. 19. The F test in each data set was also performed using this software.

A stepwise manual demonstration on the calculation of magnitude of treatment effect was done on each data set based on the omega squared algorithm (Kotrlik et al 2011; Field 2005 & Volker 2006). This was performed to show the conceptual and computational foundation in measuring effects. Two result presentations were then generated based on the conventional and the proposed presentation methods that incorporate magnitude of treatment effect.

Results and Discussion. The re-analysis of the data successfully determined the magnitude of treatment effect equivalent to 8.0 % for the average final weight of pigs and 99.11 % for the feed cost per pig fed with Madre de Agua. Both magnitudes of effects were considered large based on the descriptors reported by Kotrlik et al (2011), Kotrlik & Williams (2003) and Kirk (1996).

In the original data presentation for the average final weight of pigs, treatment means were found to have no significant difference ($F(3,8) = 1.353$, $P = 0.3246$). To the average reader, this result would not be interesting but misleading as well because it conveys absence of effect message which is contrary to fact that there was a real effect in a negative direction as visualized in figure 1A, which was not obvious in the customary treatment mean presentation in figure 1B. In the conventional NHST practice, the substantive value of this treatment effect cannot be determined because it is simply using probability in the binary decision making. Furthermore, that decision making is dependent on the statistical power of NHST which is also dependent on sample size, and small sample size would render NHST to have low power that may fail to detect real differences of treatment means. This weakness is inherent in NHST, and is widely known in literature (Aycaguer et al 2010; Balluerka et al 2009; Cumming & Fidler 2009; Fidler et al 2004; Levine et al 2008).

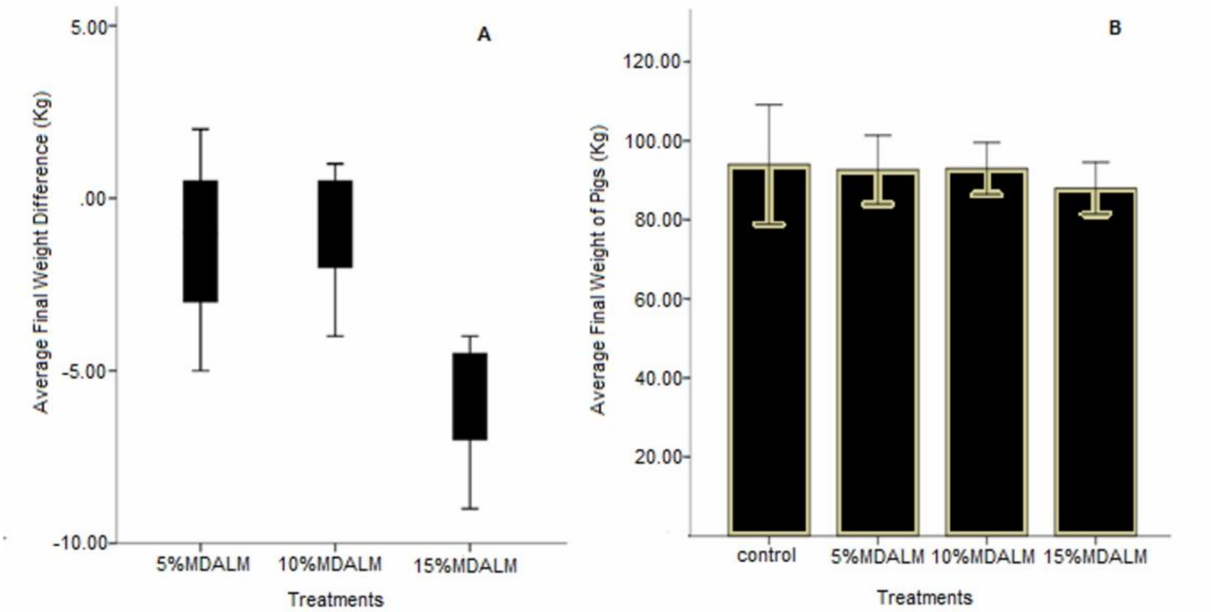

Figure 1. Two visualization alternatives showing response of pigs to different levels of Madre de Agua feed treatments. A) Confidence intervals of treatment effects (50 % bars and 95 % lines) when compared to control after reanalysis of data from Jaya (2006). B) The conventional bar graph for the average final weight of pigs with whiskers showing the standard deviations.

On the other data set, Madre de Agua feed treatment revealed to have highly significant results ($F(3,8) = 447.7$, $p = .000$) on the feed cost per pig, and that treatment mean comparison by LSD (Least Significant Difference test) also showed significant differences. Although this statistic provided corroboration to the theoretical answers to the research question (Cortina & Dunlap 1997), but this significant result in NHST was only telling us that the probability was very small for the treatment mean differences to occur by chance, based on the arbitrary cut-off of 5 % or 1 %. It did not however reveal the substantive size of effect, with respect to the original unit of measurement, which is important information to add and to complement NHST.

The proposed solution, for the readers to see the meaningfulness and practical value of the treatments, is to estimate the variability of the data accounted for by the treatments – which is called the omega (ω). Omega represents a measure of effect size, which theoretical underpinning for its calculation was reported by a number of workers (Nakagawa & Cuthill 2007; Volker 2006; Field 2005; Olejnik & Algina 2000; Cohen 1988). The square of omega (ω^2) multiplied by 100 represents the magnitude of treatment effect in percent. In cases when the omega squared is negative, Maxwell & Delaney (2004) and Olejnik & Algina (2000) suggested to consider this as zero.

Equation 1 is an algorithm to estimate effect size, which was revised from Volker (2006) and Field (2005), followed by a demonstration in figure 2 estimating the magnitude of effect of Madre de Agua treatment on the final weight response of pig and on the feed cost per pig. The needed information for the calculation to proceed (sum of squares, degrees of freedom and mean square) was taken directly from the ANOVA table 1 and table 2.

Equation 1: $$\omega^2 = \frac{SStreatment - dftreatment(MSerror)}{SStotal + MSerror}$$

$$\omega^2 \text{ (final weight)} = \frac{64.25 - 3(15.883)}{190.917 + 15.883} \quad \frac{16.601}{206.8}$$

$$\omega^2 \text{ (feed cost)} = \frac{148011 - 3(110.201)}{148893 + 110.201} \quad \frac{147680.694}{149003.201} \quad 0.9911$$

Figure 2. Demonstration for the calculation of omega squared as a measure of magnitude of Madre de Agua treatment effect on the average final weight of pigs and feed cost.

Table 1

Result of the analysis of variance for the average final weight of pigs

Source of variation	*Degrees of freedom*	*Sum of square*	*Mean square*	*F value*	*Probability (Same)*
Treatment	3	64.25	21.417	1.353	0.3246
Error	8	126.667	15.883	-	-
Total	11	190.917	-	-	-

Table 2

Result of the analysis of variance for the feed cost per pig

Source of variation	*Degrees of freedom*	*Sum of square*	*Mean square*	*F value*	*Probability (Same)*
Treatment	3	148011	49337	447.7	0.000
Error	8	881.61	110.201	-	-
Total	11	148893	-	-	-

The demonstration showed that although significant result was not apparent in the final weight of pigs (Table 1), but the 8 % effect of feeding Madre de Agua was practically large, important and cannot be ignored (Kotrlik et al 2011; Kotrlik & Williams 2003; Kirk 1996). Substantial recommendations may appropriately be given considering the fact that Madre de Agua contains Saponin that may render the feed less palatable to pigs (Cheeke & Shull 1985). This explains the downtrend pattern observed in figure 1A.

The demonstration also showed that Madre de Agua treatment had much more important effect on the feed cost per pig (99.11 %). The practical meaning and importance of this substantive value is that 99.11 % of the total variability in the response variable (i.e. feed cost) is attributed to treatment. This means further that from PhP 116.00 average differences of means visualized in figure 3B, the PhP 115.00 (PhP116.00 x .9911 = PhP 115.00) of it can be attributed directly to treatments. The PhP 1.00 excess could be due to unknown and confounding factors.

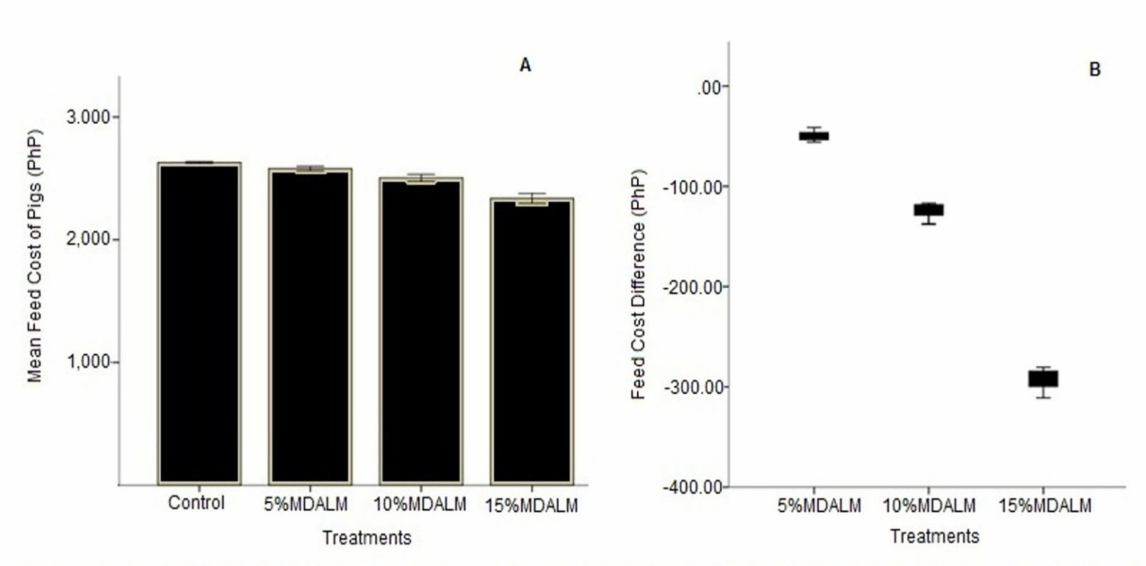

Figure 3. Two visualization alternatives for Madre de Agua effects on feed cost per pig. A) Average feed cost per pig per treatment with whiskers as standard deviations. B) Confidence intervals of treatment effects (50 % bars and 95 % lines) when means are compared.

The PhP 116.00 average differences of treatment means was actually the standard deviation which can be calculated directly from ANOVA table 2. The treatment sum of square in ANOVA table 2, when divided by the total degrees of freedom (total df = 11) gave the average sum of square - the variance. This value represents the average difference among treatment means including control, but the unit was peso squared (PhP^2) because at the outset in the computation of variance, the mean differences was squared to avoid the zero sum problem (Field 2005). To make the unit sensible, the square root of this variance was extracted to bring the original peso unit, and thus the standard deviation of PhP 166.00.

Figure 3B gives the alternative visualization method to highlight the treatment mean differences, and to bring closer the meaning of the standard deviation. The conventional method depicted in figure 3A was already good for treatment mean comparison, but the alternative visualization clearly gave better resolution. Together with the visualization technique, it is important to include in the reporting of results the followings: computed F value, degrees of freedom, the exact P value and the omega squared (ω^2) value. For example, analysis of variance of treatment effect on feed cost gave highly significant result ($F(3,8) = 447.7$, $p = .000$ and $\omega^2 = .9911$). The inclusion of ω^2 in the report will give readers the idea about the substantive value of the treatment effect.

Conclusions. A problem on estimating the magnitude of treatment effect in agricultural experiment was solved. The analysis of variance or NHST was not enough, and therefore it is proposed to bring the data analysis a step further by estimating the magnitude of treatment effect – the omega square (ω^2). This is the key value to the substantive and practical importance and meaningfulness of the treatment. The ω^2 is a unit less measure of treatment effect and therefore has practical use in the meta-analysis *i.e.,* making comparison of similar studies.

Acknowledgements. I thank to Dr. Alex F. Jaya for his wisdom and generosity by allowing the re-analysis on some part of his dissertation data. My sincere thanks to my colleague at Xavier University Prof. Nancy P. Cocamas for her time proof reading this manuscript.

References

APA - American Psychological Association, 2001 Publication manual of the American Psychological Association, 5th ed, Washington D.C., American Psychological Association.

Aycaguer L. C. S., Gil P. S., Somoano A. F., 2010 The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. BMC Med Res Methodol 10(44):1-9.

Balluerka N., Vergara A. L., Arnau G. J., 2009 Calculating the main alternatives to null-hypothesis-significance testing in between-subject experimental designs. Psicothema 21:141-151.

Cheeke P. R., Shull L. R., 1985 Natural toxicants in feed and livestock. Connecticut, AVI Publishing Inc.

Cohen J., 1988 Statistical power analysis for the behavioral sciences. 2nd ed, Hillsdale, New Jersey, Lawrence Erlbaum pp. 1-559.

Cohen J., 1994 The earth is round (P<.05). Am Psychol 49(12):997-1003.

Cortina J. M., Dunlap W. P., 1997 On the logic and purpose of significance testing. Psychol Methods 2(2):161-172.

Cumming G., Fidler F., 2009 Confidence intervals: Better answers to better questions. J Psychol 217(1):15-26.

Fidler F., Burgman M. A., Cumming G., Buttrose R., Thomason N., 2006 Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. Conserv Biol 20(5):1539-44.

Fidler F., Thomason N., Cumming G., Finch S., Leeman J., 2004 Editors can lead researchers to confidence intervals but they can't make them think: Statistical reform lessons from Medicine. Psychol Sci 15:119-126.

Field A., 2005 Discovering statistics using SPSS. 2nd ed, London, SAGE Publication, p. 357.

Hammer O., Harper D. A. T., Ryan P. D., 2001 PAST: Paleontological Statistics Software Package for Education and Data Analysis. Paleontologia Electronica 4(1):1-9.

Jaya A., 2006 Utilization of Madre de Agua (*Trichanthera gigante var. guianensis*) leaf meal as feed for growing-finishing pigs (*Sus domesticus*). PhD Thesis, Central Mindanao University, Philippines.

Kirk R. E., 1996 Practical significance: A concept whose time has come. Educ Psychol Meas 56:746-759.

Kotrlik J. W., Atherton J. C., Williams H. A., Jabor M. K., 2011 Reporting and interpreting effect size in quantitative agricultural education research. J Agric Educ 52(1):132-142.

Kotrlik J. W., Williams H. A., 2003 The incorporation of effect size in information technology, learning and performance research. Information Technology, Learning and Performance Journal 21(1):1-7.

Levine T. R, Weber R., Hullett C., Park H. S., Lindsey L. M., 2008 A critical assessment of null hypothesis significance testing in quantitative communication research. Hum Commun Res 34:171-187.

Maxwell S. E., Delaney H. D., 2004 Designing experiments and analyzing data: A model comparison perspective. 2nd ed, New Jersey, Lawrence Erlbaum.

Nakagawa S., Cuthill I. C., 2007 Effect size, confidence interval and statistical significance: a practical guide for biologist. Biol Rev Camb Philos Soc 82(4):591-605.

Olejnik S., Algina J., 2000 Measures of effect size for comparative studies: Applications, interpretations, and limitations. Contemp Educ Psychol 25:241-286.
Rinella M. J., James J. J., 2010 Invasive plant researchers should calculate effect size, not P-Values. Invasive Plant Sci Manag 3:106-112.
Sun S., Pan W., Wang L. L., 2010 A comprehensive review of effect size reporting and interpreting practices in academic journals in education and Psychology. J Educ Psychol 102(4):989-1004.
Volker M. A., 2006 Reporting effect size estimates in school psychology research. Psychol Sch 43(6):653-672.

Received: 30 January 2014. Accepted: 26 February 2014. Published online: 20 March 2014.
Author:
Dennis Alalim Apuan, Xavier University, College of Agriculture, Department of Agricultural Sciences, Philippines, Cagayan de Oro City, e-mail: dennis_apuan@yahoo.com

How to cite this article:
Apuan D. A., 2014 Measuring magnitude of treatment effects in agricultural experiments. AAB Bioflux 6(1):33-39.