Confidence Levels in Statistical Analyses. Analysis of Variances. Case Study.

Ileana Brudiu

USAMVB Timisoara, 300645, Calea Aradului 119, ROMANIA

Abstract
Applying a statistical test to check statistical assumptions offers a positive or negative response regarding the veracity of the issued hypothesis. In case of variance analysis it's necessary to apply a post hoc test to determine differences within the group. Statistical estimation using confidence levels provides more information than a statistical test, it shows the high degree of uncertainty resulting from small samples and builds conclusions in terms of "marginally significant" or "almost significant (p being close to 0,05) . The case study shows how the statistical estimation completes the application form the analysis of variance test and Tukey test.

Keywords: Analysis of variance, Tukey test, statistical estimation, confidence intervals, graphic representation

1. Introduction

Biological processes may be located at any given time under the influence of several factors, acting at the same time. To highlight the extent of how one or more factors (or a combination of factors) influence essentially a resultative characteristic, the dispersion analysis or analysis of variance, as it is also known, is used. By this method it is checked the extent to which actual values of the characteristics deviate from theoretical values, calculated usually in the form of medium or regression equations and at what extent, these changes depend or not on the grouping factor [1].
Analysis of variance is used when wishing to compare more than two environments, it is a parametric method and it protects the researcher from the "inflation of errors" that could occur by using other tests. The present survey aims to show how the use of confidence levels complete the analysis variances conclusions. Estimating parameters through intervals can be calculated better when trying to capture the parameter unlike the punctual estimation which offers as a result the one single point that is the most likely a parameter representant.

2. Materials and methods

The survey contains data from the used combined fodder (CN kg / batch), protein (g/hen/day), lysine (g/hen/day) and methionine (g/hen/day) in four batches of laying hens during 8 weeks. The lots were composed of 100 chickens Hybrid TETRA SL. Observations (measurements) were made weekly. In this situation a lot has 8 observations; they derived from an average weekly consumption. Lots are considered small. Obtaining a statistically significant result on a small lot can be considered an important result taking into consideration the correlation between the research conclusions and test results.
In order to analyse, the observed data the following were calculated:
1. The probability that all 4 lots have an equal consumption average
2. For each type of (CN, protein, lysine and methionine) consumption, the Tukey test was applied to highlight which lots (L1, L2, L3, L4) differ as an average.
3. There have been estimated the probability of 95% confidence levels of average parameter.

* Corresponding author: Ileana Brudiu,
Tel:0256277030, Email:brudiu@yahoo.com

4. Confidence intervals were plotted
5. It has been represented graphically using the Box plot type, consumption levels for the four lots.

For statistical calculations we used statistical tests implemented in Excel and Minitab [2]
Also in the graphic representation it was used type Box plot, which is the most simple and complete instrument for operating in data analysis

3. Results and discussion

Probabilities obtained by applying ANOVA: Single Factor are shown in Table 1.

Table 1. P-value Anova Single factor

Anova: Single Factor	P-value Between Groups
Consumption NC (kg / batch)	0.794802688
Protein consumption (g/hen/day)	3.0468E-05
Lysine consumption (g/hen/day)	1.64246E-11
Methionine consumption (g/hen/day)	1.73928E-14

From the analysis of the obtained data it can be observed the existence of significant differences between lots, only in the consumption of protein, lysine and methionine. CN consumption has no significant differences between analyzed lots [3] [4]. Applying the Tukey test [5], highlights the lots which have these differences. Results are shown in Table 2.

In Table 3 are shown the values obtained for the confidence levels (95.0%) of the average parameter of all the four lots for the analyzed consumption.

Table 2. Contrast value Tukey test

Consumption NC (kg/batch)		Protein consumption (g/hen/day)		Lysine consumption (g/hen/day)		Methionine consumption (g/hen/day)	
Contrast	Significant	Contrast	Significant	Contrast	Significant	Contrast	Significant
3 vs 2	No	1 vs 2	Yes	4 vs 3	Yes	4 vs 3	Yes
3 vs 1	No	1 vs 4	Yes	4 vs 2	Yes	4 vs 2	Yes
3 vs 4	No	1 vs 3	Yes	4 vs 1	Yes	4 vs 1	Yes
4 vs 2	No	3 vs 2	No	1 vs 3	Yes	1 vs 3	Yes
4 vs 1	No	3 vs 4	No	1 vs 2	No	1 vs 2	No
1 vs 2	No	4 vs 2	No	2 vs 3	Yes	2 vs 3	Yes

Table 3. Confidence Level (95%)

	L1	L2	L3	L4
Consumption NC (kg / batch)	[21.83-22.75]	[21.83-22.61]	[21.94-23.04]	[21.90-22.84]
Protein consumption (g/hen/day)	[20.21-21.06]	[18.96-19.64]	[19.05-20.02]	[19.02-19.84]
Lysine consumption (g/hen/day)	[960.48-1000.83]	[935.60-969.08]	[877.56-921.84]	[1026.08-1070.19]
Methionine consumption (g/hen/day)	[473.99-493.91]	[474.03-491.00]	[426.25-447.75]	[525.55-548.15]

We plotted the confidence levels of average consumption for the four lots in Figure 1. We noticed that confidence levels overlap only in the consumption of protein, lysine and methionine, which were marked with ***. Confidence intervals were calculated with the probability of 95%.

Confidence levels are very sensitive to sample variability. For data analysis, coefficient of variance is between 1,7 and 2,5%. But confidence levels can be quite restricted and in case of the existence of a large variability of the data only if the sample is higher, which is not the case in this example analysis. For this data, the differences are not large, the obtained significant differences for small lots reinforce the conclusion that the obtained result is important.

Confidence levels provide more information than a statistical test. A confidence level can be seen as a rejection of all null hypotheses associated with higher value or lower confidence limits, and so it can be considered as the summarizing of several statistical tests. If the purpose of the study is to reveal some significant differences regarding consumption between groups, ANOVA-test is not sufficient; it is absolutely necessary the post hoc analysis, highlighting the groups that differ on average.

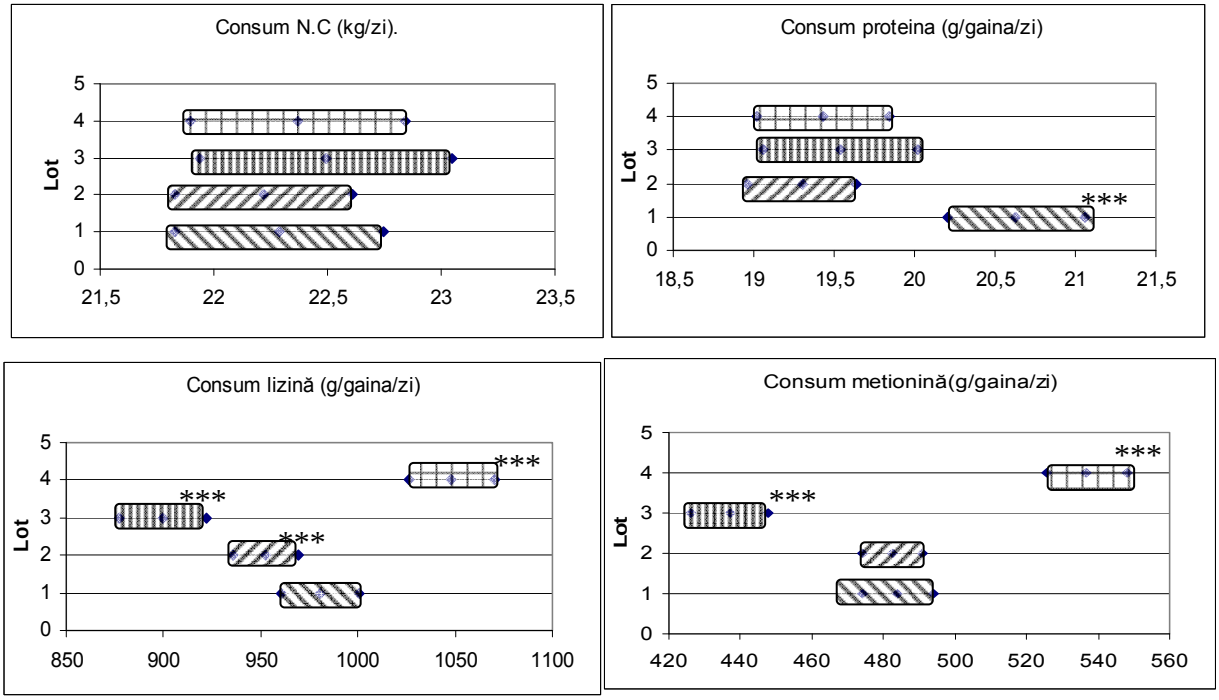

Figure 1. Confidence intervals of average consumption for the four lots.

In Figure 2 the consumptions were plotted for the 4 groups using Box plot type. This type of representation is similar to that of the confidence levels, but not conclusive in the data analysis. The diagrams of box plot type reflect the 5 values of a distribution graphically: minimum, first quartile, median, third quartile and maximum value, but cannot estimate the parameter in terms of exploring its behavior to the entire population and they cannot be used in prediction.

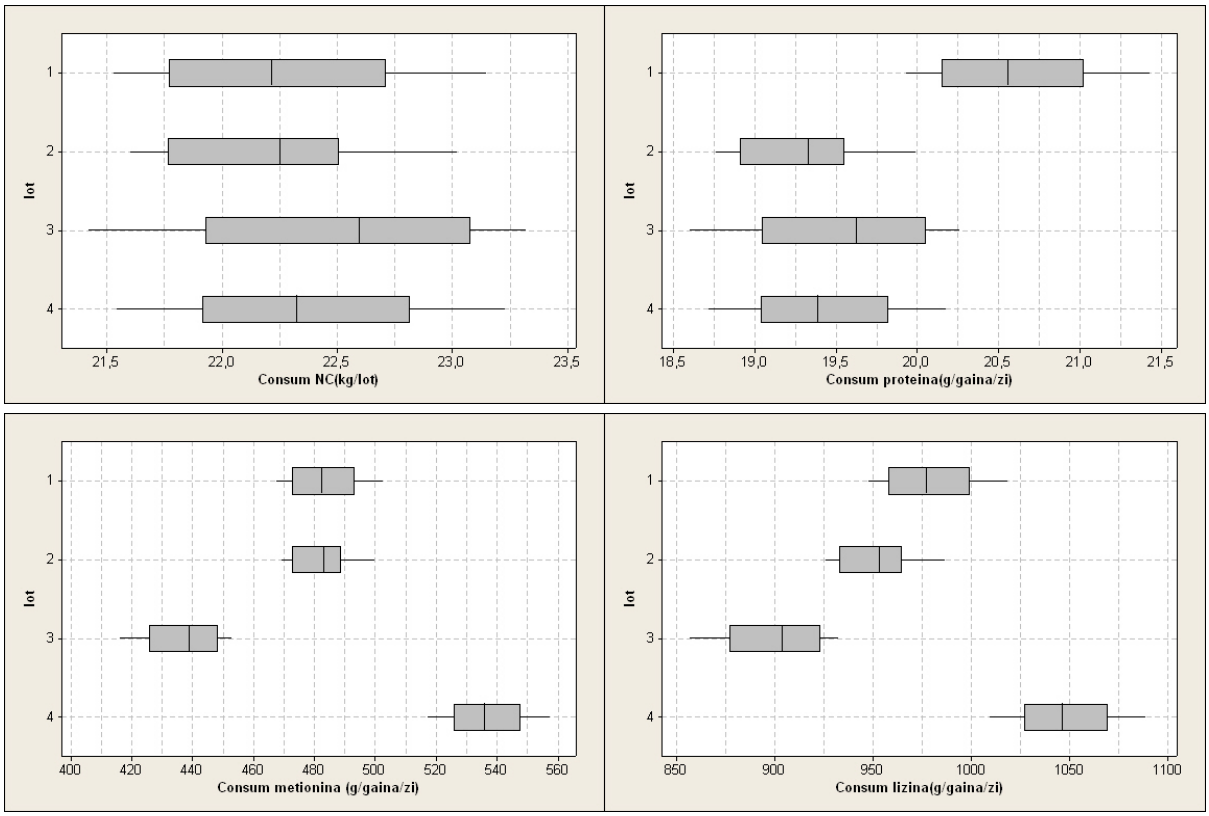

Figure 2 Fodder consumtion for four grups using Box plot type

4. Conclusions

The variance analysis is used when wishing to compare more than two environments. It provides an answer with an error margin regarding the existence of significant differences between the environments of the analyzed characteristics. To see clearly where significant differences exist, a post hoc test should be applied. Using confidence levels provide the answer. If confidence levels are plotted, no duplication draws the attention on those who create these differences. Box plot graphic charts reflect the distribution characteristics but cannot be used in prediction.

References

1. Micheal Longnecker, Rober Ott, An introduction to statistical methods and data analysis, ISBN-13: 9780495109143 ISBN-10: 0495109142, CENGAGE LEARNING (United States)
2. http://www.minitab.com/en-RO
3. De Gudmund R. Iversen,Helmut Norpoth, Analysis of variance,1987, ISBN-10:0803930011,
4. Snedecor, G.W. and Cochran, W.G. (1989), Statistical Methods, 8th Edition, Ames, IA: Iowa State University Press
5. John W. Tukey, One Degree of Freedom for Non-Additivity, *Biometrics*, Vol. 5, No. 3 (Sep., 1949), pp. 232-242, Published by: International Biometric Society, Stable URL: http://www.jstor.org/stable/3001938