Tests assessment for multiple comparisons of *in vitro* gas curves, from the root of the mean square distance

O. Jay[1], Verena Torres[2], Yoandra Marrero[2] and P. Torres[3]

[1]*Centro de Estudios de Especies Menores. Universidad de Guantánamo. Carretera al Salvador, km. 6½. Guantánamo.*
[2]*Instituto de Ciencia Animal. Apartado Postal 24. San José de las Lajas. Mayabeque. Cuba*
[3]*Centro Universitario de Ciencias Biológicas y Agropecuarias. Zapopán, Jalisco. México*
Email: ojay@ica.co.cu

In order to find a tool for comparing multiple curves of *in vitro* gas production to assess ruminants' feeds, the square mean distance between curves is proposed as statistical distance. The performance of four comparison tests of fixed range using this distance: lowest significant distance (LSD), Tukey's honest significant distance (HSD), Scheffé's significant distance (SSD) and Bonferroni-corrected significant distance (BSD) through four experiments simulated by Monte Carlo. The signification level of the tests is also analyzed when the number of treatments or the variation between the simulation parameters increases. It was proved that the Tukey test was stable and limited the inflation, the error type 1, when the number of treatments increases. Besides, it is the most adequate in multiple comparisons of this kind of experiment.

Key words: *Monte Carlo simulation, non-linear regression, post hoc tests.*

The non-linear regression curves, obtained from experiments of *in vitro* gas production, allow characterizing the dynamics of the ruminal degradation processes. This technique, among other advantages, is not so invasive, not expensive and includes a great amount of experimental units. Besides, modern technologies for obtaining data are also included (Adesogan *et al*. 2005 and Rymer *et al*. 2005).

Several models have been proposed for analyzing the results of these experiments in order to obtain characterized parameters of the degradation dynamics (France *et al*. 2005). It is of great interest to find one for comparing the treatments, once the model describing this process is obtained.

Different tests of curve homogeneity have been sugested for this purpose, through conducting hypothesis test son the equality of all treatments in one experiment. Among them, the Akaike Information Criterion has demonstrated a good performance in the biological experiments (Motulsky and Christopoulos 2003).

Once the homogeneity test indicates the rejection of the null hypothesis, a more detailed procedure is needed to exam the equality among curves, named post hoc tests. A statistical distance is needed for conducting these tests to compare using the comparison tests.

In this study, from proposing the square mean distance as a statistical one for comparing multiple curves, the performance of four comparison tests of fixed range in the formation of the groups treatments was analyzed.

Materials and Methods

The information used for the simulation by Monte Carlo comes from an experiment of *in vitro* gas production conducted in the Physiology Department of the Institute of Animal Science of Cuba and the Laboratories of Molecular Microbiology and Chemistry, of the Physiology and Animal Nutrition Program of the Colombian Corporation of Agricultural Investigation (CORPOICA) (Marrero 2005).

The original experiment had four repetitions, of 70 measurements each, conducted in intervals of 0.34 h. It was also considered that, at the beginning of the measurements (t=0), the gas production is null, representing a total of 284 experimental points. Four experiments were simulated, which treatments are made with modifications in the parameters according to the method of

Rubinstein and Kroese (2008). Each simulation was repeated 20 times. For the modification to the parameters, each variant is percent modification of i order of the original parameter, according to the following equation:

$$\theta\_i = \left(\frac{100+i}{100}\right)\theta\_0 \qquad 1$$

Where:

$\theta_{\_0}$: parameter obtained from the regression of the original data

$\theta_{\_i}$: parameter used in the simulation

The parameters variation range was taken from the values where the curve homogeneity test indicated the rejection of the null hypothesis:

Simulation 1. The parameters used in the simulation are modified varying in 5% those original (table 1).

Simulation 2. From the original values of the parameters, each of them varied 10 % (table 2).

Simulation 3. The treatments of simulations 1 and 2 were included simultaneously for seven treatments to be compared (table 3).

The distance to be proposed need to have the independence of the number of treatments to be compared as a characteristic, as well as the uniformity of the results, apart from the samples selected for the analysis.

Table 1. Parameters of simulation 1

Parameters	Variants			
	T_0	A_5	b_5	c_5
A	A_0	$1.05A_0$	A_0	A_0
b	b_0	b_0	$1.05b_0$	b_0
c	c_0	c_0	c_0	$1.05c_0$

Table 2. Parameters of simulation 2

Parameters	Variants			
	T_0	A_10	b_10	c_10
A	A_0	$1.1A_0$	A_0	A_0
b	b_0	b_0	$1.1b_0$	b_0
c	c_0	c_0	c_0	$1.1c_0$

Table 3. Parameters of simulation 3

Parameters	Variants						
	T_0	A_10	b_10	c_10	A_5	b_5	c_5
A	A_0	$1.1A_0$	A_0	A_0	$1.05A_0$	A_0	A_0
b	b_0	b_0	$1.1b_0$	b_0	b_0	$1.05b_0$	b_0
c	c_0	c_0	c_0	$1.1c_0$	c_0	c_0	$1.05c_0$

Curves comparisons were conducted for each experiment, considering, as distance measurement, the mean square distance between the regression curves, according to the formula:

$$d_{j,k} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\hat{f}_j(t_i) - \hat{f}_k(t_i)]^2} \qquad 11$$

Where:

n: number of points used for the curve adjustment

$\hat{f}_j(t_i)$ $\hat{f}_k(t_i)$: regression functions, adjusted for the treatment j and k, respectively in the point ti

t_i: i-th measurement of the independent variable (time)

Likewise, dj,k tends to zero only when there are no differences between the curves and their value is at about equal to the average of the absolute value of the differences between the adjusted values of the curves in the interval used for the adjustment.

The distance for each repetition was calculated in each experiment. The ratio between the statistical difference, difference between treatments and between the statistical distance and the number of multiple comparisons were studied.

In the experiments of gas production, introducing treatments with small variation of the parameters is frequent (Dhanoa *et al*. 2000). These variations may be due to affections of the microbial population or to changes in the feeding degradation dynamics. Therefore, the multiple comparison test used in these experiments should be as sensitive as possible in order to form the groups corresponding to these small variations.

In this type of experiments, the variability level found in the experimental units submitted to a same treatment may be considered. They may include the random dispersion of the curves obtained (Liu 2011). Therefore, an agreement between these two aspects is important when selecting the multiple comparison tests, so the proper effectiveness of the test in this type of experiment is achieved.

Those multiple comparison tests of fixed range more used in the biological researches were selected (Toutenbourg 2002 and Montgomery 2004). The signification level selected in all cases was of P < 0.05:

1. Lowest significant distance (LSD)
2. Tukey's honest significant distance (HSD)
3. Scheffé's significant distance (SSD)
4. Bonferroni-corrected significant distance (BSD)

These tests were calculated in each repetitions of the experiments simulated. Later, the mean and the standard deviation of the tests statistics were estimated and the performance of each criterion test was studied for accepting or rejecting the hypothesis equality or not of the curves.

Microsoft Excel 2007 (MicroSoft_Corporation 2008) and SPSS 15.0 (SPSS_Inc. 2006) were used for the simulation and the statistical analysis.

Results and Discussion

Figure 1 shows the box diagram of the distances estimation between the treatment curve 0 and the rest of them for the percent variations of the parameters. The percent variations of the parameters have been placed in the horizontal axis, according to the percent modification (1). The mean square distance dj,k was placed in the vertical axis (2).

For the modifications in parameter A, the highest distances between curves were found, with values from 8.87 to 11.26 for variations of 10 %, and from 3.56 to 6.01 for variations of 5 %. The parameter B had the lowest distance between curves for each percent variation of the parameter, being the lowest variation between 0.32 and 3.26, for variations of 10 %. The distances dispersion between the curves was similar between the variations of a parameter and another.

Figure 2 shows the performance of the tests LSD and HSD in the four simulations.

The estimated means of the LSD test had the lowest values for the rejection for the hypothesis of means equality. This was due to the statistical calculation of the

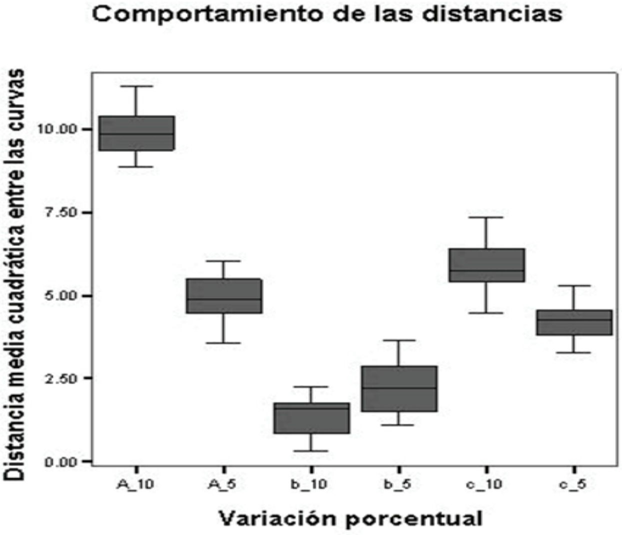

Figure 1. Performance of the distance between the treatment 0 and the rest of the curves in relation to the parameters value of variations increase.

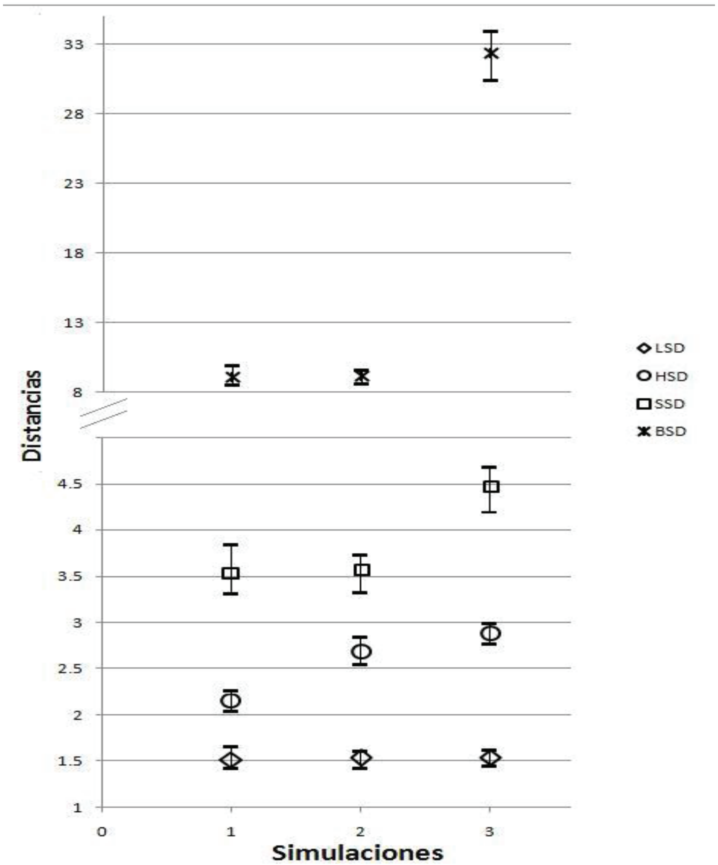

Figure 2. Performance of the statistics of the multiple comparison tests in each simulation.

test, where there is inflation of the error probability type 1, as the number of treatments increase (Yosihara 2009).

The estimated means of the HSD test had, for all simulations, higher values compared to those of LSD. This makes the HSD test more conservative, due to the coefficients of the T-student ranges in the calculation, to avoid the inflation of the error probability type 1(Toutenbourg 2002 and Gamst *et al*. 2008) .

As the SSD depends directly on the Fisher (F) statistics multiplied by the freedom degrees, as they increase, the test value will also increase. Therefore, the statistic value of the test for the rejection of the mean equality hypothesis would be higher as the amount of treatments increases. However, even when the test statistics increases in the experiments with their increases, the dispersion in the estimations keeps stable in every experiment.

The BSD test had similar performance to that of the SSD, but with higher dependence on the number of treatments, due to the multiplication for the treatments combinations. As it had the higher estimations of the rejection statistics of means equality, it was the more conservative test of those studied. It was the one used the most in experiments, where being conservative for the treatments comparison is necessary (Montgomery 2004).

The estimation values of the test means and the mean distances between the curves are following presented, calculated for each simulation. The test performance in accepting or rejecting the null hypothesis of means equality in every comparison is also analyzed.

Results of simulation 1 (percent variation of 5 %). Table 4 shows the average of the distances estimated between the curves and the mean values of the tests corresponding to simulation 1. As the distances matrix is symmetric, only the values of the superior diagonal appear.

For accepting or rejecting the null hypothesis, each distance between curves is compared to the statistics of the corresponding test. The null hypothesis was rejected, when the distance between curves was higher than the statistic value of the test. The comparisons were conducted for each pair of curves. Six comparisons were conducted in this simulation.

The value of the LSD test was inferior to the distance between each pair of curves, indicating significant differences between all curves. Meanwhile, the mean estimation of the BSD test, for being the highest of all the distances between curves, indicated the formation of only one group and was the most conservative test.

The estimation of the HSD mean indicated that there were no significant differences between treatments T_0 and b_5. This allowed the formation of three groups. The T_0 and b_5 treatments were together and the others were in different groups.

The estimation of the SSD mean was higher than the distance between the curves T_0 and b_5, T_0 and c_5 and, finally between A_5 and c_5. The null hypothesis is not rejected between these curves, so there were no significant differences between them. Therefore, the following groups were created: T_0 and b_5, T_0 and c_5 and, finally, A_5 and c_5. The results of these comparisons are shown in table 5.

Results of simulation 2 (percent variation of 10 %). Table 6 shows the results of the means estimations of the multiple comparison tests for simulation 2.

As the distances between the curves in this simulation were higher than those of simulation 1, the LSD mean was lower than all the distances calculated between the curves. Therefore, it continued forming different groups. The estimated mean of the HSD test was also inferior to the distances between curves. The estimated mean of the SSD test was higher was higher than the distance estimated between the curves T_0 and b_10, so these two treatments are gathered in one group and the rest in different groups. With the BSD test, still the most conservative, no significant differences were found between the curves corresponding to the treatments T_0 and b_10, T_0 and c_10 and A_10 and c_10. Table 7 shows the results of these multiple comparisons.

Result of simulation 3 (includes the treatments of simulations 1 and 2). The results of simulation 3 are shown in table 8. In this last simulation, the statistics of LSD test kept being the lowest value. Hence, significant differences were found in all curves, forming groups for

Table 4. Average of the statistics of the multiple comparison tests and of the distances between the curves estimated for simulation 1

Tests*	Estimated means	Standard deviation	Distances between the curves				
				T_0	A_5	b_5	c_5
LSD	1.44	0.08	T_0		4.97	2.36	3.47
HSD	2.69	0.11	A_5			6.34	3.35
SSD	4.76	0.19	b_5				5.42
BSD	8.61	0.45	c_5				

*the tests were estimated for the statistic for a signification level of $P < 0.05$

Table 5. Results of the multiple comparisons of each test for simulation 1

	Multiple comparison tests			
	LSD	HSD	SSD	BSD2
Curves to compare	T_0^{a}	T_0^{a}	T_0^{ab}	T_0^{a}
	A_5^{b}	A_5^{b}	A_5^{c}	A_5^{a}
	b_5^{c}	b_5^{a}	b_5^{a}	b_5^{a}
	c_5^{d}	c_5^{c}	c_5^{bc}	c_5^{a}

^{a,b,c} Different letters correspond to significant differences $P < 0.05$ according to the corresponding test

Table 6. Average of the distances between the estimated curves for simulation 2

Tests	Means	Standard deviation	Distances				
				T_0	A_10	b_10	c_10
LSD	1.44	0.07	T_0		9.91	3.72	6.03
HSD	2.15	0.08	A_10			12.59	7.11
SSD	3.81	0.14	b_10				9.54
BSD	8.68	0.41	c_10				

Table 7. Results of the multiple comparisons of the curves of simulation 2

	Multiple comparison tests			
	LSD	HSD	SSD	BSD
Curves to compare	T_0[a]	T_0[a]	T_0[a]	T_0[ab]
	A_10[b]	A_10[b]	A_10[b]	A_10[c]
	b_10[c]	b_10[c]	b_10[a]	b_10[a]
	c_10[d]	c_10[d]	c_10[c]	c_10[bc]

[a,b,c,d] Different letters correspond to significant differences $P < 0.05$ according to the corresponding test

each treatment, seven in total. The HSD test did not find significant differences between the curves corresponding to treatments T_0 and b_5, in agreement with the results of simulation 1. No significant differences were found either between the curves of treatments A_5 and c_10. The estimated mean of the statistics of SSD test did not find significant differences between the curves corresponding to the comparison between the treatments T_0, with b_10, A_5, b_5 and c_5; A_10 with b_10, b_5 and c_5; c_10 with A_5 and b_5 with c_5. The BSD test grouped all the treatments, because the number of combinations increases the test value which was, in this case, of 30.40. Table 9 shows the comparisons summary.

With the results, the LSD test is rejected for being so excluding and the BSD for conservative. None of them showed capacity for forming different groups in the parameter variations present in the experiments of gas production.

Table 8. Average of the distances between the curves estimated for simulation 3

Tests	Means	Standard deviation	Distances							
				T_0	A_10	b_10	c_10	A_5	b_5	c_5
LSD	1.44	0.06	T_0		9.53	3.82	6.40	5.36	2.18	3.69
HSD	2.88	0.09	A_10			4.80	6.65	8.18	4.04	4.84
SSD	5.51	0.17	b_10				11.31	12.76	8.07	7.26
BSD	30.40	1.28	c_10					2.06	5.56	8.33
			A_5						7.17	9.96
			b_5							3.01

Tabla 9. Results of the multiple comparisons of the curves of simulation 3

	Multiple comparison tests			
	LSD	HSD	SSD	BSD
Curves to compare	T_0[a]	T_0[a]	T_0[abc]	T_0[a]
	A_10[b]	A_10[b]	A_10[de]	A_10[a]
	b_10[c]	b_10[c]	b_10[a]	b_10[a]
	c_10[d]	c_10[d]	c_10[f]	c_10[a]
	A_5[e]	A_5[d]	A_5[bf]	A_5[a]
	b_5[f]	b_5[a]	b_5[ce]	b_5[a]
	c_5[g]	c_5[e]	c_5[ce]	c_5[a]

As the treatments increased, the SSD test also augmented. It became inefficient when the comparison of more than 10 treatments took place in the experiment. Therefore, it is only recommended for experiments with less than 10 treatments.

The HSD test was efficient in experiments of gas production, no matter the amount of treatments, even more than 10.

References

Adesogan, A. T., Krueger, N. K. & Kim, S. C. 2005. A novel, wireless, automated system for measuring fermentation gas production kinetics of feeds and its application to feed characterization. Animal Feed Sci. and Tech. 123-211

Dhanoa, M. S., Lopez, S., Dijkstra, J., Davies, D. R., Sanderson, R., Williams, B. A., Sileshi, Z. & France, J. 2000. Estimating the extent of degradation of ruminant feeds from a description of their gas production profiles observed *in vitro*: comparison of models. British J. Nut. 83: 131

France, J., Lopez, S., Kebreab, E., Bannik, A., Dhanoa, M. S. & Dijkstra, J. 2005. A general compartmental model for interpreting gas production profiles. Anim. Feed Sci. Tech. 123: 473

Gamst, G., Meyers, L. S. & Guarino, A. J. 2008. Analysis of variance designs. Cambridge University Press. Cambridge. 596 pp.

Liu, W. 2011. Simultaneous inference in regression. Chapman & Hall. Boca Ratón. 292 pp.

Marrero, Y. 2005. Las levaduras como mejoradores de la fermentación ruminal de dietas con alto contenido de fibra. PhD Thesis. Instituto de Ciencia Animal. La Habana. Cuba

MicroSoft_Corporation. 2008. Microsoft Office Excel 2007 SP 2.

Montgomery, D. C. 2004. Diseño y Análisis de Experimentos. Limusa-Wiley. México D. F. 692 pp.

Motulsky, H. J. & Christopoulos, A. 2003. Fitting models to biological data using linear and nonlinear regression. Graph Pad Software. San Diego. 351 pp.

Rubinstein, R. Y. & Kroese, D. P. 2008. Simulation and the Monte Carlo method. John Wiley & Sons Inc. New Jersey. 377 pp.

SPSS. 2006. Statistical Package for the Social Sciences. SPSS Inc. Chicago Ill

Toutenbourg, H. 2002. Statistical analysis of designed experiments. Springer-Verlag. New York. 517

Yoshihara, K. & Kanagawa, S. 2009. Change point problems in nonlinear regression estimation with dependent observations. Nonlinear Analysis 71: 2152

Received: March 26, 2012