

## Multivariate Analysis for Animal Selection in Experimental Research

Renan Mercuri Pinto<sup>1</sup>, Dijon Henrique Salomé de Campos<sup>2</sup>, Loreta Casquel Tomasi<sup>2</sup>, Antonio Carlos Cicogna<sup>2</sup>, Katashi Okoshi<sup>2</sup>, Carlos Roberto Padovani<sup>1</sup>

Departamento de Bioestatística - Instituto de Ciências Biológicas da Universidade Estadual Paulista (Unesp)<sup>1</sup>; Departamento de Clínica Médica - Faculdade de Medicina de Botucatu - Universidade Estadual Paulista (Unesp)<sup>2</sup>, Botucatu, São Paulo - Brazil

### Abstract

**Background:** Several researchers seek methods for the selection of homogeneous groups of animals in experimental studies, a fact justified because homogeneity is an indispensable prerequisite for casualization of treatments. The lack of robust methods that comply with statistical and biological principles is the reason why researchers use empirical or subjective methods, influencing their results.

**Objective:** To develop a multivariate statistical model for the selection of a homogeneous group of animals for experimental research and to elaborate a computational package to use it.

**Methods:** The set of echocardiographic data of 115 male Wistar rats with supra-avalvular aortic stenosis (AoS) was used as an example of model development. Initially, the data were standardized, and became dimensionless. Then, the variance matrix of the set was submitted to principal components analysis (PCA), aiming at reducing the parametric space and at retaining the relevant variability. That technique established a new Cartesian system into which the animals were allocated, and finally the confidence region (ellipsoid) was built for the profile of the animals' homogeneous responses. The animals located inside the ellipsoid were considered as belonging to the homogeneous batch; those outside the ellipsoid were considered spurious.

**Results:** The PCA established eight descriptive axes that represented the accumulated variance of the data set in 88.71%. The allocation of the animals in the new system and the construction of the confidence region revealed six spurious animals as compared to the homogeneous batch of 109 animals.

**Conclusion:** The biometric criterion presented proved to be effective, because it considers the animal as a whole, analyzing jointly all parameters measured, in addition to having a small discard rate. (*Arq Bras Cardiol.* 2015; 104(2):97-103)

**Keywords:** Multivariate Analysis; Animals; Epidemiology, Experimental; Aortic Valve Stenosis.

### Introduction

Due to lack of statistical knowledge, several researchers choose to use empirical or subjective decision-making methods, ignoring the casualization process, a basic principle for reliability of findings, which influences the results.

A very common example of that influence occurs in the process of sample homogenization, which is important for randomization in research involving animals as an experimental unit. Regarding that process, several researchers usually adopt a fragmented or intentional approach, using only a convenient parameter to classify the group as homogeneous, which results in a biased and inappropriate homogenization, in addition to favoring the possibility of discarding animals due to a simple spurious value rather than to biological dissimilarity.

From the biological viewpoint, the most interesting process of homogenization should jointly consider all parameters assessed in the experimental unit, because most of those are correlated, and the best way to understand the behavior of the animal consists in a set of numerical data representing all its biological characteristics. Thus, statistical methods that consider the animal as a whole rather than in a fragmented way are required, because the organism reacts as a whole to any intervention or treatment.

It is worth noting that, when there is dependence between variables, multivariate analysis of data should be used, because, as such, the significance index of inferential statistics is not inflated, a problem that occurs when several univariate analyses are performed simultaneously<sup>1</sup>.

This study's objective was to develop a multivariate statistical model for the selection of a homogeneous group of animals for experimental research that follows biological and statistical principles, in addition to elaborating a computational package to apply that model.

To exemplify the development and application of the criterion, the set of echocardiographic data of animals undergoing surgery for induction of supra-avalvular aortic stenosis (AoS), provided by the Research Group on Experimental

**Mailing Address: Renan Mercuri Pinto •**

Rua Thomaz Ceneviva, 117, Vila Anita. Postal Code 13484-295, Limeira, SP - Brazil

E-mail: renanmp@usp.br; renanmp@icloud.com

Manuscript received June 26, 2014; revised manuscript September 27, 2014; accepted September 30, 2014.

**DOI:** 10.5935/abc.20140219

Cardiology of the Botucatu Medical School (FMB) of the State University of São Paulo (Unesp), was used, aiming at overcoming the difficulties in selecting a homogeneous group of animals by using the simple casualization process, for later submission to treatments of interest.

## Methods

### Animals and experimental protocol

The set of echocardiographic data was obtained from male Wistar rats undergoing surgery for induction of supra-aortic AoS and provided by the Research Group on Experimental Cardiology of the FMB - Unesp. This study was approved by

the Ethics Committee for Animal Experimentation of the FMB - Unesp (protocol number 850/2010). Of the set comprising 115 experimental units, 31 parameters regarding body mass and structural and functional variables of transthoracic echocardiography were used (Table 1).

The literature has plenty of experimental studies on the development of cardiac remodeling due to pressure overload<sup>2-6</sup>. The AoS model has been widely used to promote the gradual development of left ventricular (LV) hypertrophy in rats<sup>5</sup>. In that model, pressure overload installs and progressively increases as the animals grow, partially similar to AoS in men. Some advantages of that experimental procedure are the absence of myocardial anatomical lesions and its low operational cost<sup>4</sup>.

**Table 1 – Identification of the parameters measured**

Index	Identification	Description
X <sub>1</sub>	BM (g)	body mass
X <sub>2</sub>	HR (bpm)	heart rate
X <sub>3</sub>	LVDD (mm)	left ventricular diastolic diameter
X <sub>4</sub>	LVSD (mm)	left ventricular systolic diameter
X <sub>5</sub>	PWDT (mm)	left ventricular posterior wall diastolic thickness
X <sub>6</sub>	PWST (mm)	left ventricular posterior wall systolic thickness
X <sub>7</sub>	IVSDT (mm)	interventricular septum diastolic thickness
X <sub>8</sub>	IVSST (mm)	interventricular septum systolic thickness
X <sub>9</sub>	AOD (mm)	aorta diameter
X <sub>10</sub>	LAD (mm)	left atrial diameter
X <sub>11</sub>	LAD/AOD	left atrial diameter to aorta diameter ratio
X <sub>12</sub>	LVDD/BM (mm/kg)	left ventricular diastolic diameter to body mass ratio
X <sub>13</sub>	LAD/BM (mm/kg)	left atrial diameter to body mass ratio
X <sub>14</sub>	CO (mL/min)	cardiac output
X <sub>15</sub>	CI (mL/min/kg)	cardiac index
X <sub>16</sub>	EFS	endocardial fractional shortening
X <sub>17</sub>	MFS	midwall fractional shortening
X <sub>18</sub>	LVM (g)	left ventricular mass
X <sub>19</sub>	LVMI (g/kg)	left ventricular mass index
X <sub>20</sub>	E wave (cm/s)	early diastolic velocity of mitral flow
X <sub>21</sub>	A wave (cm/s)	late diastolic velocity of mitral flow
X <sub>22</sub>	E/A	E wave to A wave ratio
X <sub>23</sub>	LVRT	left ventricular relative wall thickness
X <sub>24</sub>	LVPWSV (mm/s)	left ventricular posterior wall shortening velocity
X <sub>25</sub>	IVRT (ms)	left ventricular isovolumetric relaxation time
X <sub>26</sub>	R-R (s)	interval between two consecutive cardiac cycles
X <sub>27</sub>	IVRTn	left ventricular isovolumetric relaxation time normalized to heart rate
X <sub>28</sub>	Tei-a (ms)	isovolumetric contraction time + ejection time + IVRT
X <sub>29</sub>	Tei-b (ms)	ejection time
X <sub>30</sub>	MPI	myocardial performance index
X <sub>31</sub>	LVEF	left ventricular ejection fraction

### Induction of supravalvular aortic stenosis

Aortic stenosis was induced according to a previously described method<sup>4-7</sup>. The three- to four-week-old animals, weighing 70-90 g, underwent anesthesia with the association of ketamine chloride (60 mg/kg) and xylidine chloride (10 mg/kg), by intraperitoneal (ip) route. Then median thoracotomy was performed, the ascending aorta dissected and a silver clip (inner diameter of 0.6 mm) placed approximately 3 mm from its root (Figure 1). The thoracic wall was closed, and the sternum, muscle layers and skin were sutured with 5.0-mononylon thread. During surgery, the animals were manually ventilated with positive pressure, 100% oxygen. After the end of surgery, they received, subcutaneously, 1 mL of warm saline solution, and were placed on a warm surface to recover from anesthesia.

### Echocardiographic assessment

Echocardiographic assessment was performed six weeks after AoS induction. To undergo that test, the rats were anesthetized with the association of ketamine chloride

(50 mg/kg/ip) and xylidine chloride (10 mg/kg/ip), and put in the left lateral decubitus position. A Vivid S6 echocardiography device (General Electric Medical Systems, Tirat Carmel, Israel) equipped with a 12-MHz electronic transducer was used. To measure the cardiac structures, M-mode images were used and the ultrasound beam was guided by the two-dimensional image with the transducer in the parasternal short axis view. The monodimensional image of the left ventricle was obtained by positioning the M-mode cursor right below the mitral valve plane between the papillary muscles<sup>8</sup>. The images of the aorta and left atrium were also obtained in the parasternal short axis view with the M-mode cursor positioned at the aortic valve level. Later, the cardiac structures were measured manually with the aid of a pachymeter in at least five consecutive cardiac cycles. The LV diastolic diameter (LVDD), the LV posterior wall diastolic thickness (PWDT) and the interventricular septum diastolic thickness (IVSDT) were measured at the time corresponding to the maximum LV diameter. The LV systolic diameter (LVSD) and the LV posterior wall systolic thickness (PWST) and the interventricular septum systolic thickness

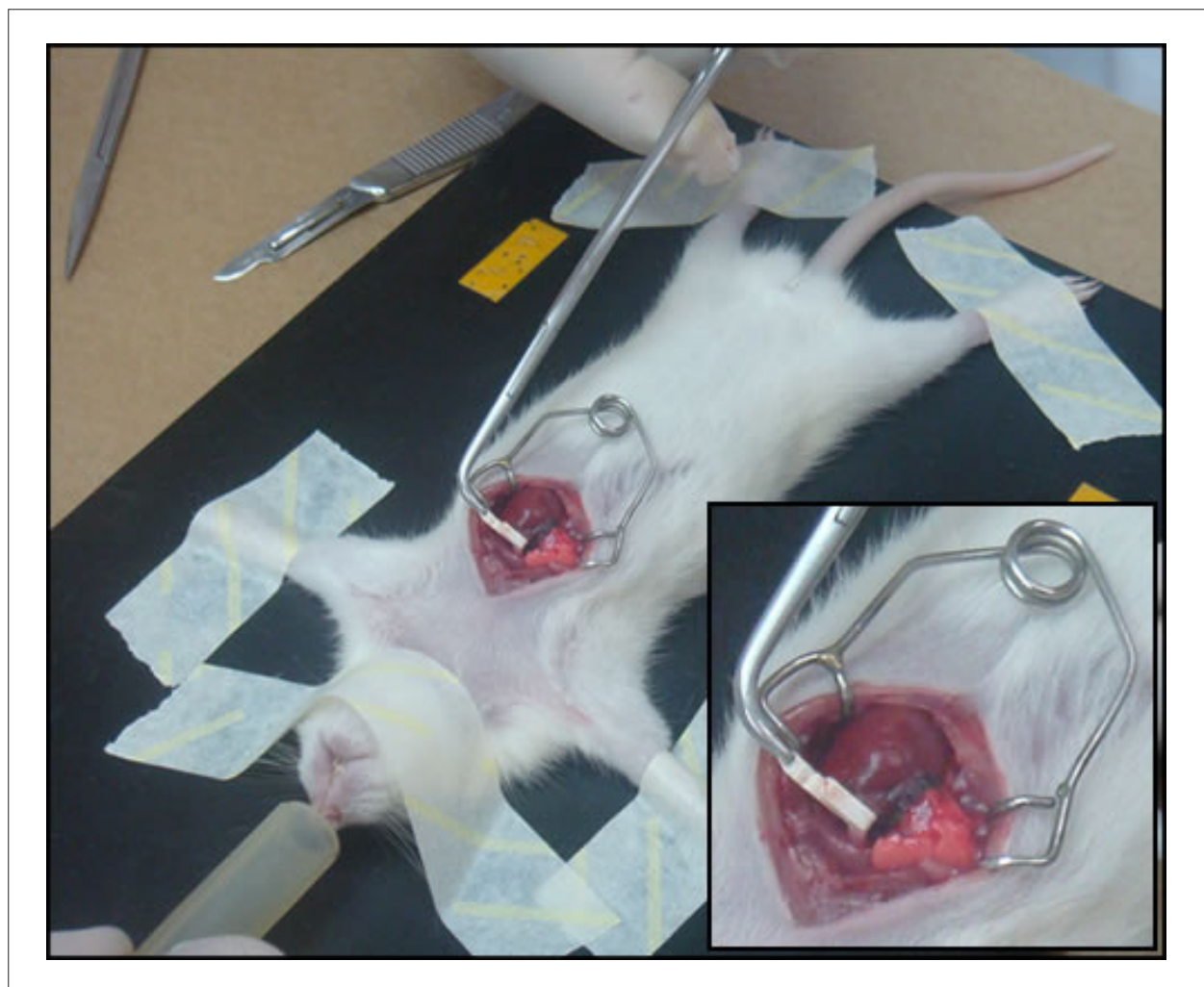


Figure 1 - Placement of the silver clip in the aortic valve.

(IVSST) were measured at the time corresponding to the minimum LV diameter. The LV systolic function was assessed by calculating the midwall fractional shortening (MFS),  $\{[(LVDD + \frac{1}{2} PWDT + \frac{1}{2} IVSDT) - (LVSD + \frac{1}{2} PWST + \frac{1}{2} IVSST)] / (LVDD + \frac{1}{2} PWDT + IVSDT)\}$ , and the LV posterior wall shortening velocity (LVPWSV), maximum tangent of the posterior wall systolic movement. Studying the LV diastolic function, the peak velocities of mitral flow corresponding to the initial filling phase (E wave) and to the late filling phase, consequent to atrial contraction (A wave), were measured, and the E wave/A wave ratio was calculated. The flows related to the diastolic function were obtained by positioning the transducer on the region corresponding to the tip of the heart in the four-chamber image; the flows were measured on the echocardiographic monitor.

### Statistical model

The statistical model developed to establish the procedure of exploratory multivariate analysis of data and assess the homogeneity of the batch involves simultaneously all variables measured and considers the entire data variation structure, that is, the variation inside the variables (intravariability) and the variation between variables (intervariability). It is worth noting that the global variation structure can be well represented by descriptive measures of data variability, more specifically, variances and covariances.

In addition, the model is elaborated to use all animals included in the research, in the cardiac remodeling group, which survived the surgical procedure previously described. Thus, there was no specific criterion to determine the number of animals that should participate in the study.

Initially, because the parameters were presented in different measure units, the data were standardized and became dimensionless, that is, each measure was presented as a value that represents how far away it is from the mean of the respective parameter. Then, to build the statistical model, named *multihomogen* criterion, the matrix structure of variances

and covariances of standardized data (variance matrix) was considered, which is equivalent to the correlation matrix (R) of the initial set<sup>9</sup>.

The multivariate technique used for data analysis consists in the principal components analysis (PCA), whose basic principle relies on the reduction of the parametric space with no loss of the data set variance structure (in the present study, biological characteristics of the animal) and no loss of the general biological information of the animal. That statistical technique, proposed by Karl Pearson in 1901, consists in transforming an initial set of interrelated variables into another set of non-interrelated variables, which is an orthogonal linear combination of the initial set<sup>10</sup>.

The principal components (PC) are presented in a decreasing order of importance for the data set variance structure, that is, the first explains the maximum possible variance, and the second explains the maximum variance still retained in the set after discounting the effect of the first, and so on, up to the last component. The greater the retention of total variance in a smaller number of linear combinations, the better the practical application of the procedure to experimental data, although that condition is not a deterrent factor for the use of the criterion to identify spurious values<sup>11</sup>.

An interesting mathematical property of the PC consists in all of them being non-interrelated and, thus, regarding data normality, independent of each other. This ensures a system of orthogonal axes for the graphic representation of animals (experimental units), which can be complemented with statistical inference to identify those that can be ruled out due to the high probability of the occurrence of spurious values as compared to the population of origin.

There are several criteria to determine the number of descriptive axes (PC) considered relevant to the orthogonal system. The present study adopted the Kaiser criterion, also known as the latent root criterion, in which a component accounts for a meaningful amount of variance when it has an

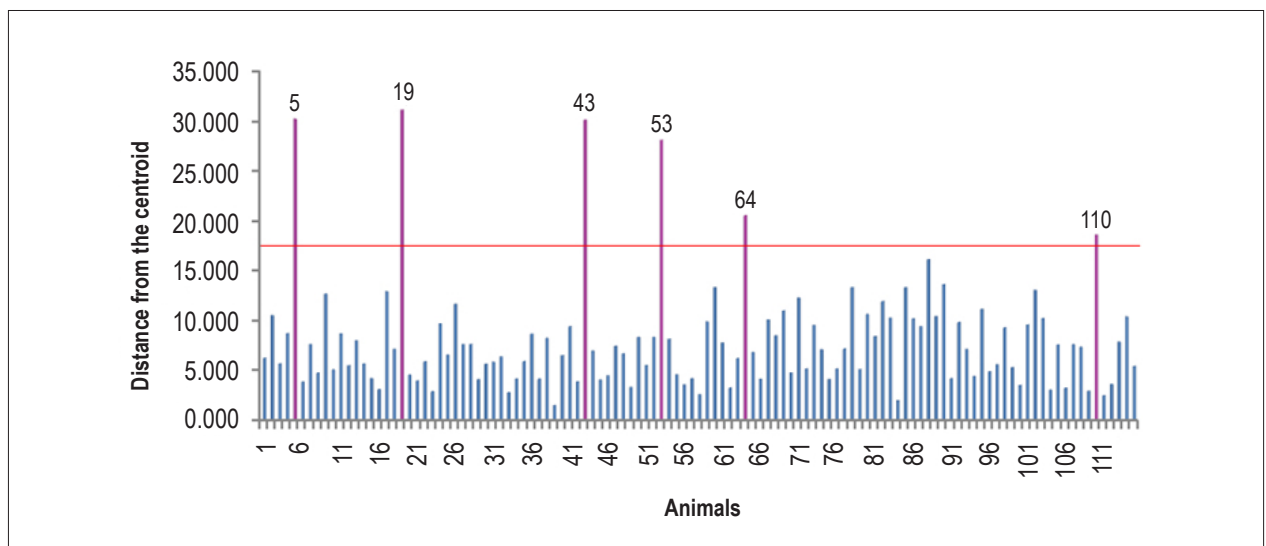


Figure 2 - Mahalanobis distance between the animals and the centroid of the group.



eigenvalue greater than 1.0. Therefore such a component is worthy of being retained<sup>12</sup>.

For the methodological practice of homogenization in experimental or observational research, the desirable situation for biological information, mainly for the graphical visualization of the procedure, consists in having a maximum number of PC (also known as descriptive axes) of three. However, for more than three dimensions (a common situation when assessing a considerable number of characteristics in the experimental unit), the technique remains the same, except for graphical visualization, which becomes unfeasible.

After assigning the animals to the Cartesian system, determined by the descriptive axes selected, the confidence (ellipsoid) region for the mean profile of the animals' homogeneous responses is built. The  $100(1 - \alpha)\%$  confidence region is constructed by using Hotelling  $T^2$  statistics transformed into Fisher-Snedecor's F distribution<sup>13</sup>. It is worth considering that the statistical technique used to establish the confidence region involves the construction of inclusion intervals, considering a jointly confidence level for all variables, that enables verifying if the animals are inside or outside the ellipsoid<sup>10</sup>.

If the response vector of the animal studied is inside the confidence ellipsoid built (that is, if the animal's generalized Mahalanobis distance from the centroid of an ellipsoid is shorter than the distance from the outline), the animal is classified as belonging to the homogeneous batch; on the contrary (greater distance), it is identified as spurious, and, thus, does not participate in the simple casualization process of treatments<sup>14</sup>.

By using the statistical procedures approached in this section, the generalized algorithm was developed, as well as the computational package, making the use of the criterion proposed possible. It was named *multihomogen package*, elaborated for the Rstudio statistical program (made available free of charge on-line: <http://www.R-project.org>), and can be directly and costlessly obtained by emailing the authors<sup>15</sup>.

## Results

To use the software developed and characterize the homogeneity of an experimental group of animals, the multihomogen criterion was applied to the set of data relating to the parameters assessed in Wistar rats undergoing AoS in the study previously mentioned. From the quadratic matrix of variance (order 31), involving the correlations, eight descriptive axes were established, corresponding to eight PC selected based on the Kaiser criterion, representing the accumulated variance of the set of data in 88.71% (Table 2).

Based on the new data structure, in the system generated by the eight descriptive axes, the Mahalanobis distances of each animal were determined in relation to the centroid of the set of all 115 animals. The multivariate statistical procedure established an animal's maximum distance from the centroid for it to be located inside the ellipsoid, built to a 95% confidence level, as being 17.419. Thus, animals whose distance values were greater than that, shown in Figure 2, were considered to be outside the confidence ellipsoid, being indicated as special (spurious) as compared to those inside that ellipsoid.

As such, spurious animals were those registered under the numbers 5, 19, 43, 53, 64 and 110, being the homogeneous batch formed by the 109 animals inside the confidence region, which can be submitted to the experimental models via the simple casualization process. Of the spurious animals, the one registered as 19 was the most dissimilar to the total group, while that registered as 110 was the least dissimilar.

## Discussion

The conclusion of the assessment of the amount of PC required to retain the relevant variability of this study was that the data set regarding the animals induced to AoS is a complex structure, from the statistical viewpoint. Although its parametric space has been reduced substantially, eight descriptive axes were necessary to form the new Cartesian system, making the graphical visualization of the method impossible.

Despite that limitation, the graph in Figure 2 can visually translate each animal's behavior regarding the group. The ordinate axis represents the Mahalanobis distance of each animal from the set centroid. Thus, knowing the distance from the centroid to the outline of the ellipsoid (represented by the red horizontal line), the animals with discrepant behavior are evidenced. They should be more deeply assessed to verify their non-homogeneous characteristics regarding the group.

Regarding the development of the statistical model to detect the non-homogeneous animals, there is no limitation to the application of the procedure in the search for spurious animals. From the practical perspective of using the results of the model proposed to new studies, it is worth noting that the process is limited to the study of animals submitted to AoS under experimental conditions similar to the one described. As such, further studies/tests would be necessary to corroborate the reliability of the model for other samples.

In the process of selecting homogeneous groups in experimental research, when using procedures with no concern about the general structure of variability, the number of experimental units (animals) discarded is relevant when compared to the amount of spurious animals identified by the criterion proposed in this study. This can be understood by the fact that the method proposed characterizes the discard by considering the animal, while those with no concern about the general structure of variability consider the discard fragmented according to the variable.

The contribution of the method is explained by the improvement in the quality of homogenization, by ensuring greater reliability to the animals' biological characteristics in the inclusion of similar animals, thus motivating a smaller discard rate, maximizing the homogeneous batch, for posterior submission to treatments by use of the simple casualization process of the animals.

## Conclusions

The criterion presented, named *multihomogen* criterion, proved to be effective in selecting homogeneous groups of animals for experimental research, because it considers the biological situation of the animal as a whole, analyzing jointly all parameters measured, in addition to having a small discard rate.

As such, this biometrical tool serves to researchers of the biological and health areas. Regarding specifically cardiac

**Table 2 – Eigenvalues corresponding to the components selected based on the Kaiser criterion**

	Eigenvalue	Explained variance (%)	Accumulated variance (%)
$\lambda_1$	9.400	30.32	30.32
$\lambda_2$	5.121	16.52	46.84
$\lambda_3$	3.995	12.89	59.73
$\lambda_4$	2.541	8.20	67.93
$\lambda_5$	2.094	6.75	74.68
$\lambda_6$	1.839	5.93	80.61
$\lambda_7$	1.456	4.70	85.31
$\lambda_8$	1.055	3.40	88.71

remodeling, in which animals are submitted to AoS, the results of this study confirm the expectations presented initially in the discussions involving the experiments previously conducted by the Research Group on Experimental Cardiology of the FMB - Unesp.

### Author contributions

Conception and design of the research: Pinto RM, Padovani CR. Acquisition of data: Campos DHS, Tomasi LC, Cicogna AC, Okoshi K. Analysis and interpretation of the data: Pinto RM, Campos DHS, Tomasi LC, Cicogna AC, Okoshi K, Padovani CR. Statistical analysis: Pinto RM, Padovani CR. Obtaining financing: Pinto RM, Padovani CR. Writing of the manuscript: Pinto RM, Campos DHS, Tomasi LC, Padovani CR. Critical revision of the manuscript for intellectual content: Pinto RM, Cicogna AC, Okoshi K, Padovani CR. Orientation / Supervision: Padovani CR.

### References

- Mesetti AV, Padovani CR. Estudo da divergência genética em girassol por meio de técnicas multivariadas. *Energ Agric Botucatu*. 2009;24(2):14-28.
- Rodrigues MA, Bregagnollo EA, Montenegro MR, Tucci PJ. Coronary vascular and myocardial lesions due to experimental constriction of the abdominal aorta. *Int J Cardiol*. 1992;35(2):253-7.
- Okoshi K, Ribeiro HB, Okoshi MP, Matsubara BB, Gonçalves G, Barros R, et al. Improved systolic ventricular function with normal myocardial mechanism compensated cardiac hypertrophy. *Jpn Heart J*. 2004;45(4):647-53.
- Bregagnollo EA, Mestrinel MA, Okoshi K, Carvalho FC, Bregagnollo IF, Padovani CR, et al. Relative role of left ventricular geometric remodeling and of morphological and functional myocardial remodeling in the transition from compensated hypertrophy to heart failure in rats with supralvalvar aortic stenosis. *Arq Bras Cardiol*. 2007;88(2):225-33.
- Bregagnollo EA, Zornoff LA, Okoshi K, Sugizaki M, Mestrinel MA, Padovani CR, et al. Myocardial contractile dysfunction contributes to the development of heart failure in rats with aortic stenosis. *Int J Cardiol*. 2006;117(1):109-14.
- Mendes OC, Sugizaki MM, Campos DS, Damatto RL, Leopoldo AS, Lima-Leopoldo AP, et al. Exercise tolerance in rats with aortic stenosis and ventricular diastolic and/or systolic dysfunction. *Arq Bras Cardiol*. 2013;100(1):44-51.
- Mendes Ode C, Campos DH, Damatto RL, Sugizaki MM, Padovani CR, Okoshi K, et al. Remodelamento cardíaco: análise seriada e índices de detecção precoce de disfunção ventricular. *Arq Bras Cardiol*. 2010;94(1):62-70.
- Litwin SE, Katz SE, Weinberg EO, Lorell HB, Aurigemma GP, Douglas PS. Serial echocardiographic-Doppler assessment of left ventricular geometry and function in rats with pressure overload hypertrophy chronic angiotensin-converting enzyme inhibition attenuates the transition to heart failure. *Circulation*. 1995;91(10):2642-54.

### Potential Conflict of Interest

No potential conflict of interest relevant to this article was reported.

### Sources of Funding

This study was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

This study was partially funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

### Study Association

This article is part of the dissertation of master submitted by Renan Mercuri Pinto, from Universidade Estadual Paulista “Júlio de Mesquita Filho”.

## Special Article

- 
9. Mingoti SA. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte: Editora UFMG; 2007.
  10. Morrison DF. Multivariate statistical methods. 3<sup>rd</sup> ed. New York: McGraw Hill, Inc; 2004.
  11. Silva NR, Padovani CR. Utilização de componentes principais em experimentação agrônoma. *Energ Agric Botucatu*. 2006;21(4):98-113.
  12. Hair JF Jr, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7<sup>th</sup> ed. New Jersey: Prentice Hall; 2010.
  13. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6<sup>th</sup> ed. New Jersey: Prentice Hall; 2007.
  14. Mahalanobis PC. Historic note on the  $D^2$  statistic. *Sankhya*. 1948;9:237-40.
  15. The R Core Team. R: a language and environment for statistical computing: reference index. Viena: Foundation for Statistical Computing; 2013.