

Scientific Electronic Archives

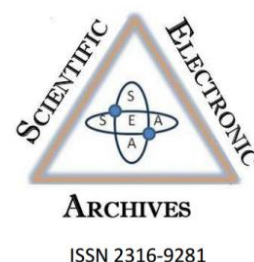
Issue ID: Sci. Elec. Arch. Vol. 10 (4)

August 2017

Article link

<http://www.seasinop.com.br/revista/index.php?journal=SEA&page=article&op=view&path%5B%5D=338&path%5B%5D=pdf>

Included in DOAJ, AGRIS, Latindex, Journal TOCs, CORE, Discoursio Open Science, Science Gate, GFAR, CIARDRING, Academic Journals Database and NTHRYS Technologies, Portal de Periódicos CAPES.



Modelagem matemática para um problema biológico

Mathematical modeling for biological problem

S. W. Hansen, R. R. Oliveira, L. Y. Sheng

Universidade Federal de Mato Grosso - Campus Sinop

Author for correspondence: leeufmt@yahoo.com.br

Resumo. Neste trabalho iremos formalizar um conjunto de procedimentos básicos de análise de dados biológicos e criar ferramentas para auxiliar nas tomadas de decisões. Tais ferramentas são métodos estatísticos, isto é, a criação de uma modelagem matemática/estatística com o intuito de prever um dado evento destinado a medir sua possibilidade de ocorrência. As estruturas foram observadas *post mortem* por análise macroscópicas, assim realizando a remodelagem cardíaca. Através da modelagem estatística espera-se criar um método de mensuração de quão "bom/melhora" teve o coração ao se aplicar determinada substância. Para isso devemos fazer as análises de estatísticas descritas, juntamente com as análises das variáveis, trabalhando com regressão linear e mínimos quadrados, possibilitando assim um melhor controle sobre as amostras.

Palavras chaves: Dados biológicos, modelagem e regressão linear.

Abstract. In this study we will formalize a set of basic procedures for analyzing biological data and create tools to assist in decision making. Such tools are statistical methods, i.e., the creation of a mathematical modeling / statistic in order to provide a given event for measuring a possibility of occurrence. The structures were observed by post-mortem macroscopic analysis, thereby performing cardiac remodeling. Through statistical modeling is expected to create a method for measuring how "good / better" had the heart to apply substance. For this we must make the statistical analyzes described, together with the analysis of the variables, and working with least squares linear regression, thus enabling better control over the samples.

Keywords: Biological data, modeling, and linear regression.

Introdução

A estatística se define por etapas de pesquisa, que exigem planejamento, levantamento de dados, questionários, medições com máxima quantidade de informações possíveis, a análise e interpretação dos dados levantados.

Atualmente o grande acesso de computadores pessoais facilita aos pesquisadores de diferentes campos de atuação a acessibilidade de trabalhar com a estatística. Os equipamentos e softwares oferecem a manipulação de grande quantidade de dados, o que dinamizou o emprego dos métodos estatísticos.

A utilização da estatística está disseminada nas universidades, empresas privadas e públicas. Com o uso de gráficos e tabelas para exposição de resultados das empresas. Censos demográficos são utilizados pelo governo para entender melhor sua população e para organizar seus gastos. Estatística apresenta um grau de muita importância, quase todos os governos possuem organizações oficiais que são destinados a realização de estudos estatísticos.

O material de ensino CNEC, edição 2011 ressalta:

"As necessidades que exigiam o conhecimento numérico dos recursos disponíveis começaram a surgir quando as sociedades primitivas se organizaram. Os Estados, desde tempos remotos, precisaram conhecer determinadas características da população, efetuarem a sua contagem e saberem a sua composição ou os seus rendimentos.

Para responder ao desenvolvimento social, surgiram as primeiras técnicas estatísticas. Classificar, apresentar e interpretar os dados recolhidos foram para os censos e são para a Estatística um aspecto essencial do método utilizado. Mas, um longo caminho havia de ser percorrido até aos dias de hoje.

O desenvolvimento do Cálculo das Probabilidades surge também no século XVII. A ligação das probabilidades com os conhecimentos estatísticos veio dar uma nova dimensão à Estatística, que, progressivamente, se foi tornando um instrumento científico poderoso e indispensável. Considera-se assim uma nova fase, a terceira, em que se começa a fazer inferência estatística: quando, a partir de observações, procurou-se deduzir relações causais, entre variáveis, realizando-se previsões a partir daquelas relações.

Na atualidade, a Estatística já não se limita apenas a Economia. O seu campo de aplicação alargou-se à análise de dados em Biologia, Medicina, Física, Psicologia, Indústria, Comércio, Meteorologia, Educação, etc., e ainda a domínios aparentemente desligados, como estrutura de linguagem e estudo de formas literárias.”

Metodos

Quando encontramos um conjunto de dados é necessário trabalhar com esses dados para transformá-los em informações, para assim poder julgar sua adequação a alguma teoria.

A inferência estatística pode ser dedutiva ou indutiva, sendo estas partes da Estatística, é a metodologia que tem por objetivo a coleta, análise e modelagem de dados. Quando fala-se da modelagem de dados estamos falando de revisões, onde a partir das quais se pode tomar decisões.

Quando é realizada uma análise de dados, é porque está sendo buscada uma forma de regularidade ou padrão ou ainda modelo, que esteja presente nas observações.

A análise de regressão descreve através de um modelo matemático, a relação existente entre duas ou mais variáveis, onde temos n observações dessas variáveis. Observaremos primeiramente o caso da relação de duas variáveis de forma geral, considerando uma como a variável Y sendo a posteriori e uma como a variável X como a priori. Ou seja, seria muito bom se para um dado valor de X , obtivéssemos uma função matemática a qual $Y = f(X)$, para determinarmos a melhor função que mais se aproxima deste fato teremos a seguinte equação:

$$Y = f(X) + \varepsilon$$

A qual f descreve a relação entre X e Y e ε são os erros aleatórios. Dizemos que Y é uma variável resposta ou dependente, e X é uma variável independente, concomitante, covariável ou variável preditora.

O modelo de regressão é chamado de simples quando envolve uma relação causal entre apenas duas variáveis. O modelo de regressão é múltiplo quando envolve uma relação causal com mais de duas variáveis. Isto é, quando o comportamento de Y é explicado por mais de uma variável independente X_1, X_2, \dots, X_n .

Vejam os exemplos de dois modelos de regressão, inicialmente vamos conhecer uma regressão linear simples. Sendo que agora com mais detalhes.

Lembrando que regressão linear é o processo matemático pelo qual derivamos os parâmetros “ a ” e “ b ” de uma função $f(X) = aX + b$. Estes parâmetros determinam as características da função que relaciona ‘ Y ’ com ‘ X ’ que no caso do modelo linear se representa por uma reta chamada de reta de regressão ou conhecida como ajustamento.

A reta ajustada (regressão linear simples) é representada por $\hat{Y} = a + bX$, onde a e b são parâmetros do modelo: a é o ponto onde a reta ajustada corta o eixo da variável Y , e b é a tangente do ângulo que a reta forma com uma paralela ao eixo da variável X .

Obtemos a reta ajustada através do método dos mínimos quadrados, pois os valores de a e b são obtidos de tal forma que é mínima a soma dos quadrados das diferenças entre os valores observados de Y e os obtidos a partir da reta ajustada para os mesmos valores de X . Onde simbolicamente teremos:

$$\sum \varepsilon = \sum (Y - \hat{Y})^2 = \sum (Y - a - bX)^2 \quad (2.1)$$

Isto é o Método dos Mínimos Quadrados, ou Mínimos Quadrados Ordinários é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados.

Para obtermos os valores a e b aplicamos a condição de mínimo à função $\sum (Y - \hat{Y})^2$. Onde temos:

$$a = \bar{Y} - b\bar{X} \quad (2.2)$$

$$b = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} \quad (2.3)$$

Os dados abaixo, retirados de Toledo e Ovalle (1985) referem-se ao volume de precipitação pluviométrica (mm) e ao volume de produção de leite tipo C (milhões de litros), em determinada região do país.

ANOS	Produção De Leite C (1.000.000 L)	índice pluviométrico (mm)
1970	26	23
1971	25	21
1972	31	28
1973	29	27
1974	27	23
1975	31	28
1976	32	27
1977	28	22
1978	30	26
1979	30	25

Objetivo: Ajustar os dados através de um modelo linear.

Passo 1: Para facilitar os cálculos necessários ao ajustamento vamos montar uma tabela auxiliar

Y	X	X ²	XY
26	23	529	598
25	21	441	525
31	28	784	868
29	27	729	783
27	23	529	621
31	28	784	868
32	27	729	864
28	22	484	616
30	26	676	780
30	25	625	750
$\sum Y=289$	$\sum X=250$	$\sum X^2=6.310$	$\sum XY=7.273$

Passo 2: Determinar o valor do Parâmetro b , conforme (2.3):

$$b = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{7.273 - \frac{(250)(289)}{10}}{6.310 - \frac{(250)^2}{10}} = \frac{48}{60} = 0,8$$

Então temos que $b=0,8$.

Passo 3: Determinar o valor do Parâmetro a , conforme (2.2):

$$a = 8,9$$

$$a = \bar{Y} - b\bar{X} = \frac{\sum Y}{n} - b \frac{\sum X}{n} = \frac{289}{10} - 0,8 \frac{250}{10} = 8,9$$

Passo 4: Equação da Reta Ajustada:

$$\hat{Y} = a + bX = 8,9 + 0,8X$$

Temos também a representação em diagrama de dispersão em que o mesmo nos apresenta a Reta Ajustada (Figura 1).

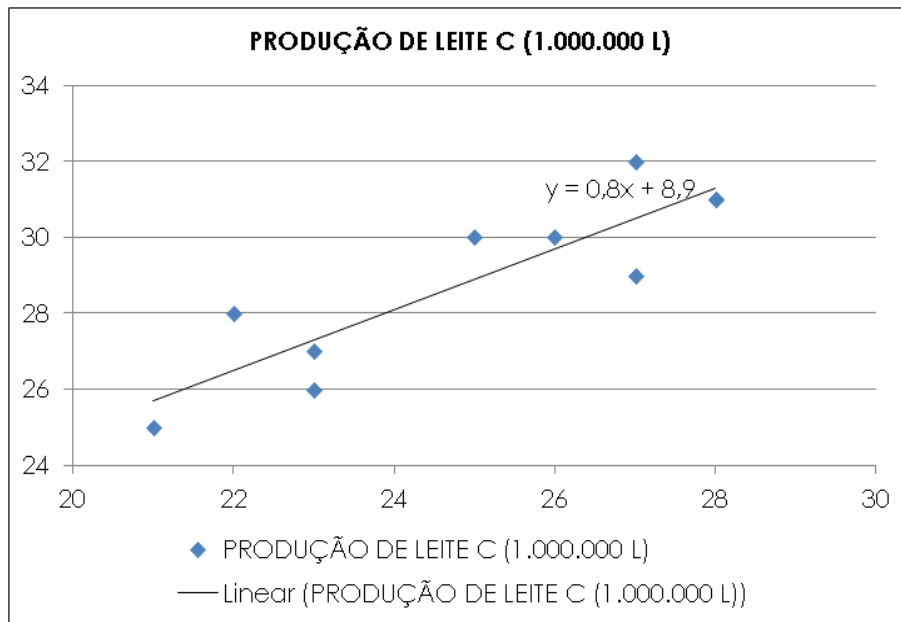


Figura 1: Reto ajustada da equação da regressão linear simples.

Como já vimos, quando fala-se em regressão linear múltipla estamos envolvendo mais de duas variáveis, ou seja, quando temos a nossa variável dependente Y em função de duas ou mais variáveis independentes X_1, X_2, \dots, X_n .

Pode-se tomar como o caso mais simples, em que temos apenas duas variáveis independentes X_1, X_2 . Devemos nesse caso, obter uma equação na forma:

$$\hat{Y} = a + b_1X_1 + b_2X_2, \quad (2.4)$$

equação do plano estimativa no espaço a três dimensões.

A imposição do critério de mínimos quadrados leva ao seguinte sistema:

$$\begin{cases} \sum_{j=1}^n x_{1j}y_j = a \sum x_{1j} + b_1 \sum x_{1j}^2 + b_2 \sum x_{1j}x_{2j}, \\ \sum_{j=1}^n y_j = na + b_1 \sum x_{1j} + b_2 \sum x_{2j}, \\ \sum_{j=1}^n x_{2j}y_j = a \sum x_{2j} + b_1 \sum x_{1j}x_{2j} + b_2 \sum x_{2j}^2. \end{cases} \quad (2.5)$$

De forma semelhante, no caso geral, em que a equação procurada é da forma:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_kX_k, \quad (2.6)$$

equação de um hiperplano no espaço $k + 1$ dimensões, chegaríamos a um sistema cuja: primeira equação seria:

$$\begin{aligned} \sum y_j &= na + b_1 \sum x_{1j} + \dots + b_k \sum x_{kj}, \\ \therefore a &= \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 - \dots - b_k\bar{X}_k. \end{aligned} \quad (2.7)$$

as demais k equações seriam da forma:

$$\left\{ \begin{array}{l} \sum x_{1j}y_j = a \sum x_{1j} + b_1 \sum x_{1j}^2 + b_2 \sum x_{1j}x_{2j} + \dots + b_k \sum x_{1j}x_{kj}, \\ \sum x_{2j}y_j = a \sum x_{2j} + b_1 \sum x_{1j}x_{2j} + b_2 \sum x_{2j}^2 + \dots + b_k \sum x_{2j}x_{kj}, \\ \vdots \\ \sum x_{kj}y_j = a \sum x_{kj} + b_1 \sum x_{1j}x_{kj} + b_2 \sum x_{2j}x_{kj} + \dots + b_k \sum x_{kj}^2. \end{array} \right. \quad (2.8)$$

Se agora, ao invés de trabalharmos com y_j e x_{ij} , usarmos $y_j - \bar{y}$ e $x_{ij} - \bar{x}_i$, teremos $a = 0$, permanecendo inalterados os b_i , pois uma simples translação não afeta os demais coeficientes do hiperplano de regressão. Podemos, portanto escrever o sistema (2.8) na forma:

$$\left\{ \begin{array}{l} S_{1y} = b_1S_{11} + b_2S_{12} + \dots + b_kS_{1k}, \\ S_{2y} = b_1S_{21} + b_2S_{22} + \dots + b_kS_{2k}, \\ \vdots \\ S_{ky} = b_1S_{k1} + b_2S_{k2} + \dots + b_kS_{kk}. \end{array} \right. \quad (2.9)$$

Retirado de Costa Neto (2002). Uma reação química foi realizada sob seis pares de diferentes condições de pressão e temperatura. Em cada caso, foi medido o tempo necessário para que a reação se completasse. Os resultados obtidos são os que seguem.

Condição	Temperatura (°C)	Pressão (atm)	Tempo (s)
1	20	1,5	9,4
2	30	1,5	8,2
3	30	1,2	9,7
4	40	1,0	9,5
5	60	1,0	6,9
6	80	0,8	6,5

Objetivo: Obter a equação da função de regressão linear de tempo (y) em relação à temperatura (x_1) e a pressão (x_2).

Temos:

x_{1j}	x_{2j}	y_j	x_{1j}^2	x_{2j}^2	y_j^2	$x_{1j}x_{2j}$	$x_{1j}y_j$	$x_{2j}y_j$
20	1,5	9,4	400	2,25	88,36	30,0	188,0	14,10
30	1,5	8,2	900	2,25	67,24	45,0	246,0	12,30
30	1,2	9,7	900	1,44	94,09	36,0	291,0	11,64
40	1,0	9,5	1.600	1,00	90,25	40,0	380,0	9,50
60	1,0	6,9	3.600	1,00	47,61	60,0	414,0	6,90
80	0,8	6,5	6.400	0,64	42,25	64,0	520,0	5,20
$\Sigma=260$	$\Sigma=7,0$	$\Sigma=50,2$	$\Sigma=13.800$	$\Sigma=8,58$	$\Sigma=429,80$	$\Sigma=275,0$	$\Sigma=2.039,0$	$\Sigma=59,64$

Passo 1: Foi usada a formulação dada pelo sistema de equações (2.9). Por analogia foi usada as fórmulas de correlação linear dada como:

$$\begin{aligned} \sigma_{00} &= \sum_{i=1}^n (\sigma_{0i} - \bar{\sigma}_0)(\sigma_i - \bar{\sigma}) = \sum \sigma_{0i}\sigma_i - \frac{\sum \sigma_{0i} \sum \sigma_i}{n}, \\ \sigma_{00} &= \sum_{i=1}^n (\sigma_{00} - \bar{\sigma}_0)(\sigma_{00} - \bar{\sigma}_0) = \sum \sigma_{i0}\sigma_{00} - \frac{\sum \sigma_{00} \sum \sigma_{00}}{n}. \end{aligned}$$

Então temos que,

$$\sigma_{10} = 2.039,0 - \frac{260,50,2}{6} \cong -136,33,$$

$$\sigma_{20} = 59,64 - \frac{7,0,50,2}{6} \cong 1,073,$$

$$\sigma_{11} = 13.800 - \frac{(260)^2}{6} \cong 2.533,33,$$

$$\sigma_{22} = 8,58 - \frac{(7,0)^2}{6} \cong 0,413,$$

$$\sigma_{12} = \sigma_{21} = 275,0 - \frac{260,7,0}{6} \cong -28,33.$$

O sistema de equações (2.9) ficou sendo, portanto,

$$\begin{cases} -136,33 = 2,533,33\sigma_1 - 28,33\sigma_2 \\ 1,073 = -28,33\sigma_1 + 0,413\sigma_2 \end{cases}$$

que, resolvido, forneceu:

$$\sigma_1 = -0,1063, \quad \sigma_2 = -4,6945.$$

Passo 2: Realizar os cálculos preliminares para a regressão linear múltipla. O valor do coeficiente a é fornecido por (2.7):

$$\sigma = \bar{\sigma} - \sigma_1\bar{\sigma}_1 - \sigma_2\bar{\sigma}_2 \cong \frac{50,2}{6} - (0,1064) \cdot \frac{260}{6} - (-4,6945) \cdot \frac{7,0}{6} \cong 18,450.$$

Passo 3: Equação:

$$\hat{\sigma} = 18,450 - 0,1063\sigma_1 - 4,6945\sigma_2$$

Aplicação

O coração é um órgão formado por musculatura estriada cardíaca. Ele divide-se em dois lados o lado esquerdo e o direito, onde respectivamente temos átrio esquerdo, ventrículo esquerdo e átrio direito, ventrículo direito.

O seu lado esquerdo bombeia o sangue para todo o corpo retirando e distribuindo substâncias como gás carbônico e oxigênio. Esse processo é chamado de grande circulação, o processo de coração-corpo-coração.

A pequena circulação também chamada circulação pulmonar, compreende o trajeto do sangue desde o ventrículo direito até o átrio esquerdo. Nessa circulação, o sangue passa pelos pulmões, onde é oxigenado. Realizando sempre o mesmo processo coração, pulmão, coração.

O coração apresenta movimentos cardíacos chamados de diástole e sístole, onde respectivamente um representa o relaxamento do músculo, é nessa fase que os ventrículos recebem sangue dos átrios já a contração ventricular força (), então, a passagem de sangue para as artérias pulmonar e aorta.

A obtenção dos dados

A obtenção dos dados foi através da parceria do Prof. Dr. Lee Yun Sheng e o Prof. Dr. Mario Mateus Sugizaki, onde o Prof. Mario obteve esses dados para o seu projeto de doutorado Efeitos do treinamento físico sobre a função mecânica do miocárdio de ratos submetidos a diferentes graus de restrição de ingestão alimentar.

Para realizar a remodelagem cardíaca, deve se levar em consideração que a obesidade pode acarretar sua remodelagem. Por isso deve se observar/avaliar a estrutura *post mortem* por análise macroscópicas.

Essa análise macroscópica permite a presença ou não de hipertrofia cardíaca em nível atrial e ventricular. Observando também o peso total do coração, do átrio, dos ventrículos esquerdo e direito.

A função foi analisada *in vitro* utilizando o músculo papilar isolado do ventrículo esquerdo, essa análise foi feita rotineiramente, avaliando a função mecânica do miocárdio *in vitro*. Permitindo detectar alterações precoces na contração e no relaxamento do músculo cardíaco, isso independente das variações da pós-carga, pré-carga, frequência cardíaca e influência hormonal.

Como sabemos que a obesidade pode causar efeitos indiretos no coração, o estudo com músculo papilar isolado tem a vantagem de possibilitar a análise direta da função miocárdica.

Para a realização dos estudos funcionais, foram utilizadas algumas técnicas. Os animais foram anestesiados com pentobarbital sódico, os corações foram removidos rapidamente e colocados em solução Krebs-Henseleit. Os músculos papilares foram cuidadosamente dissecados.

A análise do comprimento da Tensão desenvolvida TD e da Tensão de repouso TR, foi realizada por meio da construção de um modelo de regressão linear. Tal comparação foi feita através do teste do coeficiente angular e da constante de regressão.

Porem verificou-se a necessidade de se avaliar quando o músculo cardíaco é “bom” ou “ruim”, isto é, este tipo de avaliação é importante para as tomadas de decisões principalmente quando se faz experimentos com substâncias para melhorar/cura de algum mal cardíaco.

Devido a estes fatos houve a necessidade de criar uma mensuração destes fatos. Através da modelagem estatística espera-se criar um método de mensuração de quão “bom/melhora” teve o coração ao se aplicar na determinada substância.

Para isso utilizaremos o método de Regressão Linear, juntamente com o método dos mínimos quadrados.

Para todo modelo tem que ser separado a amostra em duas partes a de modelagem e a outra de validação, em geral se escolhe 75% de amostra para modelar, sendo que esta escolha das observações é de maneira aleatória e o restante é denominado o conjunto de validação.

Vejamos a seguinte tabela.

M1	M2	M3	M4	Bom(1) Ruim (0)	Delta 1	Delta 2
3,54	3,75	3,54	3,71	1	5,9322	4,8023
3,84	3,49	3,67	3,67	0	-9,1146	0,0000
3,17	3,67	3,17	3,34	1	15,7729	5,3628
3,03	2,95	2,88	2,92	0	-2,6403	1,3889
5,12	5,43	5,18	5,18	1	6,0547	0,0000
4,99	5,25	5,19	5,51	1	5,2104	6,1657
3,22	3,38	3,18	3,42	1	4,9689	7,5472
3,38	3,49	3,67	3,53	1	3,2544	-3,8147
4,55	4,4	4,55	4,31	0	-3,2967	-5,2747
4,82	5,19	4,92	4,65	1	7,6763	-5,4878
3,67	3,82	4,08	3,93	1	4,0872	-3,6765
3,93	4,39	3,98	4,24	1	11,7048	6,5327
4,51	4,63	4,51	4,29	1	2,6608	-4,8780
4,51	4,77	4,58	4,58	1	5,7650	0,0000
5,28	5,73	5,58	5,43	1	8,5227	-2,6882
4,92	4,98	4,86	5,16	1	1,2195	6,1728
2,75	3,05	2,75	2,75	1	10,9091	0,0000
5,24	5,01	5,01	4,68	0	-4,3893	-6,5868
6,7	6,83	6,02	6,66	1	1,9403	10,6312
9,08	8,41	8,22	8,41	0	-7,3789	2,3114
8,04	8,39	7,6	7,6	0	4,3532	0,0000
9,34	9,24	6,96	7,15	0	-1,0707	2,7299

6,06	5,86	5,33	5,6	0	-3,3003	5,0657
6,34	6,06	6,48	6,2	0	-4,4164	-4,3210
7,88	8,71	7,8	7,47	0	10,5330	-4,2308
8,12	8,59	8,78	8,4	0	5,7882	-4,3280
5,14	5,4	5,07	5,47	1	5,0584	7,8895
6,79	6,79	6,33	6,79	0	0,0000	7,2670
5,76	5,88	5,3	4,84	0	2,0833	-8,6792
6,32	6,04	5,56	6,56	0	-4,4304	17,9856
6,29	6,75	6,23	6,49	1	7,3132	4,1734
5,35	5,81	5,35	5,7	1	8,5981	6,5421
5,61	5,61	5,19	5,12	0	0,0000	-1,3487
8,5	8,67	8,33	8,16	0	2,0000	-2,0408
6,86	6,8	6,49	6,55	1	-0,8746	0,9245
6,86	6,46	6,53	6,06	0	-5,8309	-7,1975

Todos estes dados são referentes à Tensão Desenvolvida. Onde M1, M2, M3, M4 indicam o tempo ao qual foi coletado os dados, sendo M1 sem ação de droga, M2 solução de Cálcio, no momento M3 os corações foram

colocados na solução Krebs- Henseleit e retirados a quantidade de Cálcio presente, por esse motivo que M1 e M3 possuem quase o mesmo valor e M4 possui a ação de Isopropelenol. E a criação das variáveis delta 1 e delta 2, onde:

$$\text{Delta 1} = \frac{(M2 - M1) \cdot 100}{M1} \qquad \text{Delta 2} = \frac{(M4 - M3) \cdot 100}{M3}$$

A conceituação de Bom (1) ou Ruim (0) do desempenho muscular cardíaco é empírico até o presente momento. E representa se o comportamento do músculo cardíaco é bom ou ruim sob a presença de cálcio e sob a presença de isopropelenol, não simultaneamente. Separamos agora a amostra em treinamento e a validação de maneira aleatória.

Massa de treinamento (75% do todo).

M1	M2	M3	M4	Bom(1) Ruim (0)	Delta 1	Delta 2
3,54	3,75	3,54	3,71	1	5,9322	4,8023
3,84	3,49	3,67	3,67	0	-9,1146	0,0000
3,03	2,95	2,88	2,92	0	-2,6403	1,3889
5,12	5,43	5,18	5,18	1	6,0547	0,0000
4,99	5,25	5,19	5,51	1	5,2104	6,1657
3,22	3,38	3,18	3,42	1	4,9689	7,5472
4,55	4,4	4,55	4,31	0	-3,2967	-5,2747
4,82	5,19	4,92	4,65	1	7,6763	-5,4878
3,67	3,82	4,08	3,93	1	4,0872	-3,6765
4,51	4,63	4,51	4,29	1	2,6608	-4,8780
4,51	4,77	4,58	4,58	1	5,7650	0,0000
5,28	5,73	5,58	5,43	1	8,5227	-2,6882
4,92	4,98	4,86	5,16	1	1,2195	6,1728
2,75	3,05	2,75	2,75	1	10,9091	0,0000
5,24	5,01	5,01	4,68	0	-4,3893	-6,5868
9,08	8,41	8,22	8,41	0	-7,3789	2,3114
8,04	8,39	7,6	7,6	0	4,3532	0,0000
9,34	9,24	6,96	7,15	0	-1,0707	2,7299
6,34	6,06	6,48	6,2	0	-4,4164	-4,3210

7,88	8,71	7,8	7,47	0	10,5330	-4,2308
8,12	8,59	8,78	8,4	0	5,7882	-4,3280
5,76	5,88	5,3	4,84	0	2,0833	-8,6792
6,29	6,75	6,23	6,49	1	7,3132	4,1734
5,35	5,81	5,35	5,7	1	8,5981	6,5421
8,5	8,67	8,33	8,16	0	2,0000	-2,0408
6,86	6,8	6,49	6,55	1	-0,8746	0,9245
6,86	6,46	6,53	6,06	0	-5,8309	-7,1975

Massa de validação (25% do todo)

M1	M2	M3	M4	Bom(1) Ruim (0)	Delta 1	Delta 2
3,17	3,67	3,17	3,34	1	15,7729	5,3628
3,38	3,49	3,67	3,53	1	3,2544	-3,8147
3,93	4,39	3,98	4,24	1	11,7048	6,5327
6,7	6,83	6,02	6,66	1	1,9403	10,6312
6,06	5,86	5,33	5,6	0	-3,3003	5,0657
5,14	5,4	5,07	5,47	1	5,0584	7,8895
6,79	6,79	6,33	6,79	0	0,0000	7,2670
6,32	6,04	5,56	6,56	0	-4,4304	17,9856
5,61	5,61	5,19	5,12	0	0,0000	-1,3487

Inicialmente vejamos as seguintes modelagens (Modelo 1).

Delta1 vs Conceito, obtemos então a seguinte fórmula $y = 0,0552x + 0,3864$
onde x representa o Delta 1 e y a simulação do conceito (Bom (1) ou Ruim (0))

Delta 1	Bom(1) Ruim (0)	Simulação 1	Assertividade (Acerto = 1) (Erro = 0)
5,9322	1	0,7139	1
-9,1146	0	-0,1167	0
-2,6403	0	0,2407	1
6,0547	1	0,7206	1
5,2104	1	0,6740	1
4,9689	1	0,6607	1
-3,2967	0	0,2044	1
7,6763	1	0,8101	1
4,0872	1	0,6120	1
2,6608	1	0,5333	1
5,7650	1	0,7046	1
8,5227	1	0,8569	1
1,2195	1	0,4537	1
10,9091	1	0,9886	1
-4,3893	0	0,1441	1
-7,3789	0	-0,0209	0
4,3532	0	0,6267	1
-1,0707	0	0,3273	1
-4,4164	0	0,1426	1

10,5330	0	0,9678	1
5,7882	0	0,7059	1
2,0833	0	0,5014	1
7,3132	1	0,7901	1
8,5981	1	0,8610	1
2,0000	0	0,4968	1
-0,8746	1	0,3381	1
-5,8309	0	0,0645	1

E a análise de assertividade se o Conceito é 1 e a Simulação 1 for maior ou igual a 0,5 então acertou (o valor mínimo na assertividade é 1) caso contrário errou (isto é, o valor da assertividade é 0). Observamos que temos 85,18% de acerto na massa de treinamento.

Aplicando a Simulação 1 na Massa de validação, obtendo 88,88% de acerto.

Delta 1	Bom(1) Ruim (0)	Simulação 1	Assertividade (Acerto = 1) (Erro = 0)
15,7729	1	1,2571	1
3,2544	1	0,5660	1
11,7048	1	1,0325	1
1,9403	1	0,4935	0
-3,3003	0	0,2042	1
5,0584	1	0,6656	1
0,0000	0	0,3864	1
-4,4304	0	0,1418	1
0,0000	0	0,3864	1

Analisando Delta 2 vs Conceito, obtemos a seguinte fórmula (Modelo 2)

$$y = 0,0509x + 0,5499.$$

Delta 2	Bom(1) Ruim (0)	Simulação 2	Assertividade (Acerto = 1) (Erro = 0)
4,8023	1	0,7943	1
0,0000	0	0,5499	1
1,3889	0	0,6206	1
0,0000	1	0,5499	1
6,1657	1	0,8637	1
7,5472	1	0,9341	1
-5,2747	0	0,2814	1
-5,4878	1	0,2706	1
-3,6765	1	0,3628	1
-4,8780	1	0,3016	1
0,0000	1	0,5499	1
-2,6882	1	0,4131	1
6,1728	1	0,8641	1
0,0000	1	0,5499	1
-6,5868	0	0,2146	1
2,3114	0	0,6676	1
0,0000	0	0,5499	1
2,7299	0	0,6889	1
-4,3210	0	0,3300	1
-4,2308	0	0,3346	1

-4,3280	0	0,3296	1
-8,6792	0	0,1081	1
4,1734	1	0,7623	1
6,5421	1	0,8829	1
-2,0408	0	0,4460	1
0,9245	1	0,5970	1
-7,1975	0	0,1835	1

Observamos que agora temos 66,66% de acerto.

Aplicando a Simulação 2 na Massa de Validação, obtendo 55,55% de acerto.

Delta 2	Bom(1) Ruim (0)	Simulação 1	Assertividade (Acerto = 1) (Erro = 0)
5,3628	1	0,8229	1
-3,8147	1	0,3557	0
6,5327	1	0,8824	1
10,6312	1	1,0910	1
5,0657	0	0,8077	0
7,8895	1	0,9515	1
7,2670	0	0,9198	0
17,9856	0	1,4654	0
-1,3487	0	0,4812	1

Vejamos a representação do Modelo 1 e Modelo 2 no diagrama de dispersão.

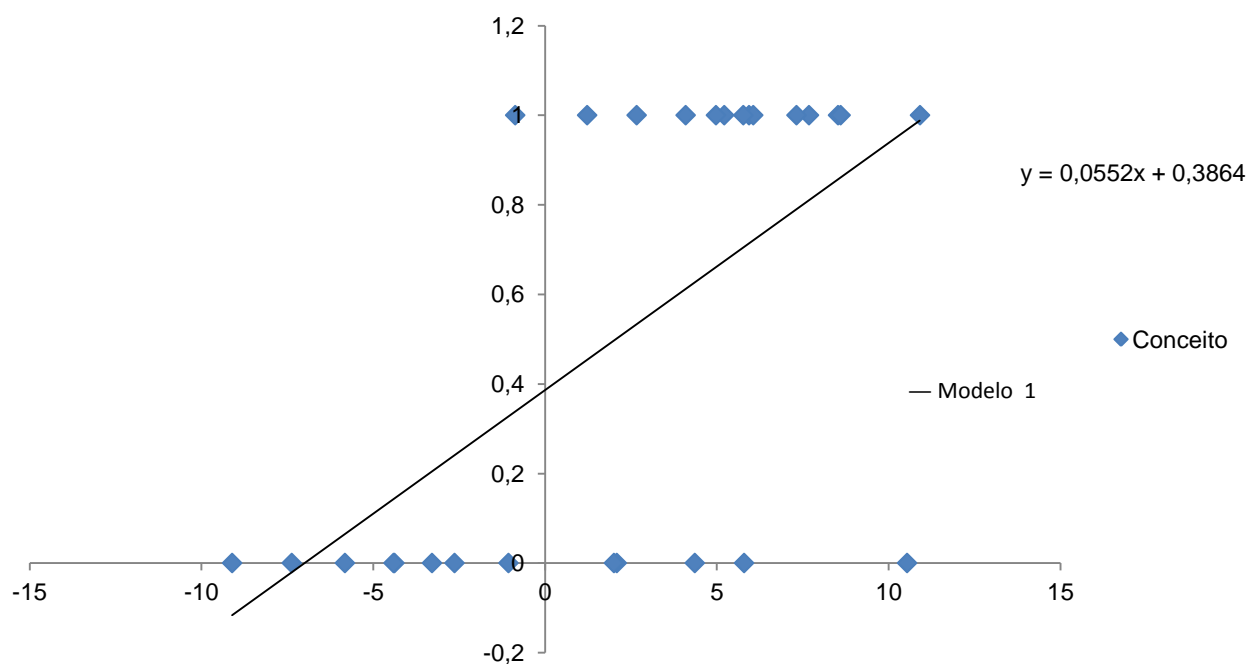


Figura 2: Gráfico de dispersão Delta 1 vs Conceito, a reta ajustada (Modelo 1).

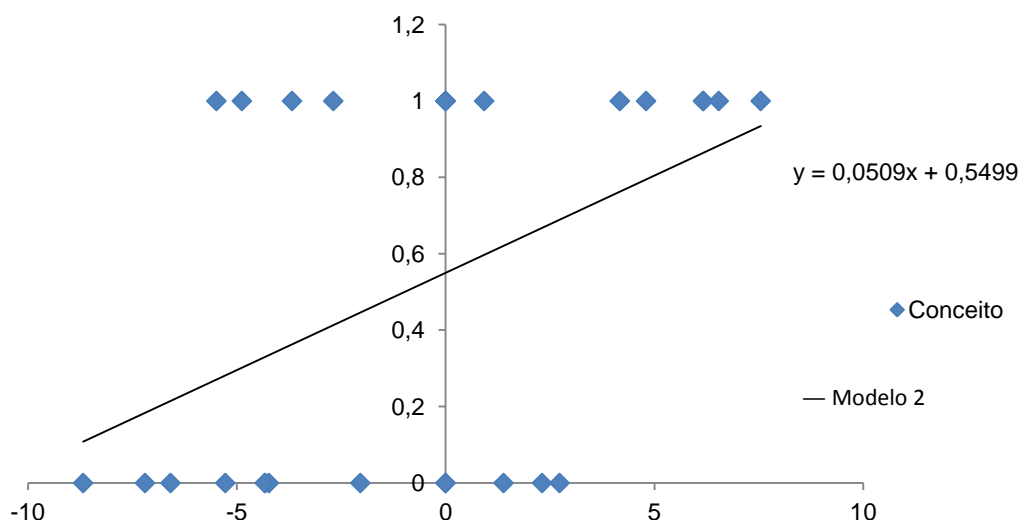


Figura 3: Gráfico de dispersão Delta 2 vs Conceito, a reta ajustada (Modelo 2).

Lembrando que esta análise ainda é falha para o presente momento, pois o objetivo é analisar se um músculo cardíaco é “bom” sob ação do cálcio e sob ação do isopropelenol isoladamente, assim modelamos o seguinte, usando Delta 1, Delta 2 para obter o conceito. Usando assim Delta 1, Delta 2 como variáveis a priori e Conceito com a posteriori. (Modelo 3)

Encontrando assim a fórmula:

$$\bar{y} = 0,0487x_1 + 0,0394x_2 + 0,4261.$$

Onde temos o índice de assertividade de 81,48%.

Delta 1	Delta 2	Bom(1) Ruim (0)	Simulação 3	Assertividade (Acerto = 1) (Erro = 0)
5,9322	4,8023	1	0,9042	1
-9,1146	0,0000	0	-0,0178	1
-2,6403	1,3889	0	0,3522	1
6,0547	0,0000	1	0,7210	1
5,2104	6,1657	1	0,9228	1
4,9689	7,5472	1	0,9654	1
-3,2967	-5,2747	0	0,0577	1
7,6763	-5,4878	1	0,5837	1
4,0872	-3,6765	1	0,4803	0
2,6608	-4,8780	1	0,3635	0
5,7650	0,0000	1	0,7069	1
8,5227	-2,6882	1	0,7352	1
1,2195	6,1728	1	0,7287	1
10,9091	0,0000	1	0,9574	1
-4,3893	-6,5868	0	-0,0472	1
-7,3789	2,3114	0	0,1578	1
4,3532	0,0000	0	0,6381	0
-1,0707	2,7299	0	0,4815	1
-4,4164	-4,3210	0	0,0408	1
10,5330	-4,2308	0	0,7724	0
5,7882	-4,3280	0	0,5375	1
2,0833	-8,6792	0	0,1856	1
7,3132	4,1734	1	0,9467	1
8,5981	6,5421	1	1,1026	1
2,0000	-2,0408	0	0,4431	1
-0,8746	0,9245	1	0,4199	0
-5,8309	-7,1975	0	-0,1414	1

A assertividade aqui não é alta, pois há uma conceituação empírica muito forte. Pode-se observar que o mesmo acontece para a Massa de validação. 66,66% de assertividade.

Delta 1	Delta 2	Bom(1) Ruim (0)	Simulação 3	Assertividade (Acerto = 1) (Erro = 0)
15,7729	5,3628	1	1,4055	1
3,2544	-3,8147	1	0,4343	0
11,7048	6,5327	1	1,2535	1
1,9403	10,6312	1	0,9395	1
-3,3003	5,0657	0	0,4650	1
5,0584	7,8895	1	0,9833	1
0,0000	7,2670	0	0,7124	0
-4,4304	17,9856	0	0,9190	0
0,0000	-1,3487	0	0,3730	1

Mas se transformarmos linearmente usando Modelo 1 (\square_1) e Modelo 2 (\square_2), os Delta 1 e Delta 2 respectivamente temos que a modelagem para obter o Modelo 4 usando estas novas variáveis afim de uma resposta otimizada.

Isso devido ao fato do conceito ser uma variável discreta e não contínua. A explicação de aplicarmos \square_1 e \square_2 é que elas são transformações lineares que preservam comportamentos, logo não se perde a generalidade no modelo.

Encontrando aqui uma assertividade de 77,77%.

Delta 1	Delta 2	Simulação 1	Simulação 2	Bom(1) Ruim (0)	Simulação 4	Assertividade (Acerto = 1) (Erro = 0)
5,9322	4,8023	0,7139	0,7943	1	0,9048	1
-9,1146	0,0000	0,0000	0,5499	0	0,0845	1
-2,6403	1,3889	0,2407	0,6206	0	0,3520	1
6,0547	0,0000	0,7206	0,5499	1	0,7213	1
5,2104	6,1657	0,6740	0,8637	1	0,9234	1
4,9689	7,5472	0,6607	0,9341	1	0,9661	1
-3,2967	-5,2747	0,2044	0,2814	0	0,0571	1
7,6763	-5,4878	0,8101	0,2706	1	0,5839	1
4,0872	-3,6765	0,6120	0,3628	1	0,4803	0
2,6608	-4,8780	0,5333	0,3016	1	0,3633	0
5,7650	0,0000	0,7046	0,5499	1	0,7072	1
8,5227	-2,6882	0,8569	0,4131	1	0,7357	1
1,2195	6,1728	0,4537	0,8641	1	0,7290	1
10,9091	0,0000	0,9886	0,5499	1	0,9581	1
-4,3893	-6,5868	0,1441	0,2146	0	0,0000	1
-7,3789	2,3114	0,0000	0,6676	0	0,1757	1
4,3532	0,0000	0,6267	0,5499	0	0,6383	0
-1,0707	2,7299	0,3273	0,6889	0	0,4815	1
-4,4164	-4,3210	0,1426	0,3300	0	0,0401	1
10,5330	-4,2308	0,9678	0,3346	0	0,7729	0
5,7882	-4,3280	0,7059	0,3296	0	0,5376	0
2,0833	-8,6792	0,5014	0,1081	0	0,1852	1
7,3132	4,1734	0,7901	0,7623	1	0,9474	1
8,5981	6,5421	0,8610	0,8829	1	1,0000	1
2,0000	-2,0408	0,4968	0,4460	0	0,4430	1
-0,8746	0,9245	0,3381	0,5970	1	0,4198	0
-5,8309	-7,1975	0,0645	0,1835	0	0,0000	1

Observamos que ao aplicarmos Modelo 4 no conjunto da Massa de validação tem assertividade de 66,66% de havendo aqui um comportamento de melhora em relação ao modelo anterior.

Delta 1	Delta 2	Simulação 1	Simulação 2	Bom(1) Ruim (0)	Simulação 4	Assertividade (Acerto = 1) (Erro = 0)
15,7729	5,3628	1,0000	0,8229	1	1,0000	1
3,2544	-3,8147	0,5660	0,3557	1	0,4342	0
11,7048	6,5327	1,0000	0,8824	1	1,0000	1
1,9403	10,6312	0,4935	1,0910	1	0,9401	1
-3,3003	5,0657	0,2042	0,8077	0	0,4649	1
5,0584	7,8895	0,6656	0,9515	1	0,9840	1
0,0000	7,2670	0,3864	0,9198	0	0,7127	0
-4,4304	17,9856	0,1418	1,0000	0	0,5587	0
0,0000	-1,3487	0,3864	0,4812	0	0,3728	1

Mas como nossos dados apresentam um índice baixo vamos tomar como melhor o Modelo 3, mas, se a quantidade de dados fossem maiores o Modelo 4 se tornaria um modelo o qual podemos utilizar para classificar o quão “bom” foi o desempenho cardíaco após a utilização da droga.

Analisando os quatro modelos inicialmente através da estatística descritiva, observando os seguintes pontos: mínimo, máximo, mediana, primeiro quartil e terceiro quartil, como mostra nas duas seguintes tabelas, uma referente a massa de treinamento e a outra referente a massa de validação:

Estatística descritiva da massa de treinamento:

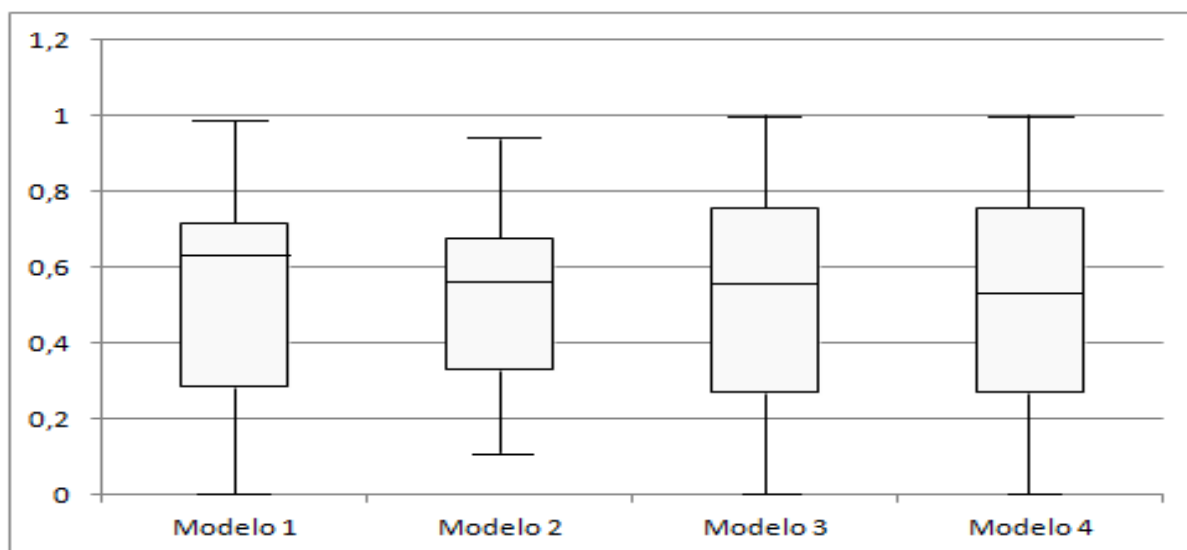
Estatística	Modelo 1	Modelo 2	Modelo 3	Modelo 4
1º quartil	0,2840	0,3298	0,2689	0,2686
mínimo	0,0000	0,1081	0,0000	0,0000
mediana	0,6120	0,5499	0,5375	0,5376
máximo	0,9886	0,9341	1,0000	1,0000
3º quartil	0,7172	0,6782	0,7538	0,7543

Estatística descritiva da massa de validação:

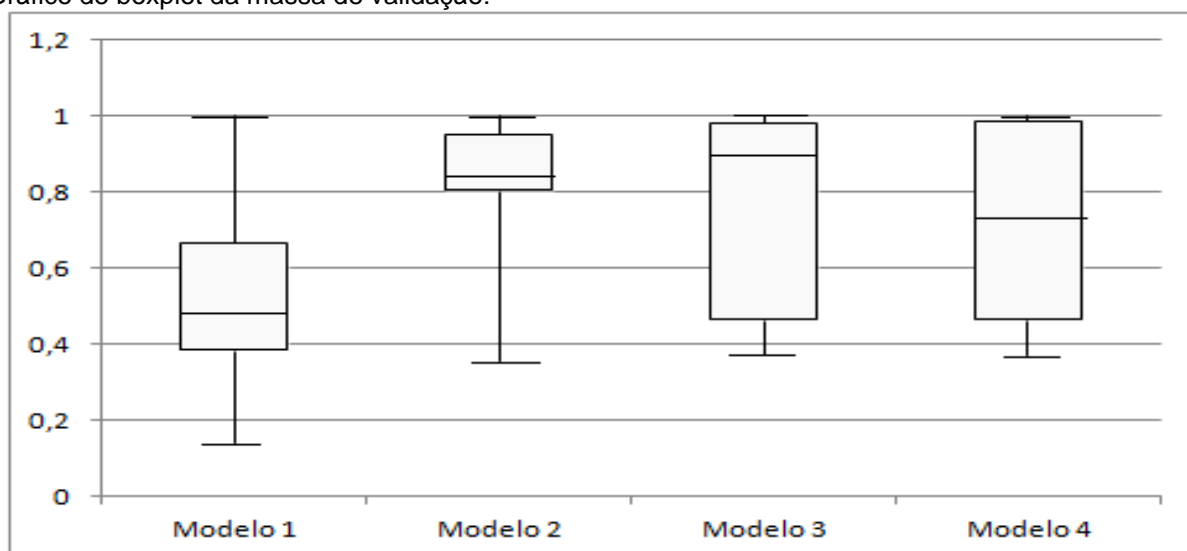
Estatística	Modelo 1	Modelo 2	Modelo 3	Modelo 4
1º quartil	0,3864	0,8077	0,4650	0,4649
mínimo	0,1418	0,3557	0,3730	0,3728
mediana	0,4935	0,8824	0,9190	0,7127
máximo	1,0000	1,0000	1,0000	1,0000
3º quartil	0,6656	0,9515	0,9833	0,9840

Obtemos os seguintes gráficos de boxplot:

– Gráfico de boxplot da massa de treinamento:



– Gráfico de boxplot da massa de validação:



Observa-se que os modelos 1 e 2 apresentam distribuições muito concentradas pois o primeiro e o terceiro quartil estão com valores próximos a mediana, uma das razões deste fato é que o conceito (variável posterior) é referente a aplicação do cálcio e do isopropelenol isoladamente para validar se o músculo cardíaco é bom ou ruim. Mas ao observarmos os modelos 3 e 4 essa distribuição já é mais dispersa, além de manter uma amplitude do primeiro ao terceiro quartil da massa de treinamento e da massa de validação o que mostra que os modelos calculados são mais condizentes e mais estáveis.

Portanto para se utilizar como um modelo de mensuração de se o músculo cardíaco é bom ou ruim sob efeito do cálcio e do isopropelenol, isoladamente, seriam os modelos 3 e 4. Pois estes modelos teriam mais êxito no auxílio de tomadas de decisões.

Conclusão

Podemos observar que com apenas um simples modelo matemático já auxilia e muito a mensuração de um determinado comportamento sem haver a necessidade do empírico que em geral é impreciso.

Além disso, pode-se observar que dependendo das variáveis que estamos trabalhando há necessidade de se fazer transformações, esta é uma teoria muito estudada pelos “desenvolvedores de modelos estatísticos”.

Os modelos elaborados são necessários para se ter uma análise precisa em relação a tomada de decisão, como se tratava de músculos cardíacos. Podendo se concluir uma melhora em relação aos métodos de mensuração da melhora do coração após a aplicação do cálcio e do isopropelenol não simultaneamente.

Referências

Apostila de Estatística. Disponível em: <
http://portal.saude.gov.br/portal/arquivos/pdf/apostila_estatistica.pdf> Acesso em 29 de março de 2012.
Apostila de Ensino CNEC- Sistema de Ensino CNEC, 1º série, Volume 1, Ensino Médio, edição 2011.

COSTA NETO, PEDRO LUIZ DE OLIVEIRA – **ESTATÍSTICA** – 2ª edição – São Paulo: Editora Blucher, 2002. Cap. 8 Correlação e Regressão, página 197 à 214.

Dissertação (mestrado), Campos, Dijon Henrique Salomé de. **Influência da oxigenação de glicose sobre a função miocárdica de ratos obesos-** Botucatu, 2010, página 31 à 41.

Importância da Estatística para tomada de decisões. Disponível em:
<http://www.ipardes.gov.br/biblioteca/docs/NT_06_importancia_estatistica_tomada_decisao.pdf> Acesso em 29 de março de 2012.

Sociedade Brasileira de Zootecnia. Disponível em: <<http://www.scielo.br/pdf/rbz/v36s0/26.pdf> > Acesso em 22 de outubro de 2012.

TOLEDO, GERALDO LUCIANO- **ESTATÍSTICA BÁSICA** - 2ª edição - São Paulo: Atlas, 1985. Cap. 8, Regressão linear simples, pag. 427 à 432.

Universidade de São Paulo. Regressão Linear. Disponível em:
<http://www.usp.br/fau/cursos/graduacao/arg_urbanismo/disciplinas/aut0516/Apostila_Regressao_Linear.pdf> Acesso em dia 27 de janeiro de 2013.

WILTON DE O. BUSSAB, A. MORETTIN- **ESTATÍSTICA BÁSICA**- 5ª edição – São Paulo: Saraiva, 2004.