



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

O MODELO YULE-SIMON: UMA DESCRIÇÃO A DADOS DE PRODUÇÃO ANIMAL

Área de concentração: Biometria e Estatística Aplicada

Orientador: Prof. Dr. Guilherme Rocha Moreira
Coorientador: Prof. Dr. Josimar Mendes de Vasconcelos

Recife - PE

13 de abril de 2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

J62m COSTA, JOÃO BATISTA FILGUEIRA COSTA
MODELO YULE-SIMON: UMA DESCRIÇÃO A DADOS DE PRODUÇÃO ANIMAL: DESCRIÇÃO DO
MODELO YULE-SIMON E APLICAÇÃO / JOÃO BATISTA FILGUEIRA COSTA COSTA. - 2021.
25 f.

Orientador: Guilherme Rocha Moreira.
Coorientador: Josimar Mendes Vasconcelos.
Inclui referências.

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em
Biometria e Estatística Aplicada, Recife, 2021.

1. Distribuição de Probabilidade Yule-Simon,. 2. Regressão, . 3. Estimação de Mínimos Quadrados,. 4.
Efetivos de Animais.. I. Moreira, Guilherme Rocha, orient. II. Vasconcelos, Josimar Mendes, coorient. III.
Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**O MODELO YULE-SIMON: UMA DESCRIÇÃO A DADOS DE PRODUÇÃO
ANIMAL**

João Batista Filgueira Costa

Orientador:

Prof. Dr. Guilherme Rocha Moreira
Orientador

Banca examinadora:

Dr. Josimar Mendes de Vasconcelos
Universidade Federal Rural de
Pernambuco

Dra. Rita de Cássia de Lima Idalino
Universidade Federal do Piauí

A Lorena e Guiga

AGRADECIMENTOS

Sou grato a todos que de forma direta ou indireta, por ter me concedido a graça da vida, e ter permitido chegar até aqui.

Aos meus pais Humberto Saturnino e Cecília Filgueira, que são a razão da minha vida e luta.

À minha Tia Marinês Barros, pelo carinho e atenção que sempre teve e tem com o bem viver da família.

Aos primo(as) João Barros e Kátia Rejane, pelo apoio e ajuda.

Ao meu Orientador Guilherme Rocha Moreira pela seu profissionalismo e dedicação.

Ao meu coorientador Josimar Mendes de Vasconcelos pela paciência e ajuda no decorrer no período como mestrando.

Ao Coordenador Prof. Moacyr Cunha Fylho pelas orientações e como Professor de Seminários.

Ao Secretário Marco Antônio dos Santos pela ajuda durante todo o decorrer deste período.

A todos meus amigos do programa de pós-graduação pela interação produtiva e harmoniosa durante nosso convívio.

Aos professores e funcionários do Departamento de Estatística e Informática pela convivência durante todo período.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*Eu meço o valor do homem pela medida que ele se liberta de seu próprio eu.
(Albert Einstein,)*

RESUMO

O MODELO YULE-SIMON, proposto por George Udne Yule, (1924) - Herbert A. Simon (1955), é uma teoria estatística aplicada para obter resultados em análises de bancos de dados que apresentem eventos discrepantes ou *outliers*. É uma distribuição de probabilidade discreta, que representa o mecanismo de probabilidade subjacente, o comportamento dos gráficos de linhas são de caudas longas, tem formato de J invertido, conhecido como processo estocástico de YULE. Esse modelo tem diversas aplicações, principalmente nas áreas: ciências sociais, biológicas, Agrárias, dentre outras. Um processo de descrição da distribuição é realizado, usando a definição da distribuição beta incompleta. Aplicando o método dos mínimos quadrados encontramos as estimativas dos parâmetros, α e β , associados a regressão, Log-Yule - Simon não linear. Assim tem-se como proposta aplicar o modelo em dados de produção animal. As análises foram realizadas em bancos de dados, em que foram classificados como classes. Nesse viés, destacamos: Búfalos, Asininos, Leite, Galinhas e Ovos, por classe de animal. Os bancos de dados mostraram-se interessantes ao modelo. As ferramentas analisadas para verificar se o modelo é razoável aos bancos de dados foram: as estatísticas descritivas, coeficiente de determinação, r^2 , o p-valor e o teste qui-quadrado. Os resultados indicam que o modelo estocástico desenvolvido por G. Udny Yule e Herbert A. Simon, como o mecanismo de probabilidade subjacente, fornece uma boa descrição dos dados empíricos na área de produção animal. O resultado prático obtido foi mostrar que o modelo Yule-Simon pode ser um possível modelo a ser aplicado em dados de produção animal.

Palavras-chave: Distribuição de Probabilidade Yule-Simon, Regressão, Estimação de Mínimos Quadrados, Efetivos de Animais.

ABSTRACT

THE YULE-SIMON MODEL, proposed by George Udne Yule, (1924) - Herbert A. Simon (1955), is a statistical theory applied to obtain results in analysis of databases that present discrepant events or outliers. It is a discrete probability distribution, which represents the underlying probability mechanism, the behavior of the line graphs is long tails, has an inverted J shape, known as the YULE stochastic process. This model has applications, mainly in the areas: social, biological, agrarian sciences, among others. A distribution description process is carried out, using an incomplete beta distribution definition. Applying the method of found polygons found as the base of the parameters, α and β , associated with regression, Log-Yule - Simon non-linear. Thus, it is proposed to apply the model to animal production data. The analyzes were carried out in databases in which they were classified as the classes that stood out were: Buffaloes, Donkeys, Milk, Chickens and Eggs. The different databases are interesting to the model. The tools analyzed to verify whether the model is reasonable to the databases were: descriptive statistics, coefficient of determination, r^2 , the p-value and the chi-square test. The results indicated that the stochastic model developed by G. Udny Yule and Herbert A. Simon, as the underlying probability mechanism, offers an excellent description of the empirical data in the area of animal production. The practical result obtained was to show that the Yule-Simon model can be applied to animal production data.

Key-words: Yule-Simon Probability Distribution, Regression, Least Squares Estimation, Effective Animals.

LISTA DE FIGURAS

Figura 1 - George Udne Yule.....	2
Figura 4.1.3 - Gráficos Box-Plot do modelo.....	20
Figura 4.1.4 - Gráficos Box-Plot condensados.....	21

LISTA DE TABELAS

Tabela 3.1 - Estatísticas descritivas.....18

Tabela 3.2 - Estimativas e estatísticas.....19

LISTA DE SIGLAS

IBGE - Instituto Brasileiro de Geografia e Estatística

fmp - Função massa de probabilidade

Log - Regressão logarítmica

ABNT - Associação brasileira de normas técnicas.

c.v - Coeficiente de variação de Pearson

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Objetivos	3
1.2.1	Geral	3
1.2.2	Específicos	3
1.3	Plataforma computacional utilizada	3
2	A distribuição Yule-Simon	4
2.1	Introdução	4
2.2	Modelo Yule-Simon	5
2.3	Inferência Estatística	7
2.4	Modelos de Regressão	7
3	Aplicação	9
3.1	Descrição dos Dados	9
3.2	Processos de Modelagem	11
4	Conclusão	12
4.1	Conclusão dos Resultados	12
4.2	Trabalhos Futuros	12
5	Referências Bibliográficas	13

1 Introdução

1.1 Introdução

A distribuição Yule ou distribuição Yule-Simon é uma distribuição de probabilidade discreta altamente distorcida, nomeada em homenagem a George Udny Yule ou modelo estocástico de Yule. Essa distribuição foi aplicado pela primeira vez em 1924 (CHUNG, Kee H.; COX, Raymond AK. 1994). O desenvolvimento da distribuição foi aplicando-a as distribuições de gêneros biológicos por número de espécies (Cattuto.C, et al, 2006).



Figura 1 – G. U. Yule, 1924

Seu pai, Sir George era administrador do Serviço Civil da Índia e membro de uma antiga família de agricultores escoceses com uma história de alguma distinção governamental, militar e literária (YATES, Frank. George Udny Yule 1952). George Udny Yule estatístico britânico, foi palestrante do Congresso Internacional de Matemáticos em Toronto em (1924).

As principais realizações de Yule na teoria estatística diz respeito a regressão, correlação, associação e séries temporais. Suas primeiras memórias sobre correlação (1897, 1907) e coeficientes de associação (1900) provaram ser fundamentais (YATES, Frank. George Udny Yule 1952).

Desde 1912, (MILLS, Terence C.2017) Yule e o major, M. Greenwood lançaram os

fundamentos da teoria da distribuição de acidentes. Em 1921, ele escreveu sobre correlação temporal e iniciou um trabalho que levou a um artigo bem conhecido sobre manchas solares, que marcou o início da teoria moderna das séries temporais oscilatórias biométrica. Como pesquisador George Undny Yule fez importantes estudos sobre a matemática da evolução biológica e as estatísticas dos ensaios agrícolas. George Udny Yule foi talvez o primeiro cientista a considerar se as correlações observadas entre pais e filhos poderiam ser explicadas pela herança multifatorial mendeliana, um problema retomado mais tarde por R. A. Fisher, (Yates, Frank 1952).

A distribuição YULE é uma função de massa de probabilidade discreta, ou seja, faz parte da família de distribuições discretas, publicada pela primeira vez em 1924 por George Undny Yule (Simon, H. A. 1960), é considerada como uma função de caudas longas ou pesadas, seus gráficos tem um formato de J invertido (SIMON Herbert A. 1955).

O trabalho tem como objetivo, estudar valores atípicos em banco de dados de produção animal. Os fatores que motivam o estudo em destaque, é a possível flexibilidade da função massa de probabilidade (f.m.p) de Yule, alguns estudos ocorrem com a distribuição, uma aplicação realizada na área de produção artística feita por Laura Spierdijk and Mark Voorneveld em Superstars without Talent? The Yule Distribution Controversy, Ela estima o efeito bola de neve, por exemplo na compra de gravações artísticas o que pode motivar os consumidores a comprar os mais bem sucedidos autores, (SPIERDIJK, VOORNEVELD, 2009). Nesse artigo os resultados indicam que existe uma aparente ausência de correlação, entre sucesso e talento, visto que essas duas variáveis são consideradas como medidas subjetivas e seus efeitos subjacentes adaptando-se ao modelo Yule-Simon, Em tese, o modelo alcança valores distantes do centro da média, mas pode ser que não atinge os valores potentes equidistante, do ponto de vista linear.

A presença de *outliers*, na base de dados, afeta diretamente nos resultados, e consequentemente nas conclusões, dessa maneira, enviesando os resultados das pesquisas. Muitos pesquisadores quando estão diante dessa situação, por falta de informações, preferem retirar o evento aberrante da sua base de dados em suas análises, mas existem técnicas estatísticas específicas, a exemplo do modelo em estudo Yule-Simon, para subsidiar pesquisadores.

Para tanto, nosso estudo inicia-se com a introdução, abordando um aspecto teórico. No segundo capítulo, temos a revisão de literatura relatando sobre o Modelo Yule-Simon e demais ferramentas estatísticas utilizadas no trabalho. Para ilustrar o modelo no contexto aplicado, têm-se no terceiro capítulo uma aplicação real. Por fim, a conclusão com as considerações gerais e trabalhos futuros.

1.2 Objetivos

1.2.1 Geral

No que tange aos questionamentos da tese, temos como o objetivo geral aplicar o modelo Yule-Simon a dados de produção animal, na qual tenha comportamento de dados *outlier*.

1.2.2 Específicos

- Estudar o modelo Yule-Simon, nos aspectos teóricos e práticos;
- Apresentar a distribuição Yule e contribuir para o desenvolvimento na área de ciências agrárias;
- Aplicar o modelo a dados de produção animal, motivando o produtor a investir e obter ganhos com o plantel.

1.3 Plataforma computacional utilizada

A Performance ou desempenho da distribuição de probabilidade, (Kee H. Chung and Raymond A. K. Cox), em que para obter bons resultados, é necessário medir o nível do desempenho dos modelos usados em análises de dados. Os ganhos podem ser, usar um software gratuito e otimizar trabalho e tempo. Na área das Ciências Estatísticas, enfatiza-se e prioriza o uso e aplicação de linguagens de programação dos softwares livres ou linguagens de códigos abertos. Neste estudo usou-se a linguagem de programação R-Stúdio versão 1.2.13. Esse software é uma linguagem de código livre e aberto para análises de dados estatísticos, R core team(2019).

2 A distribuição Yule-Simon

2.1 Introdução

Em análises de dados estatísticos existem uma gama de alternativas para tratamentos de informações, ou seja técnicas apropriadas para estudar fenômenos. Essas são utilizadas de acordo com os diferentes tipos de variáveis, as variáveis podem ser tanto de natureza discreta quanto contínua. Dentre as distribuições de probabilidade temos a distribuição de probabilidade de massa Yule que passou a ser chamada de Yule-Simon a partir de 1952, pelo pesquisador Norte Americano Herbert A. Simon. A distribuição de probabilidade de massa Yule-Simon é um modelo com caudas superiores muito longas, seus gráficos tem formato de J invertido (SIMON Herbert A.,1960) e pode ser aplicado a processos estocásticos, planejamento de experimentos e diversas áreas da estatística. As referências de Simon são sempre enfatizadas nas aplicações em processos estocásticos, na qual encontramos referências em dados biométricos, como também em econometria e estudos sobre análises em distribuição de renda. Mas, o maior número de estudos desse modelo é principalmente na área de fenômenos sociais.

A distribuição de probabilidade Yule-Simon (CHUNG, COX, 1994), pode ser usada em análises de dados que represente o mecanismo de probabilidade subjacente. Esse mecanismo de probabilidade indica valores implícitos em fenômenos ou em classes de distribuições empíricas. Valores implícitos ocorrem em situações que por exemplo ao analisar o sucesso das estrelas da música popular, não sabemos se ela é talentosa ou se o sucesso é mera vaidade do público. Nesse caso, ainda existe uma terceira variável a ser considerada, que é a capacidade individual de vendas dos vendedores em lojas de produtos artísticos. O modelo estocástico pode representar o processo de geração do fenômeno dos artistas famosos da musica Americana. Em alguns resultados da aplicação do modelo, os talentos diferenciais não existe entre indivíduos, ou seja o fenômeno das super estrelas podem existir entre pessoas com talentos iguais.

Uma generalização do modelo foi realizada por LÁNSKÝ *et al.* (LÁNSKÝ, Petr; RADIL-WEISS, 1980), sendo sugerido para análises de estudos e comportamentos de informações ou dados não equilibrados. Estudos que tratam de associação de palavras ou palavras mais freqüentes, e sugere ainda, algumas novas aplicações possíveis do modelo, a saber: na área de montagens de células neurais.

As ocorrências de dados não equilibrados ou que podem estar fora do padrão, que são valores fora da média, considerados *outliers*. A generalização do modelo não afasta a hipótese da distribuição de probabilidade de Yule (LÁNSKÝ, Petr; RADIL-WEISS, 1980) permanecer como um modelo discreto, por exemplo, nos processos de contagem de palavras mais frequentes, pode-se considerar uma nova palavra como fora do padrão, ou *outliers*.

Usando o modelo estocástico de Yule (1924) e Simon (1955) como uma representação do comportamento de escolha do consumidor, Simon sugere que uma variedade de fenômenos sociológicos, biológicos e econômicos são movidos por certos mecanismos de probabilidade Chung e Raymond (1994). Yule e Simon mostraram que a distribuição de gênero biológico por número de espécies e distribuição de cientistas por número de artigos publicados está em conformidade com uma classe de distribuições que pode ser obtida a partir de processos estocásticos semelhantes aqueles que produzem distribuições binomiais negativas ou de séries de log (CHUNG, Kee H.; COX, Raymond AK.). A relação que existe entre os estudos de Yule-Simon e nossa proposta é uma descrição de um modelo de análises de regressão a dados discretos.

A classe de distribuição representada pelos autores Yule-Simon foi descrita pela primeira vez pelo pesquisador britânico George Udny Yule (1924), a distribuição é assim chamada, em homenagem a seu nome. Em essência, é conhecido como o processo estocástico de Yule Chung e Raymond (1994).

2.2 Modelo Yule-Simon

Associada a uma variável aleatória X , a sua função de distribuição de probabilidade ou função de massa de probabilidade (f.m.p) é dada por

$$P(X = k) = \rho B(k, \rho + 1),$$

em que k ($k \geq 1$) é um inteiro, ρ ($\rho > 0$) é um número real. Além disso, temos que $B(\cdot, \cdot)$ é a função beta dada por

$$B(k, \rho + 1) = \int_0^1 x^{k-1} (1-x)^\rho dx = \frac{\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k\rho + 1)},$$

em que $\Gamma(\cdot)$ é a função gama.

O modelo Yule-Simon de fato é uma f.m.p. porque satisfaz as duas condições. A primeira que de fato é positiva e a segunda

$$\sum_{k=1}^{\infty} P(X = k) = 1.$$

Para este fato, temos:

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} \rho B(k, \rho + 1).$$

Como $B(k, \rho + 1) = \int_0^1 x^{k-1}(1-x)^\rho$, temos

$$\begin{aligned} \sum_{k=1}^{\infty} P(X = k) &= \sum_{k=1}^{\infty} \rho \int_0^1 x^{k-1}(1-x)^\rho dx \\ &= \rho \int_0^1 (1-x)^\rho \sum_{k=1}^{\infty} x^{k-1} dx, \\ &= \rho \int_0^1 (1-x)^\rho \cdot \left[\frac{1}{1-x} \right] dx. \end{aligned}$$

Como a função converge, temos que a soma da integral, é igual a integral da soma no resultado anterior. Agora seja $u = 1 - x \Rightarrow du = -dx$, ou seja,

$$\sum_{k=1}^{\infty} P(X = k) = \rho \int_0^1 u^{\rho-1} du = 1.$$

A distribuição de probabilidade Yule-Simon tem várias ferramentas estatísticas na literatura, dentre elas são os dois primeiros momentos centrais, o valor esperado e a variância. Assim, se X é uma variável aleatória discreta seguindo uma distribuição Yule-Simon, o seu valor esperado é:

$$E[K] = \frac{\rho}{\rho - 1}, \quad \forall \rho > 1.$$

Conseqüentemente, temos uma outra medida que é a variância:

$$Var[K] = \frac{\rho^2}{(\rho - 1)^2(\rho - 2)}, \quad \forall \rho > 2.$$

Seja X uma variável aleatória seguindo uma distribuição Yule-Simon, então a função geradora de momentos, assimetria e a moda são dadas por:

1. Função geradora de momentos

$$\phi_X(t) = \frac{\rho}{(\rho + 1)^2} F_1(1, 1; \rho + 2; e^t) e^t.$$

2. Assimetria

$$\phi_X(t) = \frac{(\rho + 1)^2 \sqrt{\rho - 2}}{(\rho - 3)\rho}, \quad \text{para todo } \rho > 3.$$

3. Moda

$$Mo = 1.$$

2.3 Inferência Estatística

A Inferência estatística surge a partir da necessidade para realizar análises de dados amostrais, associada a uma determinada população em estudo, normalmente usada para avaliar, planejar e projetar soluções de problemas teóricos e práticos. O pesquisador poderá testar hipóteses das possibilidades das generalizações resultados aos experimentos semelhantes, na prática estas hipóteses são comparadas, que poderão ser rejeitadas ou não. Em fenômenos estudados em estatística, existem três tipos de modelos a ser aplicados, os de natureza discretas, contínuos e misto (Casella, George, 2002).

2.4 Modelos de Regressão

Neste estudo utilizamos as técnicas de análise de regressão não linear para avaliar os efeitos de um modelo discreto. Para estimar os parâmetros aplicamos o método dos mínimos quadrados, dentre os métodos de estimação, este método é uma opção para o cálculo dos estimadores de α, β .

A relevância desta metodologia proporciona um ajuste de curvas dos gráficos mais flexível e realista (John O. Rawlings, Sastry G. Pantula, David A. Dickey, pág 236) e traz um ganho significativo para o banco de dados em análises, o modelo ajustado obtém resultados significativos incluindo os eventos discrepantes.

Seja X uma variável aleatória seguindo uma distribuição Yule-Simon, então o modelo de regressão é dada por (CHUNG e COX, (1994))

$$\log(x_i^{\rho+1}) = \alpha + \beta \log(x_i) + \xi_i, \quad \forall i = 1, \dots, n.$$

Por outro lado, temos que a soma dos erros aleatórios são iguais a zero, ou seja

$$\sum_{i=1}^n \xi_i^2 = 0.$$

Assim, temos:

$$\underbrace{\sum_{k=1}^n [(\rho + 1) \log(x_i) - \alpha - \beta \log(x_i)]^2}_{Q_i} = 0$$

Fazendo $\frac{\partial Q_i}{\partial \alpha} = 0$, temos

$$\hat{\alpha} = -\frac{1}{n} \cdot (\hat{\rho} + 1 + \hat{\beta}) \cdot \sum_{i=1}^n \log(x_i)$$

Para obter o parâmetro β , fazemos $\frac{\partial Q_i}{\partial \beta} = 0$, ou seja,

$$\hat{\beta} = -(\hat{\rho} + 1).$$

3 Aplicação

3.1 Descrição dos Dados

Ao analisar bancos de dados, é possível elencar uma série de estatísticas, que de posse delas, o pesquisador faz inferência para tirar conclusões acerca de seus estudos. Nesta seção serão apresentados algumas estatísticas descritivas e estimativas, obtidas através das análises e uma série de aplicações, em que as fontes das informações são dos bancos de dados depositados no site do IBGE, dados coletados no CENSO de 2006 (Instituto Brasileiro de Geografia e Estatística, 2006).

Recorrendo as informações sobre produção animal, nos 26 Estados e Distrito Federal, com maior efetivo de rebanho ou matéria prima da cadeia produtiva de alimentos de origem animal, foram selecionados cinco banco de dados e analisados no modelo para uso específico. Nesse contexto aplicamos o modelo Yule-Simon para essa modelagem. As informações descritas na tabela abaixo, são as estatísticas obtidas das classes búfalos, asininos, ovos, galinhas e leite, na sequência aparece a média, o desvio padrão, mediana, coeficiente de variação de Pearson (c.v.), valores mínimo (min.) e máximo (max.).

Tabela 1 – Estatísticas descritivas

Classes	Média	Desvio padrão	Mediana	c.v.	min	max
Búfalos	57,616	9862,329	21,297	172,3618	-1,22	0,63
Asininos	59,18	7703,457	12,736	148,3036	-2,11	0,91
Ovos	109,662	30337,49	34,309	158,8304	-1,66	0,855
Galinhas	38,823	3221,044	12,7895	146,1844	-2,515	0,875
Leite	1,0639	2,030	0,5798	133,9363	-0,886	0,991

Na Tabela 3.1 estão as estatísticas descritivas dos dados para os cinco banco de dados. O resultado da classe ovos possui o maior desvio padrão (30327,49), enquanto que os da produção de leite, apresenta o menor desvio padrão (2,030). O c.v. da classe búfalos apresenta uma maior dispersão (172,3618), considerada muito elevada. Os dados da classe búfalos possui uma média maior do que a mediana. A classe ovos apresentou uma maior média, variância, mediana, maior c.v. sendo a classe que teve suas estatísticas mais elevadas. Para todas as classes observadas a média é maior que a mediana, indicando que a distribuição é assimétrica positiva ou seja, os dados estão concentrados à direita. Todas as classes observadas apresentam valores para o c.v. acima de 30, indicando que os

dados estão dispersos em torno da média, tendo uma elevada dispersão. Uma observação a ser considerada sobre as informações dos *outliers* será a classe Leite, o valor médio dessa classe foi 1,0639 comparando com a média das demais classes é um valor discrepante assim as classes leite e ovos com média 109,662 apresentam valores atípicos ou *outliers*.

A seguir ilustram-se os dados por meio do *box-plot*. Essa figura resume uma parte das informações contidas por classes nos banco de dados, indica que existe uma boa distribuição dos dados e diferentes *box-plots*, por classes observadas. Os Gráficos *box-plots* apresentam mediana aproximada da média, sendo uma boa medida para indicação da boa qualidade do modelo aos dados.

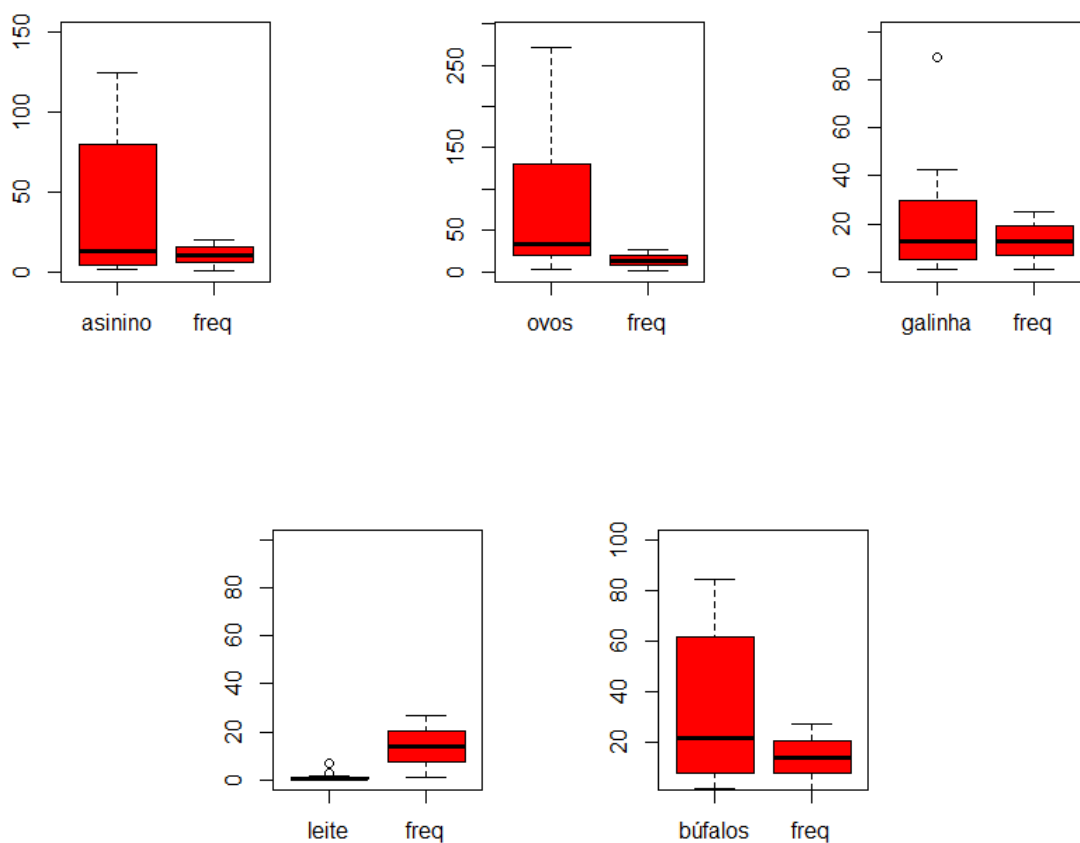


Figura 2 – Gráficos das propriedades da distribuição Yule-Simon usando o R-stúdio.

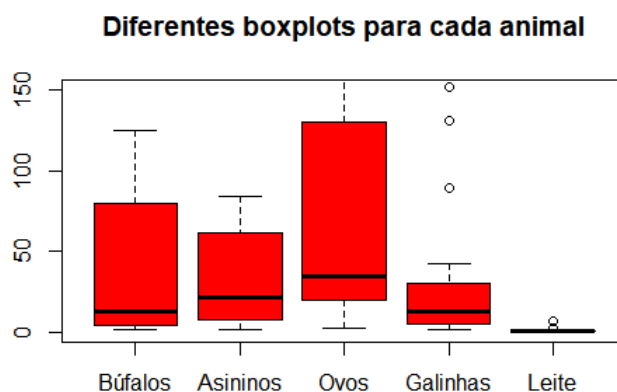


Figura 3 – Gráficos condensados Yule - Simon, no R-Stúdio.

3.2 Processos de Modelagem

O processo de modelagem da distribuição Yule–Simon, resultou na estimação dos seguintes parâmetros, α de β , da estatística de teste de hipóteses, qui-quadrado χ^2 , a informação de Fischer-F, o coeficiente de determinação, r^2 , e o valor-p. O critério de decisão usado, para justificativa da adequação do modelo, foi o coeficiente de determinação r^2 , o valor-p e a estatística de teste, qui-quadrado χ^2 .

Tabela 2 – Estimativas, estatísticas e valor-p

Classes	$\hat{\alpha}$	$\hat{\beta}$	χ^2	F	r^2	valor-p
Búfalos	0,21	-1,52	47,92	73,03	0,79	9,5e-08
Asininos	0,78	-1,75	316,13	69,18	0,78	1,4e-07
Ovos	0,393	-1,41	437,4238	84,33	0,76	1,74e-09
Galinhas	0,799	-1,443	394,766	54,86	0,68	1,2e-07
Leite	-0,6114	-0,7635	7,2604	41,38	0,60	9,79e-07

A Tabela 3.2 descreve as estimativas, estatísticas de testes, r^2 . O coeficiente de determinação da classe búfalos, indica um valor aproximadamente de 79% tendo, assim uma forte relação da adequação da distribuição Yule-Simon aos dados. Logo em seguida a classe asininos medindo 78%, ovos, galinhas e leite medindo 76%, 68% e 60%, respectivamente. No entanto, as classes galinhas que apresentam o valor 68% e leite 60% estes indicam uma média relação da adequação ao modelo, abaixo de 70%. Todas as classes de animais apresentam um valor-p, menor que 0,05, mostra que não se rejeita nenhuma delas ao modelo.

4 Conclusão

4.1 Conclusão dos Resultados

Neste trabalho analisamos os resultados da aplicação do modelo Yule-Simon em bancos de dados de produção animal (classes búfalos, asininos, ovos, galinhas e leite). Nessa aplicação fez-se uma análise descritiva dos dados e ilustrou-se os dados através de Figuras, como *box-plot*. Em seguida, Modelou-se os dados da área de produção animal com o modelo Yule-Simon, tendo como referência o coeficiente de determinação (r^2) e a estatística (χ^2). Nisso, observou-se que o modelo se ajusta bem aos dados. Neste caso recomenda-se o uso deste modelo a dados dessa natureza. Ressaltamos que mais estudos precisam ser feitos nesses contexto.

4.2 Trabalhos Futuros

Pretende-se continuar com as pesquisas com a distribuição Yule-Simon para desenvolver mais ferramentas matemáticas e estatísticas, como por exemplo: estender a regressão múltipla, métodos de diagnósticos e testes de hipóteses para modelo de regressão múltipla.

5 Referências Bibliográficas

SPIERDIJK, Laura and Voorneveld, Mark. Superstars without talent? The Yule distribution controversy. *The Review of Economics and Statistics*, 2009.

BAEZA-YATES, R., & Ribeiro-Neto, B. (2013). *Recuperação de Informação-: Conceitos e Tecnologia das Máquinas de Busca*. Bookman Editora.

SIMON, Herbert A. On a class of skew distribution functions. *Biometrika*, v. 42, n. 3/4, p. 425-440, 1955.

YATES, Frank. *George Udny Yule 1871-1951*, 1952.

HORVATH, W. J., & Foster, C. C. (1963). Stochastic models of war alliances. *Journal of conflict resolution*, 7(2), 110-116.

GARCIA Garcia, Juan Manuel (2011). "Um algoritmo de ponto fixo para estimar o parâmetro de distribuição Yule- SIMON". *Matemática Aplicada e Computação* . 217 (21): 8560- 8566. doi : 10.1016 / j.amc.2011.03.092 .

NEWEL, K. M., Slobounov, S. M., Slobounova, E. S., & Molenaar, P. C. M. (1997). Stochastic processes in postural center-of-pressure profiles. *Experimental Brain Research*, 113(1), 158-164.

CASSETTARI, Rafael-Roeck-Borges et al. COMPARAÇÃO DA LEI DE ZIPF EM CONTEÚDOS TEXTUAIS E DISCURSOS ORAIS. *El profesional de la información*, v. 24, n. 2, 2015.

SIMON, H. A. (1960). Some further notes on a class of skew distribution functions. *Information and Control*, 3(1), 80-88.

FERRARI, Pablo Augusto; GALVES, Antonio. *Acoplamento e processos estocásticos*. Rio de Janeiro: IMPA, 1997.

CATTUTO, C.; LORETO, V.; SERVEDIO, V. D. P. A Yule-Simon process with memory. *EPL (Europhysics Letters)*, v. 76, n. 2, p. 208, 2006.

ALVARADO, Rubén Urbizagástegui; OLIVEIRA, Marlene. A produtividade dos autores na antropologia Brasileira. *DataGramZero-Revista de Ciência da Informação*, v. 2, n. 6, 2001.

Referências Bibliográficas

CHUNG, Kee H.; COX, Raymond AK. A stochastic model of superstardom: An application of the Yule distribution. *The Review of Economics and Statistics*, p. 771-775, 1994.

<<https://pt.wikipedia.org/wiki/George-Udny-Yule>> Acesso em 17 de Janeiro de 2021

CASELLA, GEORGE. In *Statistical Inference* / George Casella, Roger 2nd - p cm, 2002.

MILLS, Terence C. *Uma biografia estatística de George Udny Yule: um preguiçoso do mundo* . Cambridge Scholars Publishing, 2017.

Rawlings, John O., *Applied regression analysis: a research tool*. — 2nd ed. / John, 1998.