

Comparação do desempenho da estatística tipo-Wald com o de outras estatísticas de teste em modelos lineares generalizados

Miguel Gabriel Ribeiro Miguel

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientadora: Prof. Dra. Mônica Carneiro Sandoval

São Paulo, fevereiro de 2015

Comparação do desempenho da estatística tipo-Wald com o de outras estatísticas de teste em modelos lineares generalizados

Esta é a versão original da dissertação elaborada pelo
candidato Miguel Gabriel Ribeiro Miguel, tal como
submetida à Comissão Julgadora.

Agradecimentos

Agradeço a todos que de alguma maneira me ajudaram a concluir esta dissertação de mestrado. Em especial, à minha esposa Camila Pinheiro Lesqueves e minha filha Manuela Sardenberg Lesqueves Miguel pela compreensão da minha ausência em alguns momentos importantes, pela paciência e constantes mensagens de incentivo. Sou extremamente grato à Fabiana Uchôa pela valiosa ajuda nas simulações, e às minhas orientadoras Mônica Carneiro Sandoval e Denise Aparecida Botter pelos ensinamentos, atenção e paciência. Agradeço também à Mariane Branco Alves e José Francisco Moreira Pessanha pela confiança. Aos meus pais, Maria Madalena Ribeiro Miguel e Jorge Miguel, aos meus avós, aos meus irmãos, Ana Carolina Ribeiro Miguel, Daniel Ribeiro Miguel e Gabriela Ribeiro Miguel. Aos meus tios, em especial, Aparecida Ribeiro, por ter me ensinado pelo exemplo o valor dos estudos, Cristina Maria Miguel Santos e Leonardo da Costa Santos, por terem me acolhido como pais durante a graduação. Fico grato a toda minha família pela energia positiva e carinho transmitidos mesmo que à distância. Aos amigos: Wladimir Miola, Thomaz Alexandre Coelho do Amaral, Francisco Barrera, Roland Hassler, Wagner Tovani, Giuliano Amaral, Priscila Travassos, Augusto Vieira, André Carvalho, Alessandra Rodrigues, André Martinez, Frederico Reich, Paloma Uribe, Cristiane Rodrigues, Márcia Pinto, Mariana Ferreira, Vitor Ritter, Renato Gava e Ana Paula Zerbeto.

Resumo

MIGUEL, M. G. R. Comparação do desempenho da estatística tipo-Wald com o de outras estatísticas de teste em modelos lineares generalizados. 2015. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

Este trabalho foi desenvolvido com a pretensão de apurar o desempenho de algumas estatísticas de teste de hipóteses na classe dos modelos lineares generalizados. Nesse contexto, dentre os testes existentes, há a percepção de que o teste de Wald (Wald, 1943) seja o mais utilizado pelos pesquisadores, talvez pela sua simplicidade ou mesmo por ser o que, em geral, aparece como padrão nos pacotes estatísticos. Sendo assim, a estatística tipo-Wald proposta por Cordeiro et al. (2012) baseada na estatística de Wald, teve um enfoque especial na análise e foi a referência para a comparação de desempenho com as estatísticas dos demais testes, a saber: razão de verossimilhanças (Wilks, 1938), escore (Rao, 1948) e gradiente (Terrell, 2002) em suas formas originais, bem como nas formas modificadas já propostas por outros autores. O desempenho das estatísticas foi medido pela taxa de rejeição da hipótese nula quando ela é verdadeira, sendo esta taxa calculada por simulação de Monte Carlo utilizando o software OxMetrics7. As simulações foram feitas considerando modelos de regressão com resposta gama e normal inversa. Para os dois modelos, também foi avaliado o tamanho amostral em que o valor do percentil 95 da estatística tipo-Wald se aproxima do quantil da χ^2 . Por fim, foi realizada uma aplicação com dados reais com o objetivo de comparar as estatísticas de teste estudadas neste trabalho.

Palavras-chave: correção tipo-Bartlett, estatística de Wald, estatística tipo-Wald, modelo linear generalizado.

Abstract

MIGUEL, M. G. R. Performance comparison of the Wald-type statistic with other test statistics in generalized linear models. 2015. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

This work was developed to investigate the performance of some hypothesis testing statistics in the class of generalized linear models. In this context, among existing tests, there is the perception that the Wald test (Wald, 1943) is the most widely used by researchers, perhaps because it is simple to apply and generally appears as standard in statistical packages. Thus, the Wald-type statistic proposed by Cordeiro et al. (2012) based on Wald statistics, had a special focus on the analysis and was the benchmark for performance comparison with the statistics of other tests, namely: likelihood ratio (Wilks, 1938), score (Rao, 1948) and gradient (Terrell, 2002) in their original or modified forms, proposed by other authors. The statistics performance was measured by the rejection rate of the null hypothesis when it is true. This ratio was calculated by Monte Carlo simulations using the software OxMetrics7. The simulations were performed considering gamma and inverse normal regression. Under both models, it was evaluated the sample size for which the 95 percentile value of the Wald-type statistics approaches the χ^2 quantile. Finally, it was conducted an application to real data to compare the testing statistics studied in this work.

Keywords: Bartlett-type correction, generalized linear model, Wald statistic, Wald-type statistic.

Sumário

| | |
|--|-----------|
| 1 Introdução | 1 |
| 2 Aprimoramento das estatísticas ξ_{RV}, ξ_{SR} e ξ_{ST} | 7 |
| 2.1 Aprimoramento da estatística da razão de verossimilhanças ξ_{RV} | 7 |
| 2.2 Aprimoramento da estatística escore ξ_{SR} | 8 |
| 2.3 Aprimoramento da estatística gradiente ξ_{ST} | 9 |
| 3 Aprimoramento da estatística ξ_W | 11 |
| 3.1 Construção da estatística tipo-Wald ξ_{WC} | 12 |
| 4 Simulações e resultados | 15 |
| 4.1 Simulação: modelo de regressão com resposta gama | 15 |
| 4.2 Simulação: modelo de regressão com resposta normal inversa | 25 |
| 5 Aplicação prática | 37 |
| 6 Conclusões | 41 |
| 7 Referências bibliográficas | 43 |

Capítulo 1

Introdução

Os modelos de regressão são ferramentas de grande importância que auxiliam no processo de tomada de decisão em diversas áreas do conhecimento, já que permitem explicar ou fazer previsões do valor de uma variável aleatória em função de outras. Chamamos a variável a ser explicada de variável resposta, de modo que, dependendo do tipo de sua distribuição, escolhe-se uma dentre as diversas técnicas de regressão capazes de explicá-la, tais como: linear, não linear, simples, múltipla, paramétrica, não paramétrica, etc.

Neste trabalho, estamos interessados em modelos paramétricos e lineares, sendo um dos mais simples o que se utiliza quando a distribuição da variável resposta é Normal, podendo ser escrito supondo que $Y_1, \dots, Y_n \sim Normal(\mu_i, \sigma^2)$, independentes, e assumindo que o valor esperado de Y_i , com $i = 1, \dots, n$, pode ser modelado conforme a equação $E(Y_i) = \mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$, sendo p o número de variáveis explicativas. Ocorre que, nem sempre a variável resposta seguirá uma distribuição Normal, e nesse caso, o ajuste de um modelo linear estará sujeito à utilização de uma técnica apropriada, que tem seu procedimento e suas restrições específicas.

Supondo que há o interesse em explicar uma variável Y cuja densidade pode ser escrita na forma

$$f(y_i; \theta_i, \phi) = \exp \{ \phi [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \}. \quad (1.1)$$

em que $b(\cdot)$ e $c(\cdot, \cdot)$ são funções conhecidas, dizemos que Y pertence à família exponencial (Paula, 2012). Desta maneira, $E(Y_i) = \mu_i = b'(\theta_i)$ e $Var(Y_i) = \phi^{-1} V(\mu_i)$, sendo $V(\mu_i) = V_i = d\mu_i/d\theta_i$ a função de variância, $\phi^{-1} > 0$ o parâmetro de dispersão conhecido e $b'(\theta_i) = db(\theta_i)/d\theta_i$. Se Y pertence à família exponencial, então ela pode ser modelada na classe dos Modelos Lineares Generalizados (MLGs), proposta por Nelder e Wedderburn (1972). Essa classe de modelos define uma função de ligação entre a variável resposta e as variáveis explicativas ponderadas pelos seus respectivos coeficientes. Ainda de acordo com Paula (2012), esta função de ligação é expressa por $g(\mu_i) = \eta_i$, em que $\eta_i = x_i^T \beta$ é o preditor linear, $x_i^T = (x_{1i}, \dots, x_{pi})$ é o vetor que traz o valor de cada variável candidata à explicativa para a unidade experimental i e $\beta = (\beta_1, \dots, \beta_p)^T$ é o vetor de coeficientes das p variáveis candidatas à explicativas, de modo que $p < n$. Esses

coeficientes são parâmetros que podem ser interpretados como pesos de determinada variável candidata à explicativa na função de ligação.

O logaritmo da função de verossimilhança para β , denotado por $L(\beta)$, é dado por

$$L(\beta) = \sum_{i=1}^n \{\phi[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\} . \quad (1.2)$$

A função escore é obtida derivando (1.2) em relação à β , que em notação matricial é dada por

$$U_\beta = \phi X^T W^{\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu),$$

em que, $X = \partial\eta/\partial\beta = (x_1, \dots, x_n)^T$ é uma matriz conhecida $n \times p$ de posto completo p , $W = \text{diag}\{w_1, \dots, w_n\}$, com $w_i = (d\mu_i/d\eta_i)^2/V_i$, $V = \text{diag}\{V_1, \dots, V_n\}$, $y = (y_1, \dots, y_n)^T$ e $\mu = (\mu_1, \dots, \mu_n)^T$.

A matriz de informação de Fisher é dada, em notação matricial, por

$$K_{\beta\beta} = -E \left[\frac{\partial^2 L(\beta)}{\partial\beta \partial\beta^T} \right] = \phi X^T W X.$$

Segundo Paula (2012), o processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de β é definido expandindo a função escore U_β em torno de um valor inicial $\beta^{(0)}$, tal que

$$U_\beta \cong U_\beta^{(0)} + U_\beta'^{(0)} (\beta - \beta^{(0)}),$$

em que U_β' denota a primeira derivada de U_β com respeito a β^T , sendo $U_\beta'^{(0)}$ e $U_\beta^{(0)}$, respectivamente, essas quantidades avaliadas em $\beta^{(0)}$. Desta forma, chegamos ao processo iterativo

$$\beta^{(m+1)} = \beta^{(m)} + \{(-U_\beta')^{-1}\}^{(m)} U_\beta^{(m)}, \quad m=0,1,\dots$$

Como a matriz $-U_\beta'$ pode não ser positiva definida, a aplicação do método Scoring de Fisher, substituindo a matriz $-U_\beta'$ pelo correspondente valor esperado $K_{\beta\beta}$ pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\beta^{(m+1)} = \beta^{(m)} + \{K_{\beta\beta}^{-1}\}^{(m)} U_\beta^{(m)}, \quad m=0,1,\dots$$

Ainda segundo Paula (2012), quando n cresce, é possível assumir que o estimador de máxima verossimilhança $\hat{\beta}$ converge para β e tem distribuição normal multivariada, com média β e matriz de covariâncias $K_{\beta\beta}^{-1}$.

Em seu livro “O Sinal e o Ruído”, Nate Silver (2013) lança o seguinte questionamento: “Mesmo com todo esforço e concentração, será que somos capazes de distinguir entre informações significativas e aquelas que não passam de ruído?”. Nesse contexto, para saber se uma variável candidata à explicativa deve ou não ser aceita como tal, podem ser aplicados testes paramétricos verificando a

hipótese de nulidade de cada coeficiente, isto é, H_0 (hipótese nula): $\beta_i = 0$ contra H_1 (hipótese alternativa): $\beta_i \neq 0$, sendo $i = 1, \dots, p$ e ϕ suposto conhecido. Segundo Ferrari (1991), “a grande maioria dos testes de hipóteses estatísticas em problemas envolvendo grandes amostras é baseada em um dos seguintes princípios: razão de verossimilhanças, Wald e escore”.

Neste trabalho, o interesse é testar a seguinte hipótese nula composta $H_0: \beta_1 = \beta_1^{(0)}$, contra a alternativa $H_1: \beta_1 \neq \beta_1^{(0)}$, em que $\beta = (\beta_1^T, \beta_2^T)^T$, $\beta_1 = (\beta_1, \dots, \beta_{p_1})^T$, $\beta_2 = (\beta_{p_1+1}, \dots, \beta_p)^T$ e $\beta_1^{(0)} = (\beta_1^{(0)}, \dots, \beta_{p_1}^{(0)})^T$. A dimensão do vetor β_1 de interesse ou o número de parâmetros a serem testados é p_1 , e a dimensão do vetor β_2 de perturbação é $p - p_1$.

Para apresentarmos as formas originais e aperfeiçoadas das estatísticas dos testes razão de verossimilhanças, Wald, escore e, também, da estatística gradiente, definida por Terrell (2002), é necessário efetuar a partição do vetor escore U_β , da matriz de informação de Fisher $K_{\beta\beta}$ e de sua inversa $K_{\beta\beta}^{-1}$ da mesma maneira que fizemos com β . Sendo assim, de acordo com Cordeiro (1992), as partições de U_β , $K_{\beta\beta}$, $K_{\beta\beta}^{-1}$ são dadas por:

$$\begin{aligned} U_\beta^T &= (U_{\beta_1}^T, U_{\beta_2}^T), \\ K_{\beta\beta} &= \begin{bmatrix} K_{\beta_1\beta_1} & K_{\beta_1\beta_2} \\ K_{\beta_2\beta_1} & K_{\beta_2\beta_2} \end{bmatrix}, \\ K_{\beta\beta}^{-1} &= \begin{bmatrix} K^{\beta_1\beta_1} & K^{\beta_1\beta_2} \\ K^{\beta_2\beta_1} & K^{\beta_2\beta_2} \end{bmatrix}, \end{aligned}$$

em que $U_{\beta_1}^T = U_{\beta_1}^T(\beta_1, \beta_2)$, $U_{\beta_2}^T = U_{\beta_2}^T(\beta_1, \beta_2)$, $K^{\beta_1\beta_1} = K^{\beta_1\beta_1}(\beta_1, \beta_2)$ é a matriz de covariâncias assintótica de $\hat{\beta}_1$ (estimador de máxima verossimilhança de β_1), de modo que $K^{\beta_1\beta_1} = (K_{\beta_1\beta_1} - K_{\beta_1\beta_2}K_{\beta_2\beta_2}^{-1}K_{\beta_2\beta_1}^T)^{-1}$. Quando β_1 e β_2 são ortogonais implica $K^{\beta_1\beta_1} = K_{\beta_1\beta_1}^{-1}$.

Considerando as partições apresentadas acima, as estatísticas dos testes da hipótese nula de interesse H_0 têm as seguintes formas originais:

$$(i) \quad \text{Teste da razão de verossimilhanças: } \xi_{RV} = 2 \left\{ L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1^{(0)}, \tilde{\beta}_2) \right\},$$

em que $L(\hat{\beta}_1, \hat{\beta}_2)$ é o logaritmo da função de verossimilhança considerando as estimativas de máxima verossimilhança dos parâmetros do modelo irrestrito (com p parâmetros) e $L(\beta_1^{(0)}, \tilde{\beta}_2)$ é o logaritmo da função de verossimilhança considerando as estimativas de máxima verossimilhança dos parâmetros sob a hipótese nula, ou seja, do vetor β_2 , denotadas por $\tilde{\beta}_2$.

- (ii) Teste de Wald: $\xi_W = [\hat{\beta}_1 - \beta_1^{(0)}]^T [\hat{K}_{\beta_1\beta_1}]^{-1} [\hat{\beta}_1 - \beta_1^{(0)}]$, em que $\hat{K}_{\beta_1\beta_1}$ é a matriz $K_{\beta_1\beta_1}$ avaliada em $\hat{\beta}_1$ e $\hat{\beta}_2$.
- (iii) Teste escore: $\xi_{SR} = \tilde{U}_{\beta_1}^T \tilde{K}^{\beta_1\beta_1} \tilde{U}_{\beta_1}$, em que \tilde{U}_{β_1} e $\tilde{K}^{\beta_1\beta_1}$ são, respectivamente, U_{β_1} e $K^{\beta_1\beta_1}$ avaliadas em $\beta_1^{(0)}$ e $\tilde{\beta}_2$.
- (iv) Teste gradiente: $\xi_{ST} = \phi^{1/2} \tilde{s}^T \tilde{W}^{1/2} X_1 [\hat{\beta}_1 - \beta_1^{(0)}]$, em que $s = \phi^{1/2} V^{-1/2} (y - \mu)$ é um vetor de resíduos de Pearson e \tilde{s} e $\tilde{W}^{1/2} = \text{diag}(\tilde{w}_1^2, \dots, \tilde{w}_n^2)$ são, respectivamente, s e $W^{1/2}$ avaliados em $\beta_1^{(0)}$ e $\tilde{\beta}_2$.

Uma interessante propriedade das estatísticas ξ_{RV} e ξ_{SR} é a invariância, isto é, segundo Paula (2012), mesmo sob diferentes reparametrizações, as estatísticas mencionadas não têm seus valores alterados. Já a estatística gradiente, assim como a de Wald, não são invariantes sob reparametrizações (Terrell, 2002).

Segundo Ferrari (1991), ξ_{RV} , ξ_W e ξ_{SR} possuem, sob a hipótese nula, distribuição assintótica $\chi_{p_1}^2$, isto é qui-quadrado com p_1 graus de liberdade, em que p_1 é o número de parâmetros a serem testados. De acordo com Vargas (2013), sob a hipótese nula, a estatística gradiente também tem distribuição assintótica $\chi_{p_1}^2$.

Ocorre que, em muitos casos, não se pode contar com um tamanho de amostra grande e, portanto, a distribuição dessas estatísticas pode estar afastada da distribuição $\chi_{p_1}^2$. Com isso, alguns autores propuseram modificações capazes de suportar o cenário de pequenas amostras. Por exemplo, com o objetivo de aperfeiçoar a estatística da razão de verossimilhanças, Cordeiro et al. (1982) considerou a técnica de Bartlett (1937) para escolher um fator c de correção de escala. Já, para a estatística escore, Cordeiro e Ferrari (1991) aplicaram um fator de correção tipo-Bartlett, assim como fez Vargas (2013) no aprimoramento da estatística gradiente. Santos e Cordeiro (1999) também obtiveram um fator de correção tipo-Bartlett para aperfeiçoar a estatística de Wald em modelos da família exponencial com apenas um parâmetro. Vale ressaltar que, em modelos contínuos, as estatísticas aperfeiçoadas por fatores de Bartlett ou tipo-Bartlett têm distribuição qui-quadrado até ordem n^{-1} . Esse resultado não é necessariamente válido em modelos discretos.

Cordeiro et al. (2012) propuseram, para os MLGs com parâmetro de dispersão conhecido, uma estatística tipo-Wald, baseada no estimador de máxima verossimilhança corrigido pelo viés de primeira ordem e na matriz de covariâncias de segunda ordem desse estimador, que converge assintoticamente para a distribuição $\chi_{p_1}^2$, sob a hipótese nula. Essa estatística foi comparada com a estatística de Wald original considerando um modelo exponencial em que todos os parâmetros eram testados simultaneamente na hipótese nula.

Já existem na literatura (Vargas, 2013) comparações entre as estatísticas da razão de verossimilhanças, escore, gradiente e Wald e as versões modificadas das três primeiras estatísticas para os MLGs. Esses estudos mostram que, em geral, a estatística escore aperfeiçoada tem o melhor desempenho comparada às demais estatísticas, no sentido de que o tamanho empírico do teste baseado nessa estatística se aproxima mais dos níveis de significância usuais. Por outro lado, a estatística de Wald é a que apresenta o pior desempenho.

Apesar de ter um desempenho ruim em amostras pequenas e moderadas, em geral, a estatística de Wald é muito popular, no sentido de ser a estatística presente nas saídas dos principais pacotes computacionais, tais como R, Minitab ou SAS. Além disso, intervalos de confiança para os parâmetros de um modelo de regressão baseiam-se na inversa da matriz de informação de Fisher. Desta forma, este trabalho tem como objetivo comparar o desempenho da estatística tipo-Wald com o das demais estatísticas, aperfeiçoadas ou não.

No Capítulo 2 serão apresentadas as modificações nas estatísticas dos testes de razão de verossimilhanças, escore e gradiente. O Capítulo 3 apresentará a modificação no teste de Wald utilizando o fator tipo-Bartlett, aplicável somente aos casos onde se tem um parâmetro, e a construção da estatística tipo-Wald em MLGs. Para comparar os desempenhos das estatísticas apresentadas nos Capítulos 2 e 3, no Capítulo 4, serão realizados estudos de simulação, considerando dois diferentes modelos lineares generalizados e diferentes hipóteses nulas. No Capítulo 5 será feita uma aplicação com dados reais e o Capítulo 6 conterá nossas conclusões.

Capítulo 2

Aprimoramento das estatísticas ξ_{RV} , ξ_{SR} e ξ_{ST}

2.1 Aprimoramento da estatística da razão de verossimilhanças (ξ_{RV})

A ideia de melhorar a aproximação da distribuição da estatística da razão de verossimilhanças, ξ_{RV} , pela distribuição qui-quadrado, por meio da multiplicação de ξ_{RV} por uma constante c^{-1} , teve origem em Bartlett (1937), sendo que um método geral para o cálculo do fator de correção c foi proposto por Lawley (1956), que expôs uma fórmula geral para o primeiro momento da estatística da razão de verossimilhanças. Esta fórmula envolvia expressões de momentos do logaritmo da função de verossimilhança, bem como de suas derivadas. Outro trabalho fundamental foi desenvolvido por Hayakawa (1977), que apresenta uma expansão até a ordem n^{-1} para a função de distribuição acumulada da estatística da razão de verossimilhanças.

Cordeiro (1983, 1987) utilizou o método de Lawley em generalidade quando obteve fórmulas matriciais gerais para o fator de correção de Bartlett associado à estatística da razão de verossimilhanças na classe dos modelos lineares generalizados uni e multiparamétricos considerando hipóteses nulas simples e compostas. Deste modo, a estatística da razão de verossimilhanças modificada ξ_{RV}^* é dada por

$$\xi_{RV}^* = c^{-1}\xi_{RV},$$

sendo

$$c = \frac{E(\xi_{RV})}{p_1},$$

com $E(\xi_{RV})$, calculada até a ordem n^{-1} , assumindo a expressão

$$E(\xi_{RV}) = p_1 + \hat{\epsilon} - \tilde{\epsilon},$$

em que $\hat{\epsilon}$ e $\tilde{\epsilon}$ são as estimativas de ϵ sob o modelo irrestrito e o modelo sob H_0 , respectivamente. Na classe dos modelos lineares generalizados, supondo ϕ conhecido, com base em Cordeiro et al. (1982), a quantidade ϵ é definida por:

$$\epsilon = \frac{1}{4}Tr(\phi^{-1}HZ_d^2) - \frac{1}{3}1^T\phi^{-1}GZ^{(3)}(F+G)1 + \frac{1}{12}1^T\phi^{-1}F(2Z^{(3)} + 3Z_dZZ_d)F1,$$

sendo ϕ constante e $Z = X(X^TWX)^{-1}X^T$, uma matriz com elementos z_{ij} . Já, a matriz $Z^{(3)}$ tem elementos z_{ij}^3 , em que $Z^{(3)} = Z^{(2)} \otimes Z$, sendo $Z^{(2)} = Z \otimes Z$. Vale

ressaltar que a matriz $Z^{(2)}$ tem elementos z_{ij}^2 e o símbolo \otimes denota o produto de Hadamard (produto elemento a elemento) entre duas matrizes. As matrizes H, F, G e Z_d são matrizes diagonais de ordem n obtidas da seguinte forma:

$$H = \text{diag}\left\{\frac{1}{V} \frac{d^2\mu}{d\eta^2} \left[\frac{d^2\mu}{d\eta^2} - 4w \frac{dV}{d\mu}\right] + w^2 \left[\frac{2}{V} \left(\frac{dV}{d\mu}\right)^2 - \frac{d^2V}{d\mu^2}\right]\right\},$$

$$F = \text{diag}\left\{\frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2}\right\},$$

$$G = \text{diag}\left\{\frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta}\right)^3\right\},$$

$$Z_d = \text{diag}\{z_{11}, \dots, z_{nn}\}.$$

Já, o vetor $\mathbf{1} = (1, \dots, 1)^T$ é um vetor de dimensão $nx1$.

Segundo os resultados simulados por Cordeiro et al. (1983), mesmo com um tamanho amostral relativamente pequeno, a distribuição da estatística ξ_{RV}^* tem uma boa aproximação para a distribuição $\chi_{p_1}^2$, de modo que, à medida que se aumenta o tamanho amostral, o valor da estatística de teste modificada se aproxima mais rapidamente do valor crítico real da distribuição $\chi_{p_1}^2$ em relação à estatística não modificada. Mais detalhes sobre a modificação podem ser vistos em Cordeiro et al. (1983).

2.2 Aprimoramento da estatística escore (ξ_{SR})

A alteração na estatística escore considerada neste trabalho foi introduzida por Cordeiro e Ferrari (1991), e consiste em aplicar um fator de correção tipo-Bartlett para melhorar a aproximação da estatística de teste para a distribuição χ^2 . Este fator é chamado desta forma porque, embora se assemelhe, não é exatamente o fator de correção de Bartlett, já que contém a própria estatística em sua expressão. No Capítulo 1, foi apresentada a forma original da estatística escore, ou seja, $\xi_{SR} = \tilde{U}_{\beta_1\beta_1}^T \tilde{K}^{\beta_1\beta_1} \tilde{U}_{\beta_1\beta_1}$. A forma aprimorada tem a seguinte expressão: $\xi_{SR}^* = \xi_{SR} [1 - (c + b\xi_{SR} + a\xi_{SR}^2)]$, em que os coeficientes a , b e c são apresentados a seguir:

$$a = \frac{A_3}{12p_1(p_1 + 2)(p_1 + 4)},$$

$$b = \frac{A_2 - 2A_3}{12p_1(p_1 + 2)},$$

$$c = \frac{A_1 - A_2 + A_3}{12p_1},$$

sendo p_1 o número de parâmetros testados. Segundo Cordeiro e Ferrari (1991), considerando uma hipótese nula composta $H_0: \beta_1 = \beta_1^{(0)}$ contra $H_1: \beta_1 \neq \beta_1^{(0)}$, as expressões de A_1 , A_2 e A_3 são dadas por

$$A_1 = \phi^{-1} [3 \mathbf{1}^T F Z_{2d} (Z - Z_2) Z_{2d} F \mathbf{1} + 6 \mathbf{1}^T F Z_{2d} Z_2 (Z - Z_2)_d (F - G) \mathbf{1} - 6 \mathbf{1}^T F \{Z_2^{(2)} \otimes (Z - Z_2)\} (2G - F) \mathbf{1} - 6 \mathbf{1}^T H (Z - Z_2)_d Z_{2d} \mathbf{1}],$$

$$A_2 = \phi^{-1} [-3 \mathbf{1}^T (F - G) (Z - Z_2)_d Z_2 (Z - Z_2)_d (F - G) \mathbf{1} - 6 \mathbf{1}^T F Z_{2d} (Z - Z_2) (Z - Z_2)_d (F - G) \mathbf{1} - 6 \mathbf{1}^T (F - G) \{ (Z - Z_2)^{(2)} \otimes Z_2 \} (F - G) \mathbf{1} + 3 \mathbf{1}^T B (Z - Z_2)_d^{(2)} \mathbf{1}],$$

$$A_3 = \phi^{-1} \{ 3 \mathbf{1}^T (F - G) (Z - Z_2)_d (Z - Z_2) (Z - Z_2)_d (F - G) \mathbf{1} + 2 \mathbf{1}^T (F - G) (Z - Z_2)^{(3)} (F - G) \mathbf{1} \},$$

em que as matrizes F , G e H estão definidas na seção anterior e $B = \text{diag}(b_1, \dots, b_n)$ é uma matriz diagonal $n \times n$, com elementos dados por:

$$b_i = \frac{1}{V_i^3} \left(\frac{d\mu_i}{d\eta_i} \right)^4 \left\{ \left(\frac{dV_i}{d\mu_i} \right)^2 + V_i \frac{d^2 V_i}{d\mu_i^2} \right\}.$$

Considera-se que a matriz Z está avaliada em $(\hat{\beta}_1, \hat{\beta}_2)$ e a matriz Z_2 está avaliada em $(\hat{\beta}_1^{(0)}, \tilde{\beta}_2)$. A matriz Z_{2d} representa a diagonal principal de Z_2 , ou seja, $Z_{2d} = \text{diag}\{z_{211}, \dots, z_{2nn}\}$. Maiores detalhes sobre este trabalho de aprimoramento da estatística de escore pode ser apreciado em Cordeiro e Ferrari (1991).

2.3 Aprimoramento da estatística gradiente (ξ_{ST})

A estatística gradiente tem como vantagem sobre as estatísticas de Wald e escore o fato de não depender da matriz de informação. Outra vantagem é a sua simplicidade computacional, propriedade ressaltada por Rao (2005).

A modificação na estatística gradiente foi proposta por Vargas (2013), que utilizou um fator de correção tipo-Bartlett. Sob a hipótese nula, a estatística corrigida tem distribuição $\chi_{p_1}^2$ até ordem n^{-1} . Assumindo a forma original da estatística gradiente apresentada no Capítulo 1, ou seja, $\xi_{ST} = \phi^{1/2} \tilde{s}^T \tilde{W}^{1/2} X [\hat{\beta}_1 - \beta_1^{(0)}]$, a expressão sugerida para a correção de ξ_{ST} em função de coeficientes a , b e c (idênticos aos apresentados na correção da estatística de escore) foi a seguinte:

$$\xi_{ST}^* = \xi_{ST} [1 - (c + b\xi_{ST} + a\xi_{ST}^2)].$$

Maiores detalhes sobre a modificação na estatística gradiente são encontrados em Vargas (2013).

Capítulo 3

Aprimoramento da estatística ξ_W

Já foram mencionados os aperfeiçoamentos propostos para as estatísticas da razão de verossimilhanças, escore e gradiente, que utilizaram correções de Bartlett e tipo-Bartlett. Fazer a mesma abordagem para a estatística de Wald não é tão trivial, no entanto, Santos e Cordeiro (1999) fizeram um trabalho para modelos da família exponencial com apenas um parâmetro. Este estudo utiliza um resultado obtido por Cordeiro e Ferrari (1991), que abordou o problema do desenvolvimento de correções para qualquer estatística de teste assintoticamente distribuída segundo uma qui-quadrado. Este resultado consiste em uma fórmula geral para ajuste tipo-Bartlett de ordem n^{-1} , para qualquer estatística de teste cuja expansão assintótica seja uma combinação linear finita de distribuições qui-quadrado com graus de liberdades adequados. Sendo assim, essa fórmula geral pode ser utilizada para aperfeiçoar a estatística de Wald, sendo que o resultado fica da seguinte maneira:

$$\xi_W^* = \xi_W \left\{ 1 - \frac{1}{n} (\alpha_1 + \alpha_2 \xi_W + \alpha_3 \xi_W^2) \right\} \sim \chi_1^2,$$

em que

$$\alpha_1 = \frac{-44k_{\beta\beta\beta}^2 + 120k_{\beta\beta\beta}k_{\beta\beta}^{(\beta)} - 81(k_{\beta\beta}^{(\beta)})^2 + 12k_{\beta\beta}k_{\beta,\beta,\beta\beta} - 3k_{\beta\beta}k_{\beta,\beta,\beta,\beta}}{13k_{\beta\beta}^3}$$

$$\alpha_2 = \frac{63k_{\beta\beta\beta}k_{\beta\beta}^{(\beta)} - 22k_{\beta\beta\beta}^2 - 45(k_{\beta\beta}^{(\beta)})^2}{18k_{\beta\beta}^3} + \frac{4k_{\beta\beta\beta\beta} - 4k_{\beta\beta\beta}^{(\beta)} - 4k_{\beta,\beta,\beta\beta} - 3k_{\beta,\beta,\beta,\beta}}{12k_{\beta\beta}^2}$$

$$\alpha_3 = -\frac{(3k_{\beta\beta}^{(\beta)} - k_{\beta\beta\beta})^2}{36k_{\beta\beta}^3}.$$

Para explicar melhor a obtenção dos valores dos k'_s , podemos considerar as observações independentes e identicamente distribuídas y_1, \dots, y_n e lembrar que $L(\beta)$ representa o logaritmo da função de verossimilhança de y em função de β . Sendo assim, podemos definir $U_\beta = dL(\beta)/d\beta$, $U_{\beta\beta} = d^2L(\beta)/d\beta^2$, etc. Considerando isto, temos $k_{\beta\beta} = E(U_{\beta\beta})$, $k_{\beta\beta\beta} = E(U_{\beta\beta\beta})$, $k_{\beta,\beta} = E(U_\beta^2) = -k_{\beta\beta}$, $k_{\beta,\beta\beta} = E(U_\beta U_{\beta\beta})$, $k_{\beta\beta,\beta\beta} = E(U_{\beta\beta}^2) - k_{\beta\beta}^2$, $k_{\beta\beta\beta\beta} = E(U_{\beta\beta\beta\beta})$, $k_{\beta,\beta,\beta\beta} = E(U_\beta^2 U_{\beta\beta}) - k_{\beta,\beta} k_{\beta\beta}$, $k_{\beta,\beta,\beta,\beta} = E(U_\beta^4) - 3k_{\beta,\beta}^2$ e $k_{\beta,\beta\beta\beta} = E(U_\beta U_{\beta\beta\beta})$.

A restrição é que a fórmula acima vale apenas para modelos uniparamétricos. Em Santos e Cordeiro (1999) é possível obter mais detalhes sobre essa abordagem.

Uma outra abordagem foi a construção da estatística tipo-Wald, proposta por Cordeiro et al. (2012), que, em linhas gerais substitui a matriz de informação de Fisher pela matriz de covariâncias de segunda ordem dos estimadores de máxima verossimilhança com correção de viés de ordem n^{-1} .

3.1 Construção da estatística tipo-Wald ξ_{Wc}

Em Cordeiro et al. (2012), a fórmula geral para a matriz de covariâncias de ordem n^{-2} do estimador de máxima verossimilhança corrigido pelo viés, na classe dos MLGs com ϕ conhecido, é obtida como segue.

Seja $\beta^* = \hat{\beta} - d(\hat{\beta})$ o estimador de máxima verossimilhança com correção de viés de ordem n^{-1} , sendo $d(\hat{\beta})$ o viés de ordem n^{-1} , dado, em notação matricial, por $D(\hat{\beta})$, em que, segundo Cordeiro e McCullagh (1991):

$$D(\hat{\beta}) = -(2\phi)^{-1}(X^T W X)^{-1} X^T Z_d F 1.$$

A expressão da matriz de covariâncias de segunda ordem de β^* é dada por:

$$\text{Cov}(\beta^*) = \phi^{-1}(X^T W X)^{-1} + \phi^{-2} P \Lambda P^T + 3\phi^{-2}(X^T W X)^{-1} \Delta (X^T W X)^{-1} - \phi^{-2} P D Z_d P^T$$

sendo

$$\Lambda = H Z_d + \frac{3}{2} F Z^{(2)} F + G Z^{(2)} F - G Z^{(2)} G, P = (X^T W X)^{-1} X^T, Z = X P,$$

$$H = \text{diag}\{h_1, \dots, h_n\}, \text{ em que } h_i = -\mu_i' \mu_i''' / V_i - \mu_i'^2 \mu_i'' V_i^{(1)} / V_i^2 + \mu_i'^4 V_i^{(1)2} / V_i^3, \\ \mu_i' = d\mu_i / d\eta_i, \mu_i'' = d^2 \mu_i / d\eta_i^2, \mu_i''' = d^3 \mu_i / d\eta_i^3, V_i^{(1)} = dV_i / d\mu_i,$$

$$F = \text{diag}\{f_1, \dots, f_n\}, \text{ com } f_i = V_i^{-1} \mu_i' \mu_i'',$$

$$G = \text{diag}\{g_1, \dots, g_n\}, \text{ com } g_i = V_i^{-1} \mu_i' \mu_i'' - V_i^{-2} V_i^{(1)} \mu_i'^3,$$

$Z^{(2)} = Z \otimes Z$, em que a matriz $Z^{(2)}$ tem elementos z_{ij}^2 e o símbolo \otimes denota o produto de Hadamard entre duas matrizes,

$$\Delta = \sum_{i=1}^n c_i \Delta_i, \text{ em que } \Delta_i = (f_i + g_i) x_i x_i^T, c_i = \delta_i^T Z_\beta Z_{\beta d} F 1,$$

$x_i^T = (x_{i1}, \dots, x_{ip})$ representa a i -ésima linha da matriz de covariâncias X , $Z_\beta = X(X^T W X)^{-1} X^T$, δ_i é um vetor de dimensão $(nx1)$ e o índice d indica que a matriz diagonal foi obtida a partir da matriz original. Todas essas quantidades foram definidas anteriormente, nos Capítulos 1 e 2.

A estatística tipo-Wald, proposta por Cordeiro et al. (2012), substitui na estatística de Wald original, $\hat{\beta}$ por β^* e a inversa da matriz de informação de Fisher pela matriz de covariâncias para β^* apresentada acima.

Considerando, como no Capítulo 1, as hipóteses nula e alternativa, $H_0: \beta_1 = \beta_1^{(0)}$ e $H_1: \beta_1 \neq \beta_1^{(0)}$, respectivamente, em que o vetor $\beta_1^{(0)}$ tem dimensão p_1 , temos:

$$\xi_{Wc} = \left[\beta_1^* - \beta_1^{(0)} \right]^T \{Cov_{\beta_1^* \beta_1^*}\}^{-1} \left[\beta_1^* - \beta_1^{(0)} \right]$$

em que a matriz de covariâncias $Cov(\beta^*)$ é particionada da forma

$$Cov(\beta^*) = \begin{bmatrix} Cov_{\beta_1^* \beta_1^*} & Cov_{\beta_1^* \beta_2^*} \\ Cov_{\beta_2^* \beta_1^*} & Cov_{\beta_2^* \beta_2^*} \end{bmatrix}.$$

Capítulo 4

Simulações e resultados

As simulações foram feitas pelo processo de Monte Carlo utilizando o software OxMetrics7. O objetivo é comparar a estatística tipo-Wald ξ_{Wc} com as demais estatísticas $\xi_W, \xi_{RV}, \xi_{RV}^*, \xi_{SR}, \xi_{SR}^*, \xi_{ST}$ e ξ_{ST}^* , segundo a taxa de rejeição de H_0 ou tamanho empírico do teste, isto é, o risco de rejeição da hipótese nula dado que ela é verdadeira (probabilidade estimada do erro tipo I), fixados os valores de 1%, 5% e 10% para o nível de significância (α). Além disso, para verificar o tamanho amostral que torna mais segura a utilização da estatística ξ_{Wc} , foi observada também a aproximação dos quantis amostrais de cada estatística para os quantis da distribuição qui-quadrado com graus de liberdade iguais ao número de parâmetros testados.

4.1 Simulação: modelo de regressão com resposta gama

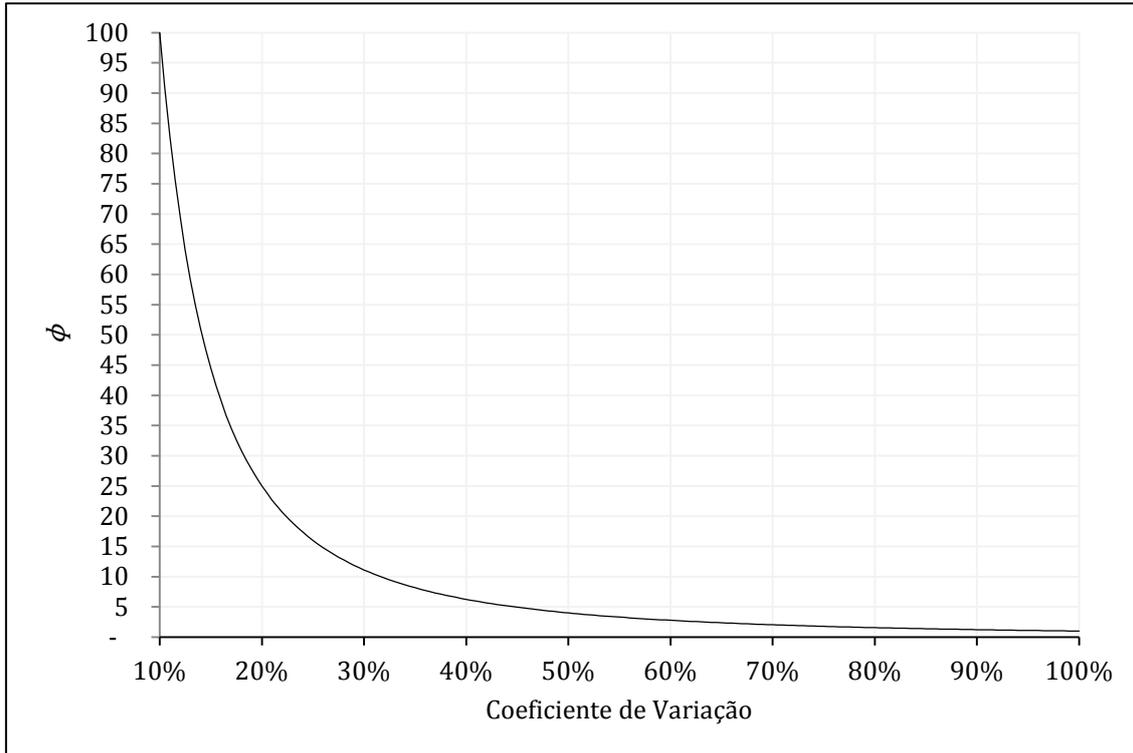
Dizemos que uma variável aleatória Y de média μ e coeficiente de variação $\phi^{-1/2}$ é distribuída segundo uma gama, que denotamos $Y \sim G(\mu, \phi)$, quando a função densidade de Y , de acordo com Paula (2012), é dada por:

$$\begin{aligned} f(y; \mu, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu}\right)^\phi \exp\left(-\frac{\phi y}{\mu}\right) d(\log y) = \\ &= \exp\left\{\phi \left[\left(-\frac{y}{\mu}\right) - \log \mu\right] - \log \Gamma(\phi) + \phi \log(\phi y) - \log y\right\}, \end{aligned}$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1} e^{-t} dt$ é a função gama. Desta maneira, para escrever a densidade da gama na forma de (1.1) basta considerar $\theta = -1/\mu$, $b(\theta) = -\log(-\theta)$ e $c(y, \phi) = (\phi - 1)\log y + \phi \log \phi - \log \Gamma(\phi)$. Vale destacar que, para a densidade da gama, $\phi = E^2(Y)/Var(Y)$, que é o inverso do coeficiente de variação de Y ao quadrado. A função de variância da gama é dada por $V(\mu) = \mu^2$.

A relação entre o coeficiente de variação e ϕ , pode ser observada no Gráfico 4.1. É possível perceber que, quanto maior for o coeficiente de variação menor será o valor de ϕ .

Gráfico 4.1: Relação entre ϕ e o coeficiente de variação do modelo gama



Para os estudos de simulação que seguem, foram arbitrados dois valores para ϕ , 11 e 2, que correspondem aos coeficientes de variação 30% e 70%, respectivamente. Considerando uma amostra y_1, \dots, y_n , com $y_i \sim G(\mu_i, \phi)$ define-se um modelo de regressão com componente sistemática

$$\mu_i = \eta_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}, i = 1, \dots, n. \quad (4.1)$$

Para cada valor fixado para ϕ , foram geradas 10.000 réplicas de n observações ($n = 20, 25, 30$), do modelo gama definido em (4.1), considerando $\beta_1 = 0,5$, $\beta_2 = \beta_3 = \beta_4 = 1$. Os valores de x_{i2} , x_{i3} e x_{i4} foram obtidos de uma distribuição uniforme no intervalo $(1, 2)$, sendo mantidos os mesmos para todas as réplicas, fixado o tamanho n da amostra. Para cada réplica, verifica-se se a hipótese nula é rejeitada. Define-se como taxa de rejeição, o percentual de vezes em que a estatística de teste simulada caiu na zona de rejeição de H_0 .

Foram consideradas as seguintes hipóteses nulas $H_0: \beta_4 = 1$, $H_0: \beta_3 = \beta_4 = 1$ e $H_0: \beta_2 = \beta_3 = \beta_4 = 1$. A taxa de rejeição de cada uma dessas hipóteses consta das Tabelas 4.1 a 4.3, respectivamente. A análise dessas tabelas mostra que, embora o desempenho da taxa de rejeição de H_0 associada à estatística de Wald usual (ξ_W) seja o pior em relação às demais estatísticas, a estatística tipo-Wald proposta por Cordeiro et al. (2012) apresenta taxa de rejeição entre 30% e 54% menor do que a associada com a estatística de Wald usual, para todo valor de ϕ e de α , o que mostra que a estatística tipo-Wald traz ganhos expressivos nos resultados em amostras pequenas. A estatística de teste com melhor desempenho foi ξ_{SR}^* , pois o tamanho empírico é o que mais se aproxima dos níveis de

significância, dentre as estatísticas consideradas, para os dois valores de ϕ e todos os valores de α . Após a estatística ξ_{SR}^* , as estatísticas com melhores desempenhos alternam-se entre ξ_{SR} , ξ_{RV}^* , ξ_{ST} , ξ_{ST}^* . O desempenho da estatística tipo-Wald é melhor do que o da estatística da razão de verossimilhanças quando se testa um único parâmetro e consideram-se os valores 5% e 10% para α .

Para as estatísticas ξ_W , ξ_{WC} , ξ_{RV} e ξ_{RV}^* , as taxas de rejeição são sempre maiores do que os valores fixados de α . Já, o número de casos em que as estatísticas ξ_{SR} , ξ_{SR}^* , ξ_{ST} , ξ_{ST}^* apresentam taxas de rejeição menores do que os valores de α , cresce à medida que aumenta o número de parâmetros testados em H_0 . Vale ressaltar que a taxa de rejeição das estatísticas de Wald e tipo-Wald se distanciam dos valores fixados para o nível de significância quando se aumenta o número de parâmetros testados em H_0 . Além disso, nota-se que o desempenho de todas as estatísticas é pior para $\phi = 2$.

Tabela 4.1: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_4 = 1$ no modelo de regressão gama com dispersão conhecida

| ϕ | n | α | ξ_W | ξ_{WC} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 11 | 20 | 10,00 | 25,21 | 16,64 | 20,04 | 13,72 | 14,82 | 10,75 | 17,46 | 13,38 |
| | | 5,00 | 17,75 | 10,88 | 12,65 | 7,76 | 7,72 | 5,35 | 9,67 | 6,82 |
| | | 1,00 | 8,99 | 4,36 | 4,06 | 1,71 | 1,36 | 1,05 | 1,84 | 1,33 |
| | 25 | 10,00 | 23,21 | 16,21 | 17,98 | 12,88 | 12,96 | 10,52 | 16,15 | 13,22 |
| | | 5,00 | 16,26 | 10,42 | 10,91 | 7,13 | 6,51 | 5,34 | 8,69 | 7,05 |
| | | 1,00 | 7,83 | 4,26 | 3,63 | 1,82 | 1,29 | 1,09 | 1,85 | 1,74 |
| | 30 | 10,00 | 20,14 | 13,99 | 15,70 | 11,67 | 12,05 | 10,15 | 14,01 | 12,30 |
| | | 5,00 | 13,48 | 8,37 | 9,21 | 6,08 | 6,07 | 5,14 | 7,53 | 6,41 |
| | | 1,00 | 5,72 | 3,36 | 2,72 | 1,50 | 1,13 | 1,05 | 1,52 | 1,39 |
| 2 | 20 | 10,00 | 30,39 | 19,04 | 22,94 | 15,25 | 13,72 | 11,42 | 17,72 | 14,64 |
| | | 5,00 | 22,37 | 12,65 | 15,25 | 8,42 | 7,11 | 5,75 | 9,36 | 7,72 |
| | | 1,00 | 12,07 | 5,55 | 5,56 | 2,03 | 1,21 | 1,05 | 1,35 | 1,18 |
| | 25 | 10,00 | 26,34 | 16,98 | 19,59 | 13,29 | 11,06 | 9,55 | 14,92 | 13,32 |
| | | 5,00 | 18,65 | 11,19 | 12,26 | 7,22 | 5,46 | 4,88 | 7,33 | 6,69 |
| | | 1,00 | 9,44 | 4,62 | 4,17 | 1,94 | 0,94 | 1,05 | 1,30 | 1,46 |
| | 30 | 10,00 | 27,23 | 18,61 | 18,86 | 13,52 | 11,41 | 10,59 | 15,49 | 14,35 |
| | | 5,00 | 19,93 | 13,07 | 12,02 | 7,21 | 6,00 | 5,41 | 7,77 | 7,05 |
| | | 1,00 | 11,21 | 6,64 | 3,93 | 1,63 | 1,07 | 1,03 | 1,28 | 1,29 |

Tabela 4.2: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_3 = \beta_4 = 1$ no modelo de regressão gama com dispersão conhecida

| ϕ | n | α | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 11 | 20 | 10,00 | 32,71 | 20,82 | 22,62 | 14,10 | 12,89 | 10,29 | 16,37 | 13,18 |
| | | 5,00 | 24,95 | 14,39 | 14,23 | 7,77 | 6,04 | 4,98 | 8,06 | 6,90 |
| | | 1,00 | 14,23 | 7,32 | 4,97 | 2,07 | 0,84 | 0,91 | 1,28 | 1,43 |
| | 25 | 10,00 | 29,94 | 19,39 | 19,57 | 13,05 | 11,89 | 10,21 | 15,18 | 13,20 |
| | | 5,00 | 21,83 | 13,11 | 12,14 | 7,09 | 5,62 | 5,17 | 7,45 | 6,92 |
| | | 1,00 | 11,67 | 6,36 | 4,10 | 1,89 | 0,85 | 0,98 | 1,31 | 1,52 |
| | 30 | 10,00 | 24,59 | 16,34 | 16,81 | 11,52 | 11,05 | 9,66 | 13,37 | 12,16 |
| | | 5,00 | 17,37 | 10,45 | 9,91 | 6,07 | 5,08 | 4,91 | 6,49 | 6,08 |
| | | 1,00 | 8,21 | 4,19 | 2,96 | 1,48 | 0,96 | 1,14 | 1,09 | 1,23 |
| 2 | 20 | 10,00 | 40,53 | 25,23 | 26,38 | 15,85 | 11,95 | 10,69 | 15,62 | 13,84 |
| | | 5,00 | 32,08 | 18,41 | 17,56 | 8,98 | 5,31 | 4,94 | 6,91 | 6,48 |
| | | 1,00 | 20,31 | 9,92 | 6,65 | 2,11 | 0,82 | 0,92 | 0,69 | 0,98 |
| | 25 | 10,00 | 34,64 | 22,03 | 21,96 | 13,70 | 10,51 | 9,90 | 13,93 | 13,26 |
| | | 5,00 | 26,78 | 15,61 | 13,96 | 7,64 | 4,76 | 4,78 | 6,60 | 6,79 |
| | | 1,00 | 15,68 | 7,83 | 4,92 | 1,84 | 0,60 | 0,91 | 0,74 | 1,04 |
| | 30 | 10,00 | 34,31 | 22,65 | 21,06 | 13,71 | 11,30 | 10,63 | 13,84 | 13,07 |
| | | 5,00 | 26,17 | 16,50 | 13,10 | 7,20 | 5,60 | 5,55 | 6,21 | 6,23 |
| | | 1,00 | 15,78 | 9,06 | 4,26 | 1,64 | 0,84 | 1,08 | 0,74 | 0,89 |

Tabela 4.3: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_2 = \beta_3 = \beta_4 = 1$ no modelo de regressão gama com dispersão conhecida

| ϕ | n | α | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 11 | 20 | 10,00 | 39,94 | 25,37 | 23,54 | 13,98 | 10,21 | 9,59 | 12,87 | 12,12 |
| | | 5,00 | 32,10 | 18,76 | 15,12 | 7,79 | 4,60 | 4,84 | 5,66 | 6,06 |
| | | 1,00 | 20,78 | 11,02 | 5,28 | 2,05 | 0,47 | 0,84 | 0,58 | 1,14 |
| | 25 | 10,00 | 35,36 | 22,73 | 20,38 | 13,18 | 10,03 | 9,89 | 12,73 | 12,59 |
| | | 5,00 | 27,35 | 16,61 | 12,81 | 7,11 | 4,46 | 4,84 | 5,74 | 6,32 |
| | | 1,00 | 16,69 | 8,89 | 4,24 | 1,72 | 0,60 | 0,97 | 0,78 | 1,23 |
| | 30 | 10,00 | 27,93 | 17,77 | 17,43 | 11,59 | 9,99 | 9,81 | 11,87 | 11,62 |
| | | 5,00 | 20,31 | 12,06 | 10,19 | 6,09 | 4,68 | 5,02 | 5,53 | 5,78 |
| | | 1,00 | 10,56 | 5,38 | 3,16 | 1,49 | 0,61 | 1,00 | 0,75 | 1,11 |
| 2 | 20 | 10,00 | 51,91 | 35,54 | 27,80 | 15,44 | 10,06 | 9,84 | 10,24 | 10,00 |
| | | 5,00 | 44,56 | 28,70 | 18,25 | 8,24 | 4,60 | 4,88 | 3,76 | 4,10 |
| | | 1,00 | 32,51 | 19,13 | 6,49 | 2,11 | 0,63 | 0,89 | 0,22 | 0,34 |
| | 25 | 10,00 | 43,49 | 28,69 | 22,21 | 13,32 | 9,83 | 9,98 | 10,27 | 10,42 |
| | | 5,00 | 35,75 | 22,79 | 13,99 | 7,50 | 3,93 | 4,44 | 4,10 | 4,65 |
| | | 1,00 | 24,42 | 14,62 | 4,98 | 1,95 | 0,50 | 0,75 | 0,32 | 0,69 |
| | 30 | 10,00 | 40,06 | 27,12 | 21,02 | 13,16 | 10,42 | 10,44 | 10,53 | 10,53 |
| | | 5,00 | 32,90 | 21,06 | 12,87 | 7,23 | 5,07 | 5,46 | 4,18 | 4,53 |
| | | 1,00 | 21,50 | 13,29 | 4,20 | 1,67 | 0,76 | 1,03 | 0,29 | 0,53 |

Adicionalmente, com o objetivo de sugerir o tamanho amostral mais seguro para a utilização da estatística tipo-Wald e analisar a proximidade do percentil 95 de sua distribuição, bem como o de cada estatística de teste presente neste estudo com relação ao quantil da χ^2 , foram consideradas as três hipóteses nulas e $\phi = 11$. A proximidade foi analisada simulando os percentis 95 de cada estatística considerando os seguintes tamanhos amostrais $n = 20, 25, 30, 35, 40, 45, 50, 55$ e 60 . Com esses percentis, foram construídos os Gráficos 4.2 e 4.3, para $H_0: \beta_4 = 1$, os Gráficos 4.4 e 4.5 para $H_0: \beta_3 = \beta_4 = 1$ e os Gráficos 4.6 e 4.7 para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$.

Por meio do Gráfico 4.2, podemos observar que, a partir de $n = 45$, o percentil 95 da distribuição da estatística ξ_{Wc} se aproxima consideravelmente do quantil da distribuição χ_1^2 . No entanto, os Gráficos 4.4 e 4.6, mostram que mesmo para uma amostra com 60 observações, o percentil 95 da distribuição da estatística ξ_{Wc} não está muito próximo do quantil da distribuição χ_2^2 e χ_3^2 , respectivamente.

Os Gráficos 4.3, 4.5 e 4.7 mostram que a performance da estatística tipo-Wald é consideravelmente melhor do que sua forma usual. Além disso, para as hipóteses $H_0: \beta_2 = \beta_3 = \beta_4 = 1$ e $H_0: \beta_3 = \beta_4 = 1$, ela apresenta um desempenho

muito semelhante ao da estatística razão de verossimilhanças usual a partir de $n = 45$ e $n = 35$, respectivamente. Para a hipótese $H_0: \beta_4 = 1$, o comportamento da estatística tipo-Wald já é extremamente semelhante ao da ξ_{RV} desde $n = 20$, apresentando valores para o percentil 95 maiores do que os da ξ_{RV} , o que fornece menores taxas de rejeição em relação a essa estatística.

Gráfico 4.2: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para $H_0: \beta_4 = 1$

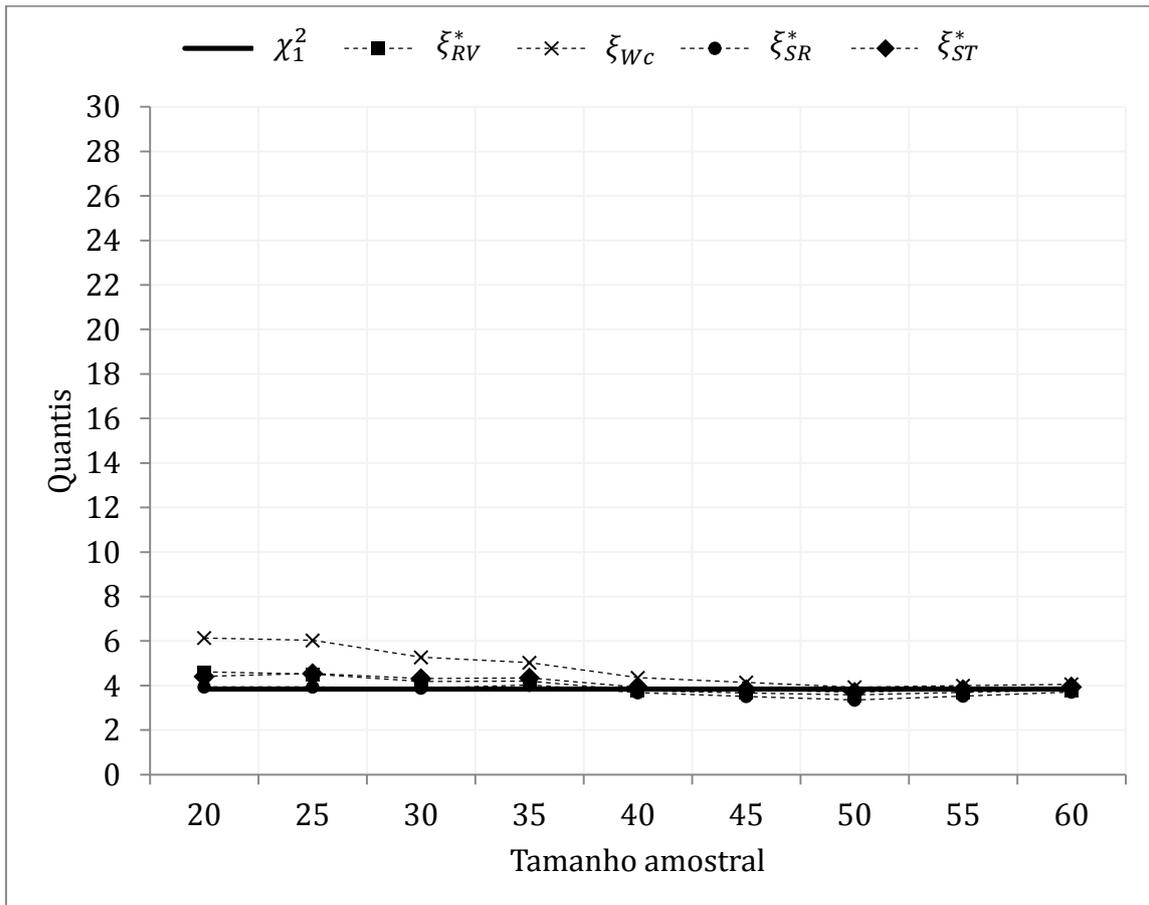


Gráfico 4.3: Percentil 95 das estatísticas ξ_{Wc} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para o cenário $H_0: \beta_4 = 1$

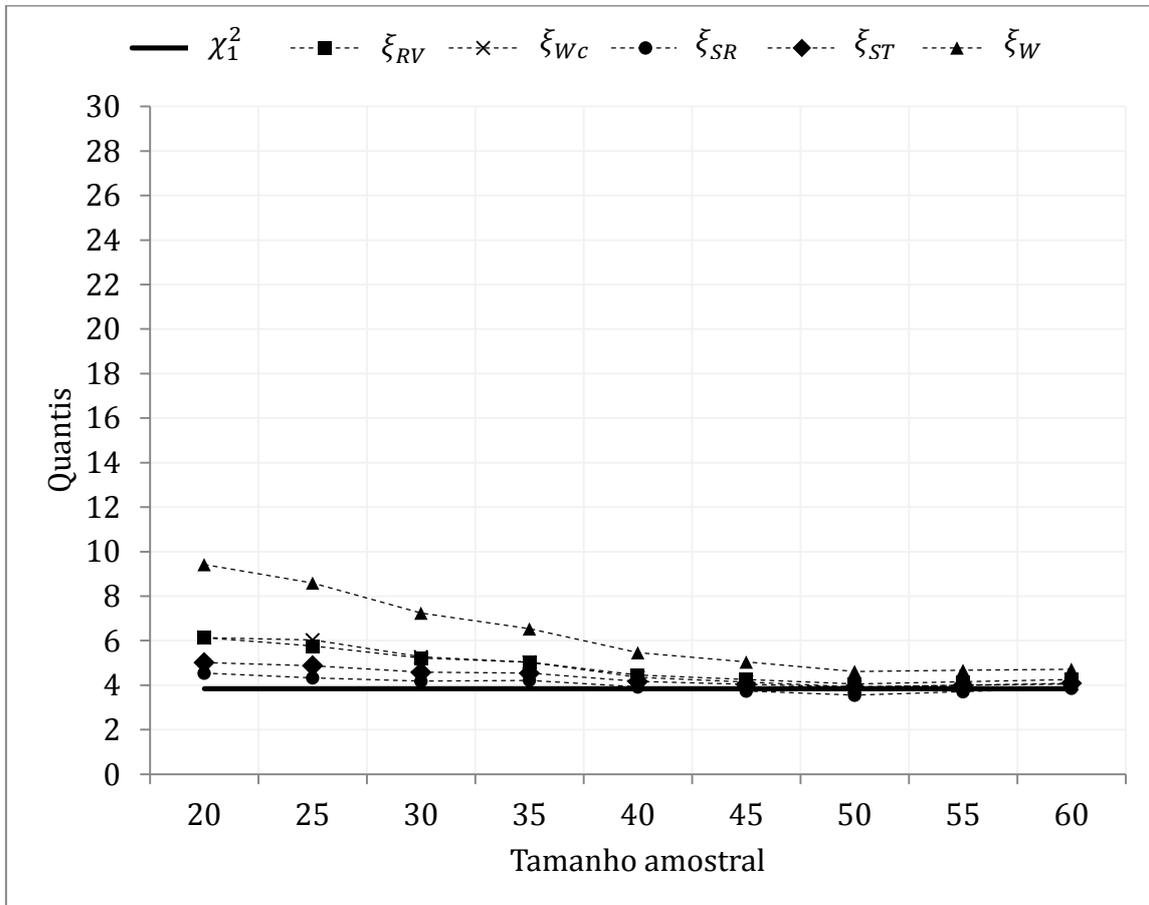


Gráfico 4.4: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para $H_0: \beta_3 = \beta_4 = 1$

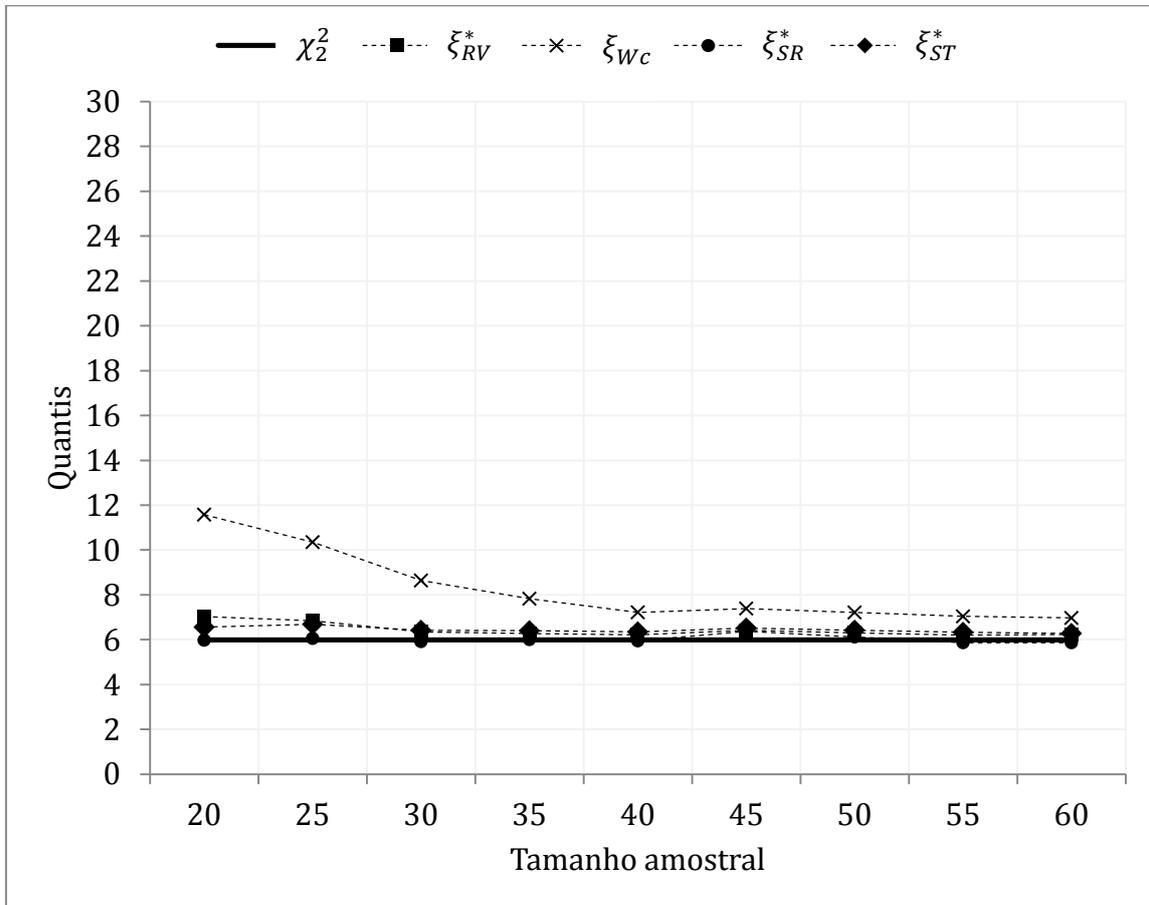


Gráfico 4.5: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para $H_0: \beta_3 = \beta_4 = 1$

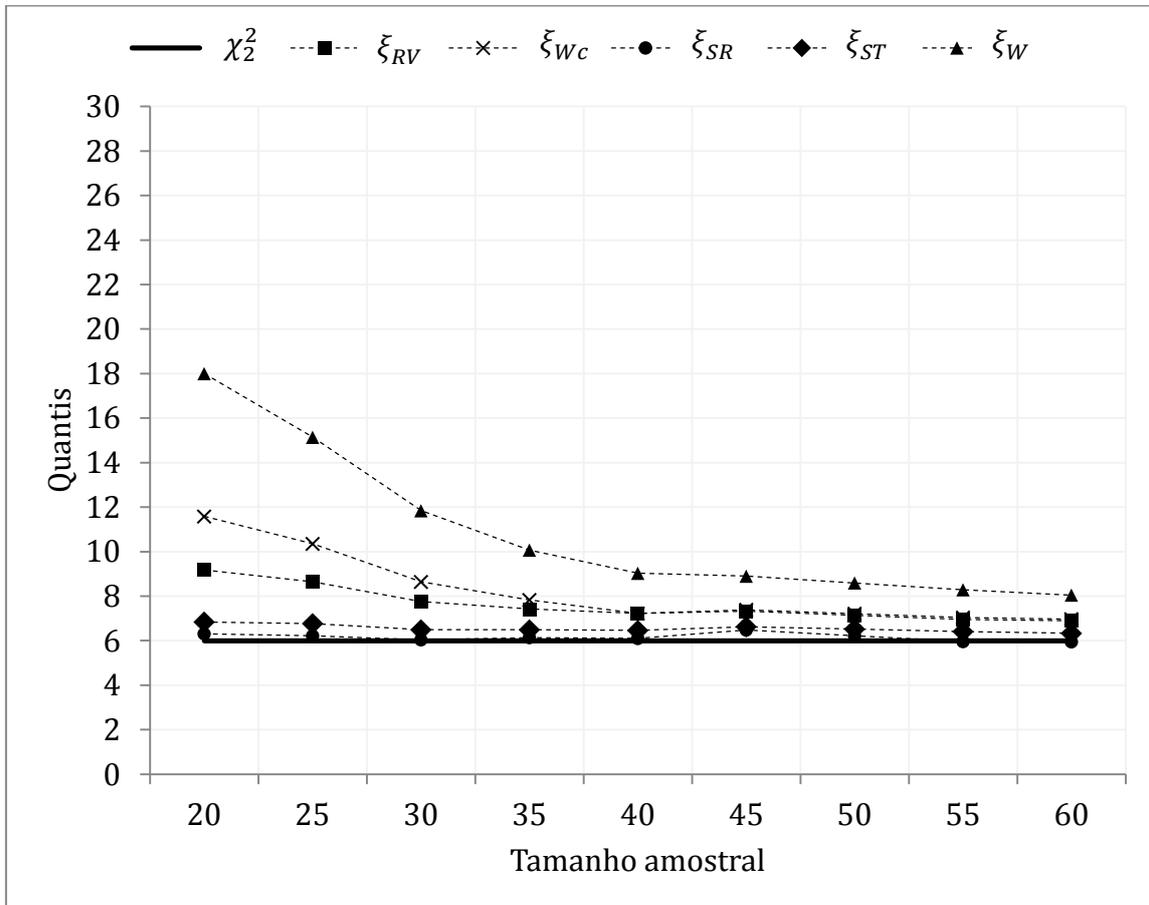


Gráfico 4.6: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$

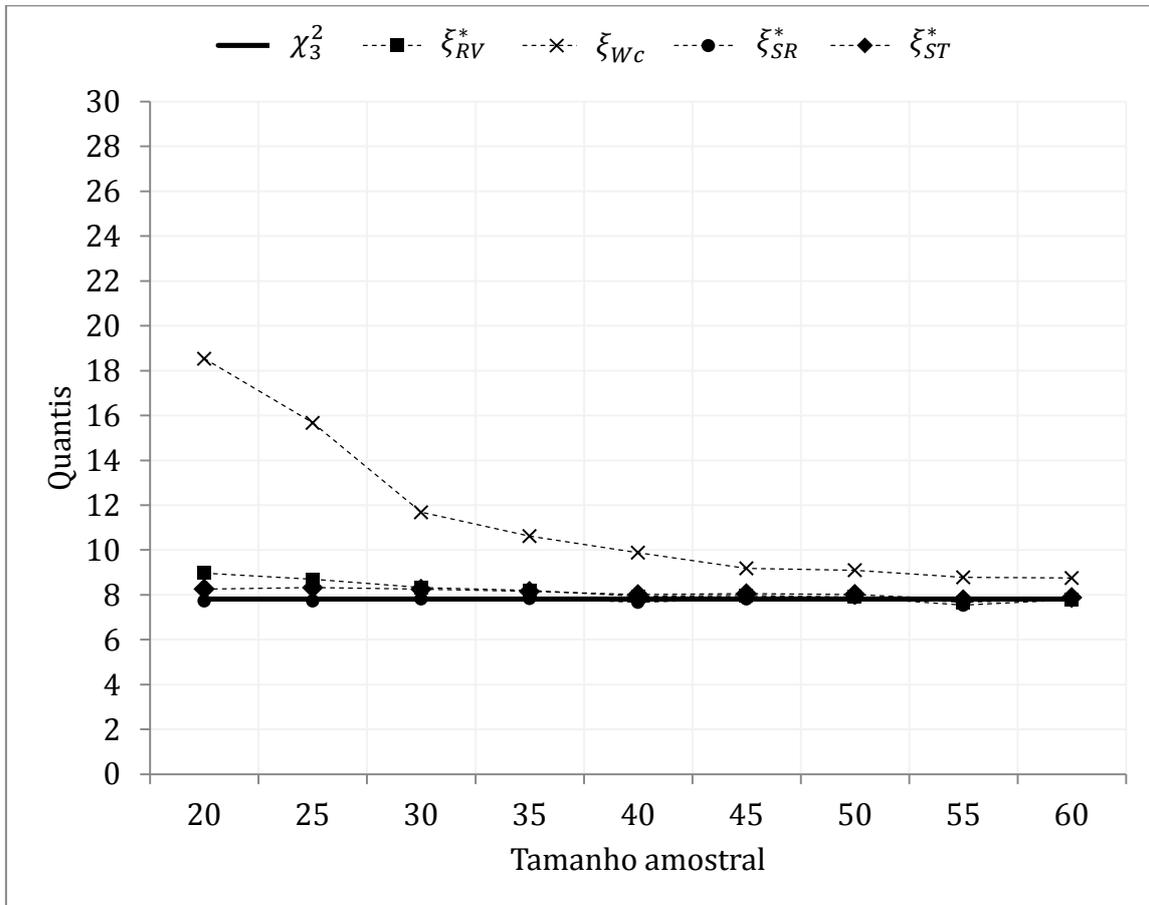
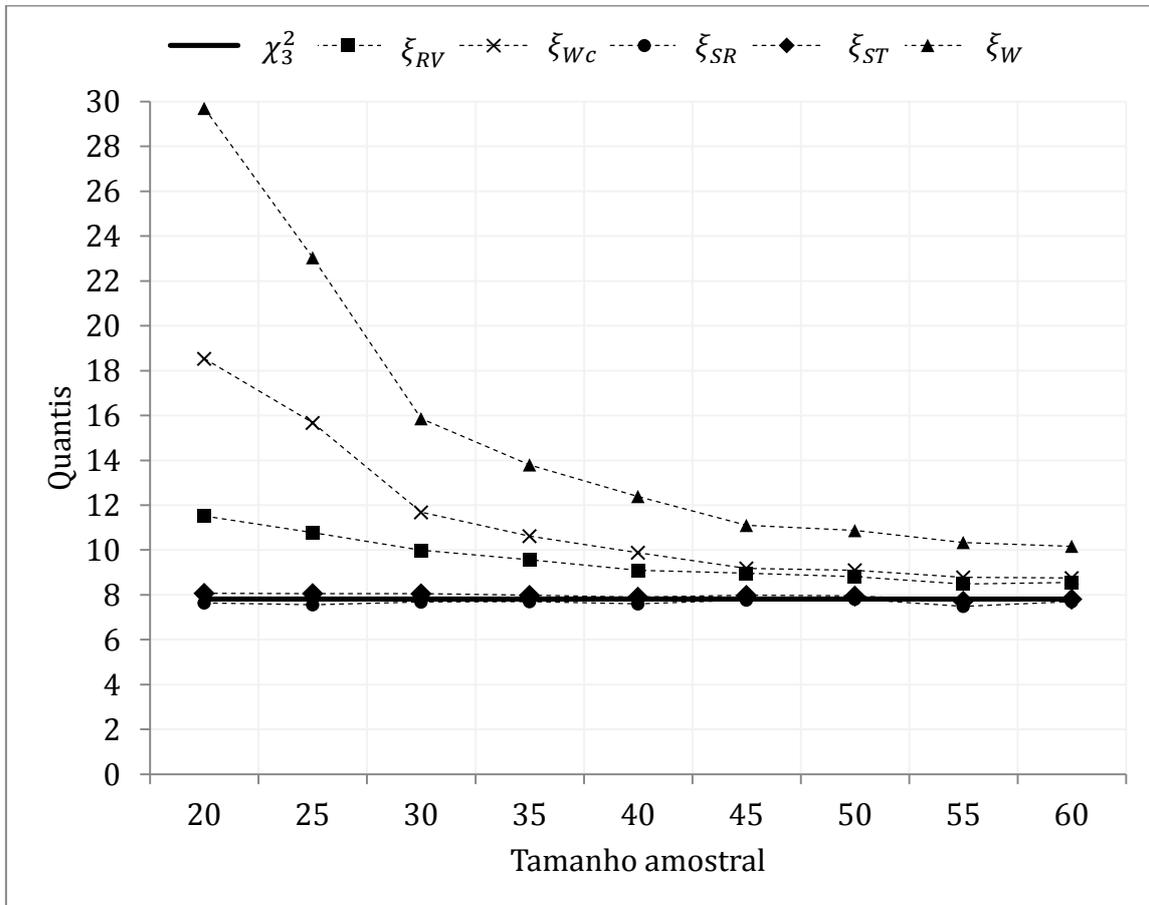


Gráfico 4.7: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão gama com dispersão conhecida ($\phi = 11$) para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$



4.2 Simulação: modelo de regressão com resposta normal inversa

Dizemos que uma variável aleatória Y de média μ e parâmetro de forma ϕ é distribuída segundo uma normal inversa, que denotamos $Y \sim NI(\mu, \phi)$, quando a função densidade de Y , de acordo com Paula (2012), é dada por:

$$f(y; \mu, \phi) = \frac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp\left\{-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right\} = \exp\left\{\phi\left[-\frac{y}{2\mu^2} + \frac{1}{\mu}\right] - \frac{1}{2}\left[\log\left(\frac{2\pi y^3}{\phi}\right) + \frac{\phi}{y}\right]\right\},$$

em que $y > 0$, $\mu > 0$. Desta maneira, para escrever a densidade da normal inversa na forma de (1.1) basta considerar $\theta = -1/2\mu^2$, $b(\theta) = -(-2\theta)^{1/2}$ e $c(y, \phi) = \frac{1}{2}\log[\phi/(2\pi y^3)] - \frac{\phi}{2y}$. A função de variância é dada por $V(\mu) = \mu^3$.

Para os estudos de simulação que seguem, foram arbitrados os valores para ϕ de 11 e 22. Considerando uma amostra y_1, \dots, y_n , com $y_i \sim NI(\mu_i, \phi)$, define-se um modelo de regressão com componente sistemática

$$\mu_i = \eta_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}, i = 1, \dots, n. \quad (4.2)$$

Para cada valor fixado para ϕ , foram geradas 10.000 réplicas de n observações ($n=20, 25, 30$), do modelo com resposta normal inversa definido em (4.2), considerando $\beta_1 = 0,5$, $\beta_2 = \beta_3 = \beta_4 = 1$. Os valores de x_{i2} , x_{i3} e x_{i4} foram obtidos de uma distribuição uniforme no intervalo (1, 2), sendo mantidos os mesmos para todas as réplicas, fixado o tamanho n da amostra.

Foram consideradas as seguintes hipóteses nulas $H_0: \beta_4 = 1$, $H_0: \beta_3 = \beta_4 = 1$ e $H_0: \beta_2 = \beta_3 = \beta_4 = 1$. A taxa de rejeição de cada uma dessas hipóteses consta das Tabelas 4.4 a 4.6, respectivamente. Nas simulações, percebeu-se que, assim como no modelo com resposta gama, o desempenho da taxa de rejeição de H_0 associada à estatística de Wald usual (ξ_W) foi o pior em relação às demais estatísticas, porém, a estatística tipo-Wald proposta por Cordeiro et al. (2012) diminuiu consideravelmente a taxa de rejeição, com reduções entre 33% e 56% em relação à taxa associada com a estatística de Wald usual, para todo valor de ϕ e de α . Isso mostra que a estatística tipo-Wald traz ganhos expressivos nos resultados em amostras pequenas também nos modelos com resposta normal inversa.

A estatística de teste com melhor desempenho foi, em geral, a ξ_{SR}^* , pois o tamanho empírico é o que mais se aproxima dos níveis de significância adotados, dentre as estatísticas consideradas, para os dois valores de ϕ e todos os valores de α . Após a estatística ξ_{SR}^* , a estatística com melhor desempenho foi a estatística *escore*. Após ela, as estatísticas alternam-se entre ξ_{ST} , ξ_{ST}^* e ξ_{RV}^* . O desempenho da estatística tipo-Wald é melhor do que o da estatística da razão de verossimilhanças quando se testa um único parâmetro e consideram-se os valores 5% e 10% para α .

Da mesma forma que no modelo gama, para as estatísticas ξ_W , ξ_{WC} , ξ_{RV} e ξ_{RV}^* , as taxas de rejeição são sempre maiores do que os valores fixados de α . Já, o número de casos em que as estatísticas ξ_{SR} , ξ_{SR}^* , ξ_{ST} , ξ_{ST}^* apresentam taxas de rejeição menores do que os valores de α , cresce à medida que aumenta o número de parâmetros testados em H_0 . Vale ressaltar que a taxa de rejeição das estatísticas de Wald e tipo-Wald se distanciam dos valores fixados para o nível de significância quando se aumenta o número de parâmetros testados em H_0 . Além disso, nota-se que o desempenho de todas as estatísticas é pior para $\phi = 11$.

Tabela 4.4: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_4 = 1$ no modelo de regressão normal inversa com ϕ conhecido

| ϕ | n | α | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 22 | 20 | 10,00 | 30,26 | 18,88 | 23,06 | 15,36 | 13,74 | 11,34 | 18,08 | 14,80 |
| | | 5,00 | 22,53 | 12,98 | 15,11 | 8,77 | 7,18 | 5,76 | 10,23 | 8,44 |
| | | 1,00 | 12,79 | 6,11 | 5,66 | 2,34 | 1,36 | 1,17 | 2,06 | 1,85 |
| | 25 | 10,00 | 26,81 | 17,13 | 20,36 | 13,82 | 12,04 | 10,70 | 15,92 | 14,18 |
| | | 5,00 | 19,12 | 11,21 | 12,39 | 7,72 | 6,15 | 5,45 | 8,21 | 7,54 |
| | | 1,00 | 9,46 | 4,56 | 4,41 | 1,91 | 1,08 | 1,13 | 1,43 | 1,47 |
| | 30 | 10,00 | 25,70 | 17,26 | 18,17 | 12,53 | 11,00 | 9,99 | 15,07 | 13,74 |
| | | 5,00 | 18,58 | 11,24 | 10,93 | 6,52 | 5,66 | 5,18 | 7,42 | 6,68 |
| | | 1,00 | 9,50 | 5,21 | 3,43 | 1,54 | 0,98 | 0,98 | 1,26 | 1,27 |
| 11 | 20 | 10,00 | 34,49 | 20,16 | 26,20 | 17,27 | 13,17 | 11,88 | 18,53 | 16,93 |
| | | 5,00 | 26,33 | 13,90 | 17,95 | 10,04 | 6,99 | 6,54 | 10,24 | 9,91 |
| | | 1,00 | 14,87 | 6,90 | 7,52 | 2,74 | 1,22 | 1,23 | 1,91 | 2,10 |
| | 25 | 10,00 | 30,89 | 18,59 | 23,12 | 15,62 | 11,32 | 10,83 | 16,35 | 15,82 |
| | | 5,00 | 22,37 | 12,36 | 15,08 | 9,04 | 5,65 | 5,64 | 8,28 | 8,58 |
| | | 1,00 | 11,63 | 5,15 | 5,92 | 2,23 | 1,11 | 1,31 | 1,35 | 1,72 |
| | 30 | 10,00 | 31,43 | 20,74 | 21,06 | 13,83 | 10,40 | 10,20 | 15,00 | 14,83 |
| | | 5,00 | 23,88 | 14,29 | 13,11 | 7,27 | 5,21 | 5,14 | 6,87 | 6,91 |
| | | 1,00 | 13,27 | 6,99 | 4,25 | 1,54 | 0,96 | 1,09 | 0,93 | 1,14 |

Tabela 4.5: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_3 = \beta_4 = 1$ no modelo de regressão normal inversa com ϕ conhecido

| ϕ | n | α | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 22 | 20 | 10,00 | 40,87 | 24,87 | 26,82 | 16,07 | 11,83 | 10,41 | 16,46 | 14,77 |
| | | 5,00 | 32,58 | 18,22 | 17,85 | 8,78 | 5,55 | 5,31 | 7,65 | 7,21 |
| | | 1,00 | 20,50 | 9,88 | 6,59 | 2,21 | 0,73 | 1,00 | 1,00 | 1,42 |
| | 25 | 10,00 | 35,78 | 22,36 | 22,79 | 14,66 | 10,65 | 10,14 | 14,95 | 14,52 |
| | | 5,00 | 27,82 | 15,93 | 14,94 | 7,82 | 4,91 | 5,11 | 6,94 | 7,24 |
| | | 1,00 | 16,17 | 7,89 | 5,02 | 1,96 | 0,60 | 0,99 | 0,80 | 1,20 |
| | 30 | 10,00 | 32,80 | 20,51 | 20,00 | 12,39 | 10,02 | 9,60 | 12,88 | 12,31 |
| | | 5,00 | 24,52 | 14,57 | 11,94 | 6,66 | 4,80 | 4,90 | 5,64 | 5,80 |
| | | 1,00 | 13,84 | 7,25 | 3,86 | 1,35 | 0,63 | 0,89 | 0,63 | 0,85 |
| 11 | 20 | 10,00 | 46,28 | 27,61 | 31,03 | 17,92 | 11,41 | 11,52 | 15,19 | 15,36 |
| | | 5,00 | 37,59 | 20,48 | 21,60 | 10,04 | 5,31 | 5,87 | 7,23 | 7,89 |
| | | 1,00 | 24,46 | 11,41 | 8,49 | 2,53 | 0,89 | 1,09 | 1,07 | 1,40 |
| | 25 | 10,00 | 40,91 | 24,67 | 26,46 | 16,04 | 10,03 | 10,58 | 13,85 | 14,82 |
| | | 5,00 | 32,10 | 17,62 | 17,61 | 8,84 | 4,54 | 5,41 | 6,27 | 7,35 |
| | | 1,00 | 19,52 | 8,93 | 6,46 | 2,06 | 0,54 | 1,04 | 0,55 | 1,12 |
| | 30 | 10,00 | 40,55 | 25,67 | 23,43 | 13,55 | 9,51 | 10,05 | 11,77 | 12,48 |
| | | 5,00 | 31,76 | 19,02 | 14,38 | 7,39 | 4,52 | 5,05 | 4,71 | 5,44 |
| | | 1,00 | 19,74 | 10,86 | 4,91 | 1,47 | 0,69 | 1,03 | 0,37 | 0,63 |

Tabela 4.6: Tamanho empírico do teste ou taxa de rejeição (%) de $H_0: \beta_2 = \beta_3 = \beta_4 = 1$ no modelo de regressão normal inversa com ϕ conhecido

| ϕ | n | α | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|--------|-----|----------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| 22 | 20 | 10,00 | 50,77 | 32,65 | 27,74 | 15,28 | 9,92 | 9,89 | 11,06 | 11,01 |
| | | 5,00 | 42,98 | 25,49 | 17,91 | 8,72 | 4,42 | 4,84 | 4,36 | 4,96 |
| | | 1,00 | 29,74 | 16,37 | 6,54 | 2,04 | 0,55 | 0,87 | 0,29 | 0,55 |
| | 25 | 10,00 | 43,60 | 28,05 | 23,82 | 14,33 | 9,86 | 10,39 | 11,48 | 12,10 |
| | | 5,00 | 35,95 | 21,51 | 15,20 | 8,01 | 4,27 | 5,02 | 4,33 | 5,21 |
| | | 1,00 | 23,72 | 12,93 | 5,00 | 1,69 | 0,50 | 0,88 | 0,32 | 0,66 |
| | 30 | 10,00 | 37,87 | 23,94 | 19,91 | 12,47 | 9,76 | 10,02 | 9,90 | 10,32 |
| | | 5,00 | 29,79 | 17,98 | 12,19 | 6,93 | 4,22 | 4,87 | 3,94 | 4,50 |
| | | 1,00 | 18,42 | 10,02 | 3,84 | 1,34 | 0,53 | 0,82 | 0,21 | 0,49 |
| 11 | 20 | 10,00 | 58,51 | 38,73 | 32,16 | 16,79 | 9,66 | 10,43 | 9,23 | 10,15 |
| | | 5,00 | 51,07 | 31,29 | 21,83 | 9,27 | 4,45 | 5,18 | 3,40 | 4,09 |
| | | 1,00 | 36,90 | 19,46 | 8,55 | 2,18 | 0,67 | 0,93 | 0,18 | 0,36 |
| | 25 | 10,00 | 51,16 | 33,75 | 27,09 | 15,69 | 9,45 | 10,43 | 9,20 | 10,40 |
| | | 5,00 | 43,20 | 27,19 | 18,10 | 8,64 | 3,96 | 4,92 | 3,10 | 4,32 |
| | | 1,00 | 30,71 | 16,97 | 6,55 | 1,86 | 0,51 | 0,87 | 0,17 | 0,48 |
| | 30 | 10,00 | 47,44 | 31,72 | 23,52 | 13,51 | 9,17 | 9,82 | 7,80 | 8,56 |
| | | 5,00 | 39,58 | 25,50 | 14,57 | 7,23 | 4,23 | 4,77 | 2,50 | 3,30 |
| | | 1,00 | 27,78 | 15,86 | 4,88 | 1,48 | 0,58 | 0,92 | 0,09 | 0,21 |

Assim como na simulação da regressão gama, foram analisadas aqui as proximidades dos percentis 95 de cada estatística para o quantil da distribuição χ^2 , considerando $\phi = 22$ e os seguintes tamanhos amostrais $n=20, 25, 30, 35, 40, 45, 50, 55$ e 60 . Os Gráficos 4.8 e 4.9, referem-se à hipótese $H_0: \beta_4 = 1$, os Gráficos 4.10 e 4.11 à $H_0: \beta_3 = \beta_4 = 1$ e os Gráficos 4.12 e 4.13 à $H_0: \beta_2 = \beta_3 = \beta_4 = 1$.

O Gráfico 4.8 mostra que, a partir de $n=50$, o percentil 95 da distribuição da estatística ξ_{Wc} se aproxima consideravelmente do quantil da distribuição χ_1^2 . No entanto, os Gráficos 4.10 e 4.12, mostram que mesmo para uma amostra com 60 observações, o percentil 95 da distribuição da estatística ξ_{Wc} não está muito próximo do quantil da distribuição χ_2^2 e χ_3^2 , respectivamente.

Os Gráficos 4.9, 4.11 e 4.13 mostram que a performance da estatística tipo-Wald é consideravelmente melhor do que o da forma usual da estatística de Wald. Para a hipótese $H_0: \beta_4 = 1$, o comportamento da estatística tipo-Wald já é extremamente semelhante ao da ξ_{RV} desde $n = 20$. Já, para a hipótese $H_0: \beta_3 = \beta_4 = 1$, esse resultado só ocorre a partir de $n = 45$ e para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$, o quantil da estatística tipo-Wald está mais distante do quantil da distribuição χ_3^2 do que o da estatística razão de verossimilhanças usual, mesmo para $n = 60$.

Gráfico 4.8: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão normal inversa com $\phi = 22$ para $H_0: \beta_4 = 1$

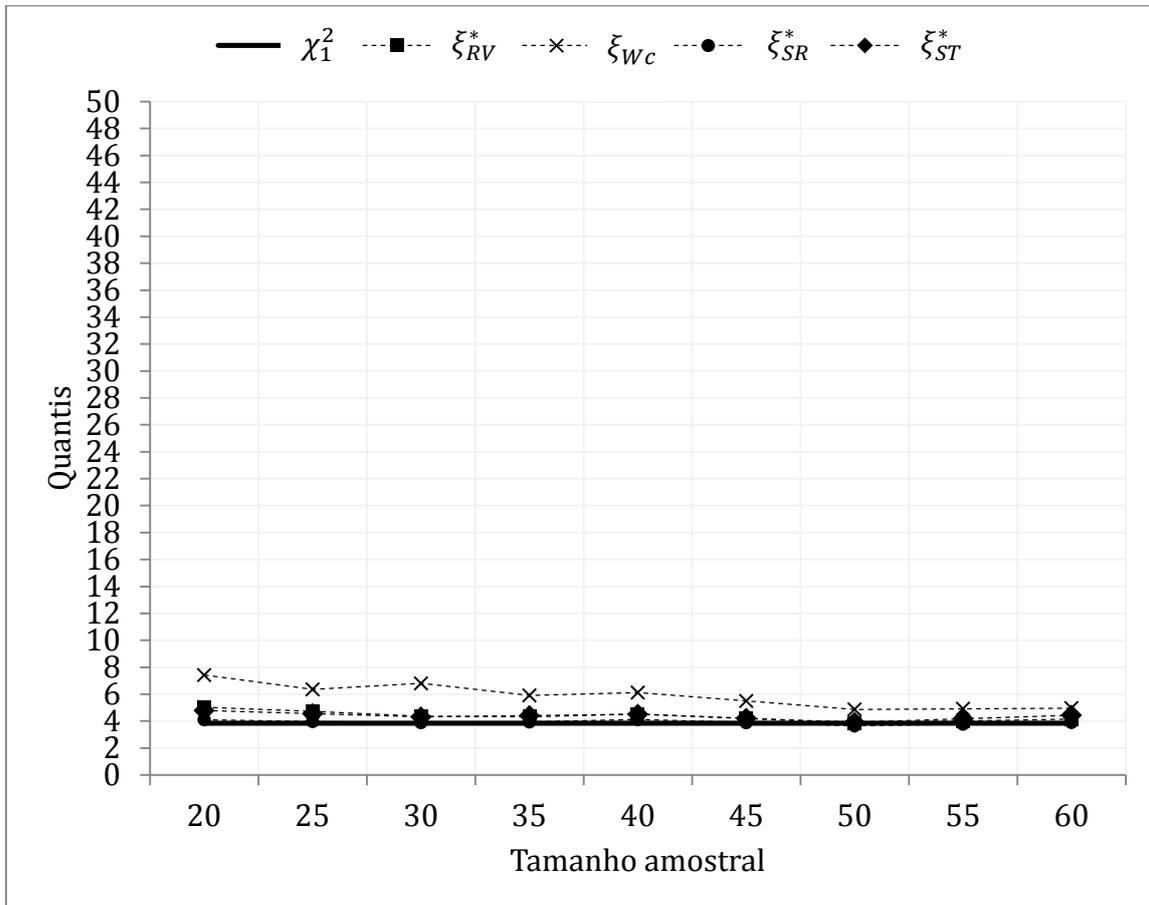


Gráfico 4.9: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão normal inversa com $\phi = 22$ para o cenário $H_0: \beta_4 = 1$

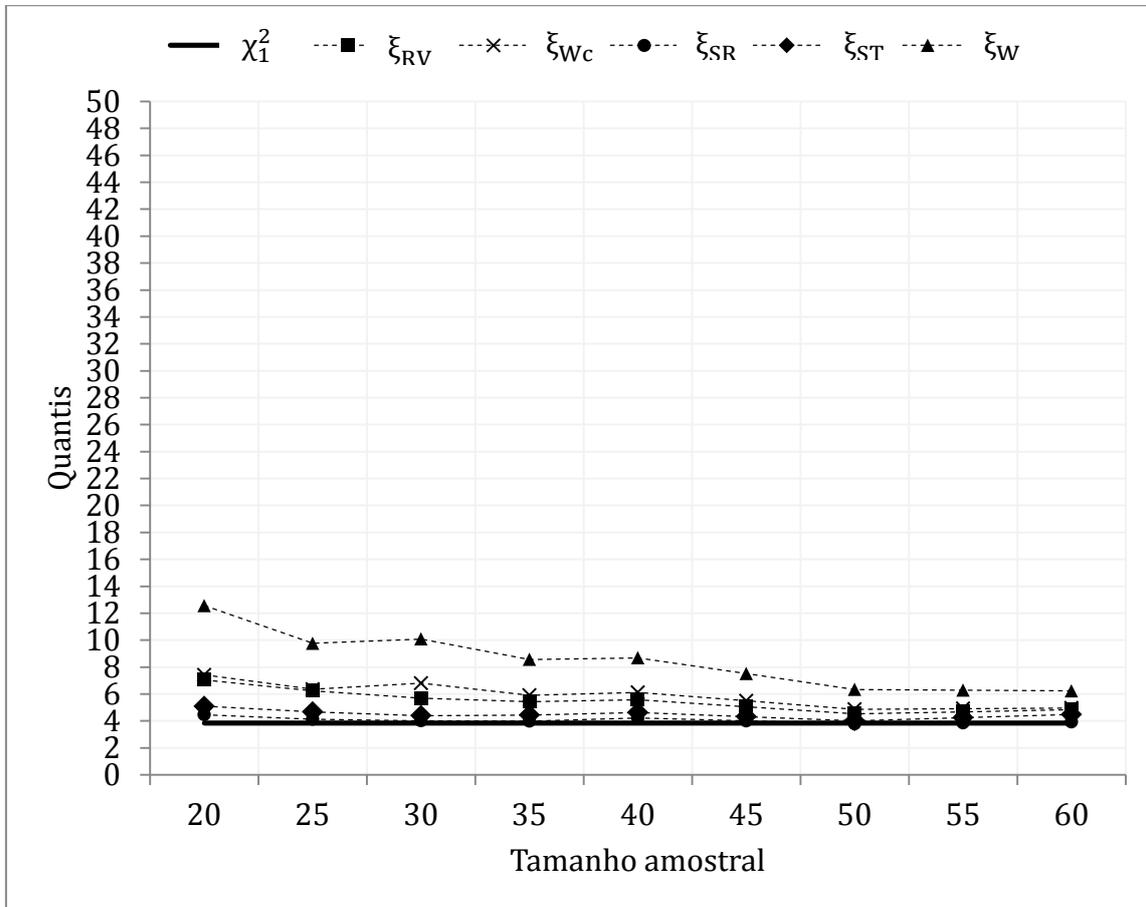


Gráfico 4.10: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão normal inversa com $\phi = 22$ para $H_0: \beta_3 = \beta_4 = 1$

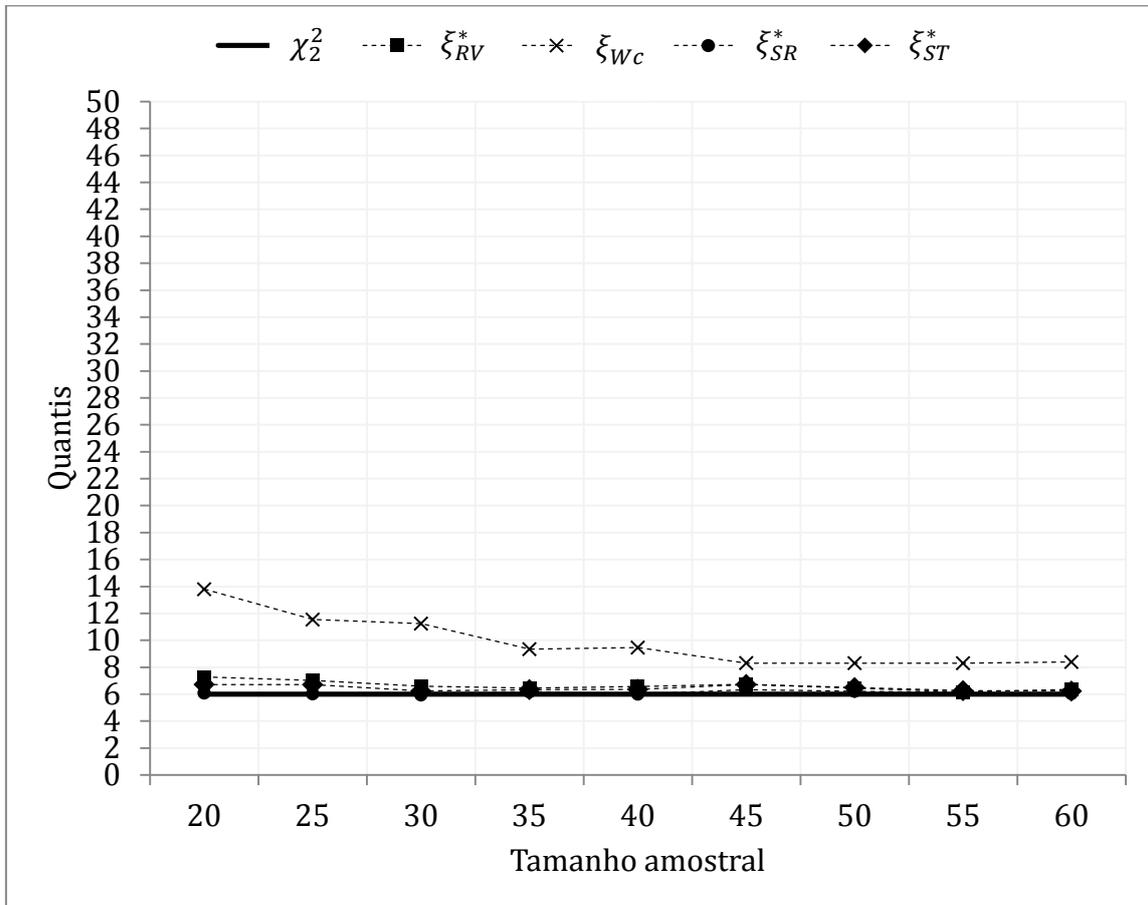


Gráfico 4.11: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão normal inversa com $\phi = 22$ para $H_0: \beta_3 = \beta_4 = 1$

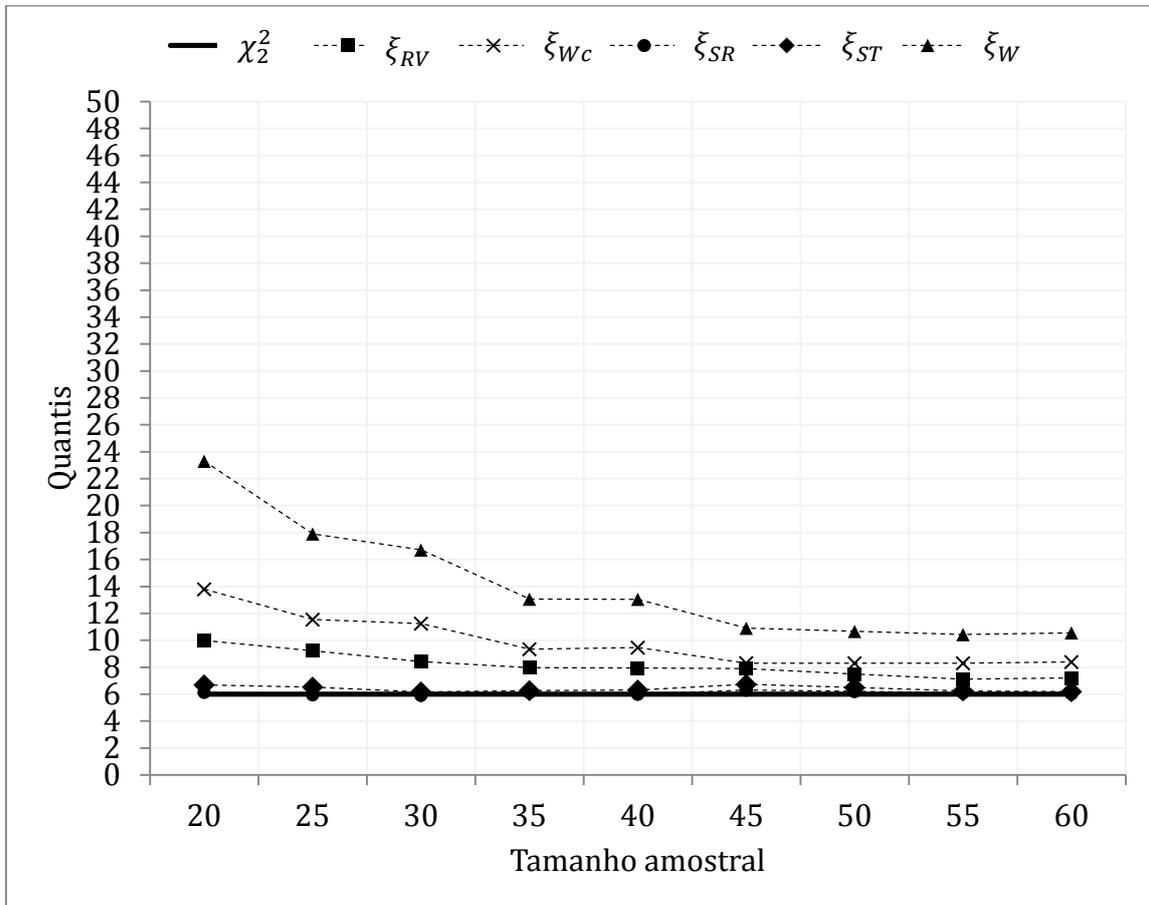


Gráfico 4.12: Percentil 95 das estatísticas ξ_{WC} , ξ_{SR}^* , ξ_{ST}^* e ξ_{RV}^* no modelo de regressão normal inversa com $\phi = 22$ para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$

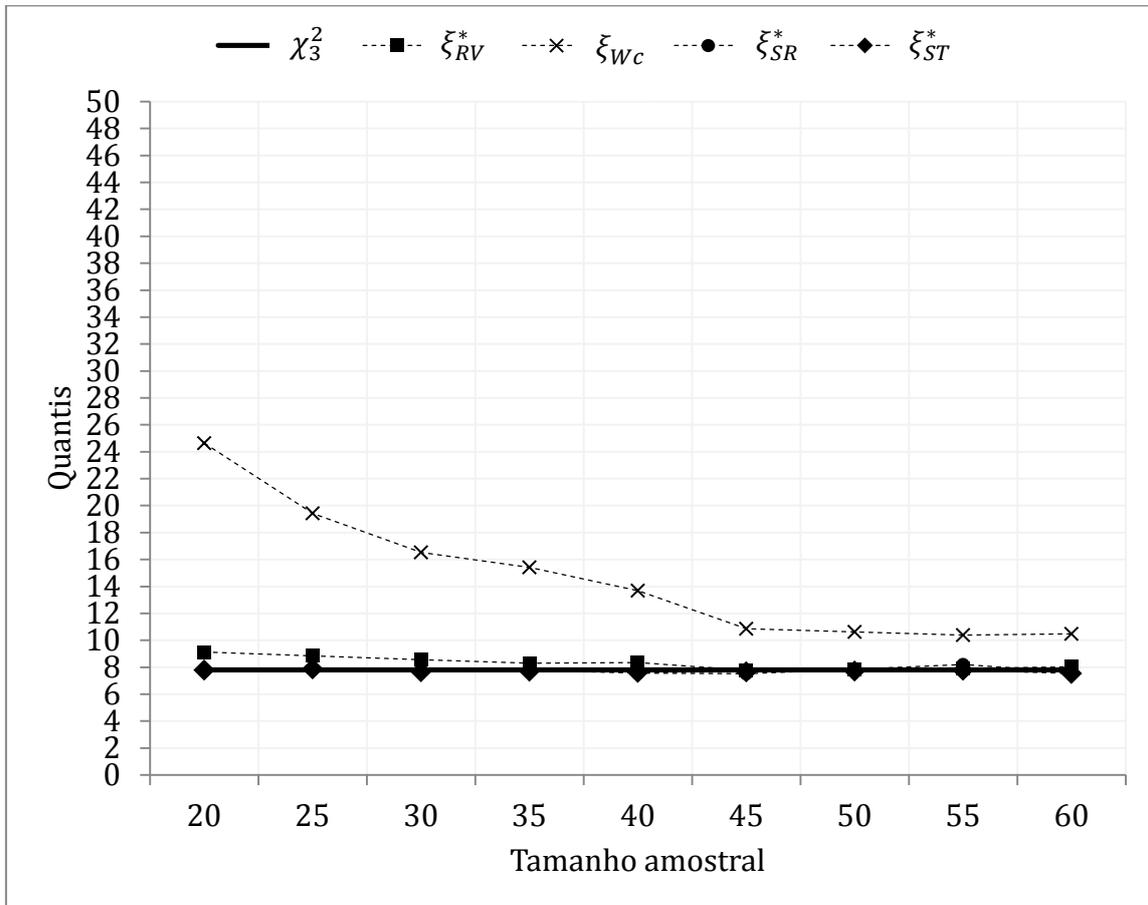
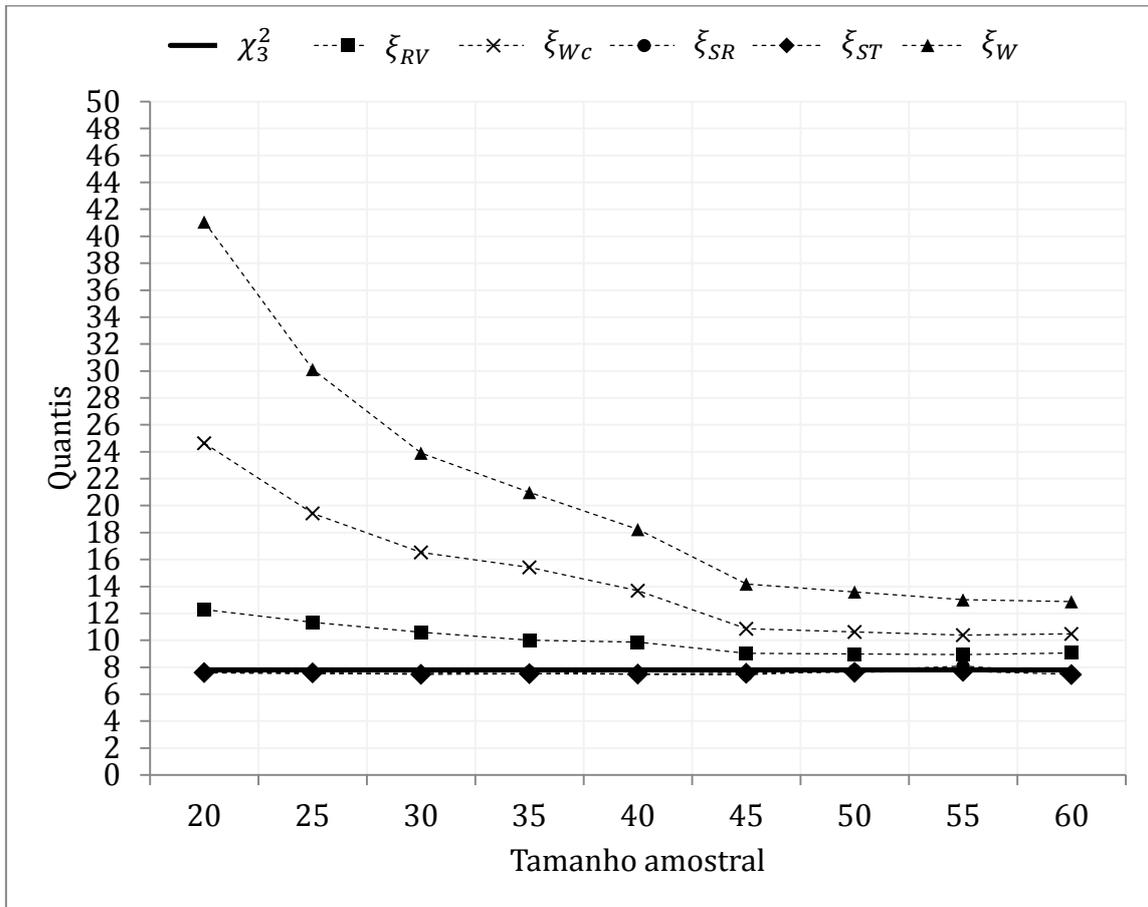


Gráfico 4.13: Percentil 95 das estatísticas ξ_{Wc} , ξ_{SR} , ξ_{ST} , ξ_{RV} e ξ_W no modelo de regressão normal inversa com $\phi = 22$ para $H_0: \beta_2 = \beta_3 = \beta_4 = 1$



Capítulo 5

Aplicação prática

Para comparar o desempenho das estatísticas apresentadas nos Capítulos 2 e 3, na prática, foi realizada uma aplicação com dados reais do atendimento telefônico de uma grande empresa. Foram selecionados 30 atendentes da central telefônica por amostragem aleatória simples (*AAS*) e medidas quatro variáveis em um intervalo de um mês. São elas: média do tempo de atendimento referente ao atendente (Y), número de ligações transferidas para outro atendente (X_1), número de ligações finalizadas em que o cliente sentiu necessidade de retornar ao atendimento em menos de 24h do primeiro contato (X_2) e número de ligações que o atendente provocou queda (X_3). O interesse é modelar a média de Y em função de X_1 , X_2 e X_3 . Desta forma, será assumido que $Y_i \sim G(\mu_i, \phi)$, $i = 1, \dots, 30$, e será fixado $\phi = 1/[CV_y]^2$, sendo que, pelos dados, $CV_y = 11,79\%$ e $\phi = 72$.

Adotando a ligação identidade, temos o seguinte modelo inicial:

$$\mu_i = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

A Tabela 5.1 apresenta as estimativas e respectivos erros padrões dos parâmetros do modelo inicial ajustado.

Tabela 5.1: Estimativas pontuais e erros padrões dos parâmetros do modelo inicial

| | Estimativa | Erro Padrão |
|-----------|------------|-------------|
| β_0 | 548,56 | 32,69 |
| β_1 | -0,32 | 0,14 |
| β_2 | -0,47 | 0,20 |
| β_3 | -0,14 | 0,67 |

A análise de resíduos é apresentada nos Gráficos 5.1 a 5.4, em que são diagnosticados os pontos de alavanca e candidatos a influentes, bem como a qualidade do ajuste. Analisando o Gráfico 5.1, a observação 24 é um ponto de alavanca. Além disso, o Gráfico 5.2 mostra que as observações 3 e 30 são fortes candidatas a pontos influentes, mas a retirada desses pontos não altera a inferência. O Gráfico 5.3 não mostra nenhum padrão nos resíduos em relação aos valores ajustados, portanto, aparentemente não há presença de

heterocedasticidade. O Gráfico 5.4 mostra que todos os resíduos estão dentro dos limites do envelope, portanto, o modelo gama está bem ajustado aos dados. Maiores detalhes sobre a medida h , distância de Cook e gráficos de envelope podem ser obtidas em Paula (2012).

Gráfico 5.1: Identificação de pontos de alavanca (medida h)

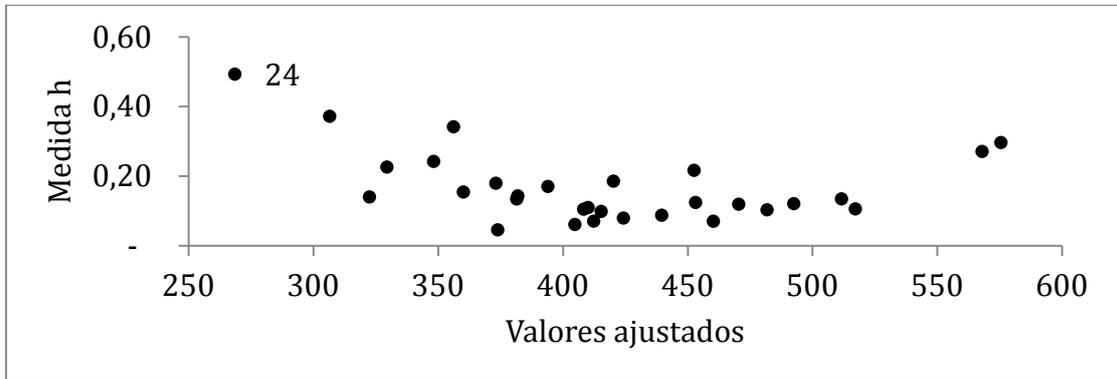


Gráfico 5.2: Identificação de possíveis pontos influentes (distância de Cook)

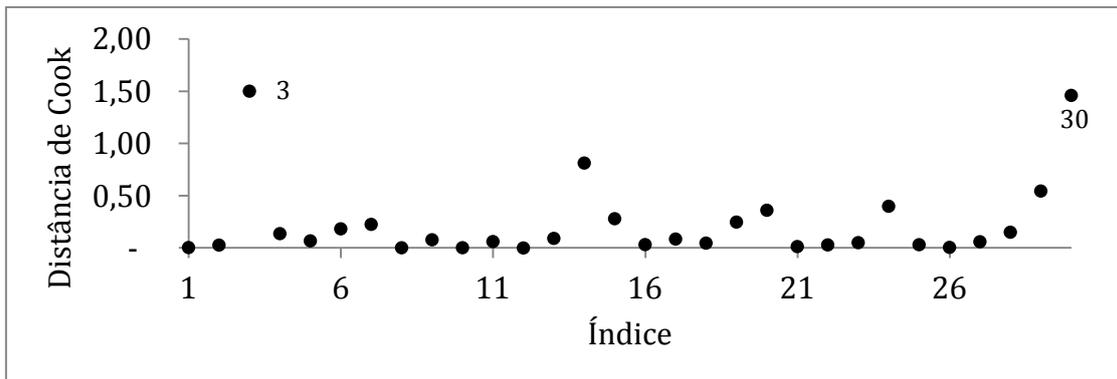


Gráfico 5.3: Dispersão da componente do desvio segundo os valores ajustados

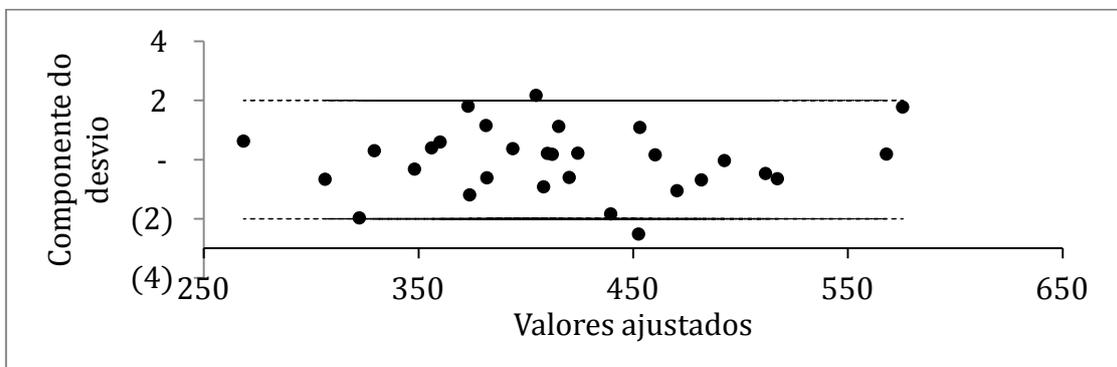
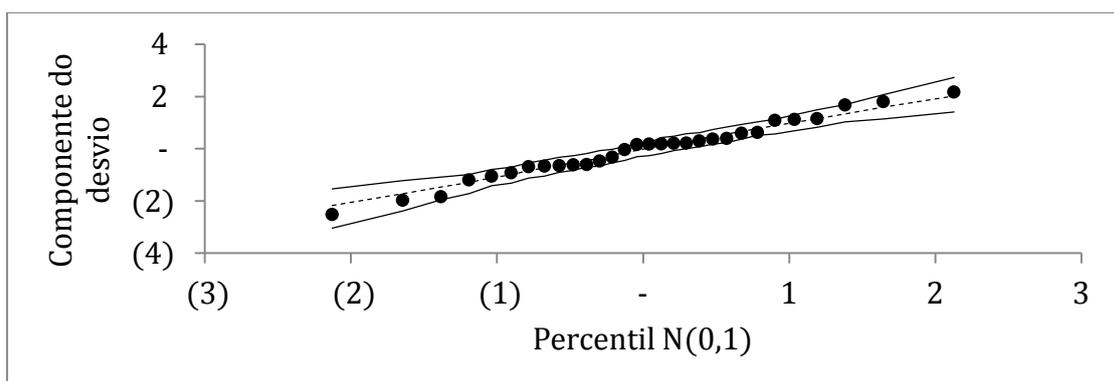


Gráfico 5.4: Qualidade do ajuste (envelope)



Por meio dos testes já apresentados neste trabalho, será avaliado se os coeficientes das variáveis candidatas a serem explicativas podem ser considerados estatisticamente diferentes de zero.

Na Tabela 5.2, é possível observar que o valor-p da estatística de Wald usual é sempre menor ou igual aos valores-p das outras estatísticas, o que aumenta o risco de rejeitar a hipótese nula sendo ela verdadeira, em acordo com o que foi visto nas simulações. Além disso, o valor-p do teste tipo-Wald está extremamente próximo ao do teste da razão de verossimilhanças usual. O teste escore aperfeiçoado é o que apresenta os maiores valores-p, o que o torna mais conservador mediante o risco de cometer o erro do tipo I. A Tabela 5.2 mostra também que a hipótese conjunta de nulidade de todos os coeficientes foi rejeitada em todos os testes. Individualmente, não rejeitamos a hipótese de que o coeficiente da variável número de quedas provocadas seja igual à zero, portanto, podemos concluir que esta variável não tem força na explicação do tempo médio até o fim da ligação. Foi rejeitada em todos os testes a hipótese de nulidade do coeficiente da transferência para outro atendente. No caso do coeficiente da quantidade de ligações com retorno do cliente em até 24h, o teste escore aperfeiçoado não rejeitou a hipótese de nulidade ao nível de significância de 5%, porém, os demais a rejeitaram.

A Tabela 5.3 apresenta as estimativas e respectivos erros padrões dos parâmetros do modelo final, considerando as variáveis explicativas X_1 e X_2 . É possível observar que todas as variáveis apresentam coeficientes negativos, o que parece fazer sentido no caso da transferência para outro atendente, pois aparentemente, ela faz o tempo médio até o fim do atendimento do primeiro atendente reduzir. Já a interpretação da relação entre a quantidade de ligações em que o cliente teve a necessidade de fazer um novo contato e o tempo médio até o fim do atendimento não parece tão direta. A dúvida de interpretação se deve à possível componente de qualidade do tratamento e solução apresentada pelo atendente ao cliente. Esta componente não é diretamente observável, porém,

podemos supor que um atendimento com um tempo consideravelmente baixo, em geral, tende a ter menos qualidade e, portanto, o cliente sente a necessidade de fazer uma nova tentativa de solução de seu problema em menos de 24h do último contato. Embora não seja garantia de que aconteça assim em todos os casos, faz sentido com a prática.

Tabela 5.2: Valores-p dos testes das hipóteses individuais e conjunta para as estatísticas Wald, tipo-Wald, razão de verossimilhanças, escore, gradiente e suas versões aperfeiçoadas

| H_0 | ξ_W | ξ_{Wc} | ξ_{RV} | ξ_{RV}^* | ξ_{SR} | ξ_{SR}^* | ξ_{ST} | ξ_{ST}^* |
|-----------------------------------|---------|------------|------------|--------------|------------|--------------|------------|--------------|
| $\beta_1 = \beta_2 = \beta_3 = 0$ | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| $\beta_1 = 0$ | 0,01 | 0,02 | 0,02 | 0,03 | 0,04 | 0,05 | 0,03 | 0,04 |
| $\beta_2 = 0$ | 0,01 | 0,03 | 0,03 | 0,04 | 0,05 | 0,06 | 0,03 | 0,04 |
| $\beta_3 = 0$ | 0,76 | 0,79 | 0,76 | 0,79 | 0,76 | 0,79 | 0,76 | 0,78 |

Tabela 5.3: Estimativas pontuais e erros padrões dos parâmetros do modelo final

| | Estimativa | Erro Padrão |
|-----------|------------|-------------|
| β_0 | 549,22 | 31,99 |
| β_1 | -0,31 | 0,13 |
| β_2 | -0,50 | 0,14 |

Capítulo 6

Conclusões

Com base nos resultados obtidos nas simulações, é possível dizer que a melhor opção dentre as estatísticas avaliadas é a estatística escore aperfeiçoada e, em segundo lugar, a sua versão usual, pois ambas apresentaram taxas de rejeição muito próximas aos respectivos níveis de significância, mesmo com tamanhos amostrais pequenos na maioria dos cenários simulados. A estatística gradiente também apareceu como uma excelente opção, tanto na forma usual quanto na forma aperfeiçoada, por sua simplicidade de cálculo e proximidade com a estatística escore no que diz respeito aos resultados. Com relação à estatística tipo-Wald, embora não tenha o mesmo desempenho da estatística escore (usual e aperfeiçoada), gradiente (usual e aperfeiçoada) e razão de verossimilhanças aperfeiçoada, foi observado que ela reduz consideravelmente a probabilidade estimada de se cometer o erro do tipo I em relação à estatística de Wald usual. Esta redução ficou entre 30% e 54% no modelo com resposta gama e entre 33% e 56% no modelo com resposta normal inversa. Além disso, em alguns casos a estatística tipo-Wald ofereceu mais assertividade do que a estatística da razão de verossimilhanças usual, por exemplo, esse resultado foi encontrado nos dois modelos simulados, num teste de hipótese para um único parâmetro, considerando-se os valores 5% e 10% para α .

Em linhas gerais, pode-se concluir que a estatística tipo-Wald representa uma valiosa contribuição para os estudos inferenciais nos MLGs e pode substituir sua versão usual nas análises de regressão, evitando equívocos na seleção das variáveis explicativas.

Referências Bibliográficas

- Bartlett, M. S. (1937) Properties of sufficiency and statistical tests. *Proceedings of Royal Society A* 160, 268–282.
- Cordeiro, G. M. (1982) Improved likelihood ratio statistics for generalized linear models, *Imperial College* v. 04, p. 01-21.
- Cordeiro, G. M. (1983) Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B* 45, 404–413.
- Cordeiro, G. M. (1987) On the corrections to the likelihood ratio statistics. *Biometrika* 74, 265–274.
- Cordeiro, G. M., Ferrari, S. L. P. (1991) A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika* v. 78, p. 573-582.
- Cordeiro, G. M., Ferrari, S. L. P. (1991) A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika* 78, 573–582.
- Cordeiro, G. M., McCullagh, P. (1991) Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* 53, 629–643.
- Cordeiro, G. M., Ferrari, S. L. P., Paula, G.A. (1993) Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B* 53, 629–643.
- Cordeiro, G. M. ; Santos, S. J. P. (1999) Corrected Wald test statistics for one-parameter exponential family models. *Communications in Statistics: Theory and Methods* v. 28, p. 1391-1414.
- Cordeiro, G. M. ; Botter, D. A. ; Cavalcanti, A. B. ; Barroso, L. P. (2012) Covariance matrix of the bias-corrected maximum likelihood estimator in generalized linear models. *Statistical Paper* v. 55, p. 643-652.
- Ferrari (1991) Silvia Lopes de Paula Ferrari. Aperfeiçoamento de testes escore, aplicações e extensões. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil.
- Hayakawa, T. (1977) The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* 29, 359–378.
- Lawley, D. (1956) A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* 43, 295–303.

- Nelder, J. A., Wedderburn, R. W. M. (1972) Generalised linear models. *Journal of Royal Statistical Society A* 135, 370–384.
- Paula (2012) Gilberto Alvarenga Paula. Modelos de Regressão com apoio computacional <http://www.ime.usp.br/~giapaula/textoregressao.htm>, Junho 2012. Último acesso em 30/06/2012.
- Rao, C. R. (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* 44, 50–57.
- Rao, C. R. (2005) Calyampudi Radhakrishna Rao *Score test: historical review and recent developments. Advances in Ranking and Selection, Multiple Comparisons, and Reliability*. N. Balakrishnan, H. N. Nagaraja, N. Kannan, 1ª edição.
- Silver (2013) Nate Silver. *O Sinal e o Ruído. Intrínseca*, 1ª edição. Tradução Ana Beatriz Rodrigues.
- Terrell, G. R. (2002) The gradient statistic. *Computing Science and Statistics* 34, 206–215.
- Vargas (2013) Tiago Moreira Vargas. Estatística gradiente: teoria assintótica de alta ordem e correção tipo-Bartlett. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil.
- Wald, A. (1943) Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54, 426–482.
- Wilks, S. S. (1938) The large-sample distribution of the likelihood ratio for testing composite hypothesis. *Annals of Mathematical Statistics* 9, 60–62.