Internacionalização da agricultura brasileira e os efeitos nos preços, disponibilidade e consumo de alimentos no Brasil

Pesquisador: Jonatan Alexandre Oliveira

Orientador: Prof. Dr. José Giacomo Baccarin

Responsáveis pela análise:

Alex Rodrigo dos Santos Souza
Chang Chiann
Gustavo de Oliveira Kanno
Rodrigo Marcel Araujo Oliveira
Victor Ribeiro Baião Decanini

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

CONTEXTO

- A agricultura no Brasil vem sofrendo mudanças ano após ano, principalmente na questão de importação e exportação.

- Essas mudanças refletem no mercado interno brasileiro, alterando:
  - preços
  - qualidade dos produtos
  - disponibilidade
  - forma de consumo

- Aumento significativo dos preços relativos da alimentação quando comparado com outras despesas familiares

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

OBJETIVO

- Procurar identificar quais cadeias de alimentos têm um maior impacto na variação de preços em função do IPCA para uma determinada cadeia de interesse

- Verificar se há correlação entre a variação do preço em função do IPCA de diferentes cadeias agropecuárias de produtos ao longo do tempo

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

TÉCNICAS UTILIZADAS

- Média dos IPCA's dos subitens para descrever o IPCA de cada cadeia
- Teste de Dickey-Fuller e Phillips-Perron para testar se há raiz unitária
- Função de Autocorrelação para identificar correlação dentro de uma única cadeia
- Função de Autocorrelação Parcial para identificar correlação dentro de uma única cadeia
- Correlação Cruzada para identificar correlação entre as cadeias
- Regressão LASSO para ponderar o impacto (pesos) de cada cadeia em uma cadeia de interesse
- Regressão clássica no contexto de séries temporais para compreender melhor o funcionamento da série
- Regressão com erros autocorrelacionados (ARIMA)
- Teste Ljung-Box para ARIMA(p,d,q)

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

TESTES DE DICKEY-FULLER e PHILLIPS-PERRON

- Hipótese nula: Afirma que há presença de raiz unitária, ou seja, série não estacionária
- Hipótese alternativa: Afirma que não há raiz unitária, ou seja, série é estacionária
- β_1 : Intercepto do modelo
- β_2 : Coeficiente de tendência
- δ : Coeficiente de presença de raiz unitária

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta y_{t-i} + \varepsilon_t$$

- Estatística de teste T de Dickey-Fuller
- Tabelados por Dickey-Fuller através de simulação Monte Carlo

$$T = \frac{\hat{\delta}}{se(\hat{\delta})}$$

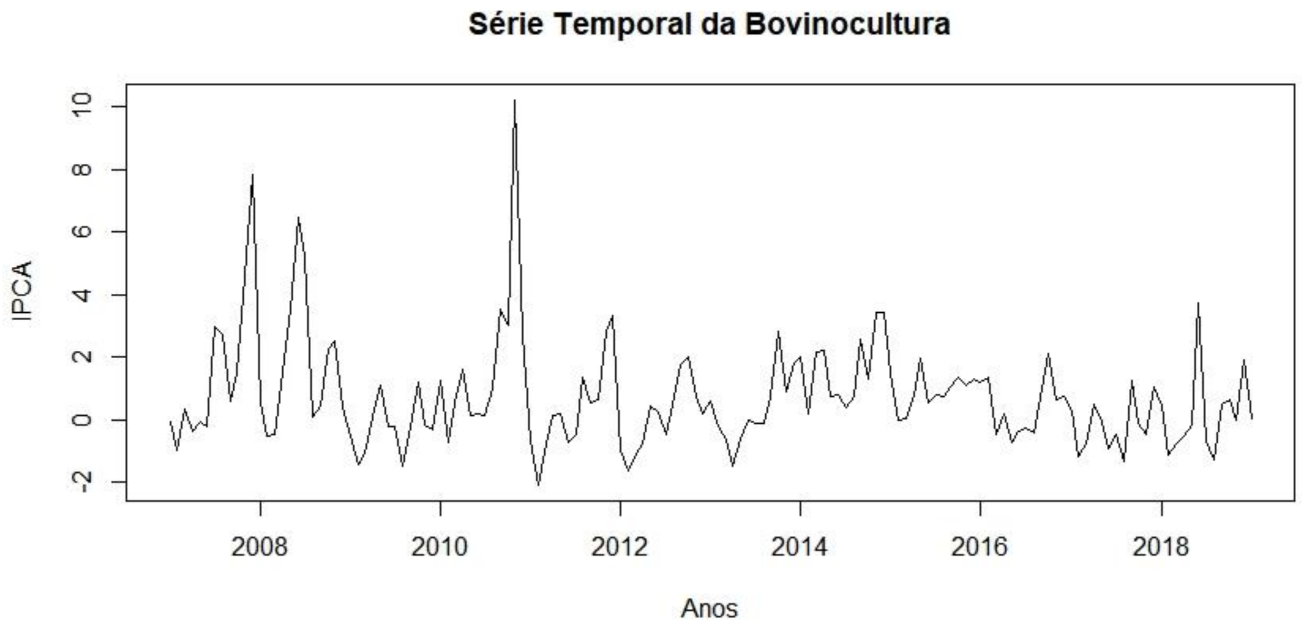
- Estatística de teste Z de Phillips-Perron
- Z é um ajuste na estatística de Dickey-Fuller

$$Z = n\hat{\delta}_n - \frac{n^2\hat{\sigma}^2}{2s_n^2}\left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}\right)$$

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- **Série Temporal**
- Correlações
- Regressões

SÉRIE TEMPORAL DA BOVINOCULTURA



- Não há indícios de componente sazonal
- Observa-se uma grande variação do IPCA em torno de uma reta positiva, principalmente no período dos anos de 2013 a 2017, com destaque para 2011, indicando assim um aumento no preço continuamente.
- Teste de Dickey-Fuller e Teste de Phillips-Perron: o p-valor é de 0,01 para a série
- Rejeitamos a hipótese nula de que a série temporal contém uma raiz unitária.

FUNÇÃO DE ACF e PACF

- A autocorrelação é definida como correlação entre duas observações em instantes diferentes. A autocorrelação de lag um caracteriza séries onde uma observação está correlacionada com a observação imediatamente anterior ou posterior (fevereiro e janeiro, ou fevereiro e março, por exemplo)

- A função de autocovariância amostral entre duas observações de lag k é definida como:

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

- A função de autocorrelação amostral de lag k é :

$$r_k = \frac{c_k}{c_0} = Corr(x_t, x_{t+k})$$

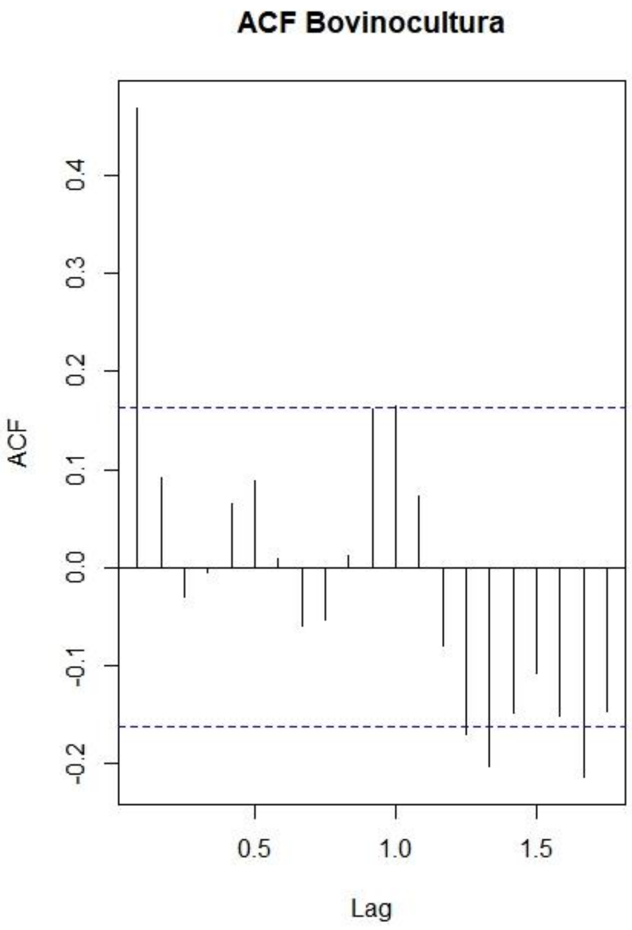
- O intervalo de confiança de aproximadamente 95% para ACF pode ser construído por:

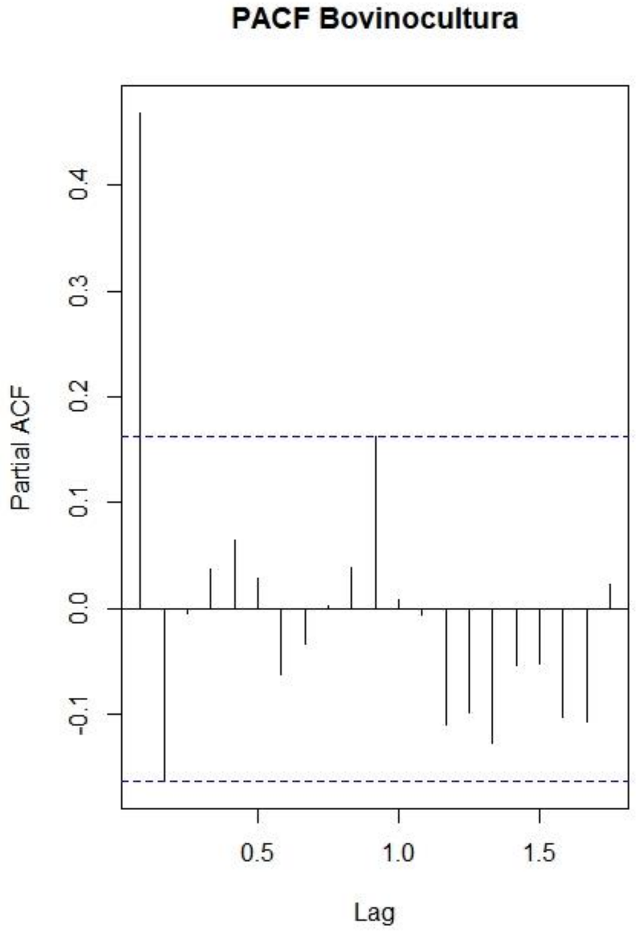
$$r_k \pm \frac{2}{\sqrt{n}}$$

- A função de autocorrelação parcial (PACF) mede a correlação entre X_(t) e X_(t-k) eliminando a influência X_(t-1), ... X_(t-k+1):

$$f_k = \begin{cases} Corr(X_t, X_{t+k}) = r_1 & \text{se k} = 1 \\ Corr(X_{t+k} - X_t^{k-1}, X_t - X_t^{k-1}) & \text{se k} > 1 \end{cases}$$

FUNÇÃO DE AUTOCORRELAÇÃO





- Notamos que temos um pico no primeiro lag do ACF e do PACF, ambos fora do intervalo de confiança, isso significa que esta observação está correlacionada positivamente com a observação imediatamente anterior.

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

CORRELAÇÃO CRUZADA

- Representa as correlações entre duas séries temporais em diferentes períodos de tempo

- Pode ajudar a entender se uma série de dados "conduz" uma outra série, e até que ponto isso acontece

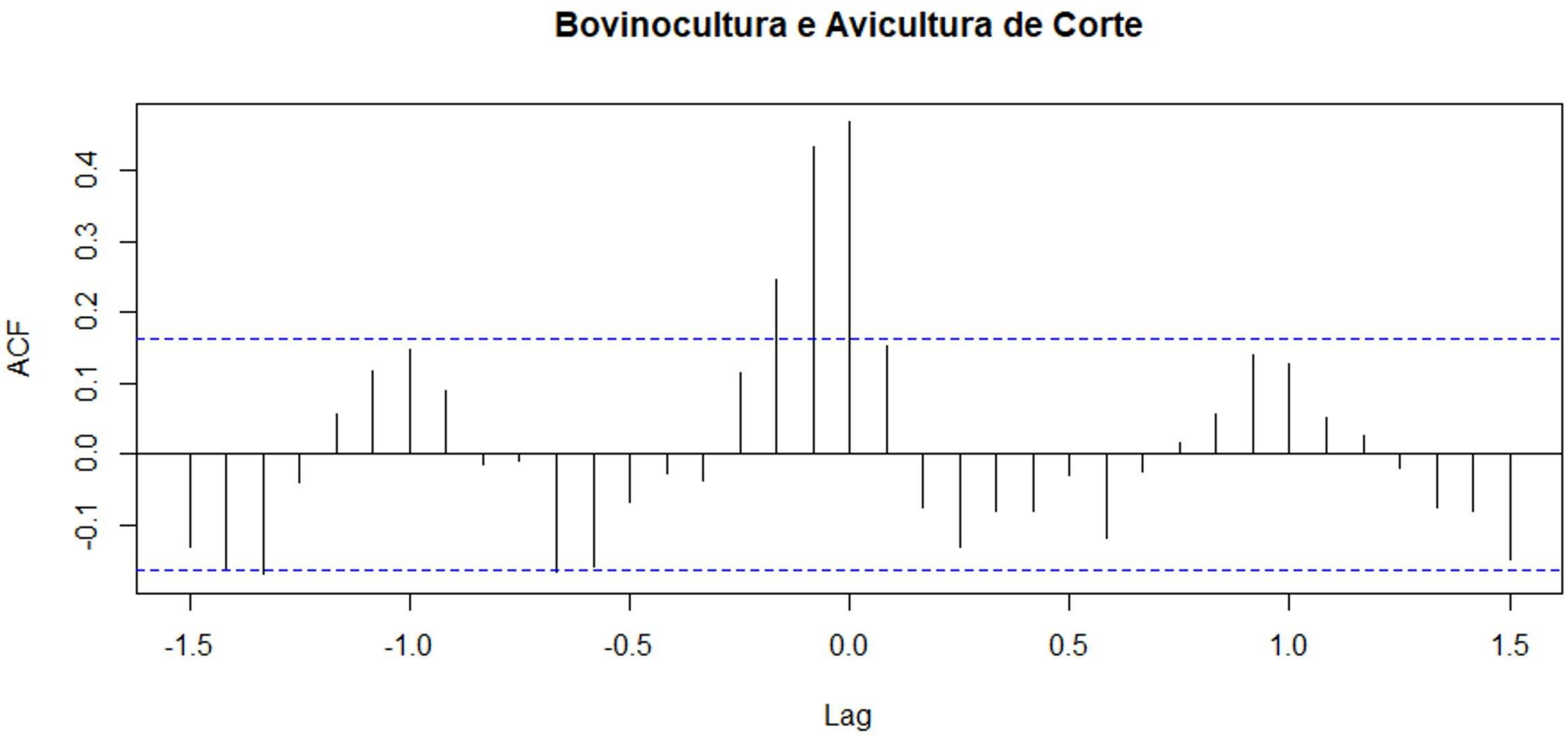
- Correlação cruzada amostral dada por:

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}.$$

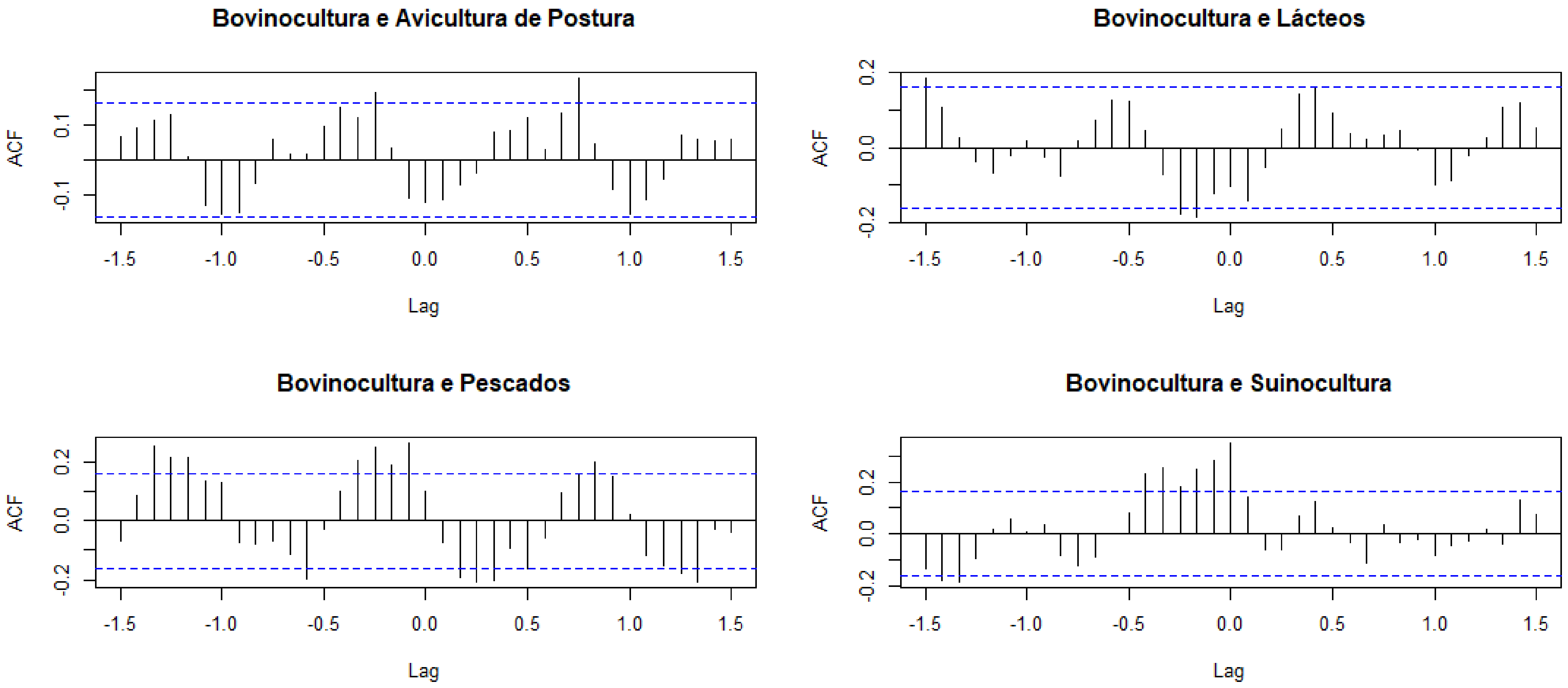
Com:

$$\hat{\gamma}_{xy}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(y_t - \bar{y}),$$

CORRELAÇÃO CRUZADA DA BOVINOCULTURA



CORRELAÇÃO CRUZADA DA BOVINOCULTURA



MAIORES CORRELAÇÕES

- Tendo analisado os gráficos de autocorrelação e correlações cruzadas da série de Bovinocultura, vê-se que as principais correlações entre as cadeias (e seus respectivos Lags) são dados por:

Cadeia	Lags
Avicultura de Corte	Lag 0
Avicultura de Postura	Lag 9
Pescado	Lag 3
Pescado	Lag 10
Suinocultura	Lag 0
Bovinocultura	Lag 1

AGENDA

- Contexto
- Objetivo
- Técnicas Utilizadas
- Testes Estatísticos
- Série Temporal
- Correlações
- Regressões

REGRESSÃO LASSO

- A Regressão Lasso tem o efeito de fazer com que alguns coeficientes (os menos importantes para o modelo) diminuam em módulo até ficar exatamente igual a zero

- A ideia aqui é inferir as correlações mais importantes entre as cadeias com base nas variáveis que restam.

REGRESSÃO LASSO

- Generalização da Regressão Linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Fixado um λ > 0, queremos β que minimize:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} |\beta_j|.$$

- O estimador será dado por:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}.$$

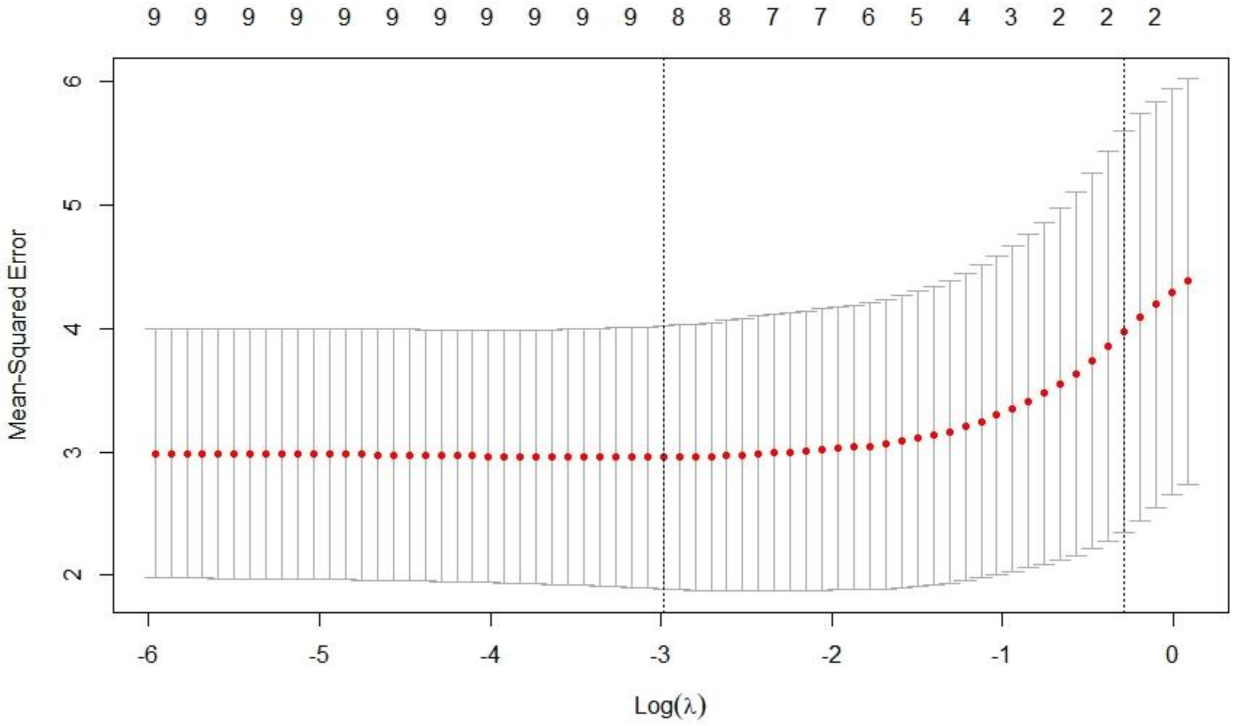
VALIDAÇÃO CRUZADA

- Antes de aplicar a Regressão Lasso, precisamos fixar um valor de λ;

- Através de um método conhecido como validação cruzada, conseguimos escolher λ de tal forma que minimizamos o erro.

VALIDAÇÃO CRUZADA

- Para nosso caso, fizemos validação cruzada com nfolds = 10



	Measure: Mean-Squared Error			
	Lambda	Measure	SE	Nonzero
min	0.0504	2.955	1.067	8
1se	0.7489	3.976	1.625	2

REGRESSÃO LASSO

- Os coeficientes obtidos para as variáveis através do modelo foram:

Library(glmnet)

Variáveis	Estimativas
(Intercept)	0.31262107
\`Avicultura de Corte\`	0.39843764
\`Avicultura de Postura\`	0
\`Pescado\`	-0.15329263
\`Lácteos\`	-0.1509418
\`Suinocultura\`	0.24605653
\`Avicultura de Postura 9\`	0.14932952
\`Pescado 3\`	-0.01311084
\`Pescado 10\`	0.01739267
\`Bovinocultura 1\`	0.35784156

REGRESSÃO LASSO

- Os coeficientes obtidos para as variáveis através do modelo foram:

Library(islasso)

Variáveis	Estimativa	Erro Padrão	Valor Z	Pr(>|Z|)
(Intercepto)	0.28555	0.20359	1.403	0.160748
`Avicultura de Corte`	0.41319	0.11348	3.641	0.000271
`Avicultura de Postura`	0.04527	0.06032	0.750	0.452995
`Pescado`	-0.25994	0.11192	-2.323	0.020202
`Lácteos`	-0.20752	0.12319	-1.685	0.092067
`Suinocultura`	0.27996	0.21147	1.324	0.185555
`Avicultura de Postura 9`	0.17970	0.05358	3.354	0.000796
`Pescado 3`	-0.0226	0.10177	-0.217	0.828397
`Pescado 10`	0.07148	0.10156	0.704	0.481548
`Bovinocultura 1`	0.37949	0.09756	3.890	0.000100

REGRESSÃO CLÁSSICA NO CONTEXTO DE SÉRIES TEMPORAIS

- A regressão simples no contexto de séries temporais pode ser expressa por:

$$x_t = \beta_0 + \beta_1 z_{t1} + \cdots + \beta_q z_{tq} + w_t = \beta' z_t + w_t .$$

- O erro de mínimos quadrados é dado por:

$$Q = \sum_{t=1}^{n} w_t^2 = \sum_{t=1}^{n} (x_t - \beta' z_t)^2,$$

- As estimativas dos coeficientes são explicadas por:

$$\hat{\beta} = \left(\sum_{t=1}^{n} z_t z_t' \right)^{-1} \sum_{t=1}^{n} z_t x_t .$$

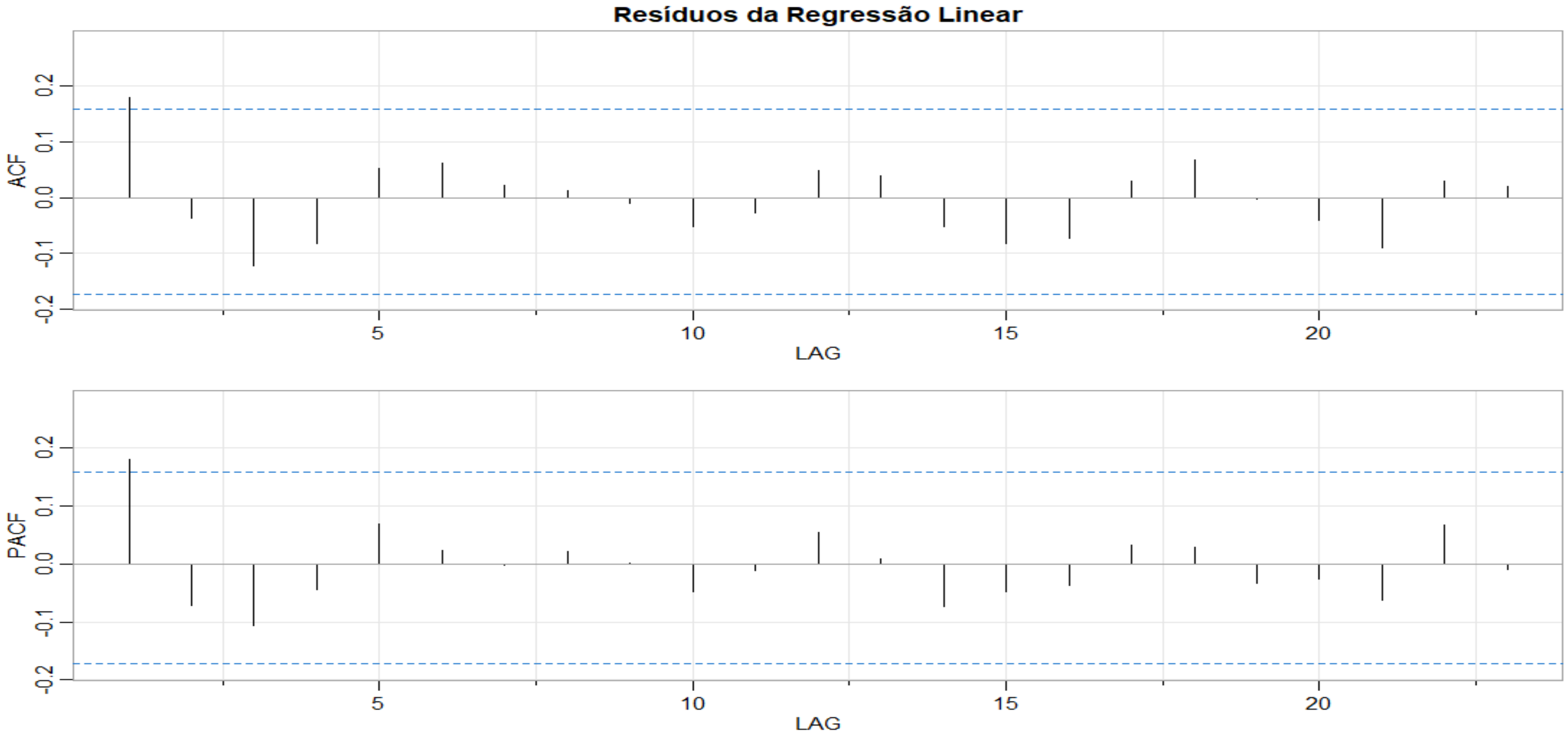
REGRESSÃO CLÁSSICA NO CONTEXTO DE SÉRIES TEMPORAIS

- Os coeficientes obtidos para as variáveis através do modelo foram:

Variáveis	Estimativa	Erro Padrão	Valor t	Pr(>|t|)
(Intercepto)	0.28536	0.20364	1.401	0.163405
`Avicultura de Corte'	0.41328	0.11349	3.642	0.000384
`Avicultura de Postura'	0.04542	0.06035	0.753	0.452982
`Pescado`	-0.26037	0.11194	-2.326	0.021498
`Lácteos`	-0.20785	0.12322	-1.687	0.093939
`Suinocultura`	0.28048	0.21162	1.325	0.187266
`Avicultura de Postura 9'	0.17980	0.05358	3.356	0.001026
`Pescado 3'	-0.02202	0.10186	-0.216	0.829147
`Pescado 10'	0.07166	0.10163	0.705	0.481954
`Bovinocultura	0.37950	0.09758	3.889	0.000157

REGRESSÃO CLÁSSICA NO CONTEXTO DE SÉRIES TEMPORAIS

- Agora precisamos analisar as funções de autocorrelação e autocorrelação parcial dos resíduos do modelo.



REGRESSÃO COM ERROS AUTOCORRELACIONADOS

- O modelo de Regressão também pode ser escrito como

$$y_t = \sum_{j=1}^{r} \beta_j z_{tj} + x_t$$

- x_t é um processo com alguma função de covariância $\gamma_x(s,t)$. Como nesse caso o resíduo do modelo é x_t o modelo com erros autocorrelacionados é dado por

$$y^* = Z^* \beta + \delta ,$$

- Essa função é composta por

$$\Gamma = \{\gamma_x(s,t)\}, \qquad y^* = \Gamma^{-1/2} y, \; Z^* = \Gamma^{-1/2} Z, \qquad \delta = \Gamma^{-1/2} x.$$

- A estimativa do método quadrados ponderados é dada por

$$\hat{\beta}_w = (Z^{*\prime} Z^*)^{-1} Z^{*\prime} y^* = (Z' \Gamma^{-1} Z)^{-1} Z' \Gamma^{-1} y,$$

REGRESSÃO COM ERROS AUTOCORRELACIONADOS

- Para o caso autoregressivo de ordem p temos

$$\phi(B)x_t = w_t, \qquad \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \qquad BX_t = X_{t-1} \quad B^m X_t = X_{t-m}$$

- Portando a regressão fica

$$\underbrace{\phi(B)y_t}_{y_t^*} = \sum_{j=1}^{r} \beta_j \underbrace{\phi(B)z_{tj}}_{z_{tj}^*} + \underbrace{\phi(B)x_t}_{w_t},$$

- O erro é minimizado em função da soma de mínimos quadrados a seguir

$$S(\phi, \beta) = \sum_{t=1}^{n} w_t^2 = \sum_{t=1}^{n} \left[\phi(B)y_t - \sum_{j=1}^{r} \beta_j \phi(B) z_{tj} \right]^2$$

- Os parametros são respectivamente

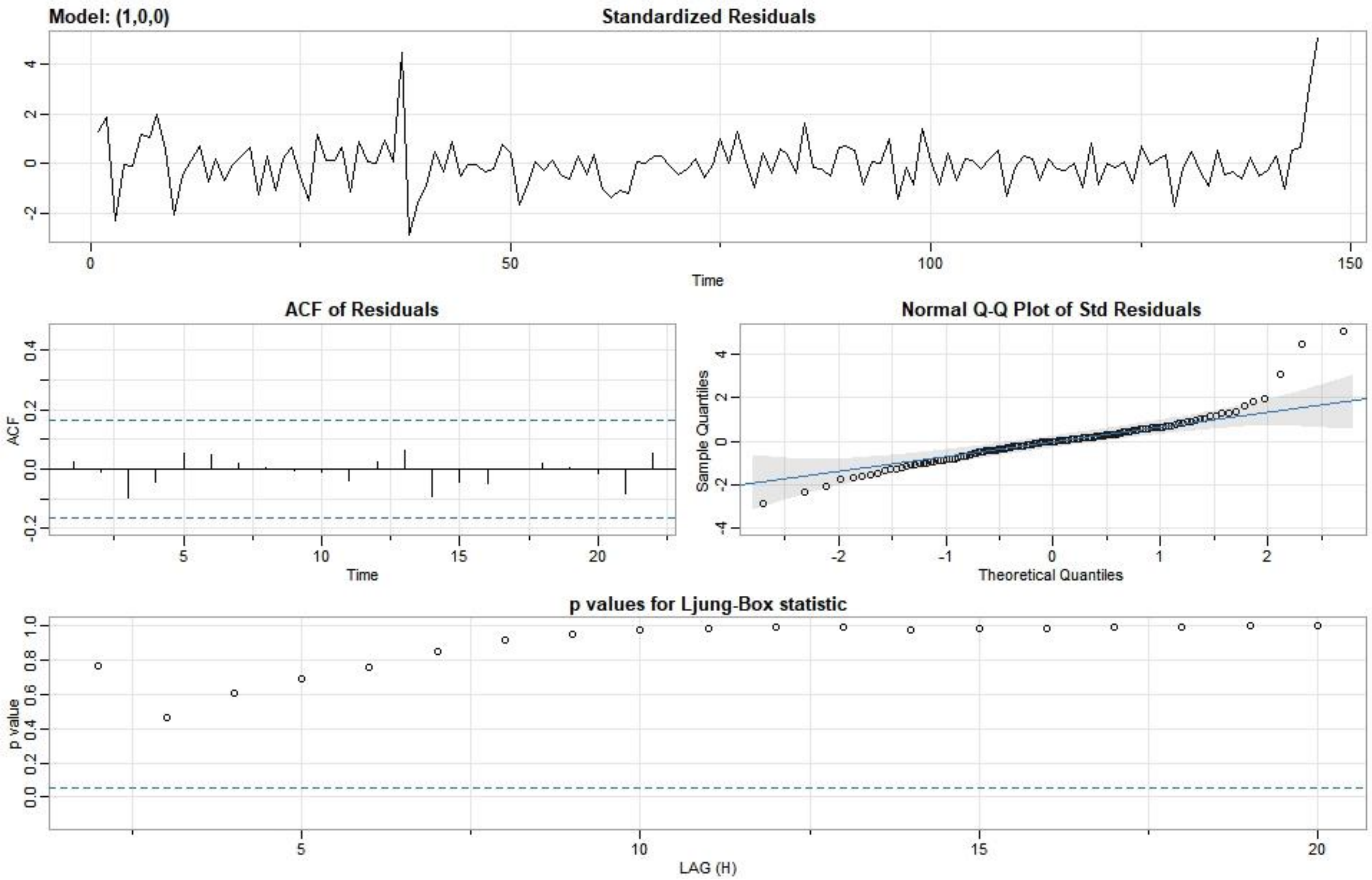
$$\phi = \{\phi_1, \ldots, \phi_p\} \qquad \beta = \{\beta_1, \ldots, \beta_r\}$$

REGRESSÃO COM ERROS AUTOCORRELACIONADOS

❖ Os coeficientes obtidos para as variáveis através do modelo foram:

Variáveis	Estimativa	Erro Padrão	Valor t	Pr(>|t|)
ar1	0.4691	0.1227	3.8246	2.00E-04
Intercept	0.4196	0.2895	1.4497	0.1495
\`Avicultura de Corte\`	0.5589	0.1131	4.9432	0
\`Avicultura de Postura\`	0.0076	0.054	0.14	0.8889
Pescado	-0.1639	0.0963	-1.7023	0.091
Lácteos	-0.1834	0.1454	-1.2614	0.2093
Suinocultura	0.3054	0.2012	1.5182	0.1313
Avicultura de Postura 9	0.1548	0.0492	3.1468	0.002
Pescado 3	0.0282	0.0856	0.329	0.7427
Pescado 10	0.1139	0.0844	1.3496	0.1794
Bovinocultura 1	0.0712	0.1194	0.596	0.5521

REGRESSÃO COM ERROS AUTOCORRELACIONADOS



TESTE DE LJUNG-BOX PARA ARIMA(p,d,q)

- H_0 : Dados são independentemente distribuídos (não tem autocorrelação)
- H_1: Dados não são independentemente distribuídos (tem autocorrelação)
- A estatística de teste é dada por:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

Com

$$Q > \chi^2_{1-\alpha,h}$$

$\hat{\rho}_k$:= autocorrelação amostral no lag k

h := número de lags sendo testados

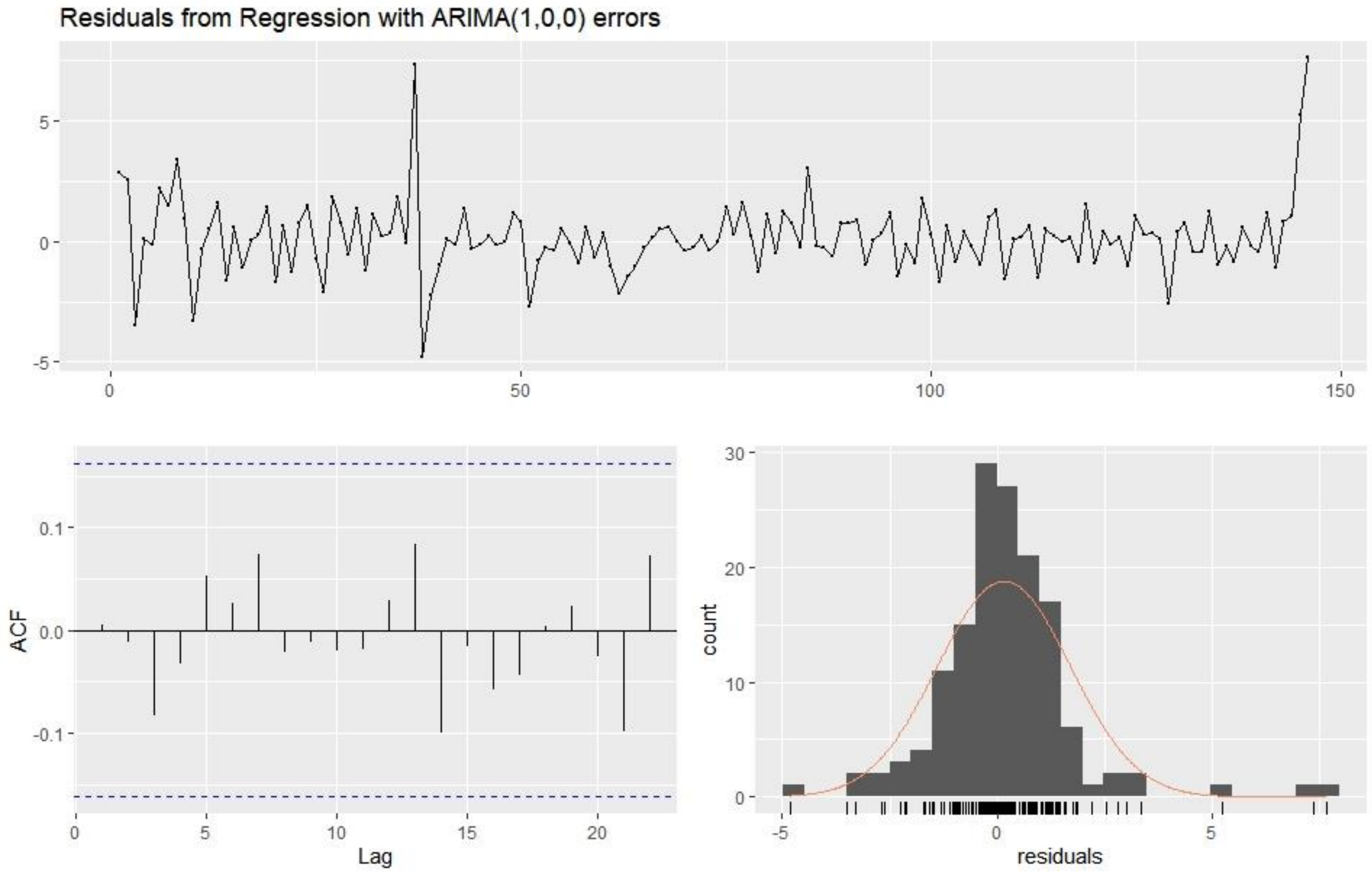
- Aplicando sobre os resíduos, passamos a testar se os erros são não correlacionados (são ruído branco). Aqui, os g.l são corrigidos para h – p – q.

REGRESSÃO COM ERROS AUTOCORRELACIONADOS

- O modelo final fica com forma:

<table>
<tr><th>Variáveis</th><th>Estimativa</th><th>Erro Padrão</th><th>Valor z</th><th>Pr(>|z|)</th></tr>
<tr><td>ar1</td><td>0,526516959</td><td>0,094723828</td><td>5,558442603</td><td>2,72192E-08</td></tr>
<tr><td>`Avicultura de Corte`</td><td>0,541743138</td><td>0,106565796</td><td>5,083649347</td><td>3,70251E-07</td></tr>
<tr><td>`Suinocultura`</td><td>0,422422157</td><td>0,194655795</td><td>2,170098032</td><td>0,02999942</td></tr>
<tr><td>`Avicultura de Postura 9`</td><td>0,133101239</td><td>0,047602027</td><td>2,796125466</td><td>0,005171932</td></tr>
</table>

REGRESSÃO COM ERROS AUTOCORRELACIONADOS



Teste Ljung-Box		
Q* = 2.7751	df = 6	p-value = 0.8365

REGRESSÃO COM ERROS AUTOCORRELACIONADOS

- Temos então para o modelo final:

$$\underbrace{\phi(B)y_t}_{y_t^*} = \sum_{j=1}^{r} \beta_j \underbrace{\phi(B)z_{tj}}_{z_{tj}^*} + \underbrace{\phi(B)x_t}_{w_t},$$

Com: $\phi(B) = 1 - 0.53B$

Sendo: $\beta_1 = \alpha = 0.54 \quad \beta_2 = \beta = 0.42 \quad \beta_3 = \gamma = 0.13$

Temos: $(1 - 0.53B)y_t = \sum \beta_j (1 - 0.53B) z_{tj} + (1 - 0.53B) x_t$

Por fim:

$$y_t = 0.54AVC_t - 0.29AVC_{t-1} + 0.42SUINO_t - 0.22SUINO_{t-1} + 0.13AVP_{t-9} - 0.07AVP_{t-10} + 0.53y_{t-1} + w_t$$

$$w_t \sim RB(0, \sigma_\epsilon^2)$$

REGRESSÃO COM ERROS AUTOCORRELACIONADOS

- Passo a passo para regressão com erros autocorrelacionados:

1. Primeiramente, rodar a regressão ordinária (clássica) para y_t com $z_{t1}, \ldots, z_{tr}$ (considerando os erros como não correlacionados);
2. Guardar os resíduos $\hat{x}_t = y_t - \sum_{j=1}^{r} \hat{\beta}_j z_{tj}$. ;
3. Identificar o(s) modelo(s) ARMA para os resíduos $\hat{x}_t$;
4. Aplicar o método dos mínimos quadrados ponderados no modelo de regressão com erros autocorrelacionados usando o modelo especificado na etapa 3. ;
5. Inspecionar os resíduos $\hat{w}_t$ para verificar se os erros formados são do tipo ruído branco e caso necessário, ajustar um novo modelo.