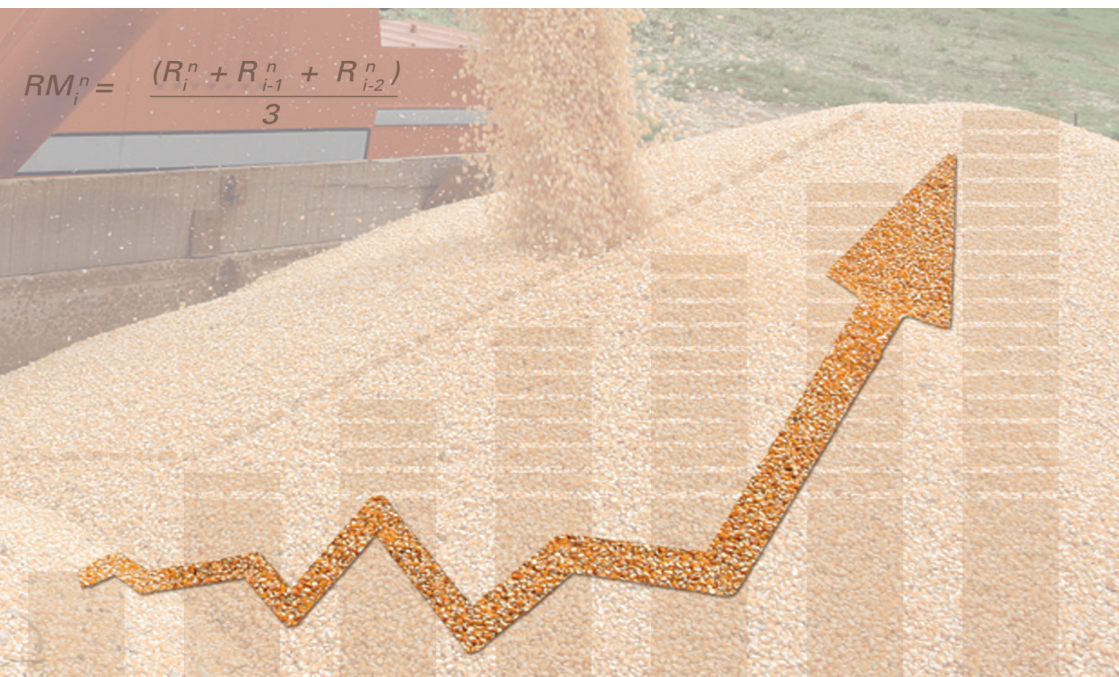


## Metodologia de análise crítica de dados estatísticos históricos sobre produção agropecuária



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Gestão Territorial  
Ministério da Agricultura, Pecuária e Abastecimento*

# ***Boletim de Pesquisa e Desenvolvimento 4***

Metodologia de análise  
crítica de dados estatísticos  
históricos sobre produção  
agropecuária

Rafael Mingoti  
Wilson Anderson Holler  
Claudio Aparecido Spadotto  
Vinícius André Velozo Lopes  
Mayra Abboudi Brasco

Embrapa Gestão Territorial  
Campinas, SP  
2014

Exemplares desta publicação podem ser adquiridos na:

Embrapa Gestão Territorial  
Av. Soldado Passarinho, 303, Fazenda Chapadão  
CEP 13070-115 Campinas, SP, Brasil  
Fone: +55 (19) 3211-6200  
www.sgte.embrapa.br  
sac.sgte@embrapa.br

Comitê de Publicações da  
Embrapa Gestão Territorial  
Presidente: *Mirian Therezinha Souza da Eira*  
Secretária-Executiva: *Rosângela Galon Arruda*  
Membros: *Alba Chiesse da Silva*  
*Helena Sicoli*  
*Ivan Sérgio Freire de Sousa*  
*Eliane Gonçalves Gomes Assunta*  
*Rosana Hoffman Câmara*  
*Chang das Estrelas Wilches*  
*Marita Féres Cardilo*  
*Otávio Valentim Balsadi*

Supervisão editorial: *Erika do Carmo Lima Ferreira*  
Normalização bibliográfica: *Daniela Maciel*  
Editoração eletrônica: *Daniela Maciel*  
Capa: *Daniela Maciel*

1ª edição  
1ª impressão (2014): versão on-line

Todos os direitos reservados  
A reprodução não-autorizada desta publicação, no todo ou em parte,  
constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)  
Embrapa Gestão Territorial

---

Metodologia de análise crítica de dados estatísticos históricos sobre  
produção agropecuária / Rafael Mingoti [et al.] ...

Campinas: Embrapa Gestão Territorial, 2014.  
26 p. : il. color. – (Boletim de Pesquisa e Desenvolvimento / Embrapa  
Gestão Territorial, ISSN 2317-8779 ; 4).

1. Análise crítica. 2. Dados estatísticos - discrepância. 3. Produção  
agropecuária. I. Mingoti, Rafael. II. Holler, Wilson A. III. Spadotto, Clau-  
dio A. IV. Lopes, Vinícius André Velozo. V. Brasco, Mayra A. VI. Embrapa  
Gestão Territorial. VII. Título. VIII. Série.

CDD 551.41

---

# Sumário

Resumo .....	5
Abstract .....	6
Introdução .....	8
Revisão de Literatura.....	10
Dados discrepantes .....	10
Teste de normalidade e transformação Box-Cox .....	12
Metodologia .....	13
Estudos de caso .....	17
Considerações Finais .....	21
Referências .....	22

# Metodologia de análise crítica de dados estatísticos históricos sobre produção agropecuária

---

Rafael Mingoti<sup>1</sup>

Wilson A. Holler<sup>2</sup>

Claudio A. Spadotto<sup>3</sup>

Vinícius André Velozo Lopes<sup>4</sup>

Mayra Abboudi Brasco<sup>5</sup>

## Resumo

O uso de dados estatísticos sobre agricultura, pecuária e silvicultura de bases oficiais tem aumentado continuamente, inclusive em estudos de suporte a gestão territorial estratégica, entretanto, são diversos os relatos acerca da dúvida quanto à precisão desses dados. Este trabalho tem como objetivo apresentar uma metodologia de análise crítica desses dados. A análise crítica proposta é feita por meio da: normalização dos dados em formato de rendimento por área ou rendimento por animal; análise da variação da média móvel aritmética de três anos do rendimento; determinação de valores discrepantes usando a técnica do box-plot e comparação dos limites

---

<sup>1</sup> Engenheiro-agrônomo, analista da Embrapa Gestão Territorial. E-mail: rafael.mingoti@embrapa.br

<sup>2</sup> Engenheiro cartógrafo, analista da Embrapa Gestão Territorial. E-mail: wilson.holler@embrapa.br

<sup>3</sup> Engenheiro-agrônomo, pesquisador da Embrapa Gestão Territorial. E-mail: claudio.spadotto@embrapa.br

<sup>4</sup> Graduando em Engenharia Agrícola, Unicamp, ex-estagiário da Embrapa Gestão Territorial. E-mail: v094546@gmail.com

<sup>5</sup> Graduanda em Geografia, Unicamp, estagiária da Embrapa Gestão Territorial. E-mail: mayra\_brasco@yahoo.com.br

inferiores e superiores com valores publicados em artigos científicos. Como estudo de caso foram analisados os dados de produção e de área colhida de laranja e trigo do cadastro de Produção Agrícola Municipal (PAM) de 1990 a 2012 e de produção de leite e vacas ordenhadas da Pesquisa Pecuária Municipal (PPM) de 1972 a 2012. Os dados históricos dos produtos agropecuários estudados mostraram-se homogêneos para praticamente toda a série histórica, mas com alguns valores discrepantes, em especial, para os dados de produção de laranja. A metodologia apresentada se mostrou eficaz na análise crítica de dados estatísticos oficiais sobre agricultura, pecuária e silvicultura, reforçando que se deve verificá-los antes de serem utilizados.

**Palavras-chave:** análise crítica, dados estatísticos oficiais, dados discrepantes.

# Methodology of critical analysis of historical data on agriculture, livestock and forestry production

---

## Abstract

*The use of official statistics on agriculture, livestock and forestry has steadily increased including studies to support strategic territorial management; however, there are many reports regarding low accuracy of these data. This work was done aiming to present a methodology of critical analysis of such data. The proposed critical analysis includes: normalization of data format in yield per area or yield per animal; variation analysis of three years arithmetic moving average of yield; determination of outliers using the box-plot technique, comparing the lower and upper bounds with values published in scientific papers. As a case study, production and harvested area data of wheat and orange from the register of the Municipal Agricultural Production (PAM) from 1990 to 2012, and milk production and milked cows data from the Municipal Livestock Survey from 1972 to 2012 were analyzed. Historical data set of agricultural products studied were homogeneous although with some outliers, particularly, to orange production data. The presented methodology was effective in critical analysis of official statistics on agriculture, livestock and forestry, stressing that one should check them before use.*

**Keywords:** *critical analysis, official statistics, outliers.*

## Introdução

A Embrapa Gestão Territorial vem desenvolvendo o Serviço de Análise Espacial para a Tomada de Decisão Estratégica (SAE<sup>1</sup>), que permite analisar a dinâmica espaço-temporal envolvendo as relações entre áreas de produção, áreas com limitações legais de uso, modais de transporte, unidades de armazenamento e possíveis áreas de expansão, facilitando, dessa forma, o direcionamento de ações de transferência de tecnologia, a implementação de barreiras sanitárias e a adoção de estratégias para o combate a doenças, pragas e plantas daninhas que afetam o agronegócio. Possibilita, ainda, estabelecer cenários de evolução territorial da agropecuária, beneficiando produtores, associações, cooperativas, sindicatos, órgãos públicos e empresas privadas do setor.

Estudos de dinâmica espaço-temporal da produção agropecuária vem ganhando cada vez mais importância para a gestão territorial estratégica. Tanto a análise histórica baseada em análises estatísticas clássicas (GARAGORRY; CHAIB FILHO, 2008) como a análise baseada em técnicas de análises espaciais (HOLLER et al, 2013) utilizam dados oficiais do governo federal.

Uma das bases de dados estatísticos oficiais disponível é a do Sistema IBGE de Recuperação Automática (SIDRA) que visa, de maneira geral, facilitar o acesso pela internet de dados relativos a 33 pesquisas, 371 variáveis agregadas, 972 tabelas e 615 milhões de variáveis. Os temas de abrangência, dessa base de dados, envolvem o território nacional, população, índices de preços, emprego, construção civil, indústria, comércio, previsão de safra, agricultura, pecuária, silvicultura, horticultura, extração vegetal, orçamento familiar, contas nacionais, registro civil, serviços e empresas. Na área agropecuária, possibilita a obtenção de dados sobre Produção Agrícola Municipal (PAM), Pesquisa Pecuária Municipal (PPM), Produção da Extração Vegetal e da Silvicultura, Censos Agropecuários, entre outros. Essa base disponibiliza, ainda, séries históricas com dados que podem abranger mais de 40 anos de pesquisa (IBGE, 2013).

---

<sup>1</sup> [www.sgte.embrapa.br/SAE](http://www.sgte.embrapa.br/SAE)



Entretanto, são diversos os relatos acerca da dúvida quanto à precisão dos dados oficiais disponibilizados pelo IBGE, em muitas das esferas para as quais se tem informações. Em trabalhos de levantamento da produção agrícola, em alguns municípios, não existem Comissões Municipais Estatísticas Agropecuárias (COMEAS), que têm grande contribuição na coleta de dados municipais. Essa ausência é motivo de alerta pelo IBGE, pois pode levar à imprecisão dos dados informados, correndo, ainda, o risco de não refletirem a realidade do município. Isso também pode acarretar na falta do repasse real de incentivos e benefícios concedidos à prefeitura e aos produtores (NUNES, 2013). Em outros trabalhos, tais como o levantamento populacional e o levantamento do total pago em tributação por empresas, as discrepâncias comparativas entre dados estatísticos e reais é, muitas vezes, evidenciada, e pode ser atribuída a causas intrínsecas à metodologia de obtenção de dados, ou à concessão de informação errônea pelo informante (YOKOTA, 2012; SANTOS, 2011).

Outras fontes de dados estatísticos oficiais são: Companhia Nacional de Abastecimento (CONAB<sup>2</sup>) ; FAOSTAT – sítio da divisão estatística da Food and Agriculture Organization (FAO<sup>3</sup>); EUROSTAT - serviço de estatística da União Europeia<sup>4</sup>; NASS – serviço nacional de estatísticas agrícolas dos Estados Unidos da América<sup>5</sup>.

A utilização dos dados oficiais tem diversas finalidades, entre elas, as análises relacionadas ao desempenho da produção agropecuária. A presença de dados incorretos nessas bases pode acarretar na produção de informações, total ou parcialmente, equivocadas. Com isso, torna-se necessária a avaliação cuidadosa dos dados, antes da geração de informações a seu respeito.

Em relação à comprovação da acurácia, ou correlação dos dados de bases oficiais com a verdade de campo, encontra-se grandes dificuldades. Por exemplo, se fosse necessário verificar a veracidade da informação disponível sobre a quantidade produzida no país,

---

<sup>2</sup> <http://www.conab.gov.br/conteudos.php?a=1252&t=2>

<sup>3</sup> <http://faostat.fao.org/>

<sup>4</sup> <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>

<sup>5</sup> <http://www.nass.usda.gov/>

de um determinado produto agropecuário, ter-se-ia que compilar a quantidade consumida (ou adquirida) desse produto em todos os seus possíveis destinos, como por exemplo: processamento industrial; exportação; armazenagem; comércio varejista etc.; o que é uma atividade de difícil execução. Entretanto, é possível analisar a qualidade dos dados, ou melhor, verificar se há valores discrepantes nesses dados, ou por meio de técnicas estatísticas que avaliem a homogeneidade dos dados, ou pela comparação desses com parâmetros específicos encontrados na literatura técnico-científica.

Devido à abrangência e à importância das estatísticas de produtos agropecuários de bases oficiais, esse trabalho apresenta a metodologia de análise crítica desses dados desenvolvida para o SAE.

## Revisão de Literatura

### Dados Discrepantes

Valores discrepantes ou *outliers* são dados que se destacam como notoriamente divergentes dos demais da amostra. Sua ocorrência tem origem ligada à variabilidade natural dos dados amostrais ou está associada a erros em procedimentos experimentais, captação dos dados ou até mesmo no momento de repassar esses dados a uma planilha (GRUBBS, 1969).

A identificação da ocorrência do dado discrepante pode ser realizada utilizando-se testes estatísticos específicos que, além da identificação, auxiliam na tomada de decisão quanto ao mais correto tratamento a ser realizado sobre os dados, seja retirando-os da amostra ou mantendo-os. Essa tomada de decisão tem origem inerente à natureza do dado discrepante que pode ser devido à variabilidade natural ou a erros. Destaca-se que, muitas vezes, os testes estatísticos, ou geram uma primeira perspectiva da ocorrência de dados discrepantes, ou dão suporte a uma constatação manual prévia. Na maioria das vezes esses mesmos testes assumem que os dados avaliados são parte de uma distribuição normal, a qual, apesar de ser muito utilizada como

base para análise e obtenção de dados discrepantes, assim como, a variância constante e a independência das observações (BOX; COX, 1964), pode ser substituída por outras distribuições unimodais simétricas com função densidade positiva (BEN-GAL, 2005; GRUBBS, 1969).

Dentre as técnicas para amostras univariadas, pode-se citar o grupo de técnicas denominadas Univariate Robust Measure, aplicadas a casos em que a média ou a variância da amostra não se mostram como bons parâmetros para avaliação da localização e forma dos dados, devido à existência de dados discrepantes (BEN-GAL, 2005). Nesse grupo de técnicas há a do box-plot, que foi inicialmente concebida para utilização na visualização da distribuição dos dados. Atualmente, é bastante utilizada, tendo como princípio a apresentação da distribuição de dados no formato “5 números”: limite inferior (LI) (Equação 1), primeiro quartil (Q1), mediana (Q2), terceiro quartil (Q3) e limite superior (LS) (Equação 2) (McGILL et al., 1978; BEN-GAL, 2005).

$$LI = Q1 - 1,5(Q3-Q1) \quad | \quad (1)$$

$$LS = Q3 + 1,5(Q3-Q1) \quad | \quad (2)$$

O método box-plot foi atribuído de maneira consolidada em alguns trabalhos e obteve, para os mesmos, resultados satisfatórios. Sua apresentação tem algumas fraquezas intrínsecas e possibilita em alguns momentos interpretações errôneas do comportamento amostral (McGILL et al., 1978).

A geração de gráficos do tipo box-plot (Figura 1) também propicia uma perspectiva visual da presença ou não de dados discrepantes, uma vez que dados discrepantes são encontrados acima do limite superior e abaixo do inferior.

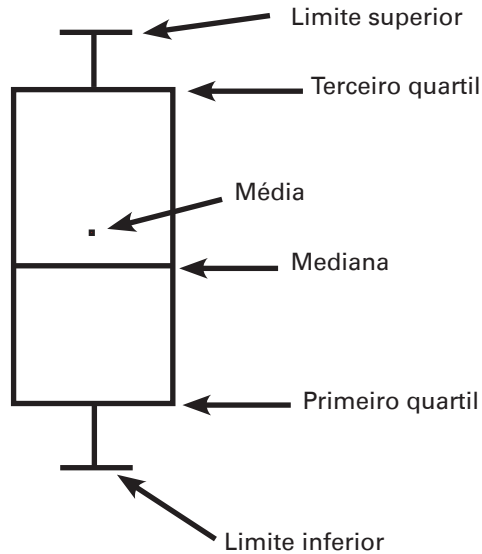


Figura 1. Representação esquemática de um gráfico do tipo box-plot.

Fonte: Elaborado pelos autores.

## Teste de normalidade e transformação Box-Cox

Existem inúmeros testes para verificação da ocorrência ou não do padrão de normalidade da amostra, como Anderson-Darling (ANDERSON; DARLING, 1952) e Shapiro-Wilk (SHAPIRO; WILK, 1965). A metodologia básica dos testes é o estabelecimento de uma curva com característica normal e posteriormente a verificação se há ou não correlação da curva normal estimada com a curva obtida a partir dos dados amostrais (GARSON, 2012).

A transformação Box-Cox é realizada visando gerar distribuição normal em dados que não apresentam essa característica, sem torná-los inadequados para posteriores análises (BOX; COX, 1964). A implementação dessa transformação se torna necessária na determinação de dados discrepantes, uma vez que, a maioria dos testes utilizados avalia dados assumindo que os mesmos fazem parte de uma distribuição normal.

## Metodologia

A metodologia de análise crítica de dados estatísticos de produtos agropecuários de bases oficiais, proposta nesse trabalho, está apresentada a seguir:

A. Obter os dados com a escala temporal desejada (por exemplo: anuais) e para a unidade territorial de interesse (por exemplo: municípios);

A.1. Verificar se houve alteração na unidade do dado informado; por exemplo, no caso da laranja e outras frutas, que tiveram a unidade de quantidade produzida alterada de mil frutos para toneladas, a partir de 2001 na base SIDRA<sup>6</sup>.

A.2. Os dados precisam estar normalizados, ou seja:

A.2.1. No caso de produção agrícola, precisam estar no formato de rendimento por área (unidade de massa por unidade de área; no sistema internacional de unidades (SI):  $\text{kg.m}^{-2}$ , ou  $\text{kg.ha}^{-1}$ );

A.2.2. No caso de produção pecuária, podem estar no formato de rendimento por área (como no caso anterior) ou podem estar no formato de rendimento por animal (unidade de massa por animal (por exemplo: produção de leite, lã, carne e ovos); no SI:  $\text{kg.animal}^{-1}$ );

B. Eliminar os dados que apresentem variação temporal abrupta. Para isso deve-se:

B.1. Calcular a média móvel aritmética, apresentada na Equação 3, adotando-se o intervalo de três anos;

$$RM_i^n = \frac{(R_i^n + R_{i-1}^n + R_{i-2}^n)}{3} \quad | \quad (3)$$

<sup>6</sup> Informação disponível em: < <http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=1613&z=p&o=26> >.

Em que:  $RM_i^n$  é o rendimento médio, em três anos, ( $\text{kg} \cdot \text{ha}^{-1}$  ou  $\text{kg} \cdot \text{animal}^{-1}$ ) do produto agropecuário em análise para a enésima unidade territorial e para a  $i$ -ésima unidade temporal; e  $R_{i-1}^n$  é o rendimento ( $\text{kg} \cdot \text{ha}^{-1}$  ou  $\text{kg} \cdot \text{animal}^{-1}$ ) do produto agropecuário em análise para a enésima unidade territorial e para a  $i$ -ésima unidade temporal;

B.2. Calcular a variação temporal dos dados pela Equação 4;

$$\Delta RM_i^n = \frac{(RM_i^n - RM_{i-1}^n)}{RM_{i-1}^n} 100 \quad | \quad (4)$$

Em que:  $\Delta RM_i^n$  é a variação (%) do rendimento médio do produto agropecuário em análise para a enésima unidade territorial e para a  $i$ -ésima unidade temporal;

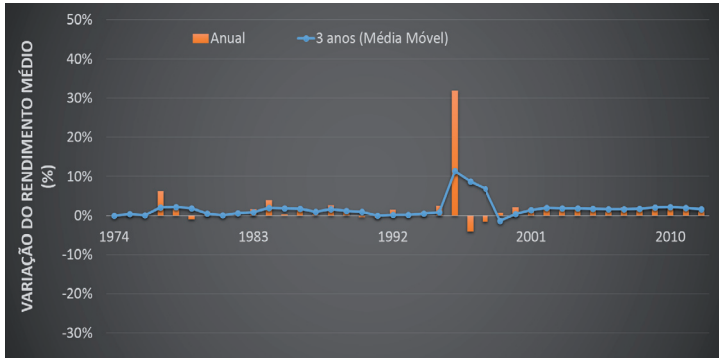
B.3. Deve-se eliminar os dados que apresentarem variação temporal acima de um valor de variação considerado possível;

B.3.1. A obtenção desse valor deve-se basear em revisão de literatura específica;

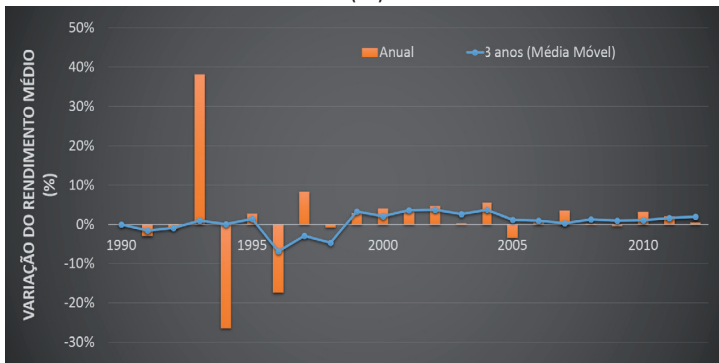
B.3.2. Para os casos em que essa informação não estiver disponível, sugere-se a utilização do valor de variação de rendimento máximo de 100% ao ano.

A média móvel constata as mudanças de tendência e acompanha o seu desenvolvimento indicando tendência de aumento, diminuição ou estagnação da produção. Essa média suaviza as oscilações na produção de modo a identificar somente grandes variações.

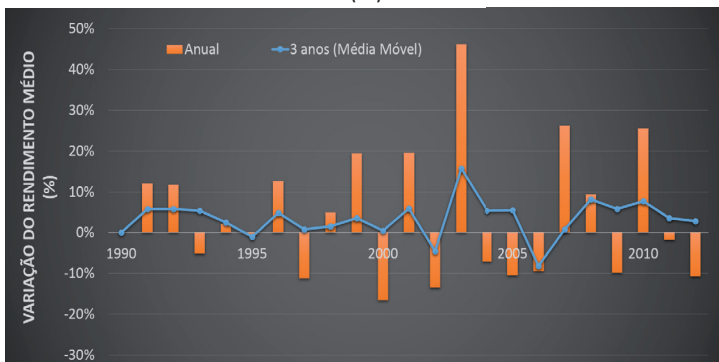
Na Figura 2, é apresentada uma comparação da variação média anual do rendimento de três produtos agropecuários.



(A)



(B)



(C)

Figura 2. Comparação da variação média anual do rendimento de três produtos agropecuários, com o cálculo anual ou com o cálculo por meio da média móvel aritmética, sendo: A = leite; B = laranja; e C = trigo.

Fonte: Elaborado pelos autores.

Como pode-se verificar na Figura 2, a utilização do cálculo por meio da média móvel aritmética promove a eliminação de valores extremos. Se fosse utilizado o rendimento anual, ao invés da média móvel aritmética de três anos desse rendimento, detectar-se-ia uma quantidade maior de grandes variações no rendimento médio do produto agropecuário, pois é comum o fato de uma região ter, em um ano, o rendimento médio de um ou mais produtos agropecuários diminuído por variações climáticas (geadas ou secas, por exemplo) ou ocorrências de doenças e pragas e, no ano seguinte, ter um rendimento normal.

### C. Remover os dados de produção igual a zero;

C.1. Realizar, nos dados filtrados (com os critérios anteriores), o teste de aderência de Anderson-Darling (ANDERSON; DARLING, 1952) ao ajuste à curva de distribuição normal;

C.2. No caso dos dados não se ajustarem à distribuição normal, aplicar o teste de Box-Cox (BOX; COX, 1964), de modo a verificar qual a melhor transformação dos dados.

### D. Elaborar, para os dados filtrados (sem valores zero e sem os de alta variação), o gráfico de box-plot e determinar os valores limites para dados discrepantes mínimos (Equação 1) e máximos (Equação 2);

D.1. Eliminar os dados que se caracterizarem como sendo discrepantes mínimos;

D.2. Em relação aos dados que se caracterizarem como discrepantes máximos, deve-se confirmar se os valores realmente o são, utilizando, por base, critérios e índices publicados em literatura científica especializada. Para isso deve-se:

D.2.1. Revisar, em literatura técnico-científica, critérios que permitam estabelecer os valores máximos de rendimento que o produto agropecuário em questão pode alcançar, considerando a escala temporal dos dados e, principalmente, a unidade territorial adotada.

D.2.2. De posse desses critérios, eliminar os dados que não atendam ao limite estabelecido.



## Estudos de caso

A metodologia apresentada foi aplicada em três estudos de caso, para os seguintes produtos agropecuários: leite, laranja e trigo. Para isso, foram utilizados os dados do cadastro de Produção Agrícola Municipal (PAM) para laranja e trigo, de 1990 a 2012, e do cadastro de Pesquisa Pecuária Municipal (PPM) para leite, de 1974 a 2012, disponíveis no SIDRA. Escolheu-se esses produtos dentre os existentes no SIDRA que pudessem representar a atividade da pecuária, o cultivo de culturas anuais e de culturas permanentes.

O processamento dos dados foi realizado utilizando os programas R e Microsoft Excel® com o suplemento *Action* (permite a realização de análises estatísticas no programa R de maneira integrada – com a interface - ao Microsoft Excel®).

Os dados dos três produtos agropecuários analisados não apresentaram aderência à distribuição normal, segundo o teste de Anderson-Darling. Após a aplicação do teste de Box-Cox, todos eles foram transformados, utilizando-se o logaritmo neperiano.

Após a eliminação dos valores zero e dos valores correspondentes à alta variação (>100%) do rendimento, considerando a média móvel de três anos, foram elaborados gráficos box-plot para os produtos em estudo (Figura 3).

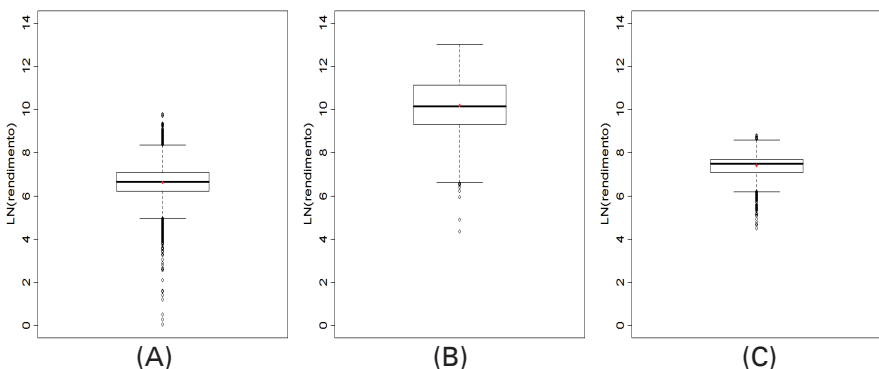


Figura 3. Gráfico box-plot dos produtos agropecuários estudados, sendo: A = leite; B = laranja; e C = trigo.

Fonte: Elaborado pelos autores.

Após a elaboração dos gráficos box-plot, foram obtidos os valores limítrofes para dados discrepantes máximos e mínimos. Em relação aos valores para dados discrepantes máximos, de modo a verificar se o valor obtido é considerado discrepante pela literatura científica, realizou-se uma revisão bibliográfica sobre o assunto para cada produto agropecuário estudado, encontrando-se os seguintes valores:

- 12.508,3 L.vaca<sup>-1</sup>.ano<sup>-1</sup> para o leite, que por sua vez foi proveniente dos seguintes critérios:
  - Produtividade de 42,2 kg.vaca<sup>-1</sup>.dia<sup>-1</sup> (BATH, 1985);
  - Período de produção: 305 dias.ano<sup>-1</sup> (MÜHLBACH, 2003);
  - Densidade do leite: 1,029 Kg.L<sup>-1</sup> (GONZÁLES et al., 2001);
- 69.000,0 kg.ha<sup>-1</sup> para a laranja, proveniente dos seguintes critérios:
  - Produtividade de 69.000,0 kg.ha<sup>-1</sup> (DUENHAS et al., 2005);
- 6.600,0 kg.ha<sup>-1</sup> para o trigo, proveniente dos seguintes critérios:
  - Produtividade média de 6.600,0 kg.ha<sup>-1</sup> para a cultivar BRS 331 (SCHEEREN et al., 2012).

Os valores limítrofes para dados discrepantes máximos, obtidos pelo método do box-plot, foram 1.881,0 L.vaca<sup>-1</sup>.ano<sup>-1</sup>, 23.338,0 kg.ha<sup>-1</sup> e 2.979,0 kg.ha<sup>-1</sup>, respectivamente, para leite, laranja e trigo. Esses valores foram menores do que os encontrados na literatura (apresentados acima), com isso, considerou-se os limites da literatura científica. Os valores limites adotados para mínimos e máximos discrepantes estão apresentados na Tabela 1.

Tabela 1 – Valores mínimos e máximos de dados discrepantes para os produtos agropecuários estudados.

Produto	Limite para dados mínimos discrepantes	Limite para dados máximos discrepantes
leite	140,0 L.vaca <sup>-1</sup> .ano <sup>-1</sup>	12.508,3 L.vaca <sup>-1</sup> .ano <sup>-1</sup>
laranja	724,0 kg.ha <sup>-1</sup>	69.000,0 kg.ha <sup>-1</sup>
trigo	483,0 kg.ha <sup>-1</sup>	6.600,0 kg.ha <sup>-1</sup>

Fonte: Elaborado pelos autores.

Considerando-se todos os critérios adotados para a análise dos dados dos produtos agropecuários em estudo, calculou-se as proporções de dados excluídos, considerando cada etapa da metodologia de análise crítica (Tabela 2).

Tabela 2 – Proporção (%) de dados excluídos em cada etapa da metodologia proposta.

Produto	Valores zero	Alta variação de rendimento em 3 anos	Mínimos discrepantes	Máximos discrepantes
leite	14,447%	0,107%	0,575%	0,001%
laranja	41,759%	0,273%	1,009%	0,011%
trigo	84,725%	0,016%	0,122%	0,001%

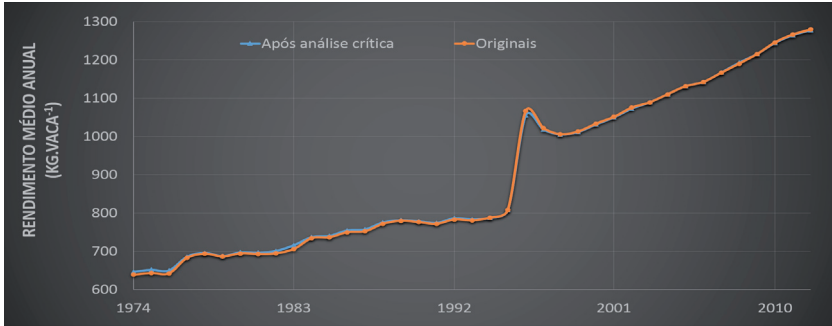
Fonte: Elaborado pelos autores.

Como pode-se observar na Tabela 2, os dados de laranja e trigo apresentam alta quantidade de valores zero, evidenciando a importância de sua remoção para a análise do conjunto de dados. A quantidade de valores zero é, aproximadamente, inversamente proporcional à distribuição da cultura no país, ou seja, ao número de municípios que produzem aquele produto agropecuário.

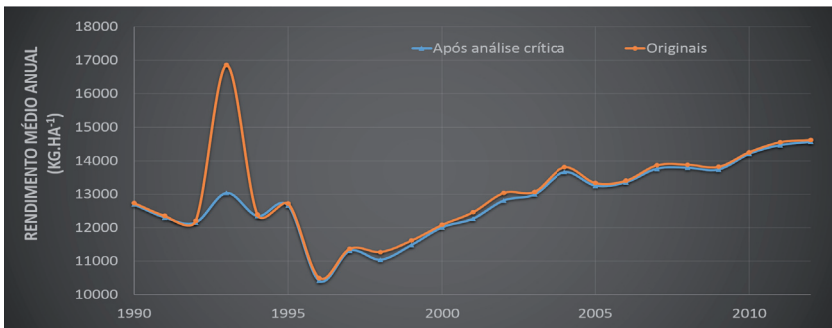
Todos os produtos agropecuários estudados apresentaram poucos dados excluídos, em relação aos três últimos critérios da Tabela 2 (dados com alta variação de rendimento em 3 anos, dados mínimos e máximos discrepantes), evidenciando assim, uma boa homogeneidade dos dados.

Após a aplicação da metodologia de análise crítica dos dados, elaborou-se uma comparação do rendimento médio anual calculado com os dados originais e com os dados filtrados (Figura 4).

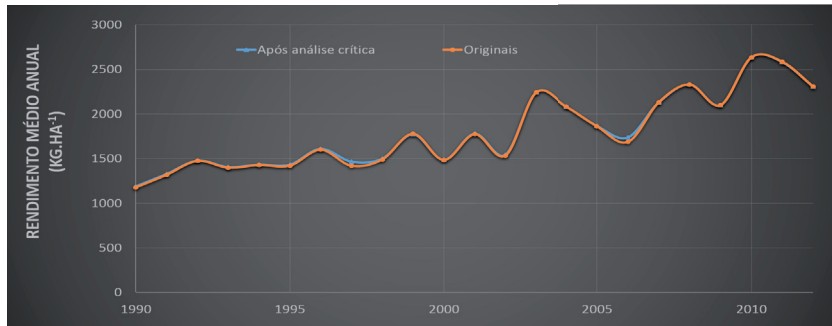
Pelos resultados apresentados na Figura 4, observa-se que, com exceção de 1993 na cultura da laranja, o rendimento médio anual dos produtos estudados foram semelhantes entre o cálculo pelos dados originais e o cálculo com dados filtrados pela análise crítica. Isso corrobora com o resultado observado na Tabela 2, ou seja, a boa homogeneidade dos dados. A diferença do rendimento médio anual da laranja em 1993 evidencia a necessidade da análise crítica dos dados antes da sua utilização.



(A)



(B)



(C)

Figura 4. Comparação do cálculo do rendimento médio anual dos produtos agropecuários estudados, com os dados originais ou com os dados filtrados pela análise crítica, sendo: A = leite; B = laranja; e C = trigo.

Fonte: Elaborado pelos autores.

## **Considerações Finais**

A metodologia apresentada mostrou-se eficaz na análise crítica de dados agropecuários oficiais.

Os dados de bases oficiais referentes aos produtos agropecuários estudados mostraram-se homogêneos. Entretanto, os dados para a produção de laranja apresentaram alguns valores extremos discrepantes sendo, com isso, necessária sua revisão antes da utilização em trabalhos futuros. Isso reforça a importância da análise crítica de dados agropecuários para a geração de informações a partir deles, como é realizada no SAE.

Existem trabalhos publicados que podem complementar a metodologia apresentada, como por exemplo o de Neves e Luiz (2006), que permite a verificação da correlação entre os valores da produção para municípios próximos.

## Referências

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, Washington, v. 23, p. 193–212. 1952.

BATH, D. L. Nutritional requirements and economics of lowering feed costs. *Journal of dairy science*, Chanpaign, v. 68, n. 6, p. 1579-1584, 1985.

BEN-GAL, I. Outlier detection. In: MAIMON, O.; ROCKACH, L. (Ed.). *Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers*. Tel Aviv: Springer, 2005. ISBN 0-387-24435-2.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. *Journal of the Royal Statistical Society, Londres, Series B (Methodological)*, v. 26, n. 2, p. 211-252. 1964.

DUENHAS, L. H.; VILLAS BÔAS, R. L.; SOUZA, C. M. P.; OLIVEIRA, M. V. A. M.; DALRI, A. B. Produção, qualidade dos frutos e estado nutricional da laranja valência sob fertirrigação e adubação convencional. *Engenharia Agrícola*, Jaboticabal, v. 25, n. 1, p.154-160, 2005.

GARAGORRY, F. L.; CHAIB FILHO, H. *Evolução da Agricultura Brasileira em um Tempo Recente: elementos de agrodinâmica*. Brasília: SGE/Embrapa. 2008. Relatório Técnico, versão preliminar.

GARSON, D. *Testing statistical Assumptions*. Asheboro: Statistical Publishing Associates, 2012. Disponível em: <http://www.unm.edu/~marcusj/testingassumptions.pdf>. Acessado em 12/11/2013.

GONZÁLEZ, F. H. D.; DÜRR, J. W.; FONTANELI, R. S. Uso do leite para monitorar a nutrição e o metabolismo de vacas leiteiras. Porto Alegre: [UFRGS], 2001. Disponível em: < <http://www.ufrgs.br/lacvet/restrito/pdf/leite%20metabolismo.pdf> >. Acesso em: 20 mar. 2014.

GRUBBS, F. E. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, Washington, v.11, n.1, p. 1-21. 1969.

HOLLER, W. A.; MINGOTI, R.; SPADOTTO, C. A.; BRASCO, M. A. Com-

paração entre metodologias de representação da dinâmica territorial da produção agropecuária. Campinas: Embrapa Gestão Territorial, 2013. 38 p. (Embrapa Gestão Territorial. Boletim de Pesquisa e Desenvolvimento, 3).

IBGE. Sistema IBGE de Recuperação Automática - SIDRA. Disponível em: < <http://www.ibge.gov.br/home/disseminacao/eventos/workshop/sidra.shtm> >. Acesso em: 12 nov. 2013.

McGILL R.; TUKEY J. W.; LARSEN, W. A. Variations of Box Plots. The American Statistician, Washington, v. 32, n. 1, p. 12-16. 1978.

MÜHLBACH, P. R. F. Produção de leite com vacas de alta produtividade. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 40., 2003, Santa Maria, RS. Otimizando a produção animal: anais. Santa Maria: Sociedade Brasileira de Zootecnia, 2003. 1 CD ROM.

NEVES, M. C.; LUIZ, A. J. B. Distribuição espacial da cultura de café no Estado de São Paulo. Jaguariúna: Embrapa Meio Ambiente, 2006. 22p. (Embrapa Meio Ambiente. Boletim de Pesquisa e Desenvolvimento, 40).

NUNES, D. Menos de 5% dos municípios têm comissões agropecuárias. Diário do nordeste, 14 ago. 2013. Disponível em: < <http://diariodo-nordeste.globo.com/m/materia.asp?codigo=1303927> >. Acesso em: 10 nov. 2013.

SANTOS, J. IBGE divulga levantamento impreciso sobre população de Paraisópolis. Espaço do povo, Paraisópolis, 23 dez. 2011. Disponível em: < <http://jornal.paraisopolis.org/ibge-divulga-levantamento-impreciso-sobre-populacao-de-paraisopolis/> >. Acesso em: 12 nov. 2013.

SCHEEREN, P. L.; CAIERÃO, E.; SILVA, M. S.; CASTRO, R. L.; CARGNIN, A. (Org.). Cultivares de trigo da Embrapa indicadas para cultivo no Brasil 2012/2013. Passo Fundo: Embrapa Trigo, 2012. 72p. Disponível em: < <http://www.cnpt.embrapa.br/culturas/trigo/cultivares/Catalogo%20TRIGO%202012-completo.pdf> >. Acesso em: 10 nov. 2013.

SHAPIRO, S. S.; WILK, M. An analysis of variance test for normality (complete samples). Biometrika, Londres, v. 52, p. 591-611. 1965.

YOKOTA, P. Imprecisão dos dados estatísticos chineses e brasileiros. Disponível em: < <http://www.asiacomentada.com.br/2012/07/impreciso-dos-dados-estatsticos-chineses-e-brasileiros/> >. Acesso em: 12 nov. 2013.



**Embrapa**

---

**Gestão Territorial**

Ministério da  
**Agricultura, Pecuária  
e Abastecimento**

GOVERNO FEDERAL  
**BRASIL**  
PAÍS RICO E PAÍS SEM POBREZA