

Métodos Estatísticos em Pesquisa Científica

Abertura

Prof. Paulo Justiniano Ribeiro Jr

Departamento de Estatística
Setor de Ciências Exatas
Transversais - PRPPG
Universidade Federal do Paraná

20 de março de 2024



Equipe e agradecimentos - UFPR

- ▶ PRPPG
- ▶ Equipe Transversais: Mirian, Larissa e Luiz.
- ▶ Prof. André Rodacki.
- ▶ Demais estruturas de apoio da UFPR.
- ▶ Coordenação da monitoria.
- ▶ Monitoria-MEPC: Bianca, Ana Carolina, Lis.
- ▶ Professores convidados.
- ▶ Departamento de Estatística, docentes e técnicos.
- ▶ Equipe do LEG/DEST/UFPR.

Equipe e agradecimentos - Instituições parceiras

Boas vindas à UFPR e agradecimentos adicionais.

Participantes de instituições públicas do Paraná.

Representantes Institucionais

- ▶ Hellen Geremias dos Santos, ICC-Fiocruz, Curitiba.
- ▶ Everaldo dos Santos, IFPR,
- ▶ Diogo Francisco Rossoni, UEM, Maringá.
- ▶ Deyse Marcia Pacheco Gerbert, UEPG, Ponta Grossa.
- ▶ Peter Löwenberg Neto, UNILA, Foz do Iguaçu.
- ▶ Lirane Elize Defante Ferreto, Unioste.
- ▶ Lucas Pioli Rehbein Kürten Ihlenfeld, UTFPR.

Representantes institucionais:

- ▶ Contato para participantes de instituições parceiras

Coordenação/administração:

- ▶ Administrativa: matrículas, acesso ao sistema e similares:
Equipe das Transversais (transversal@ufpr.br)

Apoio acadêmico:

- ▶ Acadêmica/conteúdo: monitores e professores
Equipe MEPC (mepc@ufpr.br).

Para organização, rastreamento e efetividade nas respostas:

NÃO enviar emails relacionados ao curso para endereços pessoais dos professores e monitores do curso.

Participantes

400+ inscritos

- ▶ PG's UFPR,
- ▶ PG's instituições parceiras,
- ▶ Professores, técnicos e pesquisadores,
- ▶ Outros.

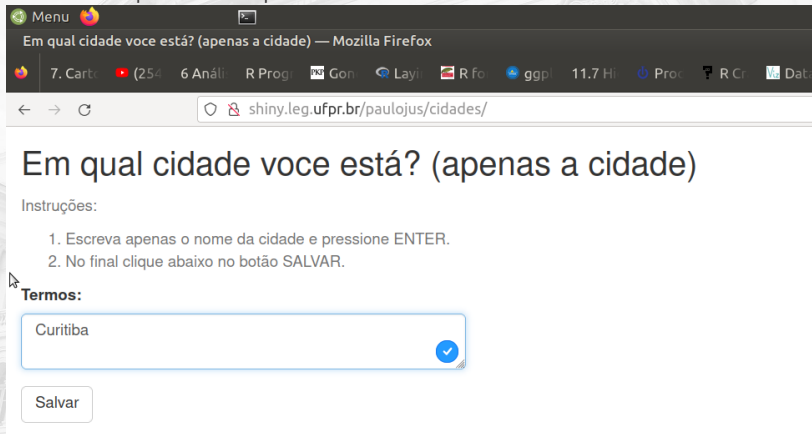
Onde estão?

Acessar

<http://shiny.leg.ufpr.br/paulojus/cidades>

e digitar o nome da cidade onde está.

Digite com maiúscula apenas no primeiro caractere.



The screenshot shows a Mozilla Firefox browser window. The address bar displays the URL `shiny.leg.ufpr.br/paulojus/cidades/`. The page title is "Em qual cidade voce está? (apenas a cidade)". Below the title, the instructions are listed: "1. Escreva apenas o nome da cidade e pressione ENTER." and "2. No final clique abaixo no botão SALVAR." A text input field contains the word "Curitiba", and a blue checkmark icon is visible to its right. At the bottom, there is a button labeled "Salvar".

Menu

Em qual cidade voce está? (apenas a cidade) — Mozilla Firefox

7. Carto (254) 6 Análi R Prog Gon Layi R fo ggpi 11.7 Hi Proc R Cr Data

← → ↻

shiny.leg.ufpr.br/paulojus/cidades/

Em qual cidade voce está? (apenas a cidade)

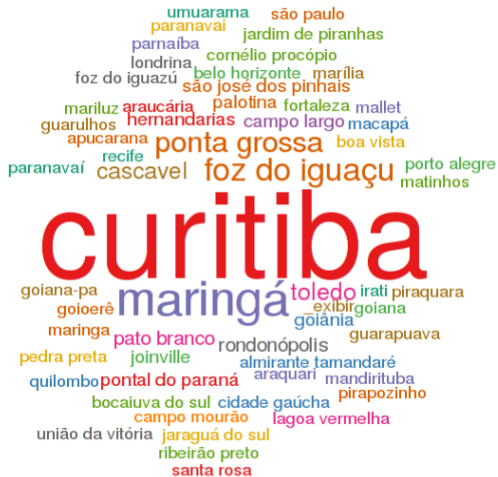
Instruções:

1. Escreva apenas o nome da cidade e pressione ENTER.
2. No final clique abaixo no botão SALVAR.

Termos:

Curitiba

Salvar



umarama são paulo
paranavai jardim de piranhas
parnaíba cornélio procópio
londrina belo horizonte marília
foz do iguaçu são josé dos pinhais
mariluz araucária palotina fortaleza mallet
guarulhos hermandarias campo largo macapá
apucarana ponta grossa boa vista
paranavai recife cascavel foz do iguaçu porto alegre
matinhos

curitiba

goiana-pa maringá toledo irati piraquara
goioerê goiânia exibir goiana
maringa pato branco guarapuava
pedra preta joinville rondonópolis
quilombo pontal do paraná almirante tamandaré
araguari mandirituba
bocaiuva do sul cidade gaúcha pirapozinho
campo mourão lagoa vermelha
união da vitória jaraguá do sul
ribeirão preto
santa rosa

Começando ...do começo

Estatística?

OK, estatística é *importante* ...mas ...
Que palavras/ideias voce associa com estatística?

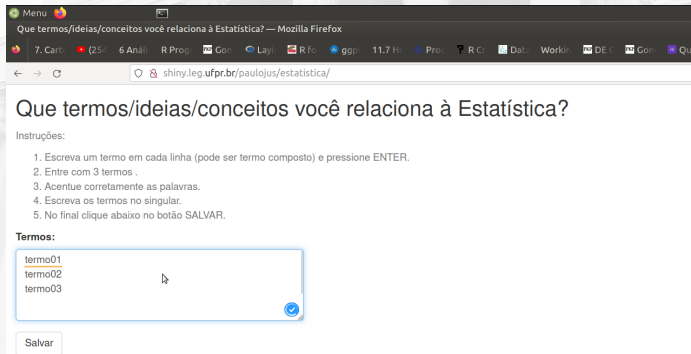
Estatística ?

Acessar:

<http://shiny.leg.ufpr.br/paulojus/estatistica>

e digite na interface até 3 termos (palavras-chave) que voce associa com estatística.

- ▶ Apenas um termo (palavra chave) por linha.
- ▶ Usar apenas letras minúsculas.
- ▶ Evite plurais, procurar usar apenas termos no singular.



The screenshot shows a web browser window with the URL `shiny.leg.ufpr.br/paulojus/estatistica/`. The page title is "Que termos/ideias/conceitos você relaciona à Estatística?". Below the title, there are instructions in Portuguese:

Instruções:

1. Escreva um termo em cada linha (pode ser termo composto) e pressione ENTER.
2. Entre com 3 termos .
3. Acentue corretamente as palavras.
4. Escreva os termos no singular.
5. No final clique abaixo no botão SALVAR.

Below the instructions, there is a section labeled "Termos:" with a text input area containing three lines of text:

```
termo01
termo02
termo03
```

A mouse cursor is positioned over the input area. At the bottom right of the input area, there is a blue checkmark icon. Below the input area, there is a button labeled "Salvar".

Temas nas respostas

Padrões? Recorrência?

Padrões? Recorrência?

- ▶ Dados I : obtenção, amostragem, pesquisa, volume, ...
- ▶ Dados II: descrição, resumos, gráficos, análises, ...
- ▶ Probabilidades.
- ▶ Inferência: incerteza, população, amostra, testes, intervalos, ...
- ▶ Modelagem e métodos.

Padrões? Recorrência?

- ▶ Dados I : obtenção, amostragem, pesquisa, volume, ...
- ▶ Dados II: descrição, resumos, gráficos, análises, ...
- ▶ Probabilidades.
- ▶ Inferência: incerteza, população, amostra, testes, intervalos, ...
- ▶ Modelagem e métodos.
- ▶ **Interpretação.**

De onde vêm os dados?

- ▶ questionários,
- ▶ amostragens,
- ▶ experimentos,
- ▶ observacionais,
- ▶ repositórios,
- ▶ bancos de dados,
- ▶ web (processamento, *web scrapping*),
- ▶ *logs*, sensores,
- ▶ ...

De onde vêm os dados?

- ▶ questionários,
- ▶ amostragens,
- ▶ experimentos,
- ▶ observacionais,
- ▶ repositórios,
- ▶ bancos de dados,
- ▶ web (processamento, *web scrapping*),
- ▶ *logs*, sensores,
- ▶ ...

Questão fundamental e nem sempre trivial:
Qual universo os dados representam?

Statistics is the *science* of collecting and interpreting data.

...

Dealing with uncertainty is the cornerstone of the statistical method.

(DIGGLE; CHETWYND, 2011)

*Statistics is the **science** of collecting and interpreting data.*

...

Dealing with uncertainty is the cornerstone of the statistical method.

(DIGGLE; CHETWYND, 2011)

- ▶ Relevante em praticamente todas áreas do conhecimento.
- ▶ Dados e geral envolvem imprecisão e incerteza.
- ▶ Métodos: matemático (dedutivo) vs estatístico (inferencial).

Programa

1. Fundamentos
 - 1.1 Estatística descritiva/exploratória
 - 1.2 Probabilidades
 - 1.3 Princípios de inferência e paradigmas
2. Métodos, modelos e extras

Conteúdo: cinco ciclos de três semanas cada.

- ▶ **Ciclo 1:** O que os dados podem contar? (Estatística descritiva, relações bi e multivariadas).
- ▶ **Ciclo 2:** Padrões no acaso. (Probabilidades, variáveis aleatórias, distribuições).
- ▶ **Ciclo 3:** Aprendendo com dados. (Inferência estatística).
- ▶ **Ciclo 4:** Introdução a modelagem estatística.
- ▶ **Ciclo 5:** Tópicos em métodos estatísticos.

Avaliação (Moodle) na última semana de cada ciclo

- ▶ Materiais pré-aula em alguns das semanas.
- ▶ Aulas remotas ao vivo às quartas (1a e 2a semana dos ciclos) ou gravadas (3a semana do ciclo).
 - ▶ Parte 1: antes do 1º intervalo: conceitual.
 - ▶ Parte 2 (*): antes 2º intervalo: conceitual.
 - ▶ Parte 3 (*): após 2º intervalo: tópicos em **R** (gravada ou ao vivo).
 - ▶ Controle de frequência online (durante as aulas - horários aleatórios).
 - ▶ CHAT (Moodle) durante a aula, perguntas/comentários, via monitores.
- ▶ Forum aberto para cada aula: forum geral, da semana e email.
- ▶ Disponibilização de material e atividades recomendadas pelo Moodle.
- ▶ Disponibilização de vídeo da aula (link pelo Moodle).
- ▶ Atividades não avaliativas: disponibilizadas a cada semana pelo Moodle.
- ▶ Avaliação: 3a semana de cada ciclo.

Tópicos de edições anteriores

- ▶ Amostragem.
- ▶ Visualização de dados.
- ▶ Tópicos em estatística e computação.
- ▶ Tópicos em aprendizado de máquina e *big data*.
- ▶ Pesquisa reproduzível.
- ▶ Planejamento de experimentos.
- ▶ Regressão.
- ▶ Extensões de regressão.
- ▶ Árvores de regressão e classificação.
- ▶ Dados Longitudinais.
- ▶ Métodos Multivariados.
- ▶ Análise de sobrevivência.
- ▶ Teoria de avaliação.
- ▶ Análise de imagens.
- ▶ Dados categóricos.
- ▶ Dados ecológicos.
- ▶ Modelos espaciais.
- ▶ Misturas.
- ▶ ...

Técnicas vs ideias: o propósito do curso

More important than learning a few methods and techniques is to understand the statistical thinking.

(BOX; HUNTER; HUNTER, 2005)

Técnicas vs ideias: o propósito do curso

More important than learning a few methods and techniques is to understand the statistical thinking.

(BOX; HUNTER; HUNTER, 2005)

- ▶ Este **NÃO** é um curso orientado a técnicas.
- ▶ Evitamos abordagem *receita de bolo*.
- ▶ Ênfase nos conceitos, no método estatístico, abordagens e paradigmas.
- ▶ Exemplos motivam a exposição e discussões das ideias centrais.
- ▶ Análise crítica e interpretações no contexto de exemplos e métodos.

Técnicas vs ideias: o propósito do curso

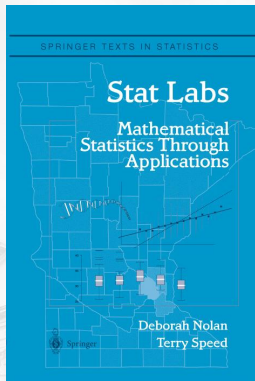


Figura 1. Nolan & Speed, 2000



Figura 2. Diggle & Chetwynd, 2011

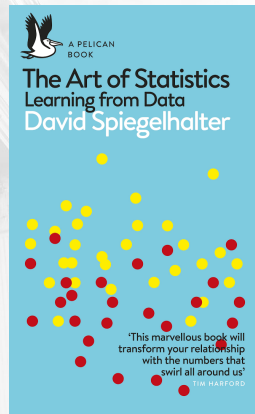


Figura 3. Spiegelhalter, 2019

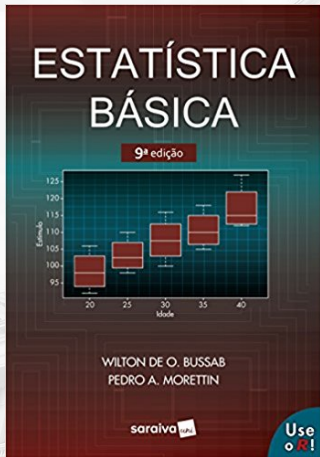


Figura 4. Bussab & Morettin



Figura 5. Magalhães & Lima

Textos de apoio - parte inicial

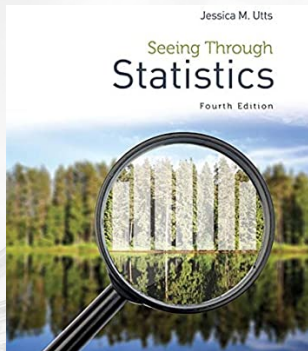


Figura 6. Utts

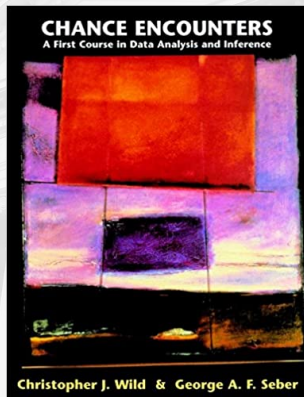


Figura 7. Wild & Seber

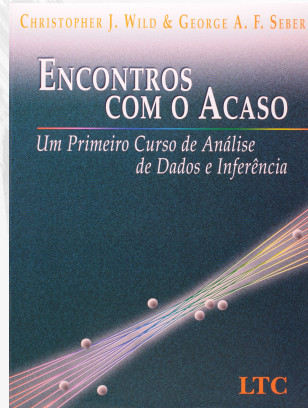


Figura 8. Wild & Seber (PT)

Mozilla Firefox

R Prog. GON Lay. R Fe ggp 11.7 H Pro R C Dat. Workin DE GON Que Ace. Paif Mer Mer

www.leg.ufpr.br/estbas

ESTATÍSTICA BÁSICA

Equipe EstBas - DEST/UFPR

Descrição do curso

Os vídeos listados abaixo compõe um curso introdutório de estatística ofertado pelo Departamento de Estatística (DEST) da Universidade Federal do Paraná (UFPR) durante um dos períodos de Ensino Remoto Emergencial (ERE) de Novembro/2020 até Março/2021.

Estas aulas resultam de um esforço dos professores do Departamento para desenvolver material didático e democratizar o acesso ao conhecimento, dando assim continuidade a formação acadêmica durante o período de ensino remoto. Os vídeos e slides disponibilizados aqui são parte do rol de materiais do curso.

Clique no botão caso deseje acessar diretamente a *playlist* do *YouTube*.

Playlist no Youtube com todas as aulas

Lista de vídeos e respectivos *slides* das aulas por conteúdo:

- 0. [Introdução \(slides\)](#).
- 1. **Métodos de amostragem**
 - 1.1. [Amostragem probabilística \(slides\)](#).
 - 1.2. [Amostragem não probabilística \(slides\)](#).
- 2. **Estatística Descritiva**
 - 2.0. [Apresentação \(slides\)](#).
 - 2.1. [Importância da estatística descritiva \(slides\)](#).
 - 2.2. [Tipos de variáveis \(slides\)](#).
 - 2.3. [Distribuições de frequência para variáveis qualitativas \(slides\)](#).
 - 2.4. [Distribuições de frequência para variáveis quantitativas \(slides\)](#).
 - 2.5. [Medidas de posição \(slides\)](#).
 - 2.6. [Medidas de dispersão \(slides\)](#).
- 3. **Probabilidades**

- Colesterol HDL (bom) associado com demência ?

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$
- ▶ Estatina e dor muscular

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$
- ▶ Estatina e dor muscular
 - ▶ Aumento de cerca de 2% pode ser de 20% (?)

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$
- ▶ Estatina e dor muscular
 - ▶ Aumento de cerca de 2% pode ser de 20% (?)
 - ▶ Risco e Risco Relativo: 85/100 vs 87/100 : $RR = 1,02$

- ▶ Colesterol HDL (bom) associado com demência ?
- ▶ Relação entre hamburger de bacon e câncer de estômago
 - ▶ Aumento de cerca de 1%, 17% ou 18% (?)
 - ▶ Risco absoluto : 1%
 - ▶ Risco e Risco Relativo: 6/100 vs 7/100 : $RR = 1,17$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$
- ▶ Estatina e dor muscular
 - ▶ Aumento de cerca de 2% pode ser de 20% (?)
 - ▶ Risco e Risco Relativo: 85/100 vs 87/100 : $RR = 1,02$
 - ▶ Chance e Razão de Chances: 6 para 94 vs 7 para 93 : $OR = 1,18$

Vamos conhecer a turma?



Vamos conhecer a turma?

...a apresentação fica por nossa conta!

Vamos ao questionário!

Vamos conhecer a turma?

A ferramenta fornece um recurso de análise gráfica,
mas não é exatamente o que queremos ...

Queremos alterar e/ou adicionar gráficos

Queremos investigar relações entre respostas de diferentes questões

- O objetivo da análise: descrição do perfil?

Alguma análise

- ▶ O objetivo da análise: descrição do perfil?
- ▶ População ou amostra?
 - ▶ (de) qual população?

- ▶ O objetivo da análise: descrição do perfil?
- ▶ População ou amostra?
 - ▶ (de) qual população?
- ▶ Que gráficos (ou análises) fazer?
- ▶ Gráficos/análises: uni, bi e multivariadas
- ▶ Tipos de variáveis
 - ▶ qualitativas (nominais ou ordinais)
 - ▶ quantitativas (discretas ou contínuas)

As dificuldades aparecem ...

- ▶ (pré) validação do questionário
- ▶ respostas inesperadas
- ▶ questões para checagem de coerência
- ▶ desenho das questões e alternativas
- ▶ tratamento (automático) de anomalias
- ▶ processamento (automático) e adequação de respostas em texto
- ▶ descarte (questão ou questionário)
- ▶ ...

Pagar na entrada ou na saída?


Removidas as dificuldades ...

- ▶ Que gráficos (ou análises) fazer?
- ▶ Gráficos/análises: uni, bi e multivariadas
- ▶ Tipos de variáveis
 - ▶ qualitativas (nominais ou ordinais)
 - ▶ quantitativas (discretas ou contínuas)

Referências bibliográficas

 BOX, G. E. P. et al. **Statistics and Scientific Method: An Introduction for Students and Researchers**. 2. ed. New Jersey: Wiley, 2005.


 BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 9. ed. São Paulo: Saraiva, 2017.

 DIGGLE, P. J.; CHETWYND, A. G. **Statistics and Scientific Method: An Introduction for Students and Researchers**. 1. ed. Oxford: Oxford University Press, 2011.

 MAGALHÃES, M. N.; LIMA, A. C. P. de. **Noções de Probabilidade e Estatística**. 7. ed. São Paulo: Edusp, 2015.


 NOLAN, D.; SPEED, T. **Stat Labs: Mathematical Statistics through Applications**. 1. ed. New York: Springer, 2000.

 SPIEGELHALTER, D. **The Art of Statistics: Learning from Data**. [S.l.]: Penguin Books Limited, 2019. (Pelican Books). ISBN 9780241258750.

 UTTS, J. M. **Seeing Through Statistics**. [S.l.: s.n.], 2005.

 WILD, C. J.; SEBER, G. A. F. **Chance Encounters: A First Course in Data Analysis and Inference**. [S.l.: s.n.], 2000.

 WILD, C. J.; SEBER, G. A. F. **Encontros Com O Acaso. Primeiro Curso De Análise De Dados e Inferência**. [S.l.: s.n.], 2004.

 ZEVIANI, W. et al. EstBas: Um curso em estatística básica. [Http://www.leg.ufpr.br/estbas](http://www.leg.ufpr.br/estbas). 2021.