

Az adatelemzés oktatása

Békés Gábor-Horn Dániel

HUN-REN KRTK Közgazdaságtudományi Intézet – *műhely*

2024. október 10.

Békés - Kézdi: Data Analysis

This textbook provides future data analysts with the tools, methods, and skills needed to answer data-focused, real-life questions, to carry out data analysis, and to visualize and interpret results to support better decisions in business, economics, and public policy.

Data wrangling and exploration, regression analysis, machine learning, and causal analysis are comprehensively covered, as well as when, why, and how the methods work, and how they relate to each other.

As the most effective way to communicate data analysis, running case studies play a central role in this textbook. Each case starts with an industry-relevant question and answers it by using real-world data and applying the tools and methods covered in the textbook. Learning is then consolidated by 340 practice questions and 120 data exercises.

Extensive online resources, including raw and cleaned data and codes for all analysis in Stata, R, and Python, can be found at www.gabors-data-analysis.com.

Cover images: the cover illustrates steps and methods of data analysis. We recognize the influence of the authors (L1), illustrate our data (C2), and describe the methods (C3) under the influence of our previous studies (B1). The main product was, and is, built with a lot of help and support from our teaching assistants (A1), and our students (C2). We thank the original images of the reviewers, and study exercise B2. Cover illustration by Agoston Hegyi 2017. Cover design: Dániel Horn.

"This exciting new text covers everything today's aspiring data scientist needs to know, managing to be comprehensive as well as accessible. Like a good confidence interval, the Gabors have got you almost completely covered!"
Professor Joshua Angrist, Massachusetts Institute of Technology

"A beautiful integration of econometrics and data science that provides a direct path from data collection and exploratory analysis to conventional regression modeling, then on to prediction and causal modeling. Exactly what is needed to equip the next generation of students with the tools and insights from the two fields."
Professor David Card, University of California-Berkeley

"This textbook is excellent at dissecting and explaining the underlying process of data analysis. Békés and Kézdi have masterfully woven into their instruction a comprehensive range of case studies. The result is a rigorous textbook grounded in real-world learning, at once accessible and engaging to novice scholars and advanced practitioners alike. I have every confidence it will be valued by future generations."
Professor Kerwin K. Charles, Yale School of Management

Békés and Kézdi
DATA ANALYSIS
FOR BUSINESS,
ECONOMICS, AND POLICY

DATA ANALYSIS
FOR BUSINESS,
ECONOMICS, AND POLICY

Gábor Békés | Gábor Kézdi

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org
978-1-108-71626-4
9 781108 716264

Békés - Kézdi: Data Analysis

- ▶ Cambridge University Press, 2021
- ▶ Egy teljes [kidolgozott ökoszisztéma](#)
- ▶ Tanításhoz [diákkal](#)
- ▶ A kidolgozott esettanulmányokhoz adatokkal ([OSF](#)) és kóddal ([Github](#))

- ▶ Minden egyben itt:
- ▶ www.gabors-data-analysis.com/


Békés - Kézdi: Data Analysis

- ▶ Cambridge University Press, 2021
- ▶ Egy teljes [kidolgozott ökoszisztéma](#)
- ▶ Tanításhoz [diákkal](#)
- ▶ A kidolgozott esettanulmányokhoz adatokkal ([OSF](#)) és kóddal ([Github](#))

- ▶ Minden egyben itt:
- ▶ www.gabors-data-analysis.com/

- ▶ *De sajnos (még?) csak angolul*
 - ▶ Magyarul az első két rész olvasható, nyomtatásban.
 - ▶ Második két rész fordítása folyamatban, de forráshiány van jelenleg.

Békés - Kézdi: Adatelemzés



Békés Gábor
Kézdi Gábor

ADATELEMZÉS

ÜZLETI,
KÖZGAZDASÁGI
ÉS SZAKPOLITIKAI
DÖNTÉSEKHEZ

BÉKÉS GÁBOR – KÉZDI GÁBOR

ADATELEMZÉS

**Üzleti, közgazdasági és szakpolitikai
döntésekhez**

*„...mindent tartalmaz, amit napjaink feltörekvő
adatelemzőinek ismerniük kell”*

Joshua Angrist (közgazdasági Nobel-díj, 2021)

Békés - Kézdi: Adatelemzés

- ▶ Alinea, 2024
- ▶ alinea.hu/adatelemzes
- ▶ gabors-data-analysis.com/magyar-kiadas

- ▶ "Gyémántkoszorú kör"
 - ▶ Facebook
 - ▶ Szótár: gabors-data-analysis.com/dictionary-hun

Horn Dániel: Egy felhasználói vélemény

Eredeti alap célok

1. Tartalmi forradalom

- ▶ Tanítsuk azt, ami gyakran kell és ne tanítsuk, ami nem olyan fontos
- ▶ Életre és ne akadémiára készítsünk, de akadémiai pályán is hasznos legyen

Eredeti alap célok

1. Tartalmi forradalom

- ▶ Tanítsuk azt, ami gyakran kell és ne tanítsuk, ami nem olyan fontos
- ▶ Életre és ne akadémiára készítsünk, de akadémiai pályán is hasznos legyen

2. A tudás demokratizálása

- ▶ Mindenki értse, de legyen érdekes
- ▶ Legyen sok példa és kevés matek, de legyen precíz

Eredeti alap célok

1. Tartalmi forradalom

- ▶ Tanítsuk azt, ami gyakran kell és ne tanítsuk, ami nem olyan fontos
- ▶ Életre és ne akadémiára készítsünk, de akadémiai pályán is hasznos legyen

2. A tudás demokratizálása

- ▶ Mindenki értse, de legyen érdekes
- ▶ Legyen sok példa és kevés matek, de legyen precíz

3. Azt adjuk, ami kell az adat elemzőnek

Eredeti alap célok

1. Tartalmi forradalom
 - ▶ Tanítsuk azt, ami gyakran kell és ne tanítsuk, ami nem olyan fontos
 - ▶ Életre és ne akadémiára készítsünk, de akadémiai pályán is hasznos legyen
2. A tudás demokratizálása
 - ▶ Mindenki értse, de legyen érdekes
 - ▶ Legyen sok példa és kevés matek, de legyen precíz
3. Azt adjuk, ami kell az adat elemzőnek
4. Technikai tudás, eredmények értelmezése és gondolkodás
 - ▶ Szerintem ebben a legerősebb

Eredeti alap célok

1. Tartalmi forradalom
 - ▶ Tanítsuk azt, ami gyakran kell és ne tanítsuk, ami nem olyan fontos
 - ▶ Életre és ne akadémiára készítsünk, de akadémiai pályán is hasznos legyen
2. A tudás demokratizálása
 - ▶ Mindenki értse, de legyen érdekes
 - ▶ Legyen sok példa és kevés matek, de legyen precíz
3. Azt adjuk, ami kell az adat elemzőnek
4. Technikai tudás, eredmények értelmezése és gondolkodás
 - ▶ Szerintem ebben a legerősebb
5. All in one - egy könyv, egy teljes program
 - ▶ Teljes ökoszisztéma
 - ▶ 4 rész, az alaptól a haladó adatelemzésig

47 esettanulmány, sokféle téma

- ▶ A diákok nagy része nem tudja még, mi érdeklí
- ▶ Business: managerek szerepe, árazás, részvényárak, mennyiség előrejelzés
- ▶ Economics: vállalati összeolvadás, otthoni munka hatása,
- ▶ Policy: étkezés, dohányzás szerepe, immunizáció
- ▶ Globális példák: 1/3 USA, 1/3 EU+UK, 1/3 egyéb

Erről fog Bárdits Anna beszélni.

47 esettanulmány, sokféle téma

480 Model Building for Prediction

Figure 14.3 Airbnb apartment price and in price distribution.

Note: Price is color right in (a) refers histograms without outliers (price < 400 US dollars). Same as price (left), including foreign of dollar, March 2017. Observations with 7 or fewer guests accommodated (N=432).

(a) Price
(b) Price in log

As usual with price data, the distribution of Airbnb apartment price is strongly skewed with a long right tail. Log price is close to normally distributed. Here we don't take log of price but rely on the prediction with price as the variable for all models. You will be invited to carry out the prediction with log price as a data exercise.

Figure 14.4 Airbnb apartment price distribution by important features

Note: Box plots to show the data don't show the extreme values.
Source: Airbnb dataset, including foreign of dollar, March 2017. Observations with 7 or fewer guests accommodated (N=432).

(a) Price by room type
(b) Price by number of people accommodated and apartment versus House in US dollar

481 Case Study

Of the other important predictor variables, some are related to the number of guests the apartment can accommodate: number of rooms, number of beds. There are some variables related to reviews: average review score, the number of reviews, the date of the first review (to indicate how long this apartment was in business). The number of apartments or rooms is left as it is, and treated as continuous.

The number of guests to accommodate, the number of beds, the average review score, the number of reviews, and the date since the first review are quantitative variables. These quantitative variables may be better converted into categories (such as number of reviews: 0, 1 to 50, 50+ or have four values to begin with) (guests to accommodate goes up to 16, but most apartments accommodate 1 through 7 guests). The log base price of Figure 14.4a and 14.4b both without values outside (whiskers) show average price by room type and size of the apartment. Single or shared rooms are cheaper, typically below 50 dollars, while whole apartments can be more than a 100 dollars. About one-third of observations had no reviews. For them, we replaced review-related variables with the sample mean and added a binary variable (flag) to capture being a missing value with regards to these quantitative variables. We made one key decision on sample design: we dropped very large apartments.

Our final original data has 4320 observations.

Property type: Apartment, House

Mean price by property type

Figure 14.5 Graphical aids of finding interactions

Note: Conditional mean plots by selected variables, 95% confidence interval on log of price.
Source: Airbnb dataset, including foreign of dollar, March 2017 (N=4320).

Regarding other predictors, we have several binary variables, which we kept as they were: type of bed, type of property (apartment, house, room), cancellation policy. Furthermore, there is a

Az adatelemzés egy folyamatának ismétlése

1. kutatási kérdés
2. adatok gyűjtése
3. adatok tisztítása és rendszerezése
4. leíró adatelemzés
5. elemzési munka
6. eredmények bemutatása
7. kérdés megválaszolása

Erről fog Békés Gábor beszélni.

Hol lehet tanítani

- ▶ Alapszak I.: Business, Economics
- ▶ Alapszak II.: Quantitative SS, PPE

- ▶ Mesterszakok I.: Economics, Applied Economics, Finance
- ▶ Mesterszakok II.: Business Analytics, Data Science

- ▶ Üzleti Phd
- ▶ MBA/EMBA

gabors-data-analysis.com/teaching-guide/undergraduate-and-graduate-programs

Sok más könyv van, miért ez?

- ▶ Diákközpontú: releváns témák, sok példa és gyakorlás
- ▶ Kimenet ökonometria és adat-tudományok irányba
- ▶ Ökoszisztéma
 - ▶ kódok és adatok
 - ▶ gyakorló kérdések, feladatok, feleletválasztós kérdések,
 - ▶ ábrák, diasorok elérhetőek (Overleaf)

Erről (is) fog Elek Péter beszéni.

Sok más könyv van, miért ez?



Giuseppe Cavaliere
@CavaliereGlu

Hi #EconTwitter! 🚩

Are you into #econometrics and **data analysis**? And looking for an amazing set of slides on the topic?

Don't miss 📌 @GaborBekes's collection of slides from his book on data analysis. 📄

Very well-crafted and clear, they cover a lot of material.

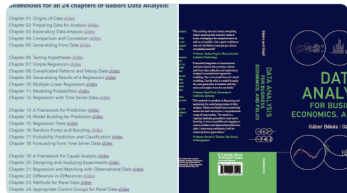
Cool stuff! 🌟

Links

Slides: gabors-data-analysis.com/chapter-detail...

Code etc: github.com/gabors-data-an...

Bejegyzés lefordítása



Utolsó szerkesztés de. 10:50 · 2024. júl. 3. · **32,2 E** megtekintés



3



103



364



395



Gabriel M Ahlfeldt
@Ahlfeldt

Gave my first Intro into Econometrics lecture at @HumboldtUni this term. It went much better than expected thanks @GaborBekes' book: cambridge.org/highereducatio... Perfect for BSc students w good training in maths and stats who just need to see why empirical research is so exciting! 🌟

Bejegyzés lefordítása



cambridge.org

Data Analysis for Business, Economics, and Policy | Higher E
Discover Data Analysis for Business, Economics, and
Policy, 1st Edition, Gabor Bekes, HB ISBN: 978110848301...

du. 7:00 · 2024. aug. 10. · **8 572** megtekintés



2



11



105



53



Válasz



Gabriel M Ahlfeldt @Ahlfeldt · aug. 10.

You won't find another textbook with so many cool case studies integrated in the textbook, lecture slides, and problem sets.

Sok más könyv van, miért ez?

"It is excellent. I plan on using it for my Business Strategy and Analytics module. I find it to be quite exhaustive and very well organized, very consistent with how I organize my module. I think they covered nearly everything that I'd want covered in a book like this." Anil Doshi (UCL)

"I like the book so much that I modified my course to better adhere to the contents of the book.[..] The case studies with code are the biggest bonus, but I also really like the scope of the contents and the intuitive explanations. Excellent book and ecosystem" Peter Fuleky (Hawaii)

"The thing that attracts me about the book is the emphasis on applied work and data management. The value of these skills is recognized. The challenge is that we and our students' employers want those skills in addition to strong applied econometric skills." John Straub (Maryland)

gabors-data-analysis.com/instructor-feedback

Hol tanítják most?



gabors-data-analysis.com/courses-using/

Amit most ajánlani tudunk

- ▶ Egy kész angol nyelvű tananyag - mindenestül
- ▶ Egy kész magyar nyelvű bevezető tankönyv
 - ▶ a "haladó rész" és a teljes "ökoszisztéma" fordítása még folyamatban (erőforráshiány!).
- ▶ Angol és magyar nyelvű tanítási "know-how"
 - ▶ aminek az átadása folyamatos (lásd a mai alkalmat)

A továbbiak

- ▶ Elek Péter: Regresszió, gépi tanulás és oksági elemzés oktatása a tankönyv segítségével
- ▶ Bárdits Anna: Esettanulmányok és gyakorlati példák
- ▶ Reguly Ágoston: Adatelemzés és Kódolás
- ▶ +1 Békés Gábor: Az adatelemzés hét lépése