**Documentation of statistics for
Book sales 2023**

1 Introduction

The purpose of the statistics Book sales is to shed light on the development of book sales in Denmark. The statistics illustrate the trends in the number of copies sold through physical bookstores and online sales channels. The statistics are new as of 2024 and include data from the third quarter of 2022 onwards.

2 Statistical presentation

Book sales is a quarterly compilation of the number of physical books, e-books and audiobooks sold through physical bookstores (including the largest supermarket chains) and online sales channels. Sales of used books (e.g. from second-hand bookstores, thrift stores, and flea markets) are not included. The statistics are categorized by genre, format/media, publication language, original language and binding type. The data is published in the newsletter Nyt fra Danmarks Statistik (in Danish only) and in Statbank Denmark. Data sources and coverage are still under development, and the figures are therefore preliminary. Future revisions may occur.

2.1 Data description

The statistics illustrate the development of book sales in Denmark by genre, format/media, publication language, original language and binding type. A sale corresponds to a single copy of a book with a 13-digit ISBN starting with 978 or 979, sold through physical bookstores and online sales channels. For example, if 10 copies of a book are sold in one transaction, this counts as 10 sold books. Each book corresponds to a unique edition. Typically, the same title has a separate ISBN for each edition, where the physical edition has a different ISBN than the e-book or audiobook editions. Additionally, a second edition of a given title may have a different ISBN than the first.

Currently, the statistics do not include sales via streaming services, meaning books accessed through subscription-based platforms. This includes, for instance, 30-day subscription plans that allow unlimited reading or listening to books via a digital platform during the subscription period.

The statistics will be continuously expanded with additional sources and content. Efforts are underway to establish a robust data foundation on book streaming, aiming to cover all forms of book sales in the market.

2.2 Classification system

The genre classification in these statistics is based on Thema codes, which are part of the international classification system: THEMA, used by the commercial book industry.

Main Genres

- Non-fiction
- Fiction
- Children's and Young Adult Literature
- Not Specified

Subgenres for Non-fiction

- Biographies, True Stories, and Non-Fiction Prose
- History
- Art and Culture

- Food and Drink
- Nature, Environment, and Geoscience
- Personal Development, Relationships, and Health
- Travel Guides
- Religion, Spirituality, and Philosophy
- Society, Social Sciences, and Debate
- Economics & Business
- Hobbies and Lifestyle: Miscellaneous
- Other Non-fiction

Subgenres for Fiction

- Poetry
- Graphic Novels
- Historical Novels
- Crime, Thrillers, and Suspense Novels
- Romance and Erotica
- Science Fiction, Fantasy, etc.
- Other Fiction

Subgenres for Children's and Young Adult Literature

- Books for Toddlers, Picture Books, Activity Books, etc.
- Non-fiction for Children and Young Adults
- Fiction for Children and Young Adults
- Educational Material for Children and Young Adults
- Miscellaneous for Children and Young Adults

The statistics also include information about media/format, publication language, original language, and binding type.

Media/Format

- Physical Books
- E-Books
- Audiobooks

Published language

- Danish
- English
- Swedish
- Norwegian
- German
- French
- Other or Multiple Languages

Original Language

- Danish
- English
- Swedish
- Norwegian

- German
- French
- Other or Multiple Languages
- Not Specified

Binding

- Hardcover or Paperback (Bound)
- Paperback and Pocket Books
- Other physical Formats (for example, cardboard books, spiral-bound books, occasion albums, etc.)
- Unknown physical formats
- Digital Format

2.3 Sector coverage

Culture.

2.4 Statistical concepts and definitions

ISBN: Stands for International Standard Book Number. It is a 13-digit identifier for a book, an edition of a book or another monographic publication by a particular publisher. Before January 1, 2007, ISBNs consisted of 10 digits.

2.5 Statistical unit

Unit sales measured in the number of books sold.

2.6 Statistical population

The population consists of books sold by booksellers, either in physical stores or through online sales.

2.7 Reference area

Denmark

2.8 Time coverage

These statistics cover the time period from third quarter of 2022 onwards.

2.9 Base period

Not relevant for these statistics.

2.10 Unit of measure

Number of books sold.

2.11 Reference period

Quarter.

2.12 Frequency of dissemination

Yearly.

2.13 Legal acts and other agreements

There is no EU regulation for statistics on book production.

2.14 Cost and burden

These statistics are based on administrative data. There is thus no direct response burden in relation to the compilation of these statistics.

2.15 Comment

Further information can be obtained by contacting Statistics Denmark directly.

3 Statistical processing

Data for this statistics is collected through monthly transfers of transaction data on book sales units from SAXO.com A/S, Indeks Retail (Bog & Idé), Gucca and the major supermarket chains (COOP, Salling and Dagrofa).

The transaction data is enriched with information from DBK, which manages the Bogportalen and Publizon, allowing the books to be classified by genre, format/media, publication language, original language and binding type. If this information is missing for certain editions, efforts are made to manually complete it using various assumptions.

3.1 Source data

The statistics are based on transaction data provided by SAXO.com A/S, Indeks Retail (Bog & Idé) and Gucca. Additionally, transaction data (barcode data) is supplied by the Prices and Consumption office in the Economic Statistics department at Statistics Denmark. The data includes all unit sales of books in stores under COOP, Salling and Dagrofa.

The transaction data includes, among other things, the ISBN number of the sold book, the sale time, and the number of copies sold at the time of the transaction, as well as a few supplementary details about the purchased book, such as whether it is a physical book, e-book, audiobook, or another form of product information, depending on the data provider.

The ISBN number is used as a key to enrich the transaction data with metadata from DBK (Bogportalen) and Publizon. The information in the metadata is used to classify the books by genre, format/media, publication language, original language and binding type.

3.2 Frequency of data collection

Monthly data transfer from data providers.

3.3 Data collection

System-to-system solution.

3.4 Data validation

The establishment of the groupings and classifications based on metadata has been validated by key stakeholders in the book industry and subject matter experts. Additionally, data is compared from quarter to quarter. In the case of significant fluctuations, the data provider or metadata provider is contacted.

3.5 Data compilation

Transaction data is cleaned by removing products that are not books, such as print paper. Additionally, ISBN numbers that deviate from the 13-digit structure corresponding to books are removed. The transaction data is enriched with metadata from DBK and Publizon. The metadata is then used to categorize the data according to the classifications that form the basis for the tables in StatBank Denmark. After the data is enriched and processed, it is aggregated for the creation of tables in the StatBank.

The metadata includes information on Thema codes, language of publication, original language, format/media and binding type (digital format for e-books and audiobooks). Thema codes are part of an international classification system where books are categorized by topic. Since September 2019, classifying books using Thema codes in the Book Portal and Publizon has been mandatory. Not all publications have been assigned Thema codes on the Book Portal, but coverage is increasing as the information on the portal is updated.

Manual adjustments are made as follows:

Genre

If Thema codes are missing in the metadata, store category information is used to manually assign a Thema code. This is done based on the distribution of Thema codes for titles in the same category, if this is relevant (e.g., Crime and Thrillers, Fantasy, Fiction, Economics, Cookbooks, Textbooks for children, etc.). This supplementation of Thema codes based on store category information accounts for approximately 1 pct. of total sales.

If metadata for a book sold in a supermarket is missing, additional information from supermarket transaction data is used. If their category information, for example, indicates "Crime/Thrillers," this information is used to place the book in a genre. This genre placement accounts for about 3 pct. of total sales.

Read more about the methodology in the document Genreopdeling ved hjælp af Themakoder (in Danish only).

Published language

If language information is available in metadata from DBK, this is used. Both direct language information and store category information (category text) from DBK data, which may contain

information about the book's language, are checked. If language information is not available in DBK or Publizon metadata, the country of publication is inferred from the ISBN number, and it is assumed that, for example, books published in England are in English, books published in France are in French, and so on.

Format/Media

The data processing assumes that the books sold are primarily physical books. To ensure correct categorization of format, ISBN numbers are validated by comparing them with metadata from DBK and Publizon. If metadata indicates that the book is an e-book or audiobook, the format is corrected accordingly. Publizon data is used as a reliable source for this, as Publizon is Denmark's largest distributor of e-books and audiobooks. Additionally, information from SAXO and Indeks Retail is used for further verification. SAXO's transaction data indicates whether the product is a physical book, e-book, or audiobook using the "TYPE" field, while Indeks Retail divides their data into categories such as "Books" and "Audiobooks."

Binding Type

In DBK metadata, direct information about binding type is incomplete, so store category information in the data (category text) is also checked, as it may contain information about the book's binding.

3.6 Adjustment

No corrections are made beyond what has already been described in sections 3.4 Data Validation and 3.5 Data Processing.

4 Relevance

The key users are stakeholders in the book industry, the Ministry of Culture and other interested parties. The statistics are used for analyses in the literature sector.

4.1 User Needs

The key users are stakeholders in the book industry, the Ministry of Culture and other interested parties. The statistics are used for analyses in the literature sector.

4.2 User Satisfaction

The statistics are new, but preliminary and general results from earlier periods, prior to those published in the Statistics Bank, have been validated and approved by stakeholders in the book industry and the Ministry of Culture.

4.3 Data completeness rate

Not relevant for this statistic.

5 Accuracy and reliability

The statistics cover the majority of book sales in Denmark, purchased new. The metadata coverage is considered satisfactory for Danish publications, although coverage is lower for foreign publications. Coverage is also lower for books/book products where the primary purpose is not reading, such as coloring books, board books, activity books, calligraphy books, books with music notation, songbooks, and puzzle books like Sudoku. ISBN errors may occur, but they are not deemed to be a widespread issue.

The figures are preliminary, and future revisions may occur as book sales data expands and metadata quality improves.

5.1 Overall accuracy

The statistics cover the majority of book sales of Danish books, but not all. For example, sales of used books (antiquarians, second-hand stores, flea markets) or small independent bookstores and self-published books online are not included.

It is possible that an ISBN number has been registered incorrectly, either in the transaction data or on the Book Portal and/or Publizon. This could potentially result in books being categorized incorrectly in the statistics, but it is not considered to be a widespread issue.

5.2 Sampling error

Not relevant for this statistic.

5.3 Non-sampling error

The statistics do not cover the sale of used books, sales from small independent bookstores or self-published books online. Additionally, the metadata coverage is lower for foreign publications and certain types of hobby books, such as activity books and sheet music.

The statistics also do not include sales through streaming services, i.e., books purchased via subscription.

The statistics will continuously be expanded with additional sources and content. Efforts will be made to establish a robust data foundation on book streaming, with the aim of covering all forms of book sales in the market.

5.4 Quality management

Statistics Denmark follows the recommendations on organisation and management of quality given in the Code of Practice for European Statistics (CoP) and the implementation guidelines given in the Quality Assurance Framework of the European Statistical System (QAF). A Working Group on Quality and a central quality assurance function have been established to continuously carry through control of products and processes.

5.5 Quality assurance

Statistics Denmark follows the principles in the Code of Practice for European Statistics (CoP) and uses the Quality Assurance Framework of the European Statistical System (QAF) for the implementation of the principles. This involves continuous decentralized and central control of products and processes based on documentation following international standards. The central quality assurance function reports to the Working Group on Quality. Reports include suggestions for improvement that are assessed, decided and subsequently implemented.

5.6 Quality assessment

The statistics are based on data from large booksellers and supermarket chains, as well as metadata from the two leading distribution platforms in Denmark. This ensures a high coverage of book sales in Denmark.

However, the statistics do not cover the sale of used books, sales from small independent bookstores or self-published books online.

The coverage is satisfactory for Danish publications but lower for foreign publications and certain types of hobby books. Since 2019, it has been mandatory for book releases to be registered with Thema codes on the Book Portal and Publizon. Therefore, it is expected that the quality of metadata for books published from 2019 onwards is higher than for publications released before 2019.

Future revisions of the figures may occur as metadata improves and data sources expand.

5.7 Data revision - policy

Statistics Denmark revises published figures in accordance with the Revision Policy for Statistics Denmark. The common procedures and principles of the Revision Policy are for some statistics supplemented by a specific revision practice.

5.8 Data revision practice

The figures are provisional, and future revisions may occur.

6 Timeliness and punctuality

The figures are provisional, and future revisions may occur. The publication of the statistics takes place at the end of November. Since the statistics are new, the punctuality of the statistics cannot yet be described.

6.1 Timeliness and time lag - final results

The figures are provisional, and future revisions may occur. The publication of the statistics takes place at the end of November.

6.2 Punctuality

Since the statistics are new, the punctuality of the statistics cannot yet be described.

7 Comparability

The statistics are not directly comparable to other published book sales statistics. There are no international guidelines for the production of book sales statistics.

7.1 Comparability - geographical

There are no international guidelines for the production of book sales statistics.

7.2 Comparability over time

The statistics are new from 2024 and contain data from the third quarter of 2022. The statistics are comparable throughout the period.

7.3 Coherence - cross domain

There is no direct correlation with other statistics.

7.4 Coherence - internal

The data is based on transaction data from major bookstores, bookstore chains, and large supermarket chains. It is assessed that the content of the data from bookstores and supermarkets is consistent across data sources. The content of metadata from DBK and Publizon is also considered comparable, as both sources contain the relevant information used in the preparation of the statistics.

8 Accessibility and clarity

These statistics are published yearly in the newsletter Nyt fra Danmarks Statistik (in Danish only) at the same time as the tables are updated in the StatBank. In the StatBank, these statistics can be found under the subject Literature.

8.1 Release calendar

The publication date appears in the release calendar. The date is confirmed in the weeks before.

8.3 User access

Statistics are always published at 8:00 a.m. at the day announced in the release calendar. No one outside of Statistics Denmark can access the statistics before they are published.

8.2 Release calendar access

The Release Calender can be accessed on our English website: Release Calender.

8.4 News release

These statistics are published yearly in the newsletter [Nyt fra Danmarks Statistik](#) (in Danish only).

8.5 Publications

Not relevant for this statistic.

8.6 On-line database

The statistics are published in the StatBank under the subject [Literature](#) in the following tables:

BOGS01: Sold books by genre and format BOGS02: Sold books by main genre and published language BOGS03: Sold books by genre, original language and published language

8.7 Micro-data access

There is no access to the Micro-data of the statistics.

8.8 Other

Not relevant for this statistic.

8.9 Confidentiality - policy

[Data Confidentiality Policy](#) for Statistics Denmark is applied.

8.10 Confidentiality - data treatment

Not relevant for this statistic.

8.11 Documentation on methodology

[Genreopdeling ved hjælp af Themakoder](#) (in Danish only).

8.12 Quality documentation

Results from the quality evaluation of products and selected processes are available in detail for each statistics and in summary reports for the Working Group on Quality.

9 Contact

The administrative placement of these statistics is in the division Science, Technology and Culture. The person responsible is Christian Törnfelt, tel. +45 21 63 60 20, e-mail: cht@dst.dk.

9.1 Contact organisation

Statistics Denmark

9.2 Contact organisation unit

Science, Technology and Culture, Business Statistics.

9.3 Contact name

Christian Törnfelt

9.4 Contact person function

Responsible for the statistics

9.5 Contact mail address

Sankt Kjelds Plads 11, 2100 Copenhagen

9.6 Contact email address

cht@dst.dk

9.7 Contact phone number

+45 21 63 60 20

9.8 Contact fax number

N/A