

# Inngangur að ályktunartölfræði

Katrín Arndís

2023-01-17



# Contents

<b>1</b>	<b>Formáli</b>	<b>5</b>
<b>2</b>	<b>Grunn hugtök</b>	<b>7</b>
<b>3</b>	<b>Tilgátur</b>	<b>9</b>
3.1	Aðaltilgáta . . . . .	9
3.2	Núlltilgáta . . . . .	9
<b>4</b>	<b>Marktekt</b>	<b>11</b>
<b>5</b>	<b>Footnotes and citations</b>	<b>13</b>
5.1	Footnotes . . . . .	13
5.2	Citations . . . . .	13
<b>6</b>	<b>Blocks</b>	<b>15</b>
6.1	Equations . . . . .	15
6.2	Theorems and proofs . . . . .	15
6.3	Callout blocks . . . . .	15
<b>7</b>	<b>Sharing your book</b>	<b>17</b>
7.1	Publishing . . . . .	17
7.2	404 pages . . . . .	17
7.3	Metadata for sharing . . . . .	17



# Chapter 1

## Formáli

Í þessu skjali fer ég yfir nokkur grunn hugtök og hugmyndir sem er gott að kunna skil á. Athugið að þetta er óyfirfarið efni svo villur gætu leynst í textanum.



## Chapter 2

# Grunn hugtök

Eftirfarandi eru grunnhugtök sem þarf að þekkja við lestur þessa skjals:

- **Þýði** er sá hópur sem við viljum læra eitthvað um; það getur endurspeglað stóran hóp einstaklinga (Evrópubúar), smærri hóp (Íslendingar) eða helður lítinn hóp (nemendur HR).
  - Þýði þarf að vera skýr, afmarkaður hópur einstaklinga sem eiga eitthvað sameiginlegt.
- **Úrtak** er sá hópur sem við mælum því við teljum hann endurspegla þýðið. Við reynum yfirleitt að mynda úrtakið úr tilviljunarkenndum hóp úr þýðinu.
- **Þýðistölur** eru þau gildi sem myndu fást ef við hefðum aðgang að öllu þýðinu.
  - Gefum okkur að við viljum vita hæð Íslendinga: Það er einhver tala, raunveruleg meðalhæð, sem fengist ef við myndum mæla hæð hvers einasta Íslending. Það er, hið raunverulega gildi sem fengist ef **öll stök þýðis** væru mæld. Það er þó sjaldnast hægt að mæla *alla* sem þýði samanstendur af – því drögum við úrtak og athugum eiginleika úrtaksins til að komast sem næst raunverulegu gildi í þýði.
- **Úrtakstölur** Eru því þau gildi sem fást frá úrtakinu okkar. Ef við höldum áfram með dæmið um hæð Íslendinga, þá myndum við draga handahófskennt úrtak af Íslendingum og myndum fá einhverja meðalhæð í úrtakinu. Þessi meðalhæð í úrtakinu er dæmi um úrtakstölu. Hún er það næsta sem við komumst því að vera hver hæðin er í *raunveruleikanum (þýðinu)*.<sup>1</sup>

---

<sup>1</sup>Hér er talað um fólk, því fólk er yfirleitt viðfangsefni rannsókna í sálfræði. Það er þó auðvitað hægt að gera rannsóknir og beita ályktunartölfræði á dýr, bíla, sveppi og jafnvel hnetusmjör. Hugmyndin stendur þó – þýði er heildin sem við viljum þekkja eða draga ályktanir um – en er gjarna of stór til að það sé raunsætt að mæla hvert einasta stak sem tilheyrir því. (Stak er því gjarna notað til að lýsa hverri einingu sem þýði samanstendur af, í þýði Íslendinga myndi hver einstaklingur vera eitt stak í þýðinu).

- **Tilgátuprófun:** Þegar við setjum upp rannsóknir eða söfnum gögnum þá höfum við einhverjar spurningar í huga sem við viljum leitast svara við.
  - Tilgátuprófun er það ferli að setja spurninguna okkar fram í formi tilgáta og nota viðeigandi tölfraðipróf til að fá einfalt já/nei við spurningunni okkar eftir því hvort niðurstöður reynist marktækar eða ómarktækar.
  - Tilgátuprófun svarar spurningunni **hvort það sé munur** – í kjölfarið viljum við auðvitað vita **hve miklu munar** og notum þá spá, þar sem úrtakstölur eru notaðar til að spá fyrir um þýðisgildin.

## Dæmi

Við viljum vita hvort aukinn menntun skili sér í hærri launum á Íslandi.

Hér er þýðið okkar afmarkað – við erum að skoða fólk með búsetu á Íslandi. Úr því þýði myndum við draga úrtak, síðan myndum við athuga fjölda ára sem hver einstaklingur úrtaksins hefur verið í námi og hver mánaðarlaun viðkomandi eru.

Spurningin okkar er einnig skýr og við gætum endurorðað hana á eftirfarandi hátt: Eru mánaðarlaun að jafnaði hærri eftir því sem fjöldi ára í námi eykst? Við getum einfaldað spurninguna enn meira og athugað hvort það sé munur á meðallaunum þeirra sem hafa lokið háskólamenntun og þeirra sem hafa ekki lokið neinni háskólamenntun.

Tölfræðiforrit skilja ekki setningar – við gætum ekki beinlínis sett gögn inn í úrvinnslu ásamt spurningunni “Eru mánaðarlaun að jafnaði hærri eftir því sem fjöldi ára í námi eykst?”. Það má því ímynda sér ferlið í grófum dráttum á eftirfarandi hátt: Við yfirfærum spurninguna okkar yfir á *tölfræðilegt form* sem forritið skilur og segjum því hvaða tölfraðiaðferð það eigi að nota við að svara spurningunni okkar. Forritið vinnur úr gögnunum fyrir okkur og spýtir út úr sér niðurstöðum. Þessar niðurstöður eru þá enn á *tölfræðilegu formi* og það er okkar verk að skilja hvað niðurstöðurnar segja okkur og geta gefið túlkun sem setur þær aftur á *mannamál*.



## Chapter 3

# Tilgátur

Við byrjum á því að setja fram **tilgátur**. Í okkar dæmi viljum við vita hvort það sé munur á launum einstaklings eftir menntun viðkomandi.

### 3.1 Aðaltilgáta

Aðaltilgáta <sup>1</sup> er sú tilgáta sem tilgreinir einhvers konar mun. Vandamálið er að hún tilgreinir enga fasta tölu sem hægt er að prófa og því ógerlegt að prófa hana beint. Í okkar dæmi gæti sannarlega verið munur á launum eftir menntun, en hve mikinn mun ætti að prófa? Fyrir hvert ár í námi væri hægt að prófa hvort laun hækki um 1kr, 10kr, 15kr, 100kr, 1.000kr, 35.000kr o.s.frv. Tilgátan “það er munur” hefur ótal möguleika en úrvinnslan krefst þess að við tilgreinum fast gildi sem hægt er að prófa. Við setjum því fram núlltilgátu.

### 3.2 Núlltilgáta

Núlltilgáta tilgreinir tiltekið tölugildi, þar sem *ef* það tiltekna gildi væri í raun rétt – þá gæti aðaltilgátan ekki líka verið rétt. Í okkar dæmi yrði núlltilgátan sú að laun séu þau sömu óháð menntunarstigi. Það er ekki það sama og að segja að *allir séu með sömu laun* heldur aðeins að ef þú berð saman meðallaun einstaklinga sem hafa ólíka menntun þá sé munurinn á meðallaunum þeirra = 0 (engin munur á meðallaunum þeirra sem hafa verið fá ár í skóla og þeirra sem hafa verið mörg ár í skóla). Ef niðurstaða er sú að núlltilgátan sé sennilega röng – þá hlýtur andstæða hennar að vera rétt. M.ö.o. ef munur á launum eftir menntun er **ekki** núll, þá er jú *einhver* munur og aðaltilgátan hlýtur því að vera rétt.

---

<sup>1</sup>Einnig kölluð gagntilgáta, rannsóknartilgáta, (e. alternative hypothesis)

Nú höfum við sett fram tilgátur og næsta skref er að athuga **hvort** við höfum rétt fyrir okkur – til þess framkvæmum við marktektarpróf.

## Chapter 4

# Marktekt

Við gætum skoðað úrtakstölurnar okkar og séð að í úrtakinu séu meðallaun þeirra sem hafa litla menntun ólík þeirra sem hafa mikla menntun. Tilgangur þess að beita ályktunartölfræði er þó ekki til að *spá fyrir* um úrtakið okkar – úrtakstölurnar eru jú beint fyrir framan okkur, lýsandi tölfræði myndi því duga til að lýsa þeim. Sá munur sem við sjáum í úrtakinu okkar gæti þó verið tilkominn af hreinni tilviljun. Ímyndum okkur tvö dæmi:

1. 20 manna úrtak þar sem launamunur reynist 10kr á milli þeirra sem hafa háskólamenntun og þeirra sem hafa ekki háskólamenntun. Í þessu úrtaki er vissulega munur en það er einnig auðvelt að ímynda okkur að m.v. úrtakstölur, þá sé í raun engin munur á launum hópana í þýði. Þ.e. þessi 10kr munur endurspeglar bara þá tilviljun sem má búast við þegar við drögum úrtak.
2. 2000 manna úrtak þar sem launamunur reynist 100.000kr á milli hópanna. Hér þætti okkur frekar ólíklegt að svo mikill munur myndist, af tilviljun, í svo stóru úrtaki **ef það væri í raun engin munur á hópunum í þýði.**

Eftir því sem munurinn er meiri og úrtakið stærra, þeim mun ólíklegra verður að þykja að þessar úrtakstölur séu einskær tilviljun. Þá situr eftir sú spurning – hversu mikill munur er nógu mikill? Og þá miðað við hversu stórt úrtak?

Marktektarpróf  $\wedge$ [Ásamt fleiri spurningum sem marktektarpróf tekur til greina við útreikning – þetta er einföldun.] tekur tillit til ýmissa upplýsinga um úrtakið okkar og athugar líkur þess að fá okkar úrtakstölur **ef þetta úrtak kæmi úr þýði þar sem núlltilgátan er í raun rétt.** Ef það er í raun **enginn** munur á launum þeirra sem hafa og hafa ekki háskólamenntun, hverjar eru þá líkurnar á því að fá 100.000kr mun í úrtaki með 2.000 manns?

Ef líkurnar eru mjög litlar, þá ályktum við að það sé ólíklegt að núlltilgátan sé rétt. Ef við náum að hafna núlltilgátunni, þá tökum við upp aðaltilgátuna og ályktum þ.a.l. að það sé í raun munur á hópunum í þýði. Ef líkurnar eru miklar

– t.d. ef niðurstöður segja okkur að það séu 50% líkur á að fá þessar úrtakstölur ef við drögum það af tilviljun úr þýði þar sem það er í raun enginn munur – þá myndum við hika verulega við að henda fram fullyrðingum.

Næsta spurning er þá – hversu litlar þurfa líkurnar að vera svo við séum sátt?

---

## Chapter 5

# Footnotes and citations

### 5.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one <sup>1</sup>.

### 5.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2022] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations:  
<https://rstudio.github.io/visual-markdown-editing/#/citations>

---

<sup>1</sup>This is a footnote.



## Chapter 6

# Blocks

### 6.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (6.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (6.1).

### 6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 6.1.

**Theorem 6.1.** *For a right triangle, if  $c$  denotes the length of the hypotenuse and  $a$  and  $b$  denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

### 6.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>





## Chapter 7

# Sharing your book

### 7.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

### 7.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

### 7.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book—all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

# Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2022. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.31.