

Inngangur að ályktunartölfræði

Katrín Arndís

2023-01-18

Contents

1	Formáli	5
1.1	Grunn hugtök	5
2	Tilgátuprófun	7
2.1	Tilgátur	7
3	Marktekt	9
3.1	Marktektarmörk	10
3.2	Marktektarpróf	10
3.3	Höfnunarmistök	11
3.4	Fastheldnimistök	11
3.5	Túlkun ómarktektar	12
3.6	Afköst	12
3.7	Loka pælingar um marktekt.	13
4	Öryggisbil	15

Chapter 1

Formáli

Í þessu skjali fer ég yfir hugtök og hugmyndir sem eru nauðsynleg undirstaða þess að skilja þau umfjöllunarefni sem koma í framhaldinu. Athugið að þetta er með öllu óyfirfarið efni, svo villur gætu leynst í textanum.

1.1 Grunn hugtök

Eftirfarandi eru hugtök 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.¹
- **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 meðalhæð sem fengist ef **öll stök þýðis** væru mæld.
 - Það er þó sjaldnast hægt að mæla *alla* sem þýði samanstendur af – við drögum því úrtak og athugum eiginleika úrtaksins til að komast sem næst raunverulegu þýðisgildi.
- **Úrtak** er sá hópur sem við mælum því við teljum hann endurspegla þýðið. Við reynum yfirleitt að draga með tilviljunarkenndum hætti úr þýði.
- **Úrtakstölur** eru þau gildi sem fást frá úrtakinu. Í úrtaki Íslendinga fæst einhvert meðaltal fyrir hæð, þetta meðaltal er dæmi um úrtakstölu og er það næsta sem við komumst því að vita hver hæðin er í *raun* (þýðinu).

¹Hér er talað um einstaklinga, það má þó beita ályktunartölfræði á hvað sem er; dýr, bíla, sveppi eða jafnvel hnetusmjör. Hugmyndin stendur þó: þýði endurspeglar heildina sem við viljum draga ályktanir um og sú heild samanstendur af stökum. Í þýði Íslendinga væri hver Íslendingur 1 stak.

Dæmi sem unnið verður með.

Við viljum vita hvort háskólamenntaðir, í samanburði við þá sem ekki hafa háskólamenntun, fái hærra laun á Íslandi.

Hér er þýðið okkar afmarkað – við erum að skoða fólk með búsetu á Íslandi. Úr þýði Íslendinga myndum við draga úrtak og athuga menntun og mánaðarlaun hvers einstaklings í úrtakinu.

Spurningin okkar er einnig skýr en hana mætti endurorða á eftirfarandi hátt: eru mánaðarlaun að jafnaði ólík á milli þeirra sem hafa háskólamenntun og þeirra sem hafa hana ekki?

Chapter 2

Tilgátuprófun

Þegar við framkvæmum rannsókn höfum við einhverjar spurningar í huga sem við viljum leitast svara við. Augljóst vandamál er að tölfræðiforrit skilja ekki setningar – við getum ekki beinlínis sett gögn inn í úrvinnslu ásamt spurningunni “eru mánaðarlaun ólík eftir menntun fólks?”.

Það mætti hugsa ferlið á eftirfarandi hátt: Við yfirfærum spurningarnar okkar yfir á *tölfræðimál* sem forritið skilur og ákveðum hvaða aðferð eigi að nota við að svara spurningunni okkar. Forritið vinnur úr gögnunum fyrir okkur og spýtir út niðurstöðum. Þessar niðurstöður eru þó á *tölfræðimáli* og það er okkar verk að skilja hvað þær segja okkur og geta gefið túlkun sem setur þær aftur á *mannamál*.

Í **tilgátuprófun** fáum við einfalt já/nei svar við spurningunum okkar - eftir því hvort niðurstöður reynist marktækar eða ómarktækar.

2.1 Tilgátur

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

2.1.1 Aðaltilgáta

Aðaltilgáta¹ er sú tilgáta sem við höfum áhuga á² og 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 verið munur á launum eftir menntun, en hvaða mun ætti að prófa? Munurinn gæti verið að laun háskólamenntaðra sé 1kr hærri, 10kr, 15kr,

¹Einnig kölluð gagntilgáta, rannsóknartilgáta, (e. alternative hypothesis)

²Undantekningar eru t.d. þegar við beitum marktektarprófum við mat á forsendum

1.000kr, 35.000kr o.s.frv. Tilgátan “það er munur” hefur því ótal möguleika en úrvinnslan krefst þess að við tilgreinum fast gildi sem hægt er að prófa. Við setjum því fram núlltilgátu.

2.1.2 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 háskólamenntun. 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 háskólamenntun og þeirra sem hafa hana ekki, þá sé munurinn á meðallaunum = 0. Það er engin munur á meðallaunum hópana.

Ef niðurstaða er sú að núlltilgátan sé sennilega röng – þá hlýtur andstæða hennar að vera rétt. M.ö.o. ef við teljum ólíklegt að það sé **engin** munur, þá hlýtur jú að vera munur - aðaltilgátan hlýtur þá að vera rétt.

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

Chapter 3

Marktekt

Þegar við skoðum úrtakið okkar gætum við séð mun á launum þeirra sem hafa og hafa ekki háskólamenntun. Tilgangur ályktunartölfræði er þó ekki *spá fyrir um* úrtakið okkar – úrtakstölurnar eru jú beint fyrir framan okkur svo lýsandi tölfræði myndi duga til að lýsa þeim. Munurinn sem við sjáum í úrtakinu okkar gæti þó verið tilkomin af hreinni tilviljun:

1. 20 manna úrtak þar sem launamunur reynist 1.000kr á milli þeirra sem hafa háskólamenntun og þeirra sem hafa ekki háskólamenntun. Í þessu úrtaki er vissulega munur en það er auðvelt að ímynda sér að hann sé ekki að endurspegla mun í þýði. Það er, að þessi 1.000kr munur endurspeglar aðeins þá tilviljun sem má búast við þegar dregið er í úrtak.
2. 2.000 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 engin munur á hópunum í þýði.**

Eftir því sem munurinn er meiri og úrtakið stærra, þeim mun ólíklegri verður að þykja að úrtakstölurnar séu tilkomnar af einskærri tilviljun. Þá situr eftir sú spurning: Hversu mikill munur er nógu mikill? Og þá miðað við hversu stórt úrtak? ¹

Marktektarpróf athugar líkur þess að fá tiltekna úrtakstölur **ef úrtakið 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 sem samanstendur af 2.000 einstaklingum?

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

¹Það er þó auðvitað fleira sem spilar inn í útreikning marktektarprófa - þetta er einföldun.

Ef líkurnar væru hins vegar miklar, þ.e. ef það væru 50% líkur á að fá okkar úrtakstölur þegar úrtakið væri dregið af tilviljun úr þýði þar sem engin munur væri, þá myndum við hika verulega við að henda fram fullyrðingum.

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

3.1 Marktektarmörk

Marktektarmörk (α) lýsa því einmitt hvar við drögum línuna, í sálfræði er nærri alltaf miðað við $\alpha = 0,05$.² Þessi tala (0,05) vísar til líkinda (5%) og við erum þá að segja að við viljum að það séu undir 5% líkur á að við höfum rangt fyrir okkur.³

3.2 Marktektarpróf

Marktektarpróf reikna líkur þess að fá tiltekna niðurstöður þegar núlltilgáta er í raun rétt og kallar þessar líkur p -gildi. P -gildi endurspeglar þannig líkur þess að við höfum rangt fyrir okkur. Ef við höfum ákveðið að miða við marktektarmörk $= 0,05$ þá munu allar niðurstöður þar sem p -gildi er yfir 0,05 þýða að niðurstöður séu ómarktækar. $P > 0,05$ þýðir að það séu yfir 5% líkur á að við höfum rangt fyrir okkur og við vorum jú búin að ákveða að draga mörkin við 5%.

Höfum samt í huga að þó 5% séu lágar líkur, þá eru þær ekki 0% – svo við gætum við haft rangt fyrir okkur. Skoðum því aðeins hvernig niðurstaða gæti verið rétt eða röng með því að gefa okkur eftirfarandi aðstæður í þýði:

Table 3.1: Tafla 1. Mögulegar niðurstöður marktektarprófs ef við þekkjum aðstæður í þýði.

	Það er ekki munur á launum í þýði	Það er munur á launum í þýði
Marktækt	Höfnunarmistök (type I error)	Rétt niðurstaða - Afköst
Ómarktækt	Rétt niðurstaða	Fastheldnimistök (type II error)

²Sumar fræðigreinar miða við 0,01 (ef röng ákvörðun er t.d. tekin í stjarnfræði geta afleiðingar verið miklar). Aðrar fræðigreinar, þar sem mistök eru ekki eins alvarleg, geta miðað við 0,1.

³Að hafa rangt fyrir okkur þýðir hér að við ályktum að það sé, eða sé ekki marktækur munur og að sú ákvörðun sé röng

3.3 Höfnunarmistök

Höfnunarmistök (type I error) er þegar við gerum þau *mistök* að *hafna* núlltilgátu sem er í raun rétt.

Skoðum hólfi 1 í töflu 1: Hér er raunverulega engin munur á launum eftir menntun en niðurstöður eru marktækar. Frá okkar sjónarhorni hugsum við eftirfarandi: P -gildi er undir 0,05 \rightarrow próf er marktækt \rightarrow ef próf er marktækt höfnum við núlltilgátunni \rightarrow ef núlltilgátan segir að það sé engin munur, og hún er röng, þá hlýtur að vera munur! \rightarrow við tökum þá upp aðaltilgátu og ályktum að það sé í raun munur á launum eftir menntun.

Hér höfum við gert þau mistök að segja að það sé munur, þegar það er í raun engin munur.

3.4 Fastheldnimistök

Fastheldnimistök (type II error) er þegar við gerum þau *mistök* að *halda fast* í núlltilgátu sem er í raun röng.

Skoðum hólfi 4 í töflu 1: Það er raunverulega munur á launum eftir menntun en niðurstöður eru ómarktækar. Við myndum hugsa: P -gildi er yfir 0,05 \rightarrow prófið er ómarktækt \rightarrow ef próf er ómarktækt þá getum við ekki hafnað núlltilgátunni \rightarrow núlltilgátan segir að það sé engin munur \rightarrow ef ég get ekki hafnað núlltilgátunni, þá get ég ekki tekið upp aðaltilgátuna \rightarrow get ekki sagt að það sé munur á launum eftir menntun.

Hér höfum við gert þau mistök að segja að það sé ekki munur, þegar það er í raun munur.

Hvort er verra; að gera höfnunarmistök eða fastheldnimistök?

Hvorugt er gott en höfnunarmistök eru oft talin valda meiri skaða heldur en fastheldnimistök. Það byggir á þeim rökum að verra sé að hafa rangar upplýsingar heldur en að missa tímabundið af réttum upplýsingum. Það er, betra að segja “ég er ekki viss” heldur en að koma með fullyrðingu sem er röng.

Það eru þó aðstæður þar sem fastheldnimistök eru verri. Dæmi um slíkt væri ef aukaverkun lyfs væri aukin hætta á heilablóðfalli. Í því tilviki myndum við frekar vilja ganga út frá því að þessi hættulega aukaverkun sé raunveruleg – og hafa rangt fyrir okkur – heldur en að hafa rangt fyrir okkur með því að segja að af lyfinu stafi engin hætta.

Íhugið eftirfarandi fullyrðingar:

- Þunglyndir einstaklingar eru ofbeldisfullir
- Tiltekinn barnabílstóll er öruggur

- Hvernig gætum við yfirfært þessi tvö dæmi á fastheldni- og höfnun-armistöð? Og hvor mistökin þætti okkur verri í hvoru dæmi fyrir sig?

3.5 Túlkun ómarktektar

Skodum hólfi 3 í töflu 1, niðurstöður eru ómarktækar og við hugsum; p -gildið er yfir 0,05 \rightarrow prófið er því ómarktækt \rightarrow við getum ekki hafnað núlltilgátu \rightarrow getum því ekki tekið upp aðaltilgátu \rightarrow við ályktum að það sé engin munur. Í þessu dæmi vitum við að það er í raun engin munur á launum eftir menntun, svo þessi ákvörðun er rétt.

Við getum þó ekki **staðfest núlltilgátu**, þó við náðum ekki að hafna henni, þá getum við heldur ekki fullyrt að hún sé endilega rétt. Ómarktækar niðurstöður segir bara að *úrtakið okkar* náði ekki að sýna fram á nógu mikinn mun á launum svo að þær niðurstöður þættu ólíklegar **ef það væri í raun engin munur**.

Fyrir það fyrsta er fjarstæðukennt að núlltilgátan sé virkilega **rétt** (þó við vorum að ímynda okkur slíkar aðstæður í hólfi 3 í töflu 1). Núlltilgáta sem segir að launamunur sé núll þýðir bókstaflega að munurinn þurfi að vera **núll**; ekki næstum því núll og ekki núll komma eitthvað – munurinn þyrfti að vera = 0,000.... Þriðju breytu áhrif útaf fyrir sig gera það að verkum að það er yfirleitt *einhver* fylgni á milli jafnvel ólíklegustu hluta.

Hugsum þetta út frá öðru sjónarhorni: Ef úrtakið okkar nær ekki að sýna fram á mun á milli hópa, þá þýðir það ekki að sá munur sé ekki raunverulega til staðar. Það þýðir bara að sá munur sem fannst í okkar úrtaki var ekki nægilega mikill miðað við úrtaksstærð til að við næðum að sýna fram á mun.

Þannig hversu mikill þarf munurinn að vera og hversu stórt þyrfti úrtakið að vera, svo að marktektarprófið ætti yfirhöfuð séns á að staðfesta mun sem væri raunverulega er til staðar?

3.6 Afköst

Hvað ef aðaltilgátan er rétt – af einhverri ástæðu vitum við fyrir vísit að það sé raunverulega munur á launum þeirra sem hafa og hafa ekki lokið háskólamenntun. Afköst eru einmitt hæfni prófs, að gefnum tileknum aðstæðum – til að geta borið kennsl á þann mun sem er í raun og veru til staðar í þýði.

Afköst eru líkur þess að fá marktækt próf þegar aðaltilgátan er í raun rétt (það er í raun munur á hópunum). Við getum þó einnig hugsað um þetta sem líkur þess að gera **ekki fastheldnimistöð**. Við getum reiknað afköst í höndunum⁴, með forritlingum á netinu, skoðað sambærilegar rannsóknir og miðað við

⁴Við ætlum ekki að gera það.

að finna a.m.k. eins mikið frávik og kemur fram í þeim ⁵, eða nota staðlaðar áhrifastærðir ⁶

3.7 Loka pælingar um marktekt.

Þegar hingað er komið höfum við framkvæmt marktektarpróf, vitum hvort niðurstöður séu marktækar eða ómarktækar og höfum því skýrt svar við spurningunni hvort munur sé á launum eftir menntun. Gefum okkur að prófið hafi verið marktækt:

Marktækur munur fannst á launum fólks eftir menntun þeirra.

Byggt á þessari staðhæfingu gæti ég hugsað “Glæsilegt – þannig ef ég mennta mig get ég búið við umtalsvert hærri launum fyrir vikið – annars væru niðurstöður ekki marktækar” Eða hvað? Þessi túlkun gæti verið rétt en það eru þó nokkur atriði til að huga að.

“þannig ef ég mennta mig.....”

- Marktækar niðurstöður þýða að það sé **að jafnaði** munur á launum hópanna – það er þó ekki þar með sagt að þær niðurstöður eigi við um **alla**. Það er munur á hópunum heilt yfir, en fyrir einn tiltekin einstakling er sennilega munur - en ekki endilega.
- Sambærilegt dæmi væru þær niðurstöður að *“fullorðnir séu að jafnaði hærri en börn”*. Við skiljum þetta vel en áttum okkur líka á því að það eru vissulega til börn sem eru hávaxnari en en margir fullorðnir.

“...get ég búið við umtalsvert hærri launum fyrir vikið...”

- Núlltilgátan okkar var *H0*: meðallaun háskólamenntaðra = meðallaun óháskólamenntaðra. Það er engin munur á launum = laun eru þau sömu fyrir báða hópa.
- Þær aðaltilgátur sem við prófum geta verið stefnutilgátur (einhliða) eða stefnulausar (tvíhliða). **Stefnulaus** aðaltilgáta segir bara “það er munur” án þess að tilgreina í hvaða átt munurinn er, svo háskólamenntaðir *gætu* þess vegna fengið lægri laun. **Stefnutilgáta** tilgreinir hins vegar í hvaða átt munurinn liggur og væri þannig að prófa þá tilgátu að “háskólamenntaðir hafa hærri laun heldur en óháskólamenntaðir” (eða öfugt).
- Við þurfum að passa að túlka marktektina í samræmi við þær tilgátur sem voru prófaðar.

“...annars væru niðurstöður ekki marktækar”

⁵Við erum þá að skoða aðrar rannsóknir, þar sem viðfangsefni er svipað og úrtaksstærð er svipuð þeirri sem við ætlum okkur að safna til að fá hugmynd af því hve mikill munurinn er sem við þyrftum að finna til að sýna fram á marktekt.

⁶Cohens *d* er dæmi um staðlaða áhrifastærð – þá er gefið viðmið um það hvaða áhrifastærð telst nógu mikil.

- Það er alltaf hættu á að við höfum rangt fyrir okkur (séum að gera höfnun-armistöð). Annað tengt vandamál er að eftir því sem úrtakið er stærra þarf munurinn að vera þeim mun minni svo marktekt náist. Í smáum úrtökum gæti munur reynst ómarktækur þó þessi sami munur næði marktekt í stærra úrtaki (sjá dæmi í upphafi marktektarkafli).

Loka þælingin snýr að orðinu *umtalsvert* en marktekt segir aðeins **hvort** það sé munur en ekki **hve mikill** hann er, sem tekur okkur í næstu umfjöllun.

Chapter 4

Öryggisbil

Niðurstöður okkar voru marktækar - við teljum því ólíklegt að munurinn sem fannst í úrtakinu sé tilkomin af tilviljun en vitum ekki hve *merkilegur* sá munur er. Næsta spurning er því “*Hve mikill er launamunur þeirra sem hafa háskólamenntun og þeirra sem hafa hana ekki?*”

Raunin gæti verið sú að ef við berum saman þá sem hafa ekki háskólamenntun við þá sem hafa háskólamenntun, þá sé marktækur munur á laununum – en að sá munur sé að jafnaði aðeins 10.000kr á mánuði. Þrátt fyrir að munurinn sé *marktækur* þá er hann vissulega ekki *merkilegur* – enda svarar það varla kostnaði að lifa við skertar tekjur til fleiri ára, í þeim tilgangi að sækja sér menntun, ef afrakstur þess eru 10.000kr aukalega á mánuði í kjölfarið.

Gefum okkur skemmtilegri niðurstöður, að munurinn sé *að jafnaði* 150.000 á mánuði. Þetta eru vissulega merkilegri niðurstöður en þær sem við fjölluðum um í dæminu á undan (þar sem munur var aðeins 10.000kr á mánuði).

Bæði dæmin lýsa **punkturþá**, það næsta sem við komumst að vita hver munurinn sé í þýði er með því að athuga hver munurinn er í úrtakinu. Við vitum þó að ekki **allir** þeir sem hafa háskólamenntun fá ákkurat 150.000kr meira í laun á mánuði – enda vissulega fleiri þættir sem spila inn í. Þælingin verður því “hversu mikið gæti þetta spágildi rokkað til eða frá?”.

Þessari þælingu er svarað með **bilspá**, öryggisbil er dæmi um bilspá auk þess að vera sennilega sú mest notaða. Öryggisbil gefa okkur **vikmörk** í kringum punktspánna. Þegar hingað er komið vitum við að “það er munur og hann er að jafnaði 150.000kr á mánuði”. Öryggisbil bætir við þessa túlkun; tökum tvö dæmi:

1. Öryggisbil = [140.000 ; 160.000]
 - Þetta myndi segja okkur að launamunur sé að jafnaði 150.000kr á mánuði - til eða frá 10.000kr.

- Hér er öryggisbilið frekar þröngt, punktspáin helst sú sama og það er lítil óvissa sem fylgir henni.
2. Öryggisbil = [10.000 ; 290.000]
- Þetta myndi segja okkur að launamunurinn sé að jafnaði 150.000kr á mánuði - til eða frá 120.000kr.
 - Hér er öryggisbilið vítt, punktspáin er sú sama en það er mikil óvissa sem fylgir henni. Munurinn gæti verið allt frá mjög ómerkilegum upp í umtalsverður. Launamunur gæti verið allt frá 10.000kr fyrir suma, upp í 290.000kr fyrir aðra.

Í seinna dæminu er óvissan töluvert meiri og niðurstöðurnar eru í raun að gefa okkur minni upplýsingar heldur en í fyrra dæminu. Seinna dæmið væri í raun að segja “það gæti verið finn munur, hann gæti líka verið mikill, eða enginn – hver veit?” Eftir stendur að ef öryggisbilið er mjög vítt, þá erum við litlu nær – *þrátt fyrir að niðurstöður séu marktækar.*