

Inngangur að ályktunartölfræði

Katrín Arndís

2023-01-22

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
4.1	Mikilvægir eiginleikar öryggisbila	16
4.2	Hvað stjórnar breidd öryggisbila?	18
4.3	Öryggið í öryggisbilum	18
5	Z-próf	21
6	Breytingar á skjali	23

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 heldur lítinn hóp (nemendur HR).
 - Þýði þarf að vera skýr, afmarkaður hópur einstaklinga sem eiga eitthvað sameiginlegt.¹
- **Þýðistölur** eru þau gildi sem fengjust 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. Við getum ekki mælt *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 verður unnið með.

Við viljum vita hvort háskólamenntaðir fái hærri laun heldur en þau sem hafa ekki háskólamenntun á Íslandi.

Hér höfum við afmarkað þýði – 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. Rannsóknarspurningin er einnig skýr en hana mætti endurorða á eftirfarandi hátt: Við viljum athuga hvort meðal mánaðarlaun á Íslandi séu að jafnaði ólík á milli tveggja hópa; þeirra sem hafa háskólamenntun og þeirra sem hafa hana ekki.²

²Spurningin gæti þó verið sett upp á ólíkan hátt eftir því hvernig við kjósum að mæla breytturnar og hvaða tölfræðiaðferð við notum

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 skýrt svar við spurningunni okkar; já eða nei – eftir því hvort niðurstöður reynast marktækar eða ómarktækar.

2.1 Tilgátur

Við byrjum á því að setja spurninguna okkar fram í formi **tilgátu**. Í okkar dæmi viljum við vita hvort það sé munur á launum einstaklinga 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 séu 1kr hærri, 15kr,

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

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

1.000kr, 35.000kr, 645.000kr, og svo framvegis. Tilgátan “það er munur” hefur svo marga möguleika að það er ekki hægt að prófa þá alla. 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.

Þetta er ekki það sama og að segja að *allir séu með sömu laun* heldur að þegar við berum saman meðallaun þeirra sem hafa háskólamenntun og þeirra sem hafa hana ekki þá sé meðaltalið hið sama, $\text{munurinn} = 0$, það er engin munur á meðallaunum hópana.

Ef niðurstaðan er að núlltilgátan sé röng – þá hlýtur andstæða hennar að vera rétt. Með öðrum orðum, 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 10.000kr 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 10.000kr munur sé tilviljun sem búast mætti við í smáu úrtaki.
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 ólíklegra 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 **engin** 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. Þegar við höfnum núlltilgátunni, þá tökum við upp aðaltilgátuna og ályktum 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 það væru hins vegar miklar líkur á úrtaksgildunum þegar það væri dregið úr þýði þar sem engin munur er, þá myndum við hika verulega við að henda fram fullyrðingum. Við færum varla að koma með staðhæfingar ef við vitum að það eru 50% líkur á að niðurstaðan sé tilkomin af tilviljun.

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

Marktektapróf 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.

Skóðum hólfi 1 í töflu 1: Hér er raunverulega engin munur á launum eftir menntun en niðurstöður eru marktækar svo við hugsum: 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.

Skóðum hólfi 4 í töflu 1: Það er raunverulega munur á launum eftir menntun en niðurstöður eru ómarktækar svo við hugsum: 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 nú eftirfarandi fullyrðingar:

- Þunglyndir einstaklingar eru ofbeldisfullir
- Tíltekinn barnabílstóll er öruggur

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

3.5 Túlkun ómarktæktar

Skodum hólfi 3 í töflu 1, niðurstöður eru ómarktækar svo við hugsum; p -gildið er yfir 0,05 \rightarrow prófið er 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 tölfræðilegri marktækt.

Þannig hversu mikill þarf munurinn að vera og hversu stórt þyrfti úrtakið að vera, svo að marktæktarprófið 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 fastheldnimistök**. 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 eru marktækar svo nú göngum við út frá því að það sé launamunur eftir menntun. Næsta spurning er þá “*Hve miklu munar á launum þeirra sem hafa háskólamenntun og þeirra sem hafa hana ekki?*”. Möguleikarnir eru ótalmargir, svo gefum okkur tvær aðstæður:

1. Munurinn er að jafnaði 10.000kr á mánuði. Þrátt fyrir að sá munur væri *marktækur* þá er hann vissulega ekki *merkilegur* – enda svarar það varla kostnaði að lifa við skertar tekjur til fleiri ára til að sækja sér menntun, ef afrakstur þess eru 10.000kr aukalega á mánuði í kjölfarið.
2. Munurinn er að jafnaði 150.000kr á mánuði. Þetta eru vissulega merkilegri (og skemmtilegri) niðurstöður heldur en í dæminu á undan.

Bæði dæmin lýsa **punkturpá**; 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 **öll** sem hafa háskólamenntun fá ákkurat 150.000kr hærri laun á mánuði – enda vissulega fleiri þættir sem spila inn í. Þælingin verður þá “hversu mikið gæti þetta spágildi rokkað til eða frá?”. Þessari þælingu er svarað með **bilspá**; öryggisbil er sennilega dæmi um þá bilspá sem er mest notuð. Öryggisbil gefa okkur **vikmörk** í kringum punktspána. Þegar hingað er komið vitum við að “það er munur og hann er að jafnaði 150.000kr á mánuði”. Öryggisbil bætir við þessa túlkun – tókum tvö dæmi:

1. Öryggisbil = [140.000 ; 160.000]
 - Það er, launamunur er að jafnaði 150.000kr – til eða frá 10.000kr.
 - Hér er öryggisbilið þröngt, punktspáin helst óbreytt og það er lítil óvissa sem fylgir henni.
2. Öryggisbil = [10.000 ; 290.000]
 - Það er, launamunur er að jafnaði 150.000kr á mánuði - til eða frá 140.000kr.
 - Hér er öryggisbilið vítt, punktspáin helst óbreytt en það er mikil

óvissa sem fylgir henni. Þessi óvissa endurspeglar í því að munurinn gæti verið allt frá mjög ómerkilegum upp í umtalsverðan; hann gæti verið 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 er sennilega finnr munur, hann gæti líka verið mikill, eða nærri enginn - hver veit?” Eftir stendur að ef öryggisbilið er mjög vítt, þá erum við litlu nær – *þrátt fyrir að niðurstöður séu marktækar*.

4.1 Mikilvægir eiginleikar öryggisbila

Það er algengast að notast við 95% öryggisbil, en hvað þýða þessi 95%? Þessi 95% vísa til þess að ef sama aðferð er notuð síendurtekið, þá munu 95% öryggisbila sem fást innihalda raunverulegt þýðisgildi.

Lengri útskýring á ofangreindri setningu:

Við drögum úrtak af tiltekinni stærð úr þýði og reiknum öryggisbil sem gefur okkur tiltekin gildi. Ef við myndum síðan draga annað úrtak af sömu stærð úr sama þýði, þá myndum við ekki búast við að gildi þess úrtaks yrðu nákvæmlega þau sömu og í fyrsta úrtakinu (við búumst við *einhverjum* breytileika á milli úrtaka) svo þegar við reiknum öryggisbil fyrir seinna úrtakið mætti einnig búast við því að gildi þess yrðu a.m.k. smávægilega frábrugðin þeim sem fengust frá fyrsta úrtakinu.

Ímyndum okkur nú að við myndum beita þessari aðferð síendurtekið; við myndum draga nýtt og nýtt úrtak af sömu stærð úr sama þýði og reikna öryggisbil fyrir hvert úrtak. Að lokum værum við komin með aragúa af öryggisbilum sem væru a.m.k. lítillega ólík hvert öðru.

Þegar við miðum við 95% öryggisbil, þá segir það í raun að af öllum öryggisbilunum sem fást, þá munu 95% þeirra innihalda raunverulega þýðistölu.

Við getum prófað okkur áfram með forritlingum á netinu, til dæmis þessum

Efst á mynd 4.1 sjáum við tvær dreifingar; **svarta** línun sýnir raunverulega dreifingu í þýði en sú **rauða** sýnir úrtaksdreifingu þegar dregið er úr þýði með tilviljunarkenndum hætti. **Græna línun** sem liggur lóðrétt niður myndina endurspeglar raunverulegt þýðis meðaltal. Á línuna hafa verið teiknuð 15 öryggisbil sem gætu fengist. Það er, þarna hafa 15 úrtök verið dregin úr þýðinu með tilviljunarkenndum hætti og öryggisbil reiknað fyrir hvert þeirra. Við sjáum að raunverulegt þýðisgildi liggur innan bilsins í 14 tilfellum af 15. Eitt bilið inniheldur þó ekki þýðisgildið, í þessu úrtaki fellur raunverulegt þýðisgildi utan bilsins.

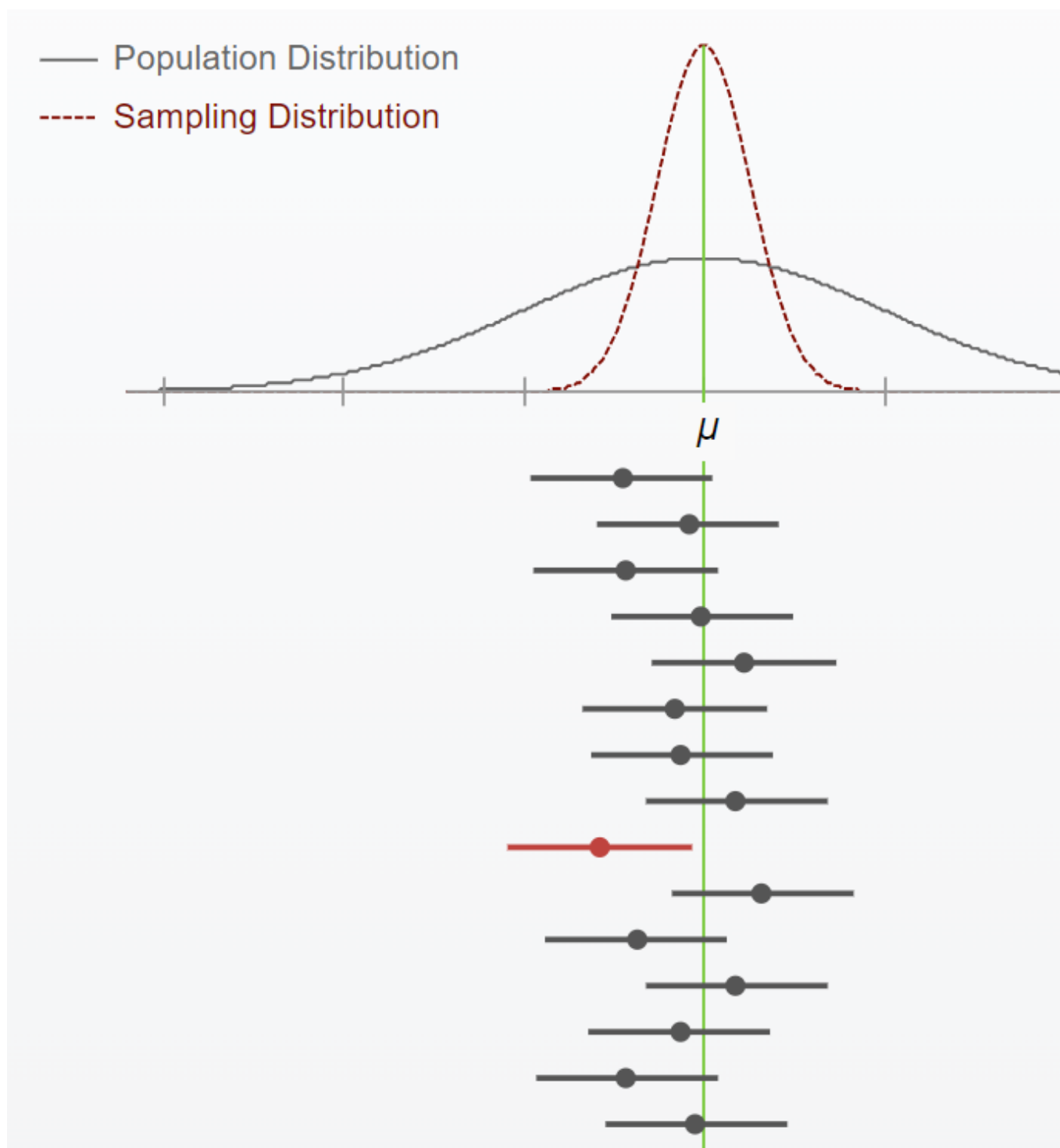


Figure 4.1: Endurtekin öryggisbil

Skoðum nú öryggisbilið sem er rautt. “*Hverjar eru líkurnar á því að þetta öryggisbil innihaldi raunverulegt þýðisgildi?*” Það eru engar líkur á því að þýðisgildið sé inn á þessu tiltekna bili. Við getum því ekki litið á eitt öryggisbil og sagt “*það eru 95% líkur á að raunverulegt þýðisgildi sé innan bilsins*” því það hreinlega liggur ekki inn á bilinu.

Mikilvægur punktur: Við getum ekki túlkað einstaka öryggisbil líkt og það séu 95% líkur á að það innihaldi þýðisgildið. Bilið ýmist inniheldur þýðisgildið, eða ekki. Við vitum ekki hvort okkar öryggisbil sé eitt af þessum 95% eða meðal þeirra 5% sem innihalda þýðisgildið ekki. Við vitum bara að *aðferðin* sem við erum að nota muni í 95% tilfella skila okkur öryggisbili sem inniheldur þýðisgildið.

Skoðum nú efstu tvö öryggisbilin. Á báðum liggur þýðisgildið inn á bilinu; á því fyrsta liggur það við efri mörk bilsins, á því seinna liggur það rétt fyrir miðju.

Mikilvægur punktur: Við vitum ekki hvar á öryggisbilinu þýðisgildið liggur. Við vitum bara að það liggur *sennilega* einhverstaðar innan þess.

4.2 Hvað stjórnar breidd öryggisbila?

Úrtaksstærð Eftir því sem úrtak er stærra \rightarrow þeim mun minni breytileiki er í gögnunum \rightarrow þeim mun minni óvissa er til staðar \rightarrow þeim mun þrengra er bilið.

- Sömuleiðis eftir því sem úrtak er smærra \rightarrow þeim mun meiri breytileiki er í gögnunum \rightarrow þeim mun meiri óvissa er til staðar \rightarrow þeim mun breiðara er bilið.

Staðalfrávik Eftir því sem staðalfrávik er lægra \rightarrow minni breytileiki í gögnunum \rightarrow minni óvissa \rightarrow þrengra bil.

- Sömuleiðis; eftir því sem staðalfrávik er hærra \rightarrow meiri breytileiki í gögnunum \rightarrow meiri óvissa \rightarrow víðara bil.

Öryggið sem við miðum við, hærra öryggi gefur breiðara bil á meðan lægra öryggi gefur þrengra bil. Skoðum hvers vegna í næsta hluta:

4.3 Öryggið í öryggisbilum

Þó langalgengast sé að nota 95% öryggisbil er þó hægt að hækka eða lækka öryggið. Hvaða afleiðingar hefur það?

90% öryggisbil: Ef við notum 90% öryggi, þá þýðir það að 90% bila munu innihalda raunverulegt þýðisgildi. Þegar við **lækkum** öryggið, þá þrengist bilið sömuleiðis. Þetta er vegna þess að kröfurnar eru ekki jafn miklar; aðeins 90% bilana þurfa að innihalda raunverulegt þýðisgildi svo bilið getur verið þrengra (í samanburði við 95%ÖB)

99% öryggisbil: Ef við notum 99% öryggi, þá þurfa 99% bílana að innihalda raunverulegt þýðisgildi. Þegar við **hækkum** öryggið, þá vikkar bilið sömuleiðis. Kröfurnar eru meiri þar sem 99% bílana þurfa að innihalda raunverulegt þýðisgildi, bilið verður víðara fyrir vikið til að tryggja að þýðisgildið muni vera innan bilsins í 99% tilfella.

Athugið að þetta þýðir ekki að **allar** úrvinnslur þar sem notast er við 99% öryggi munu allar gefa víðara bil heldur en þær úrvinnslur þar sem notast er við 95% öryggi. Ofangreindur samanburður gefur einfaldlega upp afleiðingar þess að breyta öryggi í tiltekinni rannsókn (þ.e. sömu gögn með ólíkt öryggi)

Chapter 5

Z-próf

Z-próf er einfaldasta marktektarprófið og byggir á normaldreifingu (z -dreifing). Prófið gerir ráð fyrir að staðalfrávik þýðis sé þekkt. Við þekkjum auðvitað sjaldnast staðalfrávik þýðis þó vissulega séu til dæmi um það. Til dæmis er greindarvísitala stöðluð og fylgir normaldreifingu (þar sem meðaltalið er 100 og staðalfrávik 15).

Við gætum viljað athuga hvort meðalgreind sálfræðinema við HR sé ólík þekktri meðalgreind; við myndum mæla greind í úrtaki nemenda og gætum síðan notað z -próf til að athuga hvort meðalgreind nemendana víki marktækt frá þekktri meðalgreind.

Rifjum upp það sem stóð í kafla 3:

Marktektarpróf athugar líkur þess að fá tilteknar úrtakstölur **ef úrtakið kæmi úr þýði þar sem núlltilgátan er í raun rétt.**

- Núlltilgáta: Meðalgreind sálfræðinema er sú sama og þekkt meðalgreind. Það er engin munur, $\text{munurinn} = 0$.
 - Til upprifjunar: marktektarprófið prófar núlltilgátuna!
- Aðaltilgáta: Meðalgreind sálfræðinema er **ekki** sú sama og þekkt meðalgreind. Það er munur, $\text{munurinn} \neq 0$.

Við getum vel ímyndað okkur að fá meðalgreind 103 í úrtaki af 15 manns, af tilviljun. Ólíklegra væri að fá meðalgreind 130 í 50 manna úrtaki, af tilviljun (ef núlltilgáta væri rétt og meðalgreind sálfræðinema er 100).

Marktektarprófið athugar líkur þess að fá okkar niðurstöður ef núlltilgátan væri rétt og svarar því spurningunni: “*EF núlltilgátan er rétt - það er í raun engin munur á meðaltölunum, hverjar væru þá líkur þess að fá svo mikið frávik í 50 manna úrtaki?*”

- Ef líkur þess eru **yfir** 5% ($p > 0,05$), þá tölum við um að prófið sé **ómarktækt** (gefið $\alpha = 0,05$).
- Ef líkur þess eru **undir** 5% ($p < 0,05$), þá tölum við um að prófið sé **marktækt** (gefið að $\alpha = 0,05$).
 - Litlar líkur \rightarrow ólíklegt að núlltilgáta sé rétt \rightarrow marktækt próf –
 > höfnum núlltilgátu \rightarrow höfnum því að það sé enginn munur \rightarrow
 tókum upp aðaltilgátu \rightarrow ályktum að það sé munur = meðalgreind
 sálfræðinema við HR er **ekki** 100 (þekkt meðalgreind)

Chapter 6

Breytingar á skjali

Útgáfa 1: 18 janúar

Útgáfa 2: 22 janúar

- **Breytingar:**

- Vikmörk í dæmi 2 um öryggisbil lagfært - þar voru þau 120.000 en áttu að vera 140.000

- **Bætt við:**

- Kafla 4.1: Mikilvægir eiginleikar öryggisbila
- Kafla 4.2: Hvað stjórnar breidd öryggisbila?
- Kafla 4.3: Öryggið í öryggisbilum
- Kafla 5: Z-próf