

Frá tilraun til gagna

Tölfræði frá grunni

Anna Helga Jónsdóttir og Sigrún Helga Lund

Háskóli Íslands



HÁSKÓLI ÍSLANDS

Helstu atriði:

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki
- 4 Úrtakshögun
- 5 Blindun
- 6 Orsakasamband

Yfirlit

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki
- 4 Úrtakshögun
- 5 Blindun
- 6 Orsakasamband

Úrtak og þýði

Þýði

Þýði (population) rannsóknar er safn allra viðfangsefna sem draga á ályktanir um.

Úrtak

Úrtak (sample) er safn viðfangsefna sem eru valin úr tilteknu þýði.

- ▶ Hvert úrtak getur eingöngu verið úr einu þýði.
- ▶ Það má taka mörg úrtök úr sama þýðinu.

Dæmi

Þýði: Allir Íslendingar á kjörskrá.

Úrtak: 1200 handahófsvaldar manneskjur af kjörskrá.

Yfirlit

- 1 Úrtak og þýði
- 2 **Breytur**
- 3 Bjagi og breytileiki
- 4 Úrtakshögun
- 5 Blindun
- 6 Orsakasamband

Flokkabreytur og talnabreytur

Breyta

Breyta (variable) er ákveðinn eiginleiki sem við skráum niður eða mælum á viðfangsefnunum í úrtakinu okkar.

Flokkabreytur

Flokkabreytur (categorical variables) taka ekki töluleg gildi heldur segja, eins og nafnið gefur til kynna, til um það hvaða flokki viðfangsefnið tilheyrir.

Talnabreytur

Talnabreytur (numerical variables) taka töluleg gildi sem eru mæld í tilteknum einingum.

Flokkabreytur og talnabreytur - dæmi

Dæmi

Hugsum okkur að við höfum mælt eftirfarandi breytur hjá úrtaki kjósenda. Af hvaða gerð eru þær?

Kyn (kk/kvk/annað)

Aldur (í árum)

Menntunarstig (grunnskóli, framhalds-/iðnskóli, háskóli)

Meðaltekjur á mánuði (í krónum)

Kjördæmi (Rvk N/S, SV, SU, NV, NA)

Póstnúmer (101, 103, osfrv. . .)

Fjöldi íbúa á heimili, að einstaklingi meðtöldum (1, 2, . . .)

Flokkabreytur og talnabreytur - lausn

Dæmi

Hugsum okkur að við höfum mælt eftirfarandi breytur hjá úrtaki kjósenda. Af hvaða gerð eru þær

F Kyn (kk/kvk/annað)

T Aldur (í árum)

F Menntunarstig (grunnskóli, framhalds-/iðnskóli, háskóli)

T Meðaltekjur á mánuði (í krónum)

F Kjördæmi (Rvk N/S, SV, SU, NV, NA)

F Póstnúmer (101, 103, osfrv. . .)

T Fjöldi íbúa á heimili, að einstaklingi meðtöldum (1, 2, . . .)

Raðaðar og óraðaðar flokkabreytur

Röðuð flokkabreyta

Þegar flokkabreyta er **röðuð** (ordinal categorical variable) er flokkum hennar raðað í stærðarröð.

Óröðuð flokkabreyta

Þegar flokkabreyta er **óröðuð** (categorical variable) er flokkum hennar ekki raðað í stærðarröð.

Raðaðar og óraðaðar flokkabreytur - dæmi

Dæmi

Hverjar af flokkabreytunum úr dæminu áðan eru ráðaðar og hverjar eru óraðaðar?

Kyn (kk/kvk/annað)

Menntunarstig (grunnskóli, framhalds-/iðnskóli, háskóli)

Kjördæmi (Rvk N/S, SV, SU, NV, NA)

Póstnúmer (101, 103, osfrv. . .)

Raðaðar og óraðaðar flokkabreytur - lausn

Dæmi

Hverjar af flokkabreytunum úr dæminu áðan eru ráðaðar og hverjar eru óraðaðar?

Óröðuð Kyn (kk/kvk)

Röðuð Menntunarstig (grunnskóli, framhalds-/iðnskóli, háskóli)

Óröðuð Kjördæmi (Rvk N/S, SV, SU, NV, NA)

Óröðuð Póstnúmer (101, 103, osfrv. . .)

Samfelldar og strjálar breytur

Samfelldar breytur

Þegar talnabreyta getur tekið hvaða gildi sem er á einhverju bili þá segjum við að hún sé **samfelld** (continuous). Eingöngu talnabreytur geta verið samfelldar.

Strjálar breytur

Ef breytur eru ekki samfelldar segjum við að þær séu **strjálar** (discrete). Allar flokkabreytur eru strjálar og sumar talnabreytur.

Þegar strjálar breytur geta tekið **mjög mörg gildi** (t.d. fjöldi bifreiða sem aka um gatnamót Miklubrautar og Kringlumýrarbrautar á hverjum degi) er yfirleitt unnið með þær eins og ef um **samfelldar** breytur væri að ræða.

Samfelldar og strjálar breytur - dæmi

Dæmi

Hverjar af talnabreytunum úr dæminu áðan eru samfelldar og hverjar eru strjálar?

Aldur (í árum)

Meðaltekjur á mánuði (í krónum)

Fjöldi íbúa á heimili, að einstaklingi meðtöldum (1, 2, ...)

Samfelldar og strjálar breytur - lausn

Dæmi

Hverjar af talnabreytunum úr dæminu áðan eru samfelldar og hverjar eru strjálar?

samfelld Aldur (í árum)

samfelld Meðaltekjur á mánuði (í krónum)

strjál Fjöldi íbúa á heimili, að einstaklingi meðtöldum (1, 2, ...)

Svarbreytur og skýribreytur

Svarbreytur og skýribreytur

Fyrir sérhvert viðfangsefni mun gildi **skýribreytu** (explanatory variable) þess hafa áhrif á það hvaða gildi **svarbreytan** (response variable) mun taka.

Til einnar svarbreytu geta svarað margar skýribreytur sem hafa áhrif á hana.

Dæmi

Hugsum okkur að við viljum rannsaka áhrif matarræðis á þyngd lamba. Við höfum breyturnar

- ▶ matarræði
- ▶ þyngd

Hvor breytan er svarbreyta og hvor breytan er skýribreyta?

Yfirlit

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki**
- 4 Úrtakshögun
- 5 Blindun
- 6 Orsakasamband

Breytileiki

Breytileiki

Breytileiki (variability) verður vegna þess að fyrirbærin sem við erum að skoða eru slembni háð og því geta útkomur mælinganna breyst í hvert sinn sem tilraunin er framkvæmd.

- ▶ Þessi breytileiki veldur því að niðurstöðurnar okkar geta breyst í hvert sinn sem rannsóknin er endurtekin.
- ▶ Endurtekningar gera okkur kleift að meta breytileika mælinganna.

Endurtekning

- ▶ Við gerum ráð fyrir að það sé alltaf einhver breytileiki í mælingunum okkar - þær séu slembnar (random).
- ▶ Niðurstöðurnar geta breyst í hvert sinn sem nýtt úrtak er valið og rannsóknin endurtekin.
- ▶ Um leið og við höfum mælingu á fleiri en einu viðfangsefni, það er **endurtekningu**, höfum við einhverja hugmynd um það á hvaða bili mælingarnar geta legið.
- ▶ Því fleiri mælingar sem við höfum, því betur vitum við hversu breytilegar þær geta verið.

Bjagi

Bjagi

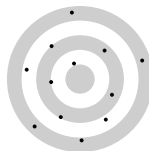
Bjagi (bias) verður þegar aðferðirnar gefa markvisst bjagaða mynd af þýðinu sem verið er að skoða.

- ▶ Viðfangsefni valin á kerfisbundið bjagaðan hátt: **Úrtaksbjagi**.
- ▶ Góð **úrtakshögun** lágmarkar úrtaksbjaga.
- ▶ Truflandi áhrif rannsakanda og viðfangsefna:
Rannsakandabjagi og lyfleysuáhrif.
- ▶ **Blindun** lágmarkar rannsakandabjaga og lyfleysuáhrif.

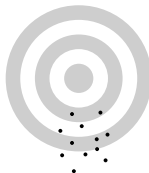
Bjagi og Breytileiki



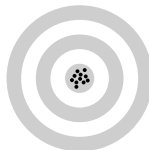
Mikill bjagi,
lítill breytileiki



Lítill bjagi,
mikill breytileiki



Mikill bjagi,
mikill breytileiki



Lítill bjagi,
lítill breytileiki

Mynd: Bjagi og breytileiki

Yfirlit

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki
- 4 Úrtakshögun**
- 5 Blindun
- 6 Orsakasamband

Slembival

Slembival

Það að velja slembið (random), eða **slembival**, þýðir að velja handahófskennt þannig að öll viðfangsefni eru jafnlíkleg til að vera valin.

Úrtak sem er valið með slembivali kallast **slembiúrtak** (random sample).

Við munum fjalla um þrjár gerðir af slembiúrtökum:

- ▶ einfalt slembiúrtak
- ▶ lagskipt slembiúrtak
- ▶ parað slembiúrtak

Einfalt og lagskipt slembiúrtak

Einfalt slembiúrtak

Þegar við framkvæmum **einfalt slembiúrtak** (simple random sample) veljum við einstaklinga af handahófi úr öllu þýðinu.

Lagskipt slembiúrtak

Þegar við framkvæmum **lagskipt slembiúrtak** (stratified random sample) er þýðinu fyrst skipt niður í nokkur lög eða hópa og síðan eru viðfangsefni valin með einföldu slembiúrtaki úr hverju lagi fyrir sig.

Fjöldi viðfangsefna sem valinn er úr hverju lagi verður að vera ákveðinn fyrirfram en hann má vera mismikill eftir lögum.

Gott að nota þegar fjöldi viðfangsefna í hverju lagi er mjög misstór.

Parað slembiúrtak

Parað slembiúrtak

Parað slembiúrtak (paired random sample) fæst þegar viðfangsefnin í þýðinu eru pörð saman tvö og tvö og síðan er ákveðinn fjöldi para valinn af handahófi í úrtakið.

Gott að nota þegar mikið af utanaðkomandi áhrifum trufla mælingarnar okkar. Þá eru pörð saman viðfangsefni sem verða fyrir svipuðum „truflunum”.

Dæmi

Af hvaða gerð eru eftirfarandi slembiúrtök?

- ▶ 1200 handahófsvaldir kjósendur
- ▶ 600 handahófsvaldir karlar og 600 handahófsvaldar konur
- ▶ 600 handahófsvalin hjón

Hvað ef slembiúrtak er ógerlegt?

Stundum valda erfiðleikar í framkvæmd því að við getum ómögulega valið slembiúrtak úr þýði. Þá er farin önnur af tveimur leiðum:

- 1 Að skilgreina þýðið upp á nýtt þannig að úrtakið verði slembiúrtak.
 - ▶ Þá eiga ályktanir okkar eingöngu við um „nýja þýðið“. Viljum við það?
- 2 Að sætta sig við bjagann.
 - ▶ Gerum grein fyrir úrtaksbjaganum í umfjöllun okkar.
 - ▶ Ræðum ítarlega hvaða afleiðingar hann getur haft í för með sér.
 - ▶ Er hægt að gera ráð fyrir að bjaginn sé léttvægur með tilliti til þess sem við erum að rannsaka?

Aðgengisúrtök

Aðgengisúrtök

Aðgengisúrtök eru fengin þegar eingöngu eru framkvæmdar mælingar á þeim viðfangsefnum sem eru (þægilega) aðgengileg rannsakendum.

Þar verður úrtaksbjagi vegna þess að ákveðin viðfangsefni þýðisins eru líklegri til að vera aðgengileg rannsakendum en önnur.

Sá bjagi getur verið mikill og valdið því að ekki er hægt að draga ályktanir út frá fengnum mælingum.

Sjálfböðaliðauúrtök

Sjálfböðaliðauúrtök

Sjálfböðaliðauúrtök eiga við þegar viðfangsefnin eru fólk og þá eru eingöngu framkvæmdar mælingar á þeim sem bjóða sig fram til þess.

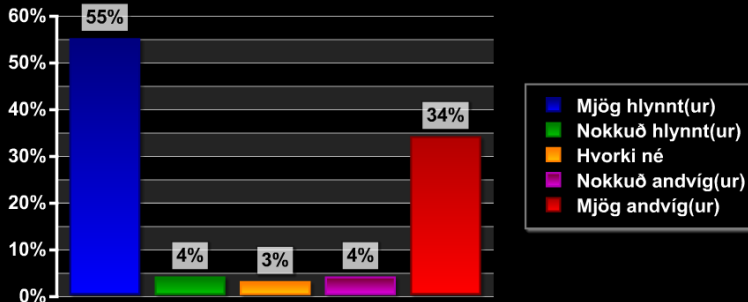
Hér verður úrtaksbjagi vegna þess að ákveðin viðfangsefni geta verið líklegri til að bjóða sig fram en önnur.

Oft getur sá bjagi orðið svo mikill að lítið er hægt að álykta út frá þeim mælingum sem fengnar eru.

Sjálfboðaliðaúrtak

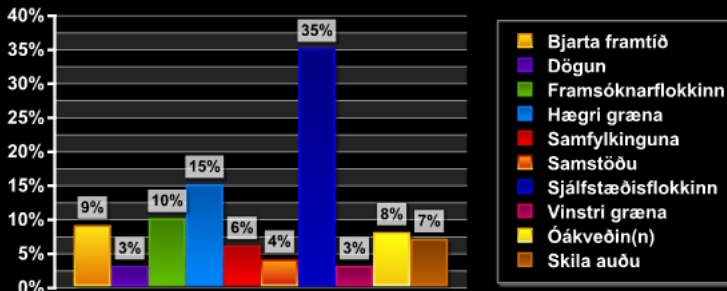
Reykjavík síðdegis spurði (23. nóvember 2011):

Hversu hlynnt(ur) eða andvíg(ur) ertu því að gera hlé á aðildarviðræðum við ESB?



Sjálfboðaliðaúrtak - slembiúrtak

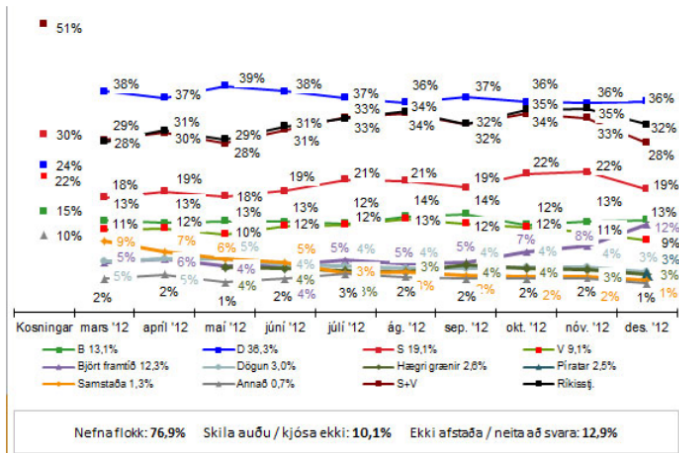
Reykjavík síðdegis spurði (13.-17. desember 2012):
Ef gengið yrði til þingkosninga nú, hvaða flokk myndirðu kjósa?



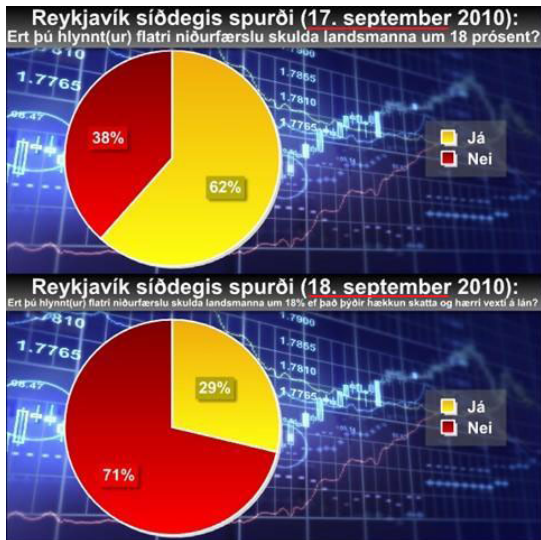
Sjálfboðaliðaúrtak - slembiúrtak

FYLGJI FLOKKA EF KOSIÐ YRÐI TIL ALÞINGIS Í DAG OG STUÐNINGUR VIÐ RÍKISSTJÓRNINA

03.01.2013



Orðalag



Vöntun mælinga

Vöntun mælinga

Oft er það svo að ekki tekst að framkvæma mælingu á öllum viðfangsefnunum sem við höfum valið í úrtakið okkar. Þá tölum við um að það **vanti mælingu** fyrir viðkomandi viðfangsefni (missing value).

- ▶ Það má ekki líta á eingöngu þær mælingar sem náðist að framkvæma sem úrtakið okkar.
- ▶ Oft er líklegra að það vanti frekar mælingar hjá ákveðnum viðfangsefnum heldur en öðrum.
- ▶ Þá er um leið líklegra að þau viðfangsefni lendi ekki í minnkaða „úrtakinu“.
- ▶ Það veldur úrtaksbjaga!

Yfirlit

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki
- 4 Úrtakshögun
- 5 Blindun**
- 6 Orsakasamband

Rannsakandabjagi og lyfleysuáhrif

Rannsakandabjagi

Rannsakandabjagi (experimenters bias) verður þegar væntingar rannsakanda um áhrif inngrips hafa áhrif á mælingarnar á viðfangsefnunum.

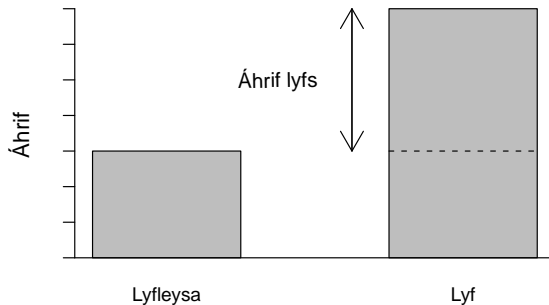
Lyfleysa

Lyfleysa (lyfleysa) er sérhvert inngrip sem viðfangsefni telur ranglega að sé inngripið sem mæla á.

Lyfleysu áhrif

Þann mun í mælingum viðfangsefna sem við sjáum fyrir og eftir lyfleysuinngrip köllum við **lyfleysuáhrif** (placebo effect).

Lyfleysuáhrif



Mynd: Lyfleysuáhrif

Einblindar og tvíblindar rannsóknir

Tvíblindar rannsóknir

Þegar rannsókn er **tvíblind** (double-blind) vita hvorki rannsakandi né viðfangsefni tilraunarinnar hvaða inngrip hvert viðfangsefni hlýtur. Athugið að inngrip getur verið lyfleysuinngrip.

Einblindar rannsóknir

Þegar rannsókn er **einblind** (single-blind) vita viðfangsefnin eða rannsakandinn ekki hvaða inngrip viðfangsefnin hljóta.

Að draga ályktanir

- ▶ Ályktunartölfræði snýst um að draga ályktanir um þýði út frá rannsóknum á úrtaki.
- ▶ Því fleiri mælingar sem við höfum, því líklegra er að við getum alhæft um þýðið út frá úrtakinu okkar.
- ▶ Meginreglan er: „Því fleiri mælingar því betra“

Yfirlit

- 1 Úrtak og þýði
- 2 Breytur
- 3 Bjagi og breytileiki
- 4 Úrtakshögun
- 5 Blindun
- 6 Orsakasamband**

Stýrð tilraun

Markmið margra rannsókna er að sýna fram á áhrif tiltekinna **inngripa** á viðfangsefnin okkar. Þá þarf að framkvæma **stýrða tilraun**.

Stýrð tilraun

Til að rannsókn geti flokkast sem **stýrð tilraun** (controlled experiment) þurfa tvö atriði að vera til staðar:

- 1 Rannsakandinn getur stýrt því hvaða viðfangsefni hljóta hvaða inngríp.
- 2 Mælingar eru framkvæmdar á viðfangsefnum bæði fyrir og eftir að inngrípinu er beitt.

Orsakasamband

Orsakasamband

Orsakasamband (causation) milli tveggja breyta er þegar gildi einnar breytu hefur áhrif á þau gildi sem önnur breyta mun taka.

Stýrðar tilraunir leyfa okkur að fullyrða um orsakasamband.

Góð tilraunahögun

Hvaða skilyrði ætti stýrð tilraun að uppfylla?

Sérhver rannsakandi ætti að leitast við að tilraun hans uppfylli eftirfarandi skilyrði:

- 1 Úrtakshögun.

Viðfangsefni eru valin úr þýðinu með slembiúrtaki og/eða skipt í hópa með slembivali.

- 2 Blindun.

Rannsóknin er tvíblind en einblind ef því verður ekki komið við.

- 3 Endurtekningar.

Inngripinu er beitt á endurtekinn fjölda viðfangsefna.

Uppfylli stýrð tilraun þessi skilyrði má fullyrða um orsakasamband.

Thursday, 16 August 2012

Sleeping With Your Shoes On Gives You Headaches



I read [this article](#) in The Daily Telegraph today. It is yet another article where the difference between correlation and causation, a difference which must be known to the journalist and editors, is blithely ignored for the sake of a better story.

<http://positionofignorance.blogspot.is/2012/08/sleeping-with-your-shoes-on-gives-you.html>