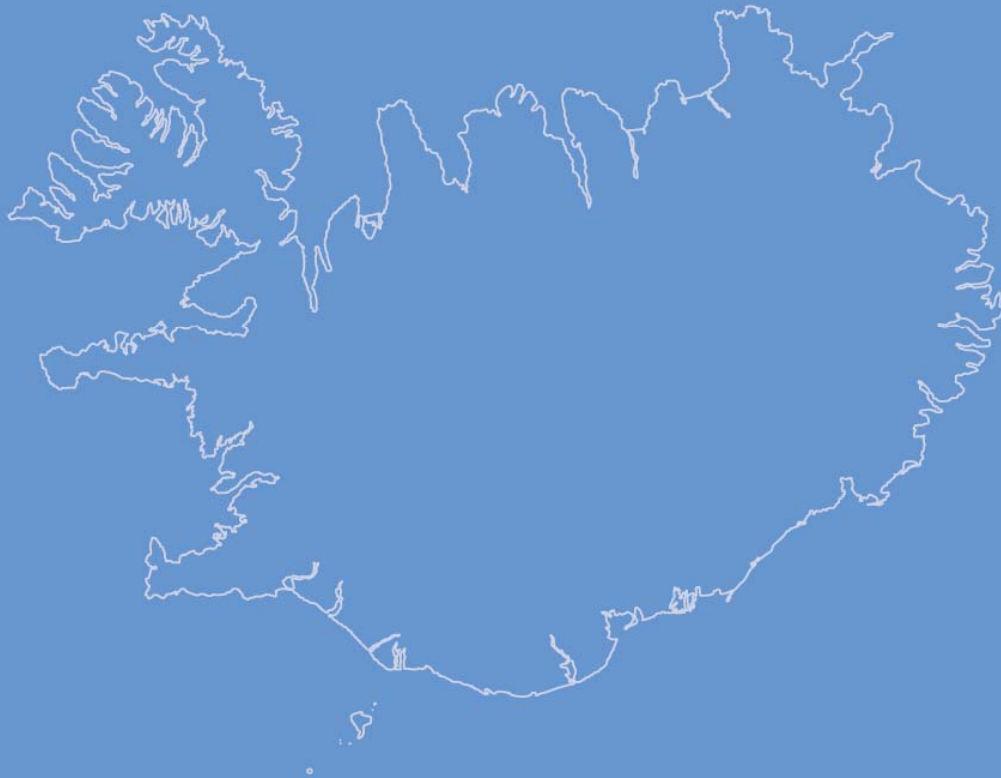




LÝÐHEILSUSTÖÐ

Heilsa og líðan Íslendinga 2007 - Framkvæmdaskýrsla -

Stefán Hrafn Jónsson
Jón Óskar Guðlaugsson
Haukur Freyr Gylfason
Dóra Guðrún Guðmundsdóttir



Febrúar 2011

Heilsa og líðan Íslendinga 2007: Framkvæmdaskýrsla

Útgefandi: Lýðheilsustöð, febrúar 2011

© Lýðheilsustöð.

Höfundar: Stefán Hrafn Jónsson, Jón Óskar Guðlaugsson Haukur Freyr Gylfason

Dóra Guðrún Guðmundsdóttir

Höfundar þakka samstarfsfólki á Lýðheilsustöð og öðrum fyrir yfirllestur og ábendingar.

Tillaga að heimildanotkun:

Stefán Hrafn Jónsson, Jón Óskar Guðlaugsson, Haukur Freyr Gylfason og Dóra Guðrún Guðmundsdóttir. (2011). *Heilsa og líðan Íslendinga 2007: Framkvæmdaskýrsla*.

Reykjavík: Lýðheilsustöð. Sótt 18. febrúar 2011 af

<http://www.lydheilsustod.is/rannsoknir/heilsa-og-lidan-2007>

Viðauki A. Úrvinnsla gagna með vigt

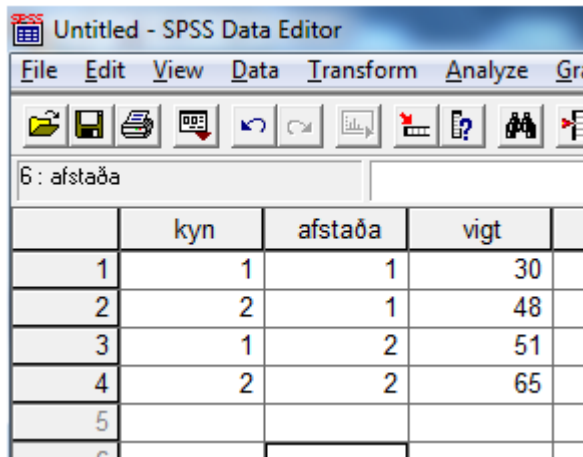
Þegar rætt er um vigt í úrvinnslu gagna í tölfræðiforritum eins og SPSS er mikilvægt að gera greinarmun á tvenns konar vigt: Annars vegar úrtaksvigt (e. sample weight) og hins vegar endurtekningarvigt (e. repetition weight). Til þess að skýra nánar þennan mun eru tekin hér tvö tilbúin dæmi.

Dæmi 1. Vigt fyrir endurtekningu

Í skoðanakönnun er afstaða karla og kvenna til stjórnar og stjórnarandstöðu mæld. Niðurstöður (fjöldatölur) eru birtar í töflu hér að neðan.

Afstaða	Karlar	Konur
Styðja stjórn	30	48
Styðja stjórnarandstöðu	51	65

Lesendur geta hæglega slegið þessar tölur í SPSS.



The screenshot shows the SPSS Data Editor interface. The title bar reads 'Untitled - SPSS Data Editor'. The menu bar includes 'File', 'Edit', 'View', 'Data', 'Transform', 'Analyze', and 'Gra'. The toolbar contains various icons for file operations and data manipulation. The active window shows a data table with the following content:

	kyn	afstaða	vigt
1	1	1	30
2	2	1	48
3	1	2	51
4	2	2	65
5			

Skipun í SPSS gæti þá verið svona:

```
weight by vigt.
```

```
crosstab table afstada by kyn /cell=col count/ stat = chi.
```

SPSS gefur þá kíkvaðrat-próf með réttum fjölda svarenda og frígráðum.

Þessi notkun endurtekningarvigtar er sérlega heppileg þegar svarfjöldi hleypur á tugum eða hundruðum þúsunda þar sem spara má innslátt verulega.

Dæmi 2. Úrtaksvigt

Tilbúið dæmi um úrtaksrannsókn um einkenni foreldra sem ættleiða annars vegar börn erlendis frá og hins vegar börn frá Íslandi.

Þýði/úrtak	Ísland	Útlönd	Samtals
Þýði	134	322	456
Einfalt tilviljunarúrtak	29	71	100
Lagskipt tilviljunarúrtak	50	50	100

Rannasakandi vinnur með þýði samtals 456 ættleiðinga frá 1990–2008. Fjármagn er takmarkað og aðeins möguleiki á að fram fari viðtalskönnun hjá 100 foreldrum. Til að hámarka möguleika á að greina mun (tölfræðilegt afl), ef einhver er til staðar, er tekið lagskipt úrtak, jafnstór hópur foreldra sem ættleiddu barn frá Íslandi og erlendis frá, 50 foreldrar úr hvorum hóp. Ef um einfalt tilviljunarúrtak hefði verið að ræða mætti búast við 29 foreldrum úr fyrri hópnum og 71 úr þeim seinni. Þegar um samanburð tveggja hópa er að ræða þarf í raun ekki að notast við vigt heldur má gera ráð fyrir tveimur tilviljunarúrtökum úr ólíku þýði (e. independent samples). Ef lýsa á einkennum allra foreldra, sem ættleiða börn, með þessu úrtaki þarf að taka tillit til þess að um er að ræða lagskipt úrtak. Hvert og eitt foreldri í þessu tveggja laga úrtaki endurspeglar ekki 4,56 önnur foreldri líkt og ef valið hefði verið einfalt tilviljunarúrtak (456/100). Lagskiptingin veldur því að hvert foreldri í 50 foreldra úrtaki, sem ættleiddir barn frá Íslandi, endurspeglar 2,68 (134/50) foreldra í þýði. Á sama hátt endurspeglar hvert foreldri í 50 foreldra úrtaki, sem ættleiða barn erlendis frá, 6,44 (322/50) foreldra í þýði.

Vogtölur fyrir tvær tegundir úrtaks

	Ísland	Útlönd
Einfalt tilviljunarúrtak	4,6	4,6
Lagskipt tilviljunarúrtak	2,7	6,4

Þegar um er að ræða einfalt tilviljunarúrtak er vogtölum oftast sleppt í skráningu gagna. Ástæðan er sú að meðaltöl eða hlutföll breytast ekki þegar allir einstaklingar eru með sömu vogtöluna. Því er nægjanlegt að margfalda prósentur með fjölda í þýði til að finna fjölda einstaklinga sem ályktað er að hafi tiltekið einkenni (e. point estimate). Á þessu eru ýmsar undantekningar. Þegar svarhlutfall er t.d. ójafnt eftir skilgreindum hópum þá er hægt að nota vigt til að koma í veg fyrir að tilteknir hópar hafi of mikið vægi í niðurstöðum.

Hefðbundnar útgáfur af SPSS greina ekki á milli vigtar fyrir endurtekningu og vigtar fyrir úrtaksgerð. Þess vegna er öll staðalvilla (þar með talið staðalvilla meðaltals og hlutfalla) rangt reiknuð ef notast er við vigt. Þetta gerist vegna þess að SPSS gerir ráð fyrir röngum fjöldatölum eins og um væri að ræða endurtekningarvigt (þ.e. reiknað með röngum fjölda einstaklinga sem svara) og því gefa tölfræðiþróf á slíkum gögnum rangar niðurstöður. Til þess að bæta úr þessu hafa verið farnar tvær meginleiðir.

Í fyrsta lagi er vogtalan stöðluð; það er gert með því að deila vogtölunni með meðalvigt. Með staðlaðri vogtölu fást sömu meðaltöl og hlutföll og með óstöðluðum vogtölum. Þannig gefa t.d. þrír einstaklingar með vogtölurnar 90, 100 og 110 sömu útreiknuðu meðallaunin ef þeir væru með vogtölurnar 0,9, 1,0 og 1,1. Fjöldatölur eru að meðaltali réttar þar sem vogtölur eru að meðaltali 1,0. Sumir svarendur eru með vogtölu hærri en einn og aðrir með lægri en einn og því er ekki tryggt að staðalvilla sé reiknuð rétt í tölfræðilegum líkönum. Með þessari aðferð eru tölfræðiþróf mun nær því að vera rétt en með óstöðluðum vogtölum.

Í öðru lagi hefur þróun í tölfræðiforritum verið mikil og unnar hafa verið viðbætur sem reikna staðalvillu metils (e. standard error of estimate) rétt. Hér er minnst á tvö tölfræðiforrit, annars vegar R-umhverfið og hins vegar SPSS. Kosturinn við R-umhverfið er að það er alveg ókeypis en þykir stundum flókið í notkun. Kosturinn við SPSS er að margir þekkja það forrit en þar þarf að greiða háar upphæðir fyrir viðbótina. Hér á eftir eru birtar vefslóðir þar sem lesa má um áður nefnd forrit.

Umfjöllun um complex survey design í R-umhverfinu:

<http://faculty.washington.edu/tlumley/survey/>

<http://faculty.washington.edu/tlumley/survey/survey-wss.pdf>

IBM SPSS (Nú PASW) Complex Sample

<http://www.spss.com/software/statistics/complex-samples/>

Umfjöllun um tölfraði flókinna úrtaka í SAS, SPSS og STATA.
<http://scialert.net/fulltext/?doi=ajms.2010.33.39&org=11>

Dæmi um úrvinnslu gagna í R

Hér er stutt dæmi um tölfraðilega úrvinnslu gagna úr Heilsu og líðan 2007 með R. Hér er brugðið á það ráð að gera gögn til í SPSS, vista út fyrir R og keyra líkön í R. Er það gert til að lágmarka skref sem nauðsynlegt er að vinna í R sem hlotið hefur minni útbreiðslu en SPSS. Forritunarkóði er skrifaður hér með `courier-lettri` til að aðgreina frá öðrum texta. Línur í R, sem byrja á #, eru athugasemdir fyrir þann sem skrifar og les kóðann. Það sama á við um línur sem byrja á * í SPSS.

Undirbúningur í SPSS:

*Hér er búin til breyta fyrir strata sem notuð er í síðari úrvinnslu.

*** 2007 gögn**

```
compute strata = hofudb*10+ faer.  
compute id2 = $casenum.
```

Sameinuð 2007 OG 2009 gögn

```
* Við skráningu gagna í framhaldsrannsókn var tekið upp nýtt kerfi við  
* nöfn á breytum. Ef notandi gagna vinnur með sameinaða skrá  
* 2007 og 2009 gagna þá þarf að skipta 2 skipanalínum hér að ofan út  
* fyrir neðangreindar 3 línur.
```

```
recode buseta2s (1 = 1) (2= 2) into hofudb.  
compute strata = hofudb*10+ faerfl.  
compute id2 = $casenum.
```

```
* Hér að neðan eru valdir þeir einstaklingar sem eru með skráð gildi  
* (valid nonmissing value) á vigt og á fylgibreytu. Þeir eru vistaðir út  
* í Portable skrá sem R getur lesið.
```

```
temp.  
sel if vigt gt 0 and offita ge 0.  
EXPORT OUTFILE= "S:\OFFITA.por".
```

Úrvinnsla í R (texta hér að neðan má afrita í textaskrá, breyta og keyra í R):

```
# ATH Það þarf að vera búið að setja inn foreign og survey-viðbætur
# áður en þær eru ræstar upp með library-skipun.

# Fyrsta skipun er að opna skipanasafnið
# foreign til að lesa spss-gögn. Foreign þarf að sækja
# í fyrsta skipti sem það er notað í R.
# Eftir það er það ræst með library(foreign).
# Það sama á við um library(survey) sem er pakki til að
# vinna tölfræði fyrir flókin úrtök sem jafnan má finna í könnunum.

library(foreign)
library(survey)

# Næst er skipun til að lesa inn SPSS portable gagnaskrá.
# Hér eru gögnin sett inn í hlutinn HL07. Notandi gagna þarf að
# sjálfsögðu að skipta um slóð og nafn á skrá til samræmis við sitt
# verkefni og skráaskipulag.
# Staðsetning á drifi hvers og eins er skipt út fyrir staðsetningu
# innan gæsalappa.

HL07 <- as.data.frame(read.spss("S:/R/OFFITA.por",
                               use.value.labels = FALSE))

# Hér er úrtakssnið skilgreint með tilvísun í gögnin, sem skilgreind eru
# hér að ofan (HL07), og nokkrar valdar breytur úr gagnaskrá.
# As.data.frame() er fall til að færa gögn á aðgengilegra format fyrir #
# úrvinnslu í R.

HL7design <-svydesign(id=~1, strata=~STRATA, weights=~VIGT2, data=HL07)

# Ágætt að sjá hvernig þetta lítur út með þessari skipun:
# summary(HL7design)
```

```

# Meðaltal o.fl. er reiknað með þessari skipun þar sem kallað
# er á HL7design:

svymean(~OFFITA, HL7design)

# Hér er skilgreint logistic aðhvarfsgreiningarlíkan með
# tilvísun í HL7design.

logisticoffita <- svyglm(OFFITA~KARLAR+EFTIRL+ ALD3039 +
  ALD4049 + ALD5059 + ALD6069 + ALD7079 + HOFUDB +
  FRAMENNT+ HAMENNT + MISMENNT +TEKJUR3 + MTEKJUR+
  EFTIRL + MEFTIRL + REGLBUR + MREGLBUR + TRADREGL +
  TRADSJ + MTRAD + SYKURMIK + MSYKUR + PAKKAAR3 ,
  family=quasibinomial, design=HL7design)

summary(logisticoffita)

# Glöggir lesendur sjá að notast er við breyturnar ID2 STRATA og VIGT2
# sem allar koma úr SPSS-skrá.

# Endir á R-skrá.

```