

Guðmundur Arnkelsson

Stoðkver

í Aðferðafræði II

Verkefni og efnisspurningar

Viðbótarefni frá 1. útgáfu

Háskóli Íslands • Félagsvísindadeild
Reykjavík 2003

© 1999–2003 Guðmundur B. Arnkelsson

Janúar 2003 Vefútgáfa

Öll réttindi áskilin.

Rit þetta má ekki afrita með neinum hætti, svo sem með ljósritun, ljósmyndun, prentun, hljóðritun eða á annan sambærilegan hátt, að hluta eða í heild, nema með skriflegu leyfi höfundar.

Nemendur í Aðferðafræði II við félagsvísindadeild Háskóla Íslands mega prenta eða láta prenta sér eintak í því skyni að nota skjalið með 1. útgáfu Stoðkvers í Aðferðafræði II.

FORMÁLI

Stoðkver í Aðferðafræði II: Verkefni og efnisspurningar, á uppruna sinn í forverum sínum, sjöttu útgáfu Dæmakvers í tölfræði eftir Þorlák Karlsson og Dæmakvers í Aðferðafræði II. Hið síðarnefnda kom út í þremur hlutum sem vefútgáfa á vefsvæði námskeiðsins Aðferðarfræði II vorið 2001. Gróft séð má áætla að fjöldi verkefna og efnisspurninga úr Aðferðafræði II hafi aukist um þriðjung í vefútgáfunni og hafi enn aukist um helming í þessari útgáfu.

Eins og í vefútgáfunni er leitast við að veita upplýsingar um úrlausnir verkefna með því að birta milliniðurstöður og jafnvel alla útreikninga. Einnig eru birt stutt svör við helmingi allra efnisspurninga og eru að jafnaði aðeins gefin svör við efnisspurningum sem eru númeruð með oddatölum.

Í 2. útgáfu Stoðkversins hefur verið bætt við 20 nýjum verkefnum og efnisspurningum auk ýmissa minniháttar leiðréttinga og breytinga.

Í samræmi við minnkaðan hlut Þorláks Karlssonar í innihaldi kversins hefur höfundarnöfnum verið víxlað frá því sem var hjá dæmakverunum tveimur.

Þetta skjal inniheldur allt það efni sem bættist við frá 1. útgáfu kversins. Auk þess má eru allar leiðréttingar við einstök verkefni og svör að finna í sérstöku skjali á heimasíðu námskeiðsins www.hi.is/pub/gba/adf2/skver/index.html

Guðmundur Arnkelsson

EFNISYFIRLIT

| | |
|--|-----|
| Formáli | i |
| Efnisyfirlit | iii |
| Dæmi og efnisspurningar | 1 |
| Svör við dæmum og efnisspurningum..... | 5 |

DÆMI OG EFNISSPURNINGAR

97. Eftirfarandi tafla sýnir niðurstöður könnunar á því hvort fólk hafi þurft að fresta eða hætta við að fara til læknis á síðustu 12 mánuðum vegna fjárskorts.

| Heimsókn til læknis | Fjölskyldutekjur | | Samtals |
|---------------------|------------------|------|---------|
| | Lágar | Háar | |
| Ekki frestað | 26 | 34 | 60 |
| Frestað | 7 | 4 | 11 |
| Samtals | 33 | 38 | 71 |

- a) Reiknaðu viðeigandi mælitölu fyrir tengsl fjölskyldutekna og því hvort læknisheimsókn hafi verið frestað.
- b) Rökstyddu val þitt á mælitölu.
- c) Túlkaðu tölfræðilega niðurstöðuna samkvæmt mælitölunni og túlkaðu efnislega tengsl breytanna.
98. Eftirfarandi tafla sýnir fjölda heimasíðna eftir skorum kennara í félagsvísindadeild miðað við febrúar 2002. Heimasíðu getur vantað (Engin), hún verið takmörkuð (t.d. án tenginga við aðrar síður eða lokuð almenningi) eða fullkomin opin síða með tengingum.

| Form á heimasíðu | Skorir í félagsvísindadeild | | | | | | Samtals |
|------------------|-----------------------------|--------|--------|-----------|---------|-----------|---------|
| | Bókasfr. | Sálfr. | Uppfr. | Félagsfr. | Mannfr. | Stjórnfr. | |
| Engin | 0 | 1 | 2 | 4 | 0 | 3 | 10 |
| Takmörkuð | 2 | 1 | 1 | 1 | 0 | 0 | 5 |
| Opin | 1 | 5 | 5 | 4 | 6 | 2 | 23 |
| Samtals | 3 | 7 | 8 | 9 | 6 | 5 | 38 |

- a) Reiknaðu viðeigandi mælitölu fyrir tengsl skorar og heimasíðna. Miðaðu við að skor sé frumbreyta og form á heimsíðu sé fylgibreyta.
- b) Rökstyddu val þitt á mælitölu.
- c) Túlkaðu tölfræðilega niðurstöðuna samkvæmt mælitölunni og túlkaðu efnislega tengsl breytanna.
- d) Ef það er ósamræmi milli mælitölnnar og tengslanna sem þú sérð í töflunni, þarf þú að útskýra hvernig ósamræmið er tilkomið.
99. Sveinn keyrir reglulega heim til sín til Akureyrar. Hann hefur komist að því að ef hann keyrir á 60 km hraða þá kemst hann á einum tanki og hefur nóg bensín til að stússast innanbæjar. Ef hann keyrir á 80 km hraða nær hann alla leið heim en þarf að taka bensín um leið og hann kemur inn í bæinn. Ef hann keyrir á 100 km hraða þarf hann að taka bensín í Varmahlíð (94 km

frá Akureyri) og ef hann keyrir á 120 þarf hann að taka bensín á Blönduósi (145 km frá Akureyri).

Valgeir benti honum á að þetta væri skýrt dæmi um beinlínutengsl (*linear relationship*) milli eigindlegrar rofinnar frumbreytu (þ.e. fullur eða tómur tankur) og eigindlegrar fylgibreytu (nafnið á bænum þar sem tankurinn er fylltur).

Taktu rökstudda afstöðu til fullyrðingar Valgeirs og leiðréttu hana ef hún er röng að einhverju leyti.

100. Í hvaða tilvikum væri rétt að nota mælitöluna t - c ?
101. Komdu með skýrt dæmi um samvirknitengsl (*interaction*) sem er hvorki komið úr fyrirlestrum né námsefni námskeiðsins. Útskýrðu, með því að beita meðal annars hugtakinu áhrif (*effect*), í hverju þessi samvirknitengsl felast.
- Athugaðu: Þú þarft að koma með raunhæft dæmi um samvirknitengsl; skilgreining ein og sér nægir ekki.
102. Í hvaða tilvikum gæfi raðfylgni Spearman réttari eða eðlilegri mynd af tengslum tveggja samfelldra megindlegra breyta heldur en Pearson fylgnistuðull?
103. Ég er að skoða tengsl tveggja samfelldra megindlegra breyta. Fyrst met ég beinlínutengsl (*linear association*). Þar sem sveigbogalína passar gögnunum betur snýst mér hugur og nota hana í staðinn. Hvaða áhrif hefur þessi ákvörðun mín á (a) mat mitt á áhrifum frumbreyttunnar á fylgibreytuna, (b) stærð skilyrtu dreifingarinnar (*conditional distribution*) og (c) skýrða dreifingu (*explained variance*)?
104. Mælitölurnar Pearson r^2 , gamma, t og lambda skýrða dreifingu (*explained variance*) eða hlutfallslega minnkun villu (*proportional reduction in error*) eftir því hvaða hugtök við kjósum að nota.
- Útskýrðu stuttlega hvernig þetta getur verið, þ.e. hvað mælitölurnar eru að meta hver um sig og hvernig megi í öllum tilvikum líta á það sem mat á minnkun villu.
105. Hvaða upplýsingar veitir mælitalan f og að hvaða leyti er túlkun hennar eins eða ólík túlkuninni á Pearson r ?
106. Bæði Pearson r og raðfylgni Spearman (r) er hægt að túlka ýmist sem áhrif (*effect*) eða sem skýrða dreifingu (*explained variance*). Útskýrðu hvernig þessar túlkanir eru hvor um sig svipaðar og á hvern hátt þær eru ólíkar fyrir þessar tvær mælitölur.

107. DV gerði könnun á viðhorfum til Reykjavíkurlugvallar daginn fyrir viðhorfskönnun á vegum Reykjavíkurborgar. Af þeim 535 sem tóku afstöðu til spurningarinnar voru 297 hlyntir því að flugvöllurinn væri kyrr en 238 vildu að hann hyrfi brott. Miðaðu við að þetta úrtak sé rétt dregið tilviljunarúrtak (*random sample*) atkvæðisbærra Reykvíkinga.
- Taktu öll nauðsynleg skref til að ákvarða á hvaða bili þýðishlutfallið liggur, þ.e. hve hátt hlutfall Reykvíkinga eru hlyntir því að völlumurinn verði kyrr. Miðaðu við 95% öryggi í útreikningum.
 - Túlkaðu niðurstöðuna bæði með almennri efnislegri túlkun en einnig með því að tilgreina nákvæmlega hvað niðurstaðan þýðir tölfræðilega.
 - Á vef Morgunblaðsins (www.mbl.is) var ályktað að niðurstaða könnunarinnar sýndi að meirihluti Reykvíkinga vildi að flugvöllurinn væri kyrr. Er þetta rétt ályktun? Rökstyddu.
108. Gerum ráð fyrir að sykurneysla sé að meðaltali 100 kg á mann á Norðurlöndum. Í rannsókn á úrtaki 98 Íslendinga reyndust þeir neyta að meðaltali 107,0 kg af sykri á mann. Staðalfrávik sykurneyslu var 29,7 í þessu úrtaki.
- Settu fram *formlegar* tilgátur með *viðeigandi táknum*. Rannsóknarspurningin er hvort munur sé á sykurneyslu Íslendinga og Norðurlandabúa almennt.
 - Prófaðu tilgáturnar með viðeigandi tölfræðiprófi. Miðaðu við $\alpha = 0,05$.
 - Túlkaðu niðurstöðu prófsins bæði tölfræðilega (hvað niðurstaðan þýðir nákvæmlega skv. hugtökum tölfræðinnar) og efnislega (hver niðurstaðan er á hversdagslegri íslensku).
109. Ef marktektarpróf (*significance test*) reynist ómarktækt, hvaða ályktanir má þá draga um núlltilgátuna? Rökstyddu.
110. Guðmundur Hrafnkelsson hefur átt stjórnuleik og varið 10 af 15 markskotum eða 67% allra skota. Við höfum hins vegar ástæðu til að ætla að hann verji að jafnaði um 33% allra markskota. Miðaðu við að sú forsenda sé rétt þegar þú svarar eftirfarandi liðum.
- Ef við gerum ráð fyrir því að við höfum rétt fyrir okkur, hversu miklar líkur eru þá á því að hann verji 16. markskotið?
 - Hver væri samkvæmt því ástæðan fyrir því að hann hefur varið svona vel í leiknum?
 - Hvaða ályktanir getum við dregið um markvörslu Guðmundar í næstu leikjum? Hver verður markvarslan að jafnaði og hver verður breytileiki hennar frá einum leik til annars?

111. Hvað er staðalvilla meðaltals (*standard error of the mean*) og hvaða upplýsingar veitir það?
112. Hvað felst í marktæktarstiginu α ? Ef t.d. $\alpha = 0,05$, hvað merkir það nákvæmlega?
113. Í framhaldi af marktæku tölfræðiprófi reikna ég öryggisbil. Öryggisbilið reynist vera mjög vítt og nær næstum alveg niður að viðmiðsgildinu en einnig mjög langt í hina áttina. Hvaða upplýsingar gefur þetta öryggisbil í viðbót við það að niðurstöðurnar séu marktækar?
114. Í ályktunartölfræði er mikilvægt að nota óskekktar (*unbiased*) mælitölur. Hvers vegna er það mikilvægt og hvað nákvæmlega felst í því að mælitala sé óskekkt?
115. Ef ég prófa núlltilgátuna að kynjahlutfall sé jafnt í námskeiði, í hverju fælust fastheldnimistök (β -error) og í hverju fælust höfnunarmistök (α -error)?
116. Ólafur, Jónína og Ásta gerðu rannsókn á símreikningum námsmanna. Þau reiknuðu öryggisbil fyrir þýðismeðaltal og komust að því að meðalsímreikningur háskólanema væri á bilinu 1.500 til 3.500 krónur miðað við 95% öryggi.
- Ólafur túlkaði þetta þannig að ef rannsóknin væri endurtekin, væru 95% líkindi á því að meðaltal þeirrar rannsóknar væri á bilinu 1.500–3.500 krónur. Jónína sagði hins vegar að þetta þýddi að 95% háskólanema væru með símreikninga sem væru á bilinu 1.500–3.500 krónur. Ásta fullyrti hins vegar að ef rannsóknin væri endurtekin og reiknað nýtt öryggisbil á grundvelli hennar væru 95% líkindi á því að öryggisbil þeirrar rannsóknar innihéldi meðalsímreikning háskólanema.
- Hver þeirra, eitt eða fleiri, hafa rétt fyrir sér? Rökstyddu hvers vegna réttu fullyrðingarnar eru réttar og hvað sé rangt við röngu fullyrðingarnar.

SVÖR VIÐ DÆMUM OG EFNISSPURNINGUM

Verkefni 97

- a) $\phi = 0,147 \approx 0,15$
- b) Þetta eru tvær tvískiptar breytur, sbr. ítarlegan rökstuðning í svari við verkefni 7

$$\phi = \frac{34 \cdot 7 - 26 \cdot 4}{\sqrt{(60)(11)(33)(38)}} = \frac{238 - 104}{\sqrt{827.640}}$$
$$= \frac{134}{909,747} = 0,147 \approx 0,15$$

- c) Hlutfallslega fleiri fresta læknisheimsókn af þeim sem hafa lágar tekjur heldur af þeim sem hafa háar. Áhrif tekna á frestun læknisheimsókna eru mikil þar sem líkur á frestun læknisheimsóknar eru helmingi meiri hjá þeim sem eru tekjulágir ($7/33 = 0,21$) en hjá þeim sem eru tekjuháir ($4/38 = 0,11$). Eftir sem áður er styrkur tengslanna fremur lítill, þar sem aðeins er hægt að skýra um 2% ($0,147^2 = 0,022$) af dreifingu læknisheimsókna með fjölskyldutekjum.

Verkefni 98

- a) $\lambda = 0,133 \approx 0,13$
- b) Frumbreytan Skor er margskipt eigindleg breyta. Því koma ekki aðrar mælitölur en lambda til greina.

$$\lambda_{yx} = \frac{\sum L_{yx} - L_y}{N - L_y}$$
$$= \frac{(2+5+5+4+6+3) - 23}{38 - 23}$$
$$= \frac{25 - 23}{15} = \frac{2}{15} = 0,133... \approx 0,13$$

- c) Ef spá á fyrir um hvaða hvers konar heimasíðu (engin, takmörkuð, opin) tiltekinn kennari er með, má fækka mistökum um 13% með því að taka tillit til þeirrar skorar sem hann er í. Þetta miðast við að ætíð sé spáð algengasta flokknum, ýmist miðað við deildina í heild eða viðkomandi skor.

Almennt séð virðist sem mannfræði, sálfræði og uppeldisfræði séu komin lengst í heimasíðugerð, þ.e. hátt hlutfall heimasíðna og flestar án takmarkana, en félagsfræði og stjórnmálafræði einna styst.

- d) Ósamræmi er á milli tiltölulega lágs lambda og þess að töluverð tengsl sjást í töflunni. Áhrif skora á heimasíðugerð virðast umtalsverð, t.d. eru allir eða nær allir kennarar með heimasíður í mannfræði og sálfræði en aðeins um helmingur í félagsfræði og stjórnmálafræði.

Fullkomnar opnar heimasíður eru algengastar í öllum skorum og því bæta upplýsingar um skorir forspána lítið. Þetta er velþekktur veikleiki lambda sem undirstrikar að það er mælikvarði á forspá fremur en tengsl breyta.

Verkefni 99

Hér er margt sem er athugavert við fullyrðingu Valgeirs. Hér er reynt að telja upp það helsta.

Það er ekkert sem kemur fram sem gefur til kynna að þetta séu beinlínutengsl. Við gætum búist við að bensínueyðsla væri með sveiglínutengsl við hraða, en okkur vantar upplýsingarnar til að skera úr um það.

Eðlilegast er að líta á hraðann sem frumbreytu og bensíneyðsluna sem fylgibreytu en ekki öfugt.

Fullur eða tómur tankur er megindleg breyta en ekki eigindleg.

Það er hæpið að líta á nafn bæjarins sem fylgibreytu; eðlilegast er að líta á eyðsluna eða akstursvegalengd á einum tanki sem fylgibreytuna í þessu dæmi.

Nafnið á bænum er tæpast nafnbreyta í þessu tilviki, þar sem það er mælikvarði á hversu langt er komist á einum tanki. Við hljótum að gera ráð fyrir að bæjarnöfnin gefi til kynna hversu langt Sveinn kemst á einum tanki og það sé viðfangsefnið en ekki nafnið á bænum. Í daglegu tali tengist þetta tvennt eðlilega saman.

Fylgibreytan samkvæmt ofangreindu er samfelld, megindleg breyta.

Verkefni 100

[Svar ekki gefið; skoðuðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 101

Feitir fá fremur hjartaáfall en grannir en áhættan er minni ef þeir stunda líkamsrækt.

Þetta er samvirkni vegna þess að áhrif þess að vera feitur á hjartaáföll eru minni ef stunduð er líkamsrækt heldur en ef hún er ekki stunduð.

Verkefni 102

[Svar ekki gefið; skoðuðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 103

Áhrif verða breytileg eftir því hvert gildi frumbreytunnar er, skilyrta dreifingin minnkar og skýrð dreifing eykst.

Verkefni 104

[Svar ekki gefið; skoðuðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 105

Fí í öðru veldi (φ^2) metur skýrða dreifingu eins og Pearson r^2 . Ólíkt Pearson r hefur það enga merkingu að túlka fí sem áhrif frumbreytu á fylgibreytu mæld í staðalfrávikum.

Verkefni 106

[Svar ekki gefið; skoðaðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 107

- Á bilinu 51 til 60% Reykvíkinga voru hlynntir því að flugvöllurinn væri kyrr miðað við 95% öryggi.
- Meirihluti Reykvíkinga virðist hlynntur því að Reykjavíkflugvöllur sé kyrr. Sá meirihluti gæti verið mjög tæpur (51%) en þetta gæti einnig verið nokkuð öruggur (60%) meirihluti.
Tölfræðilega þýðir niðurstaðan að ef könnunin hefði verið margendurtekin, alltaf nákvæmlega eins, og 95% öryggisbil reiknað fyrir hverja könnun, þá myndu 95% þeirra öryggisbila innihalda rétt hlutfall Reykvíkinga sem vildu flugvöllinn kyrran. Öryggisbil í dæminu er eitt þessara ótalmörgu hugsanlegu öryggisbila.
- Þetta er rétt niðurstaða svo langt sem það nær. Morgunblaðinu sést yfir að meirihlutinn gæti verið mjög tæpur—aðeins eitt prósentustig—og enn einhverjir dagar í kosningarnar sjálfar.

Verkefni 108

- $T_1: \mu \neq 100$ kg; $T_1: \mu = 100$ kg.
- $t(97) = 2,33$, $p < 0,05$. Prófið er marktækt og því get ég hafnað núlltilgátunni og tekið upp aðaltilgátuna.
- Ef núlltilgátan er rétt, eru innan við 5% líkur á þetta miklu eða meira frávik frá núlltilgátunni. Íslendingar neyta að jafnaði meiri sykurs en almennt gerist á Norðurlöndum.

Verkefni 109

Ekki er hægt að hafna núlltilgátunni en hún er ekki studd heldur. Niðurstaðan er líkleg ef þessi núlltilgáta er rétt en gæti samræmst óendanlegum fjölda annarra núlltilgáta.

Verkefni 110

[Svar ekki gefið; skoðaðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 111

Staðalvilla meðaltals er mælikvarði á dreifingu meðaltala yfir mörg úrtök og því mælikvarði á breidd úrtakadreifingarinnar.

Verkefni 112

[Svar ekki gefið; skoðaðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 113

Öryggisbilið gefur til kynna sennilegt gildi þýðistölunnar og því þýðir þetta víða öryggisbil að lítið er vitað um þýðistöluna þrátt fyrir að hægt sé að fullyrða að núlltilgátan sé röng.

Verkefni 114

[Svar ekki gefið; skoðaðu Spurt og svarað, námsefni og Orðgnótt.]

Verkefni 115

Fastheldnimistök fælust í því að telja kynjahlutfallið jafnt þegar það er í reynd ójafnt. Höfnunarmistök fælust í því að telja kynjahlutfallið ójafnt þegar það í reynd væri jafnt.

Verkefni 116

[Svar ekki gefið; skoðaðu Spurt og svarað, námsefni og Orðgnótt.]