Fervikagreining

Tölfræði frá grunni

Anna Helga Jónsdóttir og Sigrún Helga Lund

Háskóli Íslands

Yfirlit

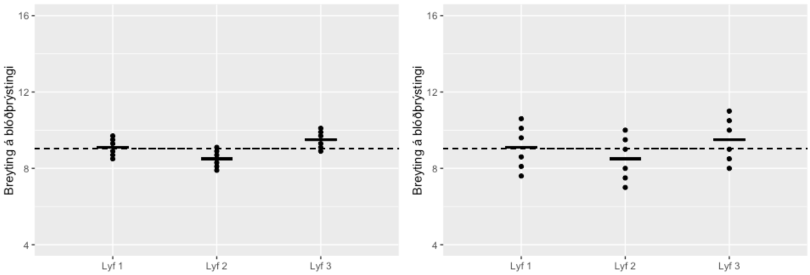
Inngangur

- Í 8. kafla sáum við hvernig við gátum borið saman meðaltöl tveggja þýða.
- Núna munum við skoða aðferð sem við getum beitt ef við viljum bera saman meðaltöl tveggja **eða fleiri** óháðra þýða.
- Við köllum þýðin oft hópa.
- Aðferðin ber heitið fervikagreining (analysis of variance - ANOVA).

Fervikagreining

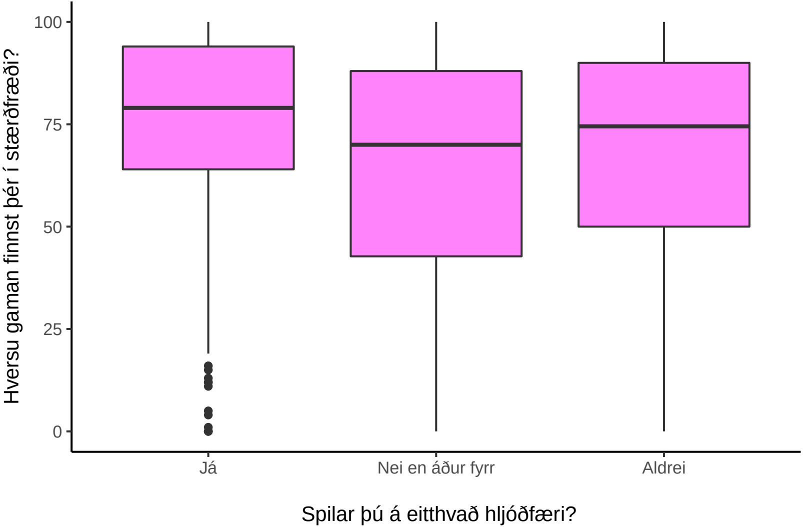
- Fervikagreining er gífurlega mikið notuð tölfræðiaðferð sem má aðlaga að ólíkum aðstæðum.
- Við munum einungis skoða eitt tilbrigði hennar sem kallast *einþátta fervikagreining* (one-sided ANOVA).
- Aðferðin gengur út á að bera saman breytileika á gildum mælinga **milli hópa** annars vegar og **innan hópa** hins vegar.
- Fervikagreining gerir ráð fyrir að mælingar innan hvers hóp fylgi normaldreifingu og að dreifnin sé sú sama í öllum hópum.
- Eins og alltaf gerum við auk þess ráð fyrir að úrtökin séu slembiúrtök.

Samanburður á mismunandi (tilbúnum) gögnum



- Hver er munurinn á meðaltölum hópanna þriggja á myndunum tveimur?
- Í hverju liggur munurinn á myndunum tveimur?
- Á hvorri myndinni þætti manni sennilegra að meðaltölin væru í raun ólík?

Einþátta fervikagreining - dæmi um notkun



Einþátta fervikagreining - dæmi um notkun

Er munur á því hversu fólki finnst gaman í stærðfræði eftir því hvort það spilar/spilaði á hljóðfæri?

Spilar á hljóðfæri	meðaltal	staðalfrávik
Já	73.73778	25.72741
Nei en áður fyrr	63.94375	28.80825
Aldrei	66.61983	29.05661

Yfirlit

Ritháttur notaður í fervikagreiningu

y_{ij} : Við notum vísinn i til að tákna númer hóps og vísinn j til að tákna númer mælingu innan hóps.
y_{ij} er því mæling númer j úr hópi i.

a : Við notum a til að tákna fjölda hópa.

n_i : Við notum n_i til að tákna fjölda mælinga í hópi i.

N : Við notum N til að tákna heildarfjölda mælinga

$$N = n_1 + n_2 + ... + n_a.$$

Ritháttur notaður í fervikagreiningu - frh.

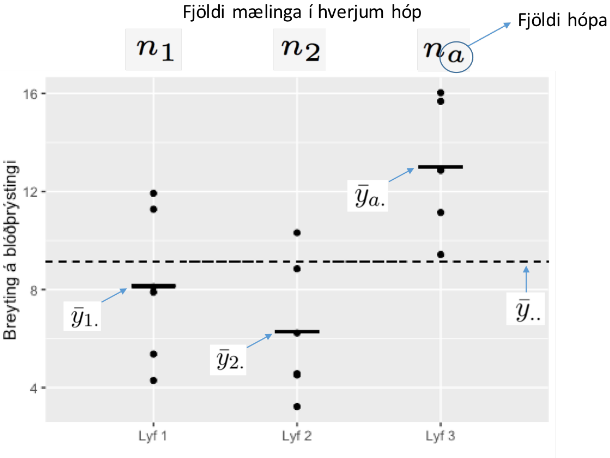
$\bar{y}_{i.}$: Við notum $\bar{y}_{i.}$ til að tákna meðaltal fyrir hóp i

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}.$$

$\bar{y}_{..}$: Við notum $\bar{y}_{..}$ til að tákna meðaltal allra mælinga (úr öllum hópum)

$$\bar{y}_{..} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}}{N}.$$

Ritháttur



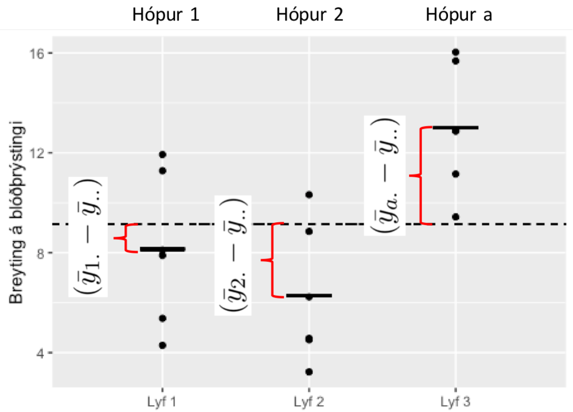
Fjöldi mælinga í hverjum hóp
n_1
n_2
n_a
Fjöldi hópa
Breyting á blóðþrýstingi
16
12
8
4
$\bar{y}_{a.}$
$\bar{y}_{1.}$
$\bar{y}_{2.}$
$\bar{y}_{..}$
Lyf 1
Lyf 2
Lyf 3

Yfirlit

Fervikasummur

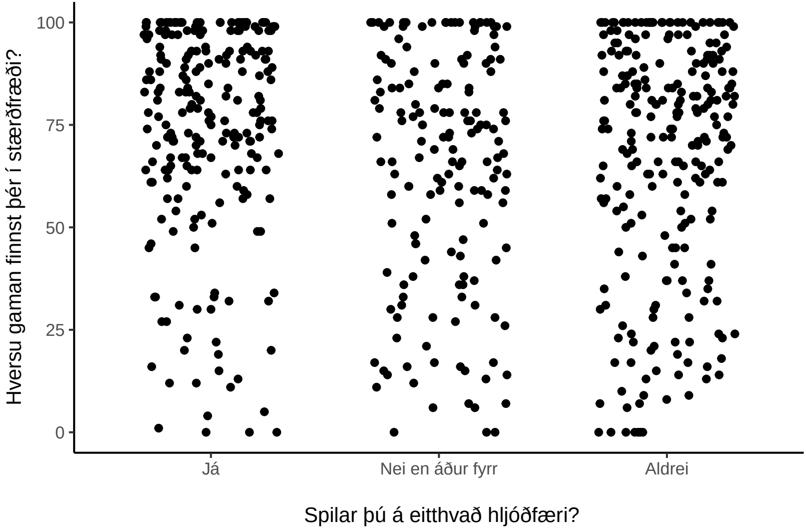
- Við þurfum að reikna þrjár fervikasummur og eru þær táknaðar með SS_{Tr}, SS_E og SS_T.
- SS_{Tr} er mælikvarði á breytileika **milli** hópanna:
  Hversu breytileg eru meðaltöl hópanna?
- SS_E er mælikvarði á breytileika **innan** hópanna:
  Hversu mikið víkja mælingar innan hvers hóps frá meðaltali hópsins?
- SS_T er heildarbreytileikinn:
  Hversu mikið víkja mælingarnar frá heildarmeðaltalinu yfir alla hópa?

Fervikasumma (breytileiki) milli hópa (SS_{Tr})

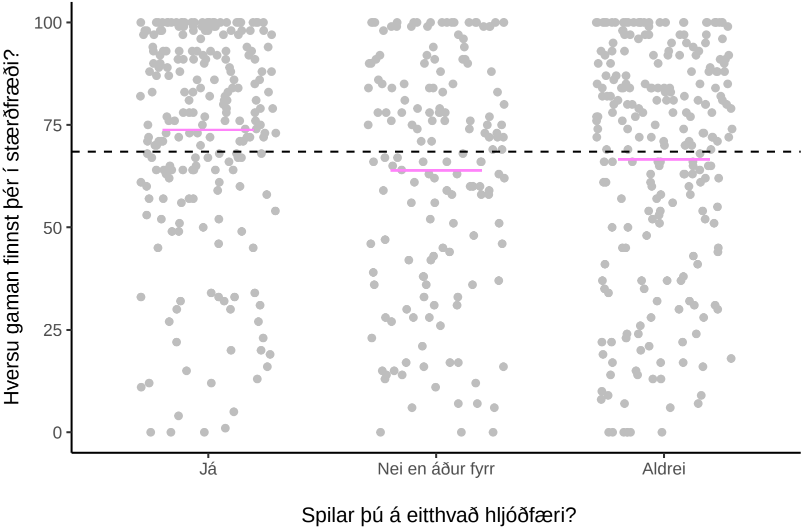


$$SS_{Tr} = n_1(\bar{y}_{1.} - \bar{y}_{..})^2 + n_2(\bar{y}_{2.} - \bar{y}_{..})^2 + \cdots + n_a(\bar{y}_{a.} - \bar{y}_{..})^2$$

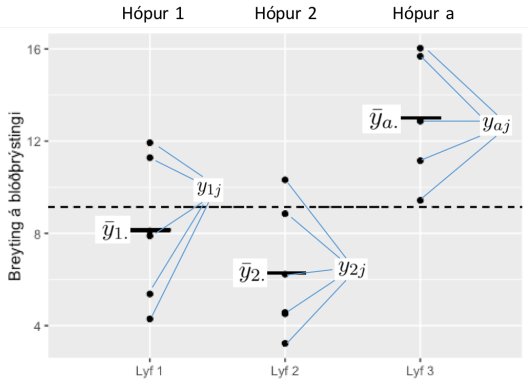
Gögnin


Hversu gaman finnst þér í stærðfræði?
100
75
50
25
0
Já
Nei en áður fyrr
Aldrei
Spilar þú á eitthvað hljóðfæri?

SSTr: Hversu breytileg eru meðaltöl hópanna?

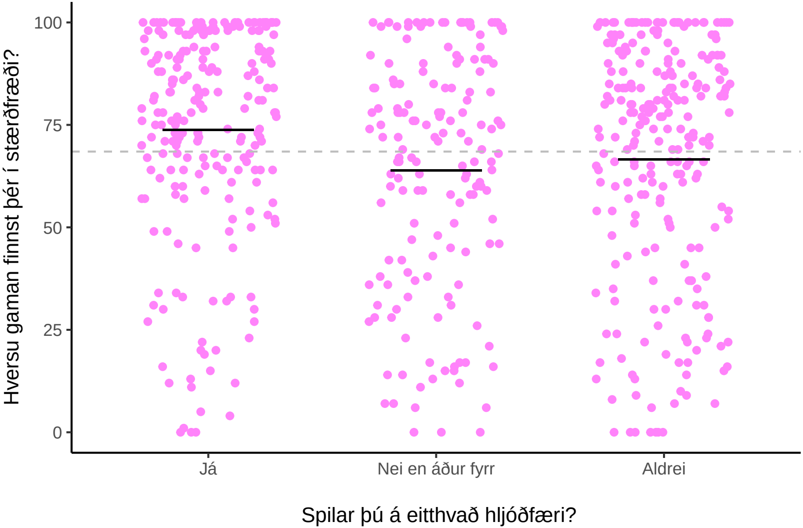


Fervikasumma (breytileiki) innan hópa (SS_E)

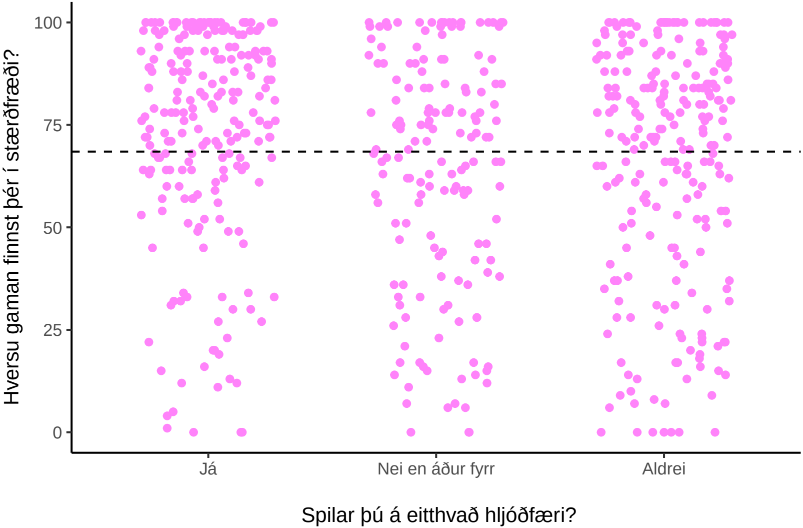


$$SS_E = \underbrace{\sum_{j=1}^{n_1}(y_{1j} - \bar{y}_{1\cdot})^2}_{\text{breytileiki í hóp 1}} + \underbrace{\sum_{j=1}^{n_2}(y_{2j} - \bar{y}_{2\cdot})^2}_{\text{breytileiki í hóp 2}} + \cdots + \underbrace{\sum_{j=1}^{n_a}(y_{aj} - \bar{y}_{a\cdot})^2}_{\text{breytileiki í hóp a}}$$

SSE: Hversu mikið víkja mælingarnar frá sínum meðaltölum?



SST: Hversu mikið víkja mælingarnar frá heildarmeðaltalinu?



Fervikasummur í einhviða fervikagreiningu

Fervikasummurnar eru reiknaðar með

$$SS_{Tr} = \sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2$$

Heildarbreytileikanum má skipta upp í breytileika milli hópanna og breytileika innan hópanna eða

$$SS_T = SS_{Tr} + SS_E.$$

Fervikasummutafla

- Algengt er að setja kvaðratsummurnar upp í svokallaða *fervikagreiningartöflu* (ANOVA table).
- Taflan samanstendur af þremur dálkum:
  - Fervikasummurnar
  - Fjöldi frígráða fyrir hverja fervikasummu fyrir sig
  - Meðalfervikasummur
- Meðalfervikasummur reiknum við með því að deila viðkomandi fervikasummu með fjölda frígráða sem henni tilheyra.

Fervikasummutafla

Fervikasummur	Frígráður	Meðalfervikasummur
SS_{Tr}	$a-1$	$MS_{Tr} = \frac{SS_{Tr}}{a-1}$
SS_E	$N-a$	$MS_E = \frac{SS_E}{N-a}$
SS_T	$N-1$	

Yfirlit

Tilgátupróf fyrir einhliða fervikagreiningu

Tilgátupróf fyrir einhliða fervikagreiningu

Tilgátan sem við viljum kanna er almennt

$$H_0 : \mu_1 = \mu_2 = ... = \mu_a$$

á móti gagntilgátunni

$$H_1 : \text{Að minnsta kosti eitt meðaltal er frábrugðið hinum}$$

Prófstærðin er

$$F = \frac{SS_{Tr}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Tr}}{MS_E}.$$

Sé núlltilgátan sönn fylgir prófstærðin F-dreifingu með $a-1$ og $N-a$ fjölda frígráða, eða $F \sim F_{(a-1,N-a)}$, þar sem a er fjöldi hópa og N er heildarfjöldi mælinga.
Hafna skal H_0 ef F $> F_{1-\alpha,(a-1,N-a)}$

Tilgátupróf fyrir einhliða fervikagreiningu

- Gagntilgátan er sú að að minnsta kosti eitt meðaltal sé frábrugðið hinum, það eru því einu upplýsingarnar sem við fáum þegar núlltilgátunni er hafnað.
- Við vitum ekki hvert meðaltalanna er frábrugðið hinum eða hvort þau séu mögulega öll frábrugðin hvort öðru.
- Það þarf að framkvæma frekari greiningu til að komst að því. Algeng próf eru t.d. Tukey's próf en ekki verður fjallað um þau hér.

Tilgátupróf fyrir fervikagreiningu - Dæmi

Hér má sjá fervikasummutöflu fyrir samband þess hversu gaman nemendum finnst í stærðfræði og hvort þeir spili á hljóðfæri. Er meðalánægjan sú sama í hópunum þremur?

Fervikasummur	Frígráður	Meðalfervikasummur
10349.65	2	5174.8272
483695.05	624	775.1523

Tilgátupróf fyrir fervikagreiningu - Dæmi

Hér má sjá fervikasummutöflu fyrir samband þess hversu gaman nemendum finnst í stærðfræði og hvort þeir spili á hljóðfæri. Er meðalánægjan sú sama í hópunum þremur?

Fervikasummur	Frígráður	Meðalfervikasummur
10349.65	2	5174.8272
483695.05	624	775.1523

Prófstærðin er

$$F = \frac{MS_{Tr}}{MS_E} = \frac{5174.8272}{775.1523} = 6.6759.$$

Tilgátupróf fyrir fervikagreiningu - Dæmi

Hér má sjá fervikasummutöflu fyrir samband þess hversu gaman nemendum finnst í stærðfræði og hvort þeir spili á hljóðfæri. Er meðalánægjan sú sama í hópunum þremur?

Fervikasummur	Frígráður	Meðalfervikasummur
10349.65	2	5174.8272
483695.05	624	775.1523

Prófstærðin er

$$F = \frac{MS_{Tr}}{MS_E} = \frac{5174.8272}{775.1523} = 6.6759.$$

Viðmiðunargildið er $F_{(a-1,N-a)} = F_{(2,624)} = 3.0102$.

Tilgátupróf fyrir fervikagreiningu - Dæmi

Hér má sjá fervikasummutöflu fyrir samband þess hversu gaman nemendum finnst í stærðfræði og hvort þeir spili á hljóðfæri. Er meðalánægjan sú sama í hópunum þremur?

Fervikasummur	Frígráður	Meðalfervikasummur
10349.65	2	5174.8272
483695.05	624	775.1523

Prófstærðin er

$$F = \frac{MS_{Tr}}{MS_E} = \frac{5174.8272}{775.1523} = 6.6759.$$

Viðmiðunargildið er $F_{(a-1,N-a)} = F_{(2,624)} = 3.0102$.

6.6759 er stærra en 3.0102 svo við getum ályktað að meðaltölin þrjú séu ekki öll jöfn.