

Fervikagreining - ANOVA

Tölfræði frá grunni - kafli 13

Anna Helga Jónsdóttir
Sigrún Helga Lund

Háskóli Íslands

Helstu atriði:

- 1 Inngangur
- 2 Einhliða fervikagreining
 - Fervikasummur
 - Tilgátupróf í fervikagreiningu

Hvert erum við komin...

- 1 Inngangur
- 2 Einhliða fervikagreining
 - Fervikasummur
 - Tilgátupróf í fervikagreiningu

Inngangur

- Í kafla 11 var fjallað um ályktanir fyrir meðaltal þýðis (μ).
- Í fyrri hluta kalfa 12 var fjallað um ályktanir tveggja óháðra þýða ($\mu_1 - \mu_2$).
- Í seinni hluta kalfa 12 var fjallað um ályktanir fyrir paraðar mælingar (μ_D).
- Núna munum við skoða aðferð sem við getum beitt viljum við bera saman meðaltöl fleiri en tveggja óháðra þýða. Aðferðin ber heitið ferkagreining (analysis of variance - ANOVA).

Fervikagreining

- Fervikagreining er ein af mest notuðu aðferðunum innan tölfræðinnar og til eru mörg tilbrigði hennar sem má laga að gífurlega mörgum ólíkum tilfellum.
- Við munum einungis skoða eitt tilbrigði hennar sem kallast *einhlíða fervikagreining* (one-sided ANOVA).
- Henni beitum við á gögn sem eru úrtök úr tveimur eða fleiri þýðum og er algengt að nota orðið hópar þegar talað er um úrtökin.
- 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 úrtökin séu slembiúrtök, að þau séu valin úr þýðum sem fylgja normaldreifingu og að dreifnin sé sú sama í öllum þýðum.

Framkvæmd tilgátuprófa

Framkvæmd tilgátuprófa

- 1 Ákveða hvaða tilgátupróf er viðeigandi fyrir gögnin okkar.
- 2 Ákveða hæstu ásættanlegu villulíkur.
- 3 Setja fram núlltilgátu og ákveða um leið áttun prófsins (einhliða/tvíhliða).
- 4 Reikna prófstærðina sem svarar til tilgátuprófsins.
- 5a Kanna hvort prófstærðin falli á höfnunarsvæði tilgátuprófsins.
- 5b Kanna p-gildi tilgátuprófsins.
- 6 Draga ályktun.

Hvert erum við komin...

- 1 Inngangur
- 2 Einhliða fervikagreining
 - Fervikasummur
 - Tilgátupróf í fervikagreiningu

Einhliða fervikagreining - dæmi um notkun

Lyfjafyrirtæki nokkurt er að þróa ný blóðþrýstingslyf og í því samhengi var lítil tilraun framkvæmd. Átján einstaklingar tóku þátt í tilrauninni og var þeim skipt tilviljunarkennt upp í þrjá hópa. Hópur eitt fékk lyf 1, hópur tvö lyf 2 og hópur þrjú fékk lyf 3. Blóðþrýstingur fólksins var mældur fyrir inntöku lyfsins og aftur eftir inntöku. Breytan sem við höfum áhuga á er breyting á blóðþrýstingi fyrir og eftir inntöku lyfsins. Meðalbreyting á blóðþrýstingi í hópunum þremur var reiknaður. Í öllum hópunum hafði blóðþrýstingurinn lækkað að meðaltali.

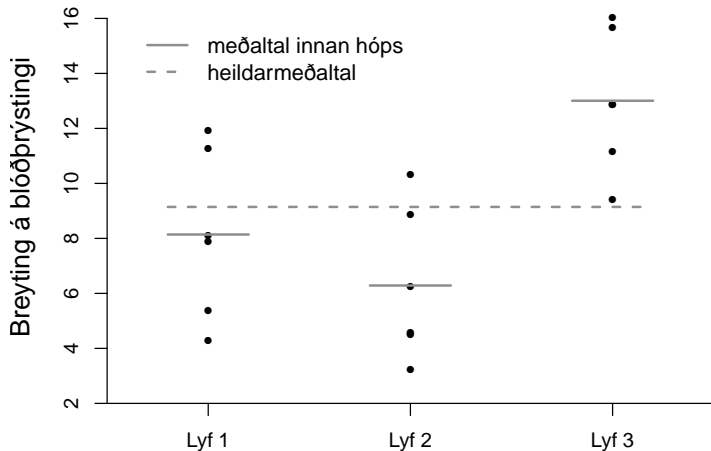
Meðalbreyting hópur 1: $\bar{y}_1 = 8.14$

Meðalbreyting hópur 2: $\bar{y}_2 = 6.28$

Meðalbreyting hópur 3: $\bar{y}_3 = 13.01$

Spurningin er nú, lækka lyfin blóðþrýsting mis mikið?

Gögnin



Mynd: Gögn fyrir fervikagreiningu

Ritháttur

Ritháttur notaður í fervikagreiningu

Eftirfarandi ritháttur er algengur í kennslubókum og ritum sem fjalla um 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

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

\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)

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

Fervikasummur

- Við þurfum að reikna þrjár fervikasummur og eru þær táknaðar með SS_T , SS_{Tr} og SS_E .
- SS_T er heildarfervikasumman og er hún mælikvarði á heildarbreytileika gagnanna (total variation).
- SS_{Tr} er mælikvarði á breytileika milli hópanna (between treatments) þ.e. hversu breytileg eru meðaltöl hópanna.
- SS_E er mælikvarði á breytileika innan hópanna (within treatments eða error) það er að segja hversu mikið víkja mælingar innan hvers hóps frá meðaltali hópsins.

Fervikasummur

Fervikasummur í einhviða fervikagreiningu (Sums of squares in one-sided ANOVA)

Fervikasummurnar eru reiknaðar með

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

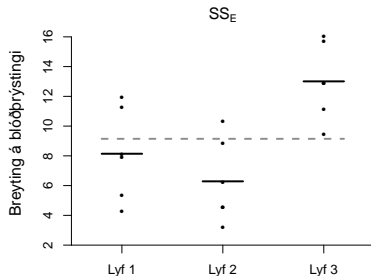
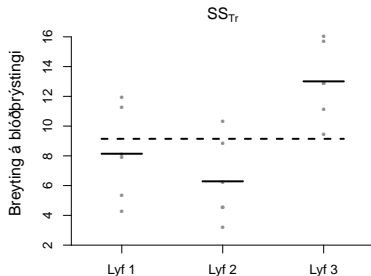
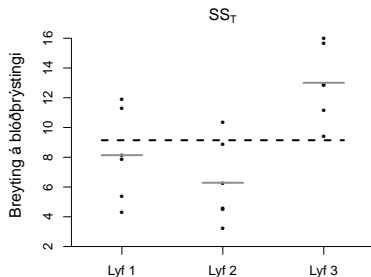
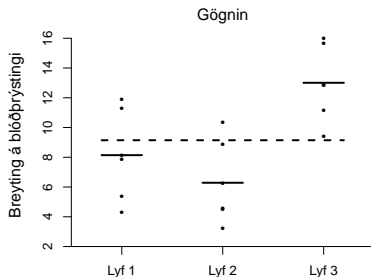
$$SS_{Tr} = \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

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

Heildarbreytileikanum má skipta upp í breytileika milli hópanna annars vegar og breytileika innan hópanna hins vegar eða

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

Fervikasummur - myndrænt



Fervikasummutafla

- Algengt er að setja kvaðratsummurnar upp í svokallaða *fervikagreiningartöflu* (ANOVA table).
- Taflan samanstendur af þremur dálkum og þremur línum.
- Fyrsti dálkurinn inniheldur fervikasummurnar, annar dálkurinn inniheldur fjölda *frígráða* fyrir hverja fervikasummu fyrir sig og þriðji dálkurinn inniheldur svokallaðar meðalfervikasummur.
- Meðalfervikasummur reiknum við með því að deila viðkomandi fervikasummu með fjölda frígráða sem henni tilheyrir (í sömu línu).

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	$N - 1$	

Tilgátupróf fyrir einhliða fervikagreiningu

Tilgátupróf fyrir einhliða fervikagreiningu

Tilgátan sem við viljum kanna er almennt

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

á móti gagn­tilgátunni

H_1 : 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 fríráða, 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 Tukey's próf og Duncan's próf en ekki verður fjallað um þau hér.

Fervikagreining - dæmi

Dæmi

Fervikagreining Kannið tilgátuna að munur sé á blóðþrýstingslyfjunum þremur sem við fjölluðum um í upphafi fyrirlestursins. Gögnin má sjá hér að neðan:

Lyf 1	Lyf 2	Lyf 3
4.29	10.32	12.89
11.28	3.23	15.68
5.37	4.51	16.03
7.89	4.57	9.43
8.10	8.85	12.86
11.93	6.23	11.15

Einnig er gefið að $SS_T = 262.16$ og $SS_E = 117.63$.