

Kendalls τ og skyldar mælitölur

Fyrirlestur í Aðferðafræði II

© 1998, 2000, 2001 Guðmundur Arnkelsson

All rights reserved. Copying or distribution prohibited without explicit permission. Students in Methodology II at the University of Iceland may print a copy for their own private use.

Tengsl sem samræmi mæligilda

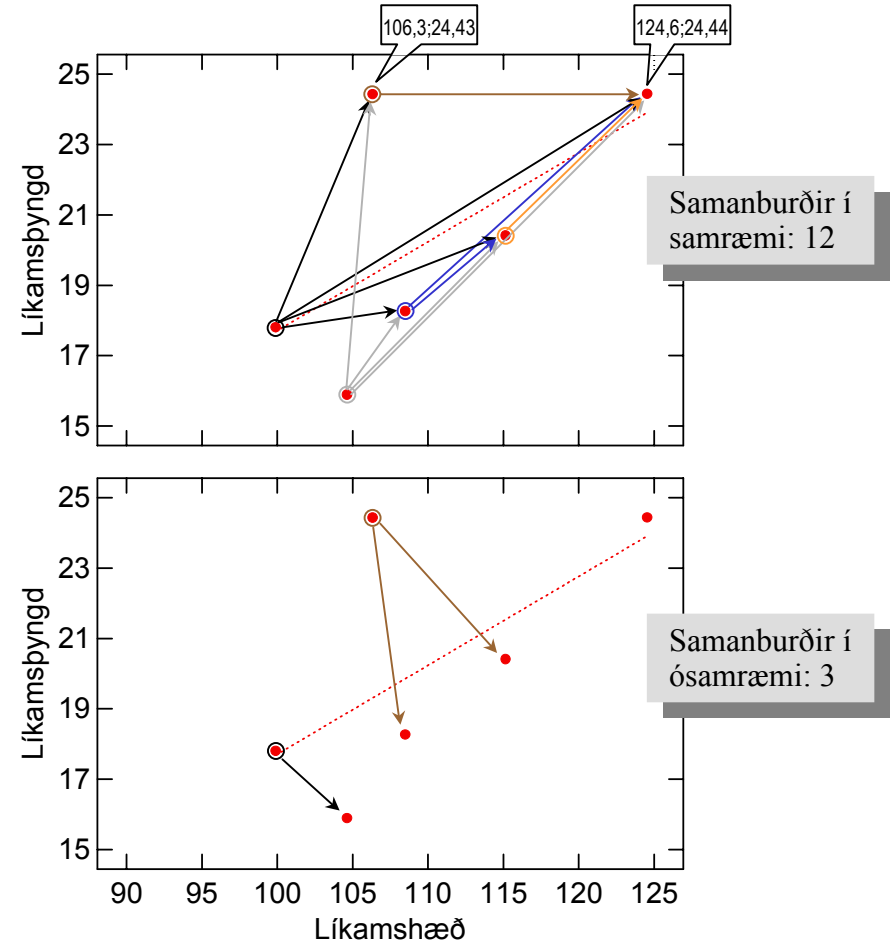
Myndin sýnir tengsl þyngdar og hæðar hjá sex börnum

Tengslin má meta með Pearson r eða raðfylgni ($r(4) = 0,62$; $r_s(4) = 0,77$).

Önnur leið er að bera saman tvö og tvö mæligildi og sjá hvort röð þeirra er sú sama á frum- og fylgibreytu.

Á myndinni eru 12 pör í samræmi og 3 í ósamræmi. Mismunurinn er 9 sem er 60% af heildarfjölda para ($((12-3)/(12+3)) = 9/15 = 0,6$).

Hlutfallið 0,6 má bera saman við $r^2 = 0,38$ og $r_s^2 = 0,59$.



Hversu oft í bíó?

Myndin sýnir tengsl aldurs og bíóferða

Báðar breytur eru meginlegar flokkabreytur.

Innan hvers flokks eru mæligildin öröðuð, en flokkarnir raðast.

Þetta eru því að hluta raðaðar (*partially ordered*) breytur.

Áhrif aldurs er greinileg á bíóferðir

16–24 ára fara mjög oft í bíó, 25–44 ára fara heldur sjaldnast og þeir elstu fara mjög lítið í bíó.

Hve oft í bíó	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	6	2	0	8
10–39	7	10	3	20
0–9	3	26	28	57
Samtals	16	38	31	85

Eftir því sem fólk eldist, því sjaldnar fer það í bíó. Það samsvarar því að neikvæð tengsl séu á milli aldurs og fjölda bíóferða.

Við viljum gjarnan hafa einhverja mælitölu fyrir raðaðar flokkabreytur sem gefur til kynna hve mikil tengslin eru og hver stefna þeirra er.

Fjöldi samanburða í samræmi og ósamræmi gefur slíka mælitölu.

Skematísk tafla

Hve oft í bíó	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	••	•		3
10–39	••	••	•	5
0–9	••	•••	•••••	10
Samtals	6	6	6	18

Hér er ný tafla með neikvæðum tengslum aldurs og bíóferða. Til hægðarauka hef ég minnkað fjöldann í töflunni og táknað stökin (einstaklingana) með punktum (•) í stað fjöldatalna.

Sérhver einstaklingur er táknaður með einum punkti. Samtals eru 18 einstaklingar í töflunni og því 18 punktar.

Með því að sýna einstaklingana sem punkta verður auðveldara að sýna samanburði í samræmi (röðin sú sama á frum- og fylgibreytu) og í ósamræmi (ólík röð á frum- og fylgibreytu).

Samræmi

Hve oft í bíó	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	••	•		3
10–39	••	••	•	5
0–9	••	•••	•••••	10
Samtals	6	6	6	18

Við berum einstaklingana saman tvo og tvo. Samræmi (*agreement*) er samanburður þar sem munur á frumbreytu er samfara sams konar muni á fylgibreytu. Þetta samsvarar því að röð einstaklinganna tveggja (parsins) sé sú sama á frum- og fylgibreytu, þ.e. sá eldri fari oftari í bíó.

Samanburður upp til hægri er í samræmi. Myndin sýnir samræmi fyrir neðsta hólf til vinstri. Hver ör stendur fyrir tvo samanburði, því tveir einstaklingar eru í þessu hólf töflunnar. Örvunar sýna því samtals 8 samanburði í samræmi.

Samtals eru 13 samanburðir í samræmi á myndinni (fimm samanburðir eru ekki sýndir með örvum).

Ósamræmi

Hve oft í bío	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	••	•		3
10–39	••	••	•	5
0–9	••	•••	•••••	10
Samtals	6	6	6	18

- Samanburðir í ósamræmi eru á ská til hægri í töflunni
 - Örvarnar sýna 22 samanburði í ósamræmi
 - Samtals eru 54 samanburðir í ósamræmi í töflunni
 - Taflan einkennist greinilega af ósamræmi milli frum- og fylgibreytu

Goodmans-Kruskal gamma

- γ gefur staðlaðan mælikvarða á samræmi umfram ósamræmi
 - $\gamma = 1,0$ ef eingöngu samræmi
 - $\gamma = 0,0$ ef jafnmikið af hvoru
 - $\gamma = -1,0$ ef eingöngu ósamræmi
- γ sem fækkun mistaka (óvissu)
 - Hve mikið mistökum fækkar við að þekkja frumbreytuna

$$\gamma = \frac{A - D}{A + D}$$

A : Samræmi

D : Ósamræmi

Útreikningur á γ

Hve oft í bíó	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	6	2	0	8
10–39	7	10	3	20
0–9	3	26	28	57
Samtals	16	38	31	85

- $A = 3 \cdot (10 + 2 + 3 + 0) + 7 \cdot (2 + 0) + 26 \cdot (3 + 0) + 10 \cdot (0) = 3 \cdot 15 + 7 \cdot 2 + 26 \cdot 3 + 10 \cdot 0 = 137$
- $D = 6 \cdot (10 + 26 + 3 + 28) + 7 \cdot (26 + 28) + 2 \cdot (3 + 28) + 10 \cdot (28) = 6 \cdot 67 + 7 \cdot 54 + 2 \cdot 31 + 10 \cdot 28 = 1.122$
- $\gamma = (137 - 1122) / (137 + 1122) = -985 / 1259 = -0,782$

Jafnir samanburðir

Hve oft í bíó	Aldur			Samtals
	16–24	25–44	45–75	
40–∞	••	•		3
10–39	•• T_x	••	•	5
0–9	••	•••	••••	10
Samtals	6	6	6	18

Pör geta ýmist verið jöfn á frum- eða fylgibreytu. Á myndinni eru sýnd dæmi um hvoru tveggja.

T_x : Fjöldi samanburða sem eru jafnir á frumbreytunni

T_y : Fjöldi samanburða sem eru jafnir á fylgibreytunni

Meðhöndlun jafnra samanburða

Jafna samanburði má meðhöndla á ýmsan hátt

Mælitalan gamma einfaldlega sleppir þeim úr útreikningunum.

Þetta samsvarar því að ákveða að ef einstaklingar eru jafnir á frum- eða fylgibreytu, þá teljist slík tilvik ekki með við forspá.

Ef tveir einstaklingar eru jafnir á frumbreytunni, einfaldlega vel ég að spá engu um mun á fylgibreytu.

Ef þeir eru ójafnir á frumbreytu en jafnir á fylgibreytu, tel ég þá heldur ekki með.

Mælitalan t_{ab} tekur tillit til jafnra samanburða á frum- og fylgibreytu

Ef einstaklingar eru ólíkir á frumbreytu, miðast t_{ab} við það að þeir eigi að vera ólíkir í sömu átt á fylgibreytu.

Ef þeir eru jafnir á frumbreytu, gerir t_{ab} ráð fyrir að þeir eigi þá einnig að vera jafnir á fylgibreytu.

T_{ab} gerir því strangari kröfur til tengsla flokkabreytanna tveggja heldur en gamma gerir. Niðurstaðan getur því orðið ólík eftir því hvor mælitalan er notuð.

Kendalls τ_b

- τ_b leiðréttir fyrir jöfn pör á bæði frum- og fylgibreytu
 - Ef munur á fylgibreytu þá er munur á frumbreytu
 - Ef munur á frumbreytu þá er munur á fylgibreytu
 - Ef jafnt á annarri breytunni þá einnig jafnt á hinni
- Samhverf mælitala

$$\tau_b = \frac{A - D}{\sqrt{(A + D + T_X)(A + D + T_Y)}}$$

A : Samræmi

D : Ósamræmi

T_Y : Jafnt á Y en ójafnt á X

T_X : Jafnt á X en ójafnt á Y

Kendalls τ_c

- τ_c er notuð þegar taflan hefur ólíkan fjölda dálka og lína
 - τ_b getur ekki orðið 1,0 í slíkum töflum
 - Því er leiðrétt fyrir ólíkan fjölda dálka og lína

$$\tau_c = \frac{2(A - D)}{N^2 \left(\frac{m - 1}{m} \right)}$$

N : Fjöldi staka

m : Fjöldi dálka eða lína
eftir því hvort er lægra