

# Raðfylgni Spearman

## Fyrirlestur í Aðferðafræði II

© 1999–2001, 2003 Guðmundur Arnelsson

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.

# Utah Jazz: Röð leikmanna

Taflan sýnir frammistöðu leikmanna Utah Jazz árið 1997; tölurnar sýna röð leikmanna.

Skýrð dreifing er um 50%, miðað við beina línu, og tengslin því tiltölulega sterk.

Aðfallsstuðullinn er um 0,7, þannig að hækkun um eitt sæti skilar sér að jafnaði sem 0,7 sæti á fylgibreytunni.

$$r_s(12) = 0,69; r_s^2 = 0,48$$

$$\text{Stigaröð} = 2,31 + 0,69 \cdot \text{Frákastaröð}$$

Leikmaður	Stig í leik	Fráköst í leik
Malone	1,0	1,0
Hornacek	2,0	5,0
Stockton	3,0	7,0
Russell	4,0	3,0
Carr	5,0	6,0
Ostertag	6,0	2,0
Anderson	7,0	8,0
Howard	8,0	10,0
Morris	9,0	9,0
Nembhard	10,5	14,0
Eisley	10,5	13,0
Keefe	12,0	4,0
Foster	13,5	11,0
Watson	13,5	12,0

# *Spearman's* $r_s$ sem hallastuðull

Pearson  $r$  má túlka sem staðlaðan hallastuðul.

Þegar frumbreytan breytist um eitt staðalfrávik, breytist fylgibreytan að jafnaði um  $r$  staðalfrávik.

Ef báðar breytur eru raðir, er staðalfrávik þeirra eins

Því er raðfylgni Spearman's jöfn hallastuðli

Pearson  $r$  fyrir raðtölur kallast Spearman's  $r_s$  og fær við það nýja eiginleika.

$$\text{Ef } r_{xy} = b_{yx} \frac{s_y}{s_x}$$

$$\text{og } s_y = s_x$$

$$\text{þá : } r_{xy} = b_{yx}$$

Tvær raðbreytur innihalda sömu tölugildi, sætisgildin, og því hafa þær sama meðaltal og sama staðalfrávik.

# Formúla fyrir raðfylgni

Í tölfræði—ólíkt ípróttum—fær lægsta mæligildið einnig lægstu raðtöluna (raðtöluna 1).

Ef tvo eða fleiri gildi eru jöfn, fá þau öll meðaltal raðtalna sinna. Eisley og Nembhard eru í 4–5 sæti og fá því raðtöluna 4,5.

Reikna má Spearman's  $r_s$  með eftirfarandi formúlu.

$$r_s = 1 - \frac{6 \cdot (\sum d^2)}{N^3 - N}$$

Leikmaður	Stig í leik	Fráköst í leik
Malone	14	14
Hornacek	13	10
Stockton	12	8
Russell	11	12
Carr	10	9
Ostertag	9	13
Anderson	8	7
Howard	7	5
Morris	6	6
Eisley	4,5	2
Nembhard	4,5	1
Keefe	3	11
Foster	1,5	4
Watson	1,5	3

# Útreikningur: Fyrstu skref

Fyrst er mismunur sætistalna reiknaður,  $d$ , sbr. fjórða dálk töflunnar.

Þá eru frávikin sett í annað veldi; niðurstöður eru í aftasta dálki.

Að síðustu er lagt saman niður dálkinn. Sú tala fer í formúluna.

$$r_s = 1 - \frac{6 \cdot (\sum d^2)}{N^3 - N}$$

Leikmaður	Stig í leik	Fráköst í leik	$d$	$d^2$
Malone	14	14	0,0	0,00
Hornacek	13	10	3,0	9,00
Stockton	12	8	4,0	16,00
Russell	11	12	-1,0	1,00
Carr	10	9	1,0	1,00
Ostertag	9	13	-4,0	16,00
Anderson	8	7	1,0	1,00
Howard	7	5	2,0	4,00
Morris	6	6	0,0	0,00
Eisley	4,5	2	2,5	6,25
Nembhard	4,5	1	3,5	12,25
Keefe	3	11	-8,0	64,00
Foster	1,5	4	-2,5	6,25
Watson	1,5	3	-1,5	2,25
Summa				<b>139,00</b>

$$(\sum d^2)$$


# Útreikningur: Niðurstaða

$$r_s = 1 - \frac{6 \cdot \left( \sum d^2 \right)}{N^3 - N} = 1 - \frac{6 \cdot 139}{14^3 - 14} =$$
$$1 - \frac{834}{2744 - 14} = 1 - 0,304 = 0,69$$

# Raðtölur og upprunaleg gögn

Lægsta gildið fær lægstu (1) og hæsta gildið hæstu raðtöluna.

Jöfn mæligildi fá meðaltal raðtalna, sbr. Nembhard og Eisley.

Raðtölurnar jafna út bil milli mæligilda.

Það munar fleiri stigum á Malone og Hornacek heldur en Hornacek og Stockton; í raðtölum munar í báðum tilvikum einum.

Leikmaður	Stig	Röð	Fráköst	Röð
	í leik		í leik	
Malone	26,1	14	10,7	14
Hornacek	15,3	13	3,2	10
Stockton	14,0	12	2,6	8
Russell	10,1	11	4,4	12
Carr	7,0	10	2,7	9
Ostertag	6,7	9	6,6	13
Anderson	5,0	8	2,5	7
Howard	4,8	7	2,2	5
Morris	4,7	6	2,3	6
Nembhard	4,6	4,5	1,1	2
Eisley	4,6	4,5	1,0	1
Keefe	3,7	3	3,5	11
Foster	2,8	1,5	1,8	4
Watson	2,8	1,5	1,5	3

# Frumgögn og raðgögn

Efri myndin sýnir línuleg tengsl frumgagna en sú neðri sýnir tengsl raðtalna.

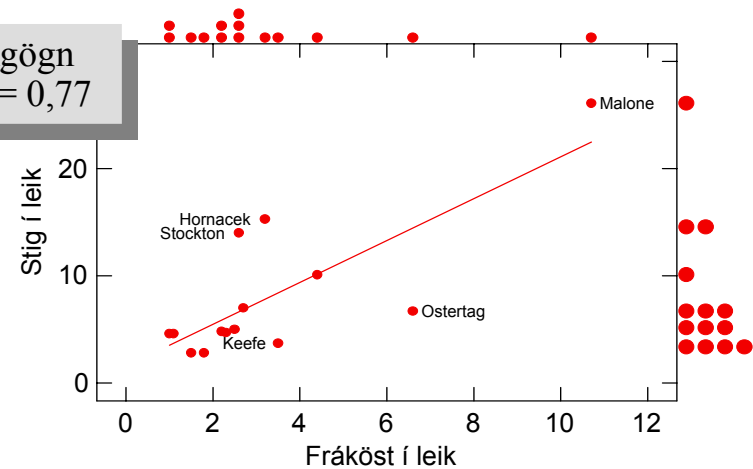
Jaðardreifingin á efri myndinni sýnir áberandi frávillinga; skilyrt dreifing sýnir einnig tvíbreytufrávillinga.

Einbreytufrávillingar hverfa á neðri myndinni en ekki tvíbreytufrávillingar.

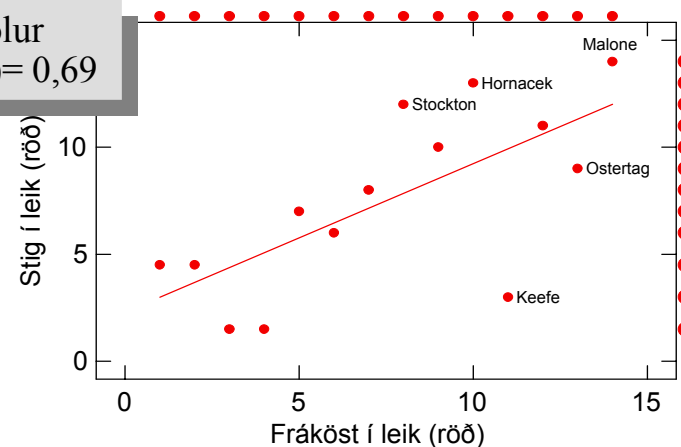
Raðfylgni er traust gagnvart einbreytu- en ekki tvíbreytufrávillingum.

Pearson fylgni er hærri en raðfylgni; ástæðan er Malone sem en hann er einbreytufrávillingur á báðum breytum. Algengast er að frávillingar minnki fylgni en þetta er óvenjuleg undantekning.

Frumgögn  
 $r(12)=0,77$



Raðtölur  
 $r_s(12)=0,69$





# Áhrif frávillinga

Stærð punktanna á myndunum sýnir áhrif á fylgnistuðulinn

Stórir holar hringir hækka fylgnina, stórir fylltir hringir lækka fylgnina.

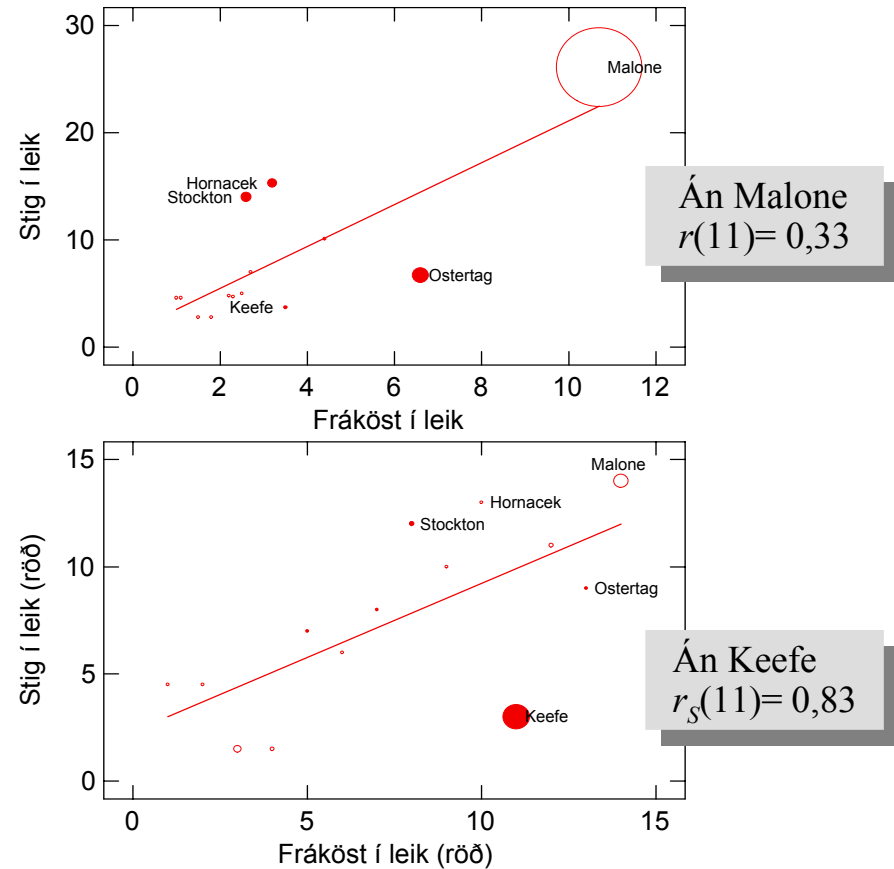
Malone hækkar Pearson  $r$  mikið

Ef Malone er með virðast áhrif vera mjög mikil og tengslin sterk.

Án hans kemur í ljós að áhrifin eru mjög lítil og veik tengsl.

Keefe veikir raðfylgnina

Tvíbreytufrávillingar draga úr bæði raðfylgni og Pearson fylgni. Hér birtist tvíbreytufrávillingurinn fyrst þegar gögnunum er breytt í raðtölur.



# Form tengsla

Tengsl tveggja breyta getur haft mismunandi form

Algengast er að miða við beina línu.

Þetta er einfalt form tengsla þar sem áhrifin eru þau sömu yfir allt svið frumbreytunnar.

Tengsl taka oft önnur form

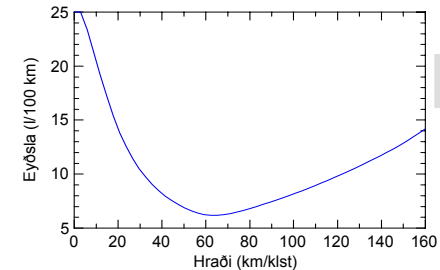
Einhallatengsl fela í sér að áhrifin aukast (eða minnka) eftir því sem frumbreytan hækkar.

Tvíhallatengsl fela í sér að áhrifin aukast fyrst en minnka svo (eða öfugt).

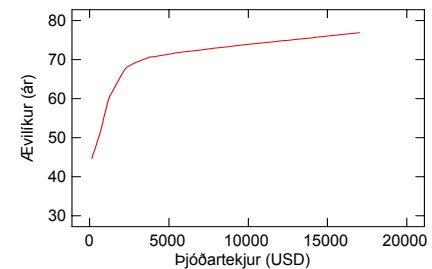
Mælitölur eru næmar fyrir ólíkum formum

Pearson  $r$  metur beinlínutengsl.

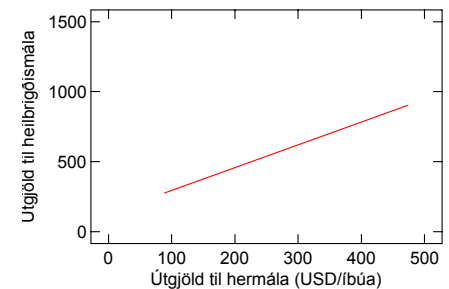
Aðrar mælitölur eru næmar á önnur form tengsla.



Tvíhallatengsl



Einhallatengsl



Beinlínutengsl

# $r_s$ og einhallatengsl

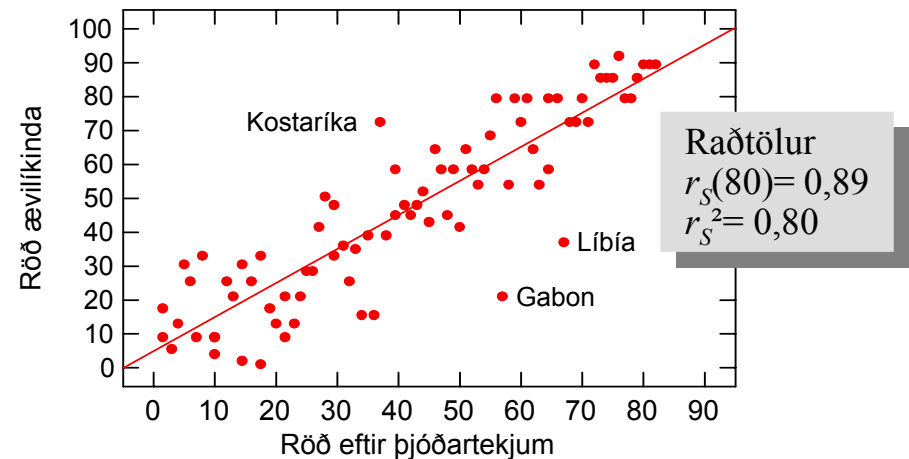
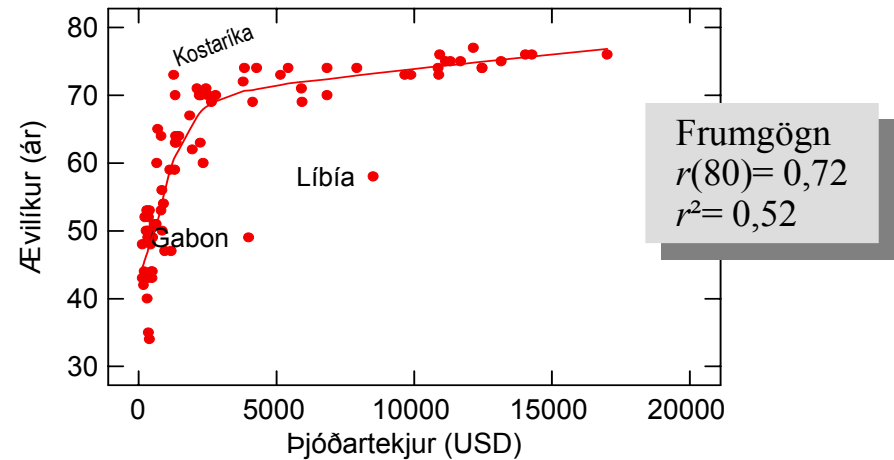
Raðfylgni metur einhallatengsl (*monotonic*), sveigbogategsl þar sem hallinn er alls staðar jákvæður eða alls staðar neikvæður.

Raðfylgni má því nota ef einhalli lýsir gögnunum betur en bein lína.

Einhallatengslin sjást með því að bera saman mælitölurnar tvær.

Myndin sýnir að raðfylgni skýrir mun meira en Pearson  $r$  af dreifingu fylgibreytunnar sökum sterkra einhallatengsla.

Tveir tvíbreytufrávillingar hafa áhrif á báðar mælitölur; auk þess bætist Kostaríka við sem frávillingur.



# Túlkun frávillinga

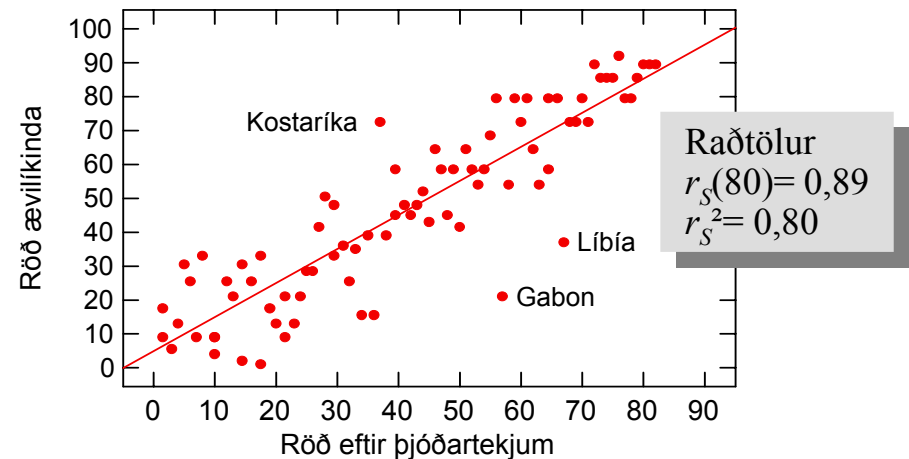
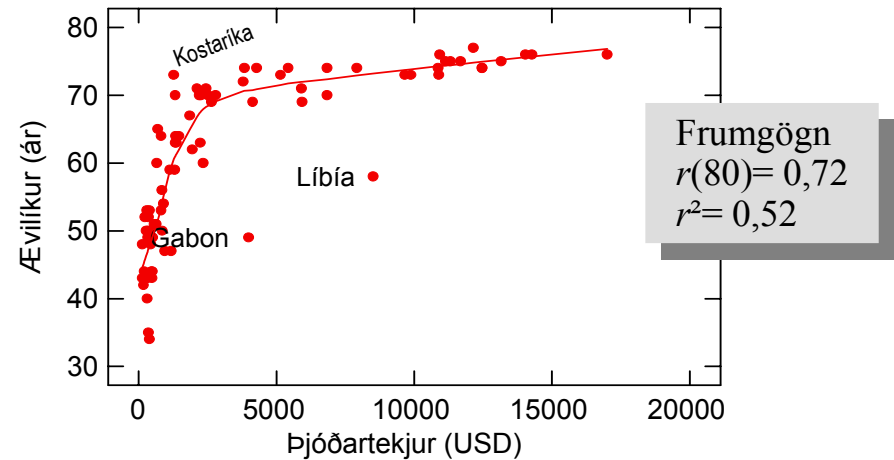
Frávillingar eru óvenjuleg stök sem er illa lýst og geta skekkt líkanið

Kostaríka, Líbía og Gabon víkja langt frá línunni miðað við raðtölur og hafa því mikla leif.

Raðfylgni felur í sér líkan af tengslum þjóðartekna og ævilíkinda, þ.e. að breyting um eitt sæti í þjóðartekjum skili sér sem um 0,9 sæti í ævilíkum.

Þetta líkan á við um öll löndin nema frávillingana þrjá; líkanið lýsir þeim löndum illa. Það veur því spurningar um það hvort þessi lönd séu á einhvern hátt eðlisólík hinum.

Frávillingar skekkja oft niðurstöður, en ekki hér þar sem þrjú lönd af 82 nægja ekki til að valda umtalsverðri skekkju.



# Túlkun á $r$ og $r_s$

## Pearson fylgnistuðull

$r$  er næmt fyrir bæði einbreytu- og tvíbreytufrávillingum

$r^2$  gefur hlutfallslega skýringu beinnar línu

$r$  samsvarar hallatölu ( $b$ ) staðlaðra breyta

$r$  metur aðeins beinlínutengsl

Hátt  $r$  þarf ekki að þýða orsakatengsl

- fölsk (*spurious*) tengsl
- ýmis orsakalíkön

## Spearman raðfylgnistuðull

$r_s$  er ónæmur fyrir einbreytu- en næmt fyrir tvíbreytufrávillingum

$r_s^2$  gefur hlutfallslega skýringu raðtalna

$r_s$  samsvarar hallatölu raðtalna

$r_s$  metur einhallatengsl (*monotonic relationship*)

Hátt  $r_s$  þarf ekki að þýða orsakatengsl

- fölsk (*spurious*) tengsl
- ýmis orsakalíkön