MERE SKLADNOSTI

doc.dr. ***Gaj Vidmar***

Univerzitetni rehabilitacijski inštitut Republike Slovenije - Soča

Univerza v Ljubljani Medicinska fakulteta

ibmi

Opisni podatki (nominalni)

- Najpreprostejše: % skladnosti
- Prve statistike
  - Bennnet et al. S (1954), “ponovno izumljen” kot
    - G (Holley & Guilford, 1964)
    - R.E. (random error, Maxwell, 1970)
    - C (Janson & Vegelius, 1979)
    - κ_n (Brennan & Prediger, 1981)
    - I_r (intercoder reliability, Perreault & Leigh, 1989)
- Scottov Π (1955) → Cohenov κ (1960)

Opisni podatki (nominalni)

- Cohenov koeficient κ:

$$\kappa = \frac{\text{opaženi delež skladnosti} - \text{pričakovani delež skladnosti}}{[1 - \text{pričakovani delež skladnosti}]}$$

  kar je računsko enako kot

$$\kappa = \frac{o - p}{[N - p]}$$

  pri čemer je
  - $o=\Sigma$ opaženih frekvenc na diagonali, ki predstavlja skladnost
  - $p=\Sigma$ frekvenc na diagonali, ki predstavlja skladnost, pričakovanih po naključju
  - N=velikost vzorca (št. cenjevanih/razvrščanih oseb/predmetov)
- Povezava z McNemarovim testom
  - 2. oc. <različna>, torej $p \downarrow \Leftrightarrow \kappa \rightarrow 0$
  - 2. oc. <enaka>, torej $p \uparrow \Leftrightarrow \kappa \rightarrow 1$

Opisni podatki (nominalni)

- Fleiss' κ is the multi-rater generalization of Scott's Π rather than Cohen's κ; it assumes marginal homogeneity when calculating expected frequencies, while Cohen's κ doesn't
  - in Cohen's κ, Eii = Pi. × P.i
  - in Scott's Π, Eii = $((\text{Pi.} + \text{P.i})/2)^2$
- Razširjena κ (več ocenjevalcev)
- Utežena κ (poseben primer je ICC)

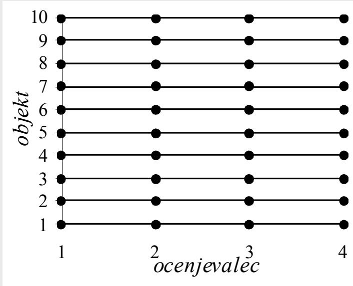
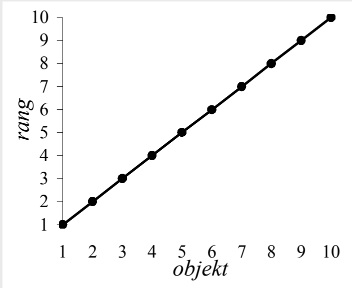
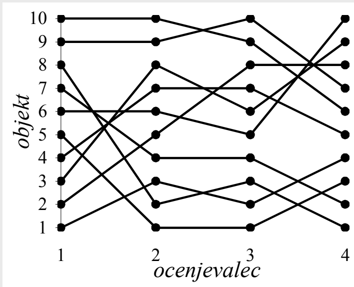
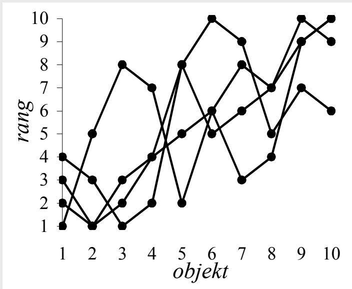
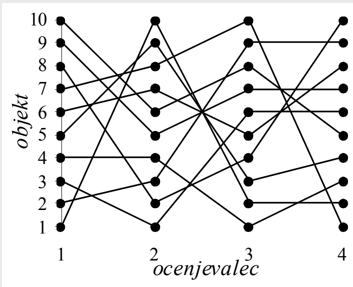
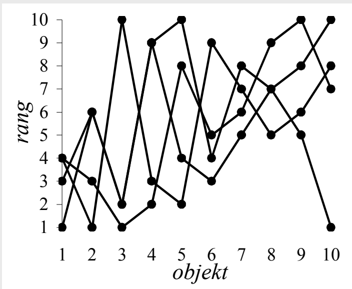
Urejenostni podatki (ordinalni)

- Kendallov *W* in sorodne mere
- metoda Svenssonove (rehabilitacija)
- Legendreov pristop
  - eksaktni permutacijski *p*-ji za *W* in a-posteriori τ-e (Holm)
  - združevanje ocenjevalcev v skupine

KonkD – čemu

- naraščajoč pomen analize urejenostnih (ordinalnih) podatkov (~Nelson in Pepe, 2000)
- številne raziskovalne hipoteze so v resnici urejenostne narave (~Cliff, 1996)
- prikaz [opisnih] podatkov doživlja razcvet (~InfoVis; Friendly, 2000); vizualizacije konkordance še ni [bilo]
- VK dopolnjuje primerjavo K med skupinami (~Schucany & ..., 1973-84; Vidmar & Černigoj, 2004; Legendre, 2005)
- Vidmar G., Rode N. (2007). Visualizing Concordance. *Computational Statistics*, 2007, *22*(4), 499-509

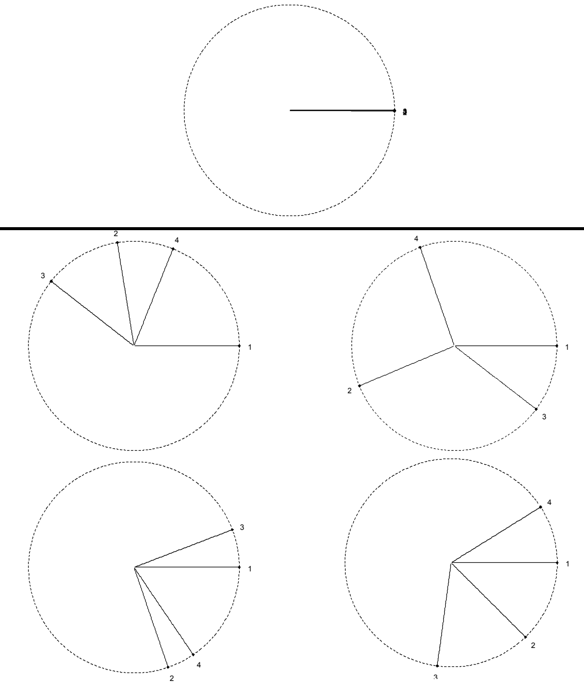
KonkD – kaj že obstaja

	ureditev "po ocenjevalcih"	ureditev "po objektih"
popolna konkor-danca *W*=1		
zmerna konkor-danca *W*=0,57		
nizka konkor-danca *W*=0,23		

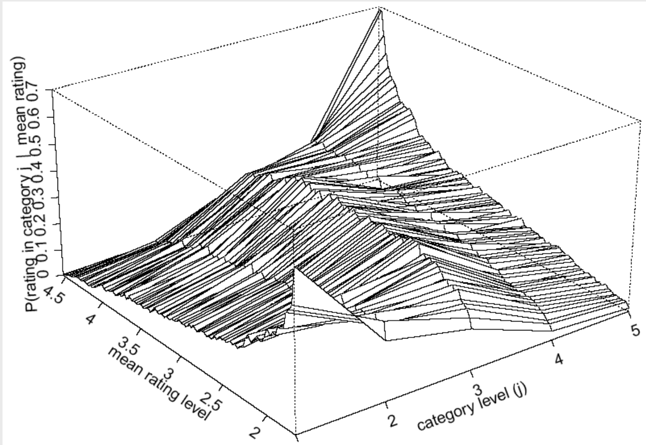
***W*=1**

k = *m* = 4

***W* = 0**



Trosset (2005)



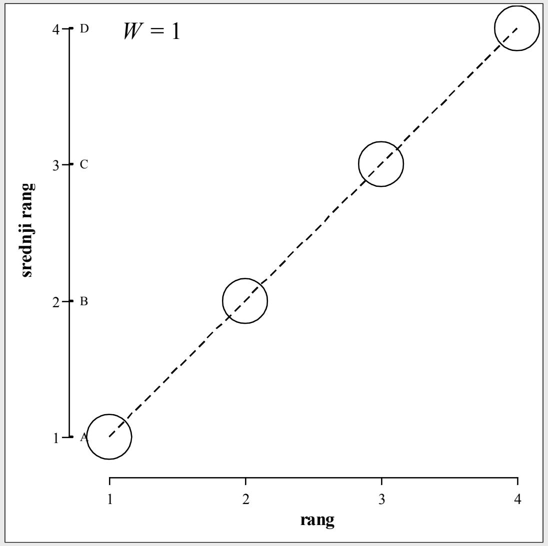
Nelson & Pepe (2000)

KonkD – 1. možnost

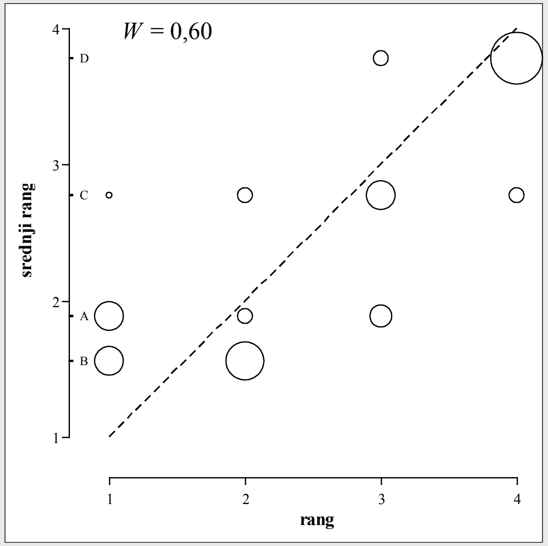
- konkordanca … ***m* rangiranj *k* objektov**
- često primerjava med skupinami

- 1. možnost: za vsak objekt **narišemo vse range**, ki so mu bili dodeljeni
  - **mehurčni konkordančni diagram**
    - frekvenca rangov kot funkcija povprečnega ranga
    - velikost kroga $\propto$ št. dodeljenih rangov
    - prikaz povprečnih rangov za objekte
    - popolna konkordanca $\Leftrightarrow$ vsi krogi na glavni diagonali (---)
    - vprašanje merila za polmer krogov
      - psihofizika!
      - pomen konteksta!
    - alternativa razsevni diag. z raztresenjem (jittered scatter-plot)
      - nižje razmerje podatki/črnilo
      - pri veliko ocenjevalcih nastopi prekrivanje točk med sosednjimi

KonkD – mehurčni



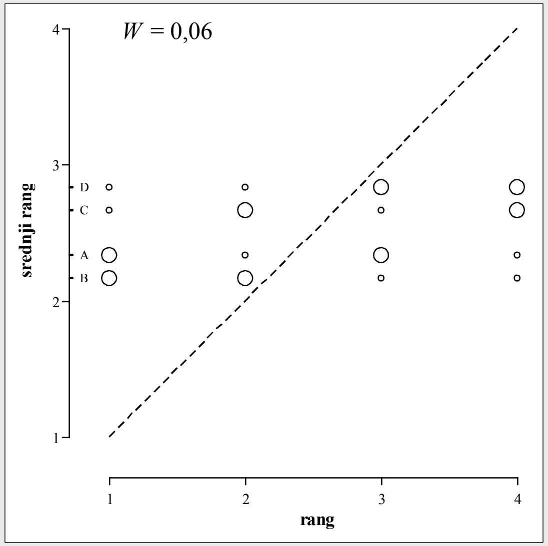
← umetni podatki



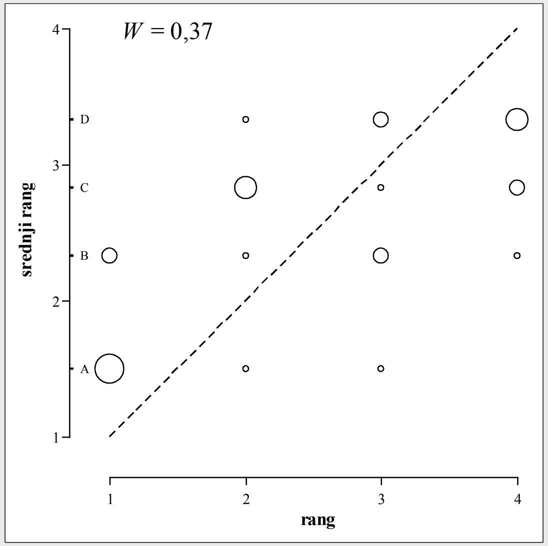
k=4

m=6↗

m=9

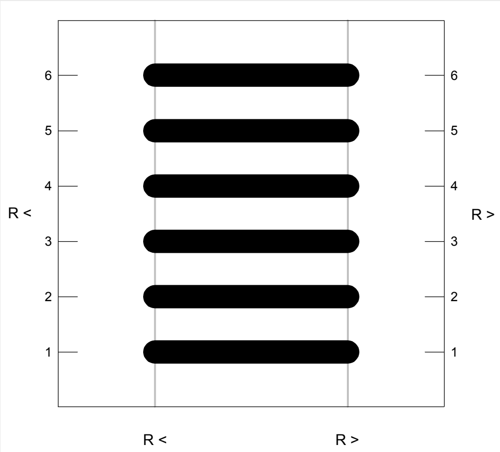


Schucany & Frawley (1973) – US in FRA ocenjevalci vin →

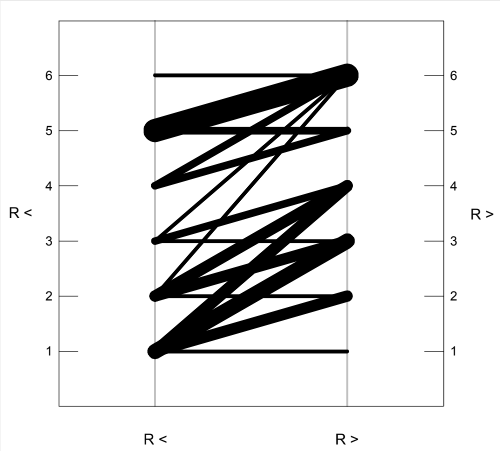


KonkD – z vzp. osema

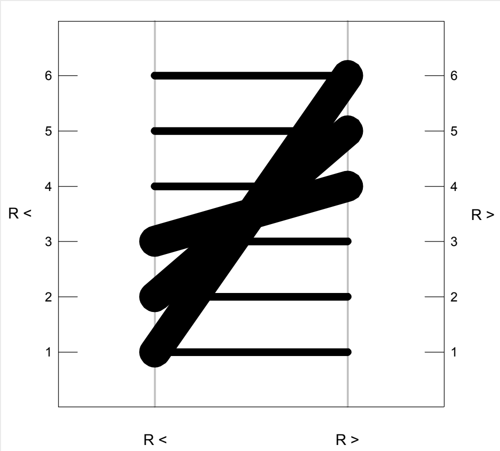
- **konkordančni diagram z vzporednima osema**
  - vzporedne osi dolgo znane, a šele od 1990-ih popularne (podatkovno rudarjenje, InfoVis)
  - vse dvojice rangov, v vsaki manjši ($R_<$) in večji rang ($R_>$)
  - število dvojic $\propto$ debelina črte
  - k=6, m=4 (umetni podatki)



W=1



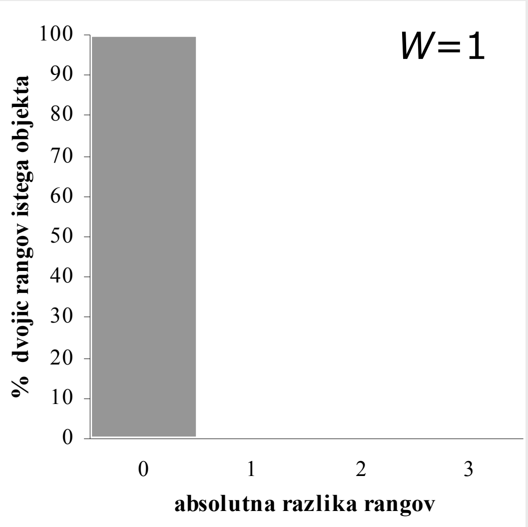
W=0,63

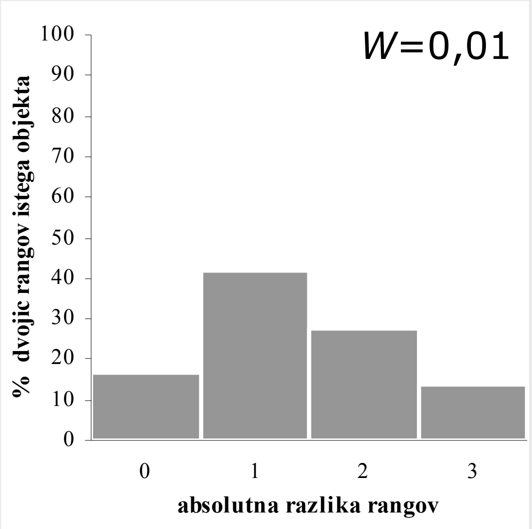


W=0

KonkD – 2. možnost

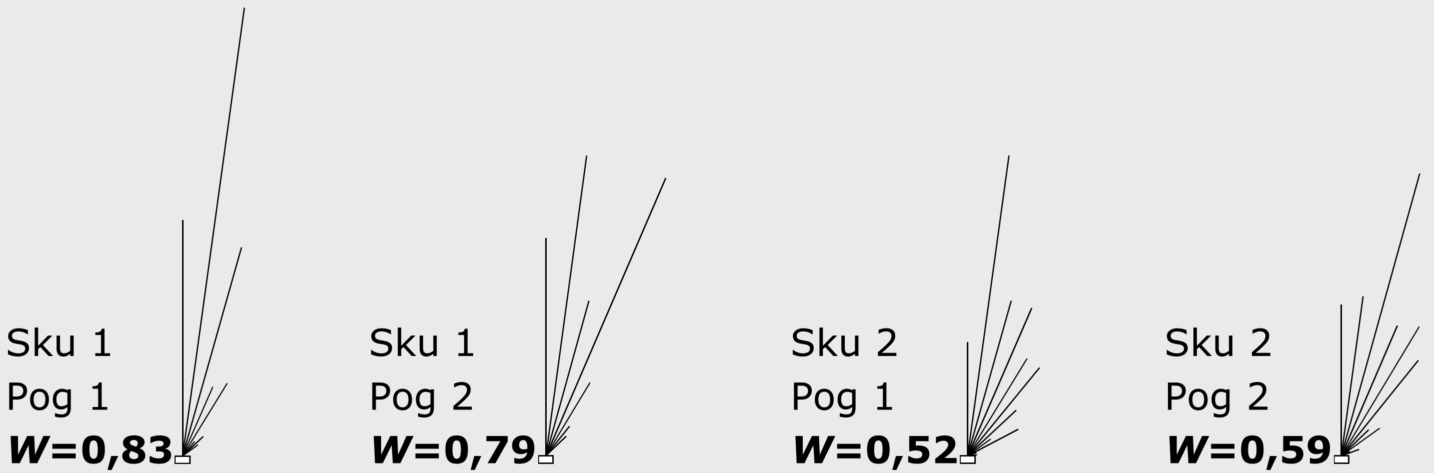
- 2. možnost: **rišemo razlike znotraj parov rangov**
  - na njih temeljita za Spearmanov ρ in Kendallov τ
  - $\forall |d_i| \in [0\ ,\ k-1]$
  - $\{|d_i|\}$ ima $m\ (m - 1) / 2$ elementov
  - **konkordančni stolpčni diagram**
    - abs. razlika rangov je urejenostna spremenljivka
    - ↓ konkordanca ⇔ ↑ E(|*d*|) ⇔ daljši rep porazdelitve
    - *k*=4, *m*=9 (umetni podatki)





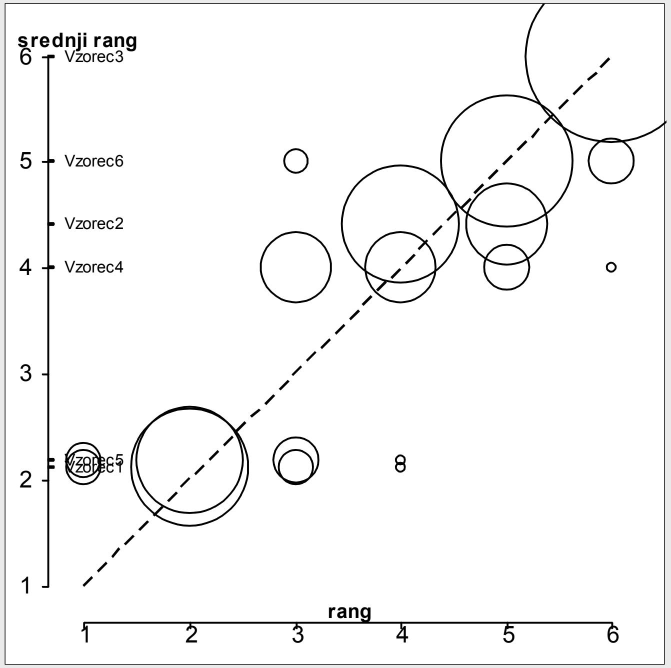
KonkD – blazinice z bucikami

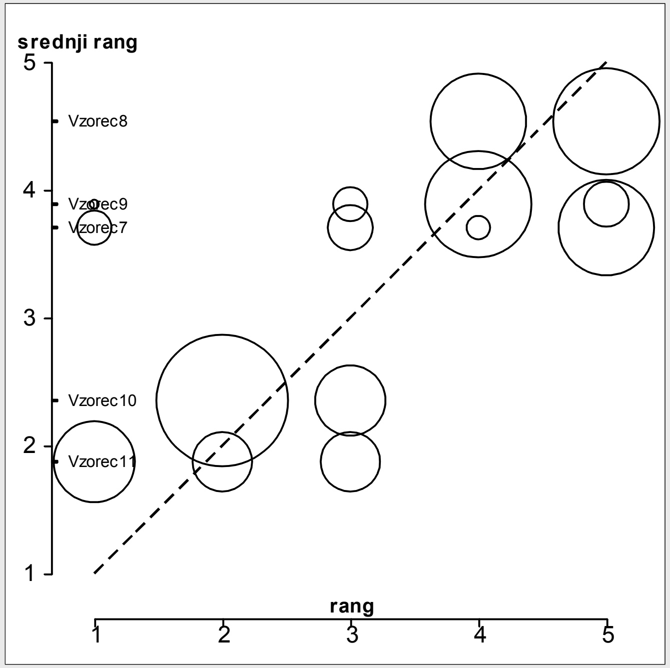
- **diagram blazinice z bucikami**
  - polarne koordinate
    - razlika 0 ∝ navpična črta
    - kot "v desno" = 90° {1 – [|*d*| / (*k* – 1)]}
    - razlike ni v podatkih ⇒ daljice ne rišemo
    - dolžina daljice ∝ št. razlik
  - v izhodišču "blazinica" (širina=2, višina=1)
  - *k*=6, *m*=4 (Vidmar & Černigoj, 2004, socialne norme)



KonkD – dod. primer 1

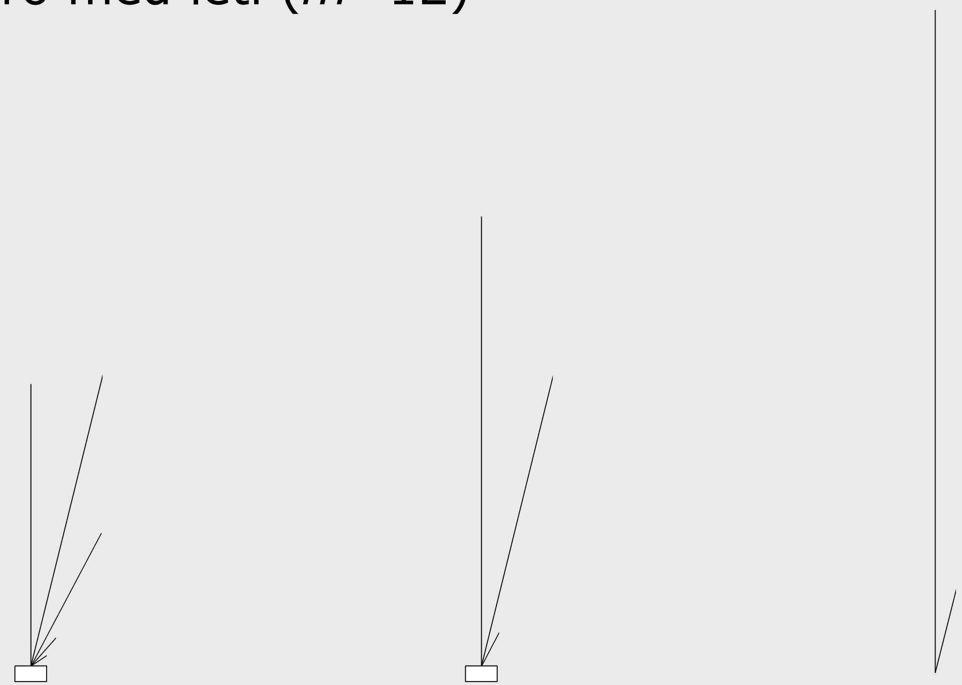
- **mehurčni konkordančni diagrami (medicina)**
  - vse pogostejše študije soglasja med ocenjevalci ali diagnostičnimi postopki glede urejenostnih ocen
  - Netto *et al*. (2006): *m*=17 patologov, *k*=6 biopsijskih vzorcev
    - stadij kroničnega hepatitisa C po Battsu in Ludwigu (*W*=0,85)
    - globalna stopnja akutne celične zavrnitve po Banhoffu (*W*=0,57)





KonkD – dod. primer

- **diagrami blazinice z bucikami (scientometrija)**
  - JCR 1994-2005 (IZUM) – skladnost rangiranja glede IF
  - k=7 področij (agr, fiz, kem, med, psi, sta, rač)
    - za vsako leto med merami (m=4: $\max_{IF}$, M_{20}, Me_{10}, $No_{čzIF}$)
    - za vsako mero med leti (m=12)



***W*=0,82** (min. med merami)

***W*=0,95** (min. med merami)

***W*=0,99** (maks. med leti)

Skladnost kot vidik zanesljivosti

- intervalni podatki
- koeficient intraklasne korelacije (ICC)
  - ICC gre proti največji možni vrednosti 1, ko ni variabilnosti med ocenjevalci, t.j. ko vsakega ocenjevanca vsi ocenjevalci ocenijo enako, kar pomeni, da je vsa variabilnost zgolj posledica razlik med ocenjevanci
  - Če so podatki urejeni v matriko, kjer vrstice predstavljajo ocenjevance, stolpci pa ocene posameznih ocenjevalcev, bo ICC torej višji, bolj ko si bodo ocene v dani vrstici podobne
  - splošna def. ICC =

$$\frac{\textbf{pojasnjena varianca}\text{ (t.j. varianca, ki jo pojasnjujejo dejanske razlike v merjeni lastnosti)}}{\textbf{skupna varianca}\text{ (t.j. pojasnjena varianca + varianca med ocenjevalci + variance napake)}}$$

Modeli oz. oblike ICC

- the judges all judges of interest or conceived as a random sample of possible judges?
- all targets rated or only a random sample?
- reliability to be measured based on individual ratings or on mean ratings of all judges?
- **6 oblik**
  - ICC(1,1) enosmerni model za posamezno meritev
  - ICC(1,k) enosmerni model za povprečje meritev
  - ICC(2,1) dvosmerni slučajni model za pos. meritev
  - ICC(2,k) dvosmerni slučajni model za povp. mer.
  - ICC(3,1) dvosmerni mešani model za pos. meritev
  - ICC(3,k) dvosmerni mešani model za povp. mer.

Modeli oz. oblike ICC

- One-way random effects model
  - Judges/raters are conceived as being a random selection of possible raters/judges, who rate all targets of interest. This model applies even when the researcher cannot associate a particular subject with a particular rater because information is lacking about which judge assigned which score to a subject. This would happen if the columns were first rating of a subject, second rating, third rating, etc., but a given rating (ex., the first rating) for one subject might be by a different judge than the first rating for another subject, etc.
- Two-way random effects model
  - Judges are conceived as being a random selection from among all possible judges, and targets/subjects are conceived as being a random factor too. Raters rate all n subjects/targets chosen at random from a pool of targets/subjects and it is known how each judge rated each subject. The ICC is interpreted as being generalisable to all possible judges.
- Two-way mixed model
  - All judges of interest rate all targets, which are a random sample. This is a mixed model because the judges are seen as a fixed effect (not as a random sample of all possible raters/judges) and the targets are a random effect. The ICC is interpreted as not being generalisable beyond the given judges.

Modeli oz. oblike ICC

- Predvidena prihodnja raba meritev:
  - **Single measure** reliability: individual ratings constitute the unit of analysis. Use this if further research will use the ratings of a single rater.
  - **Average measure** reliability: the mean of all ratings is the unit of analysis. This estimates reliability of the mean of the ratings of all raters. Use this if the research design involves averaging multiple ratings for each item.
- Dodatno pri vseh štirih dvosmernih modelih ločimo, ali želimo ocenjevati **absolutno strinjanje** (ko nam pomeni pristranost – bias, t.j. različno strog kriterij med ocenjvalci – vir nezanesljivosti) ali **konsistentnost** (ko variance med ocenjevalci, ki izhaja iz pristranosti, ne štejemo v varianco napake). Na ta način dobimo skupaj 10 oblik ICC.
- Cronbachov α je posebna oblika ICC

Razmernostni pod. (racionalni)

- Skladnost med metodami
  - postopek Blanda in Altmana (meje skladnosti)
    - navaden
    - za ponovljene meritve
  - regresija
    - navadna skozi izhodišče (v prim. z diagonalo)
    - Deming ↗Three Ways of Calculating a Regression
      - Model II regression refers to errors also in X (as opposed to only in Y in Model I), whereby MA and RMA are two possible approaches)
      - RMA line is not quite in the middle btw. y-to-x and x-to-y regression lines (all these lines are pinned at the {X-bar, Y-bar} point), because that is yet another approach, but it is quite close to that, since RMA slope is the geometric mean of b_yx and b_xy -- which is why RMA regression is also called Geometric Mean (GM) regression
      - two further synonims for RMA regression are least products regression (by analogy to least squares; and it also goes by the name of least rectangles) and standardised principal component regression (SPCR)
      - RMA regression is a particular case of the general Deming model (i.e., his solution to the Model II problem), albeit the most widely used one; c.f. also York (physics, 196*-200*)
    - Passing in Bablock (neparametrična)
  - Linov koef. skladnostne korelacije (CCC)
    - ↗Wikipedia

Splošne mere skladnosti

- Robustne in posplošene oblike CCC
  - neparametrična, vklj. krnjenje
  - robustno ocenjena razdalja (Huber, Winsor)
  - posplošena verzija za ponovljena merjenja (King, tudi za binarne podatke)
- Krippendorfova α
  - ↗Hayes A.F., Krippendorf K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods And Measures, 1*(1), 77-89

    (vključuje pregled celotnega področja)

PROGRAMJE [VLADA!]

- Linovs CCC
  - MedCalc, GenStat
  - makroji/koda Stata, SAS, SPSS (tudi za Kripp.α), R
- MedCalc tudi vse ostalo (κ, ICC, B&A in diagram gore, regresiji za prim. metod)
- SPSS/PASW
  - "specialist" za ICC, seveda Cronbachova α in Kendallov τ
  - κ Cohenova
  - popoln komplet makrojev (M.García-Granero, 2009)
  - makro za Krippendorfovo α (A.F.Hayes)
- Minitab ↗FAQ – Quality – Gage (4 zasloni!)
  - κ Cohenova in Fleissova
  - Kendallov τ
  - pristopi QC/SPC (Gage R&R, CC etc.)
- R (paket MethComp)