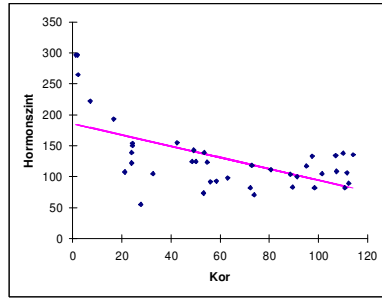


Pl.: Korreláció és görbeillesztés

Regressziós statisztika	
r értéke	0.593819
r-négyszet	0.352621
Korrigált r-négyszet	0.335585
Standard hiba	45.31186
Megfigyelések	40

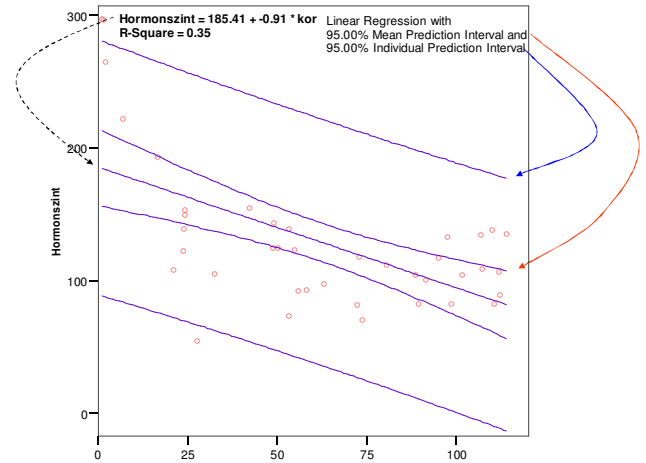


	Koefficiensek	Standard hiba	t érték	p-érték	Alsó 95%	Felső 95%
Tengelymetszet	185.4129	14.28497	12.97958	1.53E-15	156.4945	214.3313
Kor	-0.9101	0.200043	-4.54953	5.36E-05	-1.31507	-0.50514

Varga J.

Statisztika: gyakori módszerek és hibák

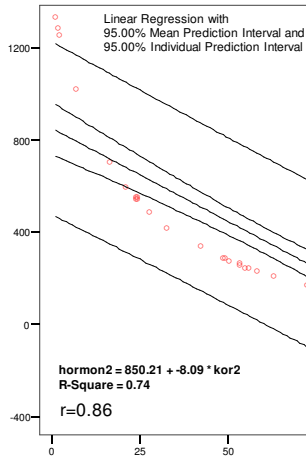
84



Varga J.

Statisztika: gyakori módszerek és hibák

85



Rank correlations		
Kendall's tau_b	Correlation Coefficient	-0.94
	Sig. (2-tailed)	1E-06
	N	40
Spearman's rho	Correlation Coefficient	-0.99
	Sig. (2-tailed)	1E-06
	N	40

Varga J.

Statisztika: gyakori módszerek és hibák

86

Tanulság:

- A Pearson-féle korrelációs együttható a lineáris kapcsolat szorosságát méri
- A kapcsolat megléte a meredekségre végzett t-próbával ellenőrizhető (0-hipótesis: meredekség=0)
- Az illesztett egyenes és az egyedi pontok konfidencia-intervalluma különbözik
- Nemlineáris kapcsolat szorosságát jobban mérik a rangsorolás utáni (nemparaméteres) korrelációs mutatók

Varga J.

Statisztika: gyakori módszerek és hibák

87

Korreláció-számítás

- Függetlenek az esetek?
Homogének a csoportok?
- Az X és Y adat független?
Nem „előtte-utána” típusú adatpárok?
- Mért adat van az x-tengelyen?
(Nem: idő, koncentráció, dózis ⇒ lin. regr.)
- Lineáris a kapcsolat?
- X és Y normális eloszlású?

Varga J.

Statisztika: gyakori módszerek és hibák

88

r² értelmezése korreláció-számításból

- A két változó szórásnégyzetének az a hányada, amit megosztanak egymással.
- Pl.: r²=0.62
- ⇒ az X szórásnégyzetének 62%-a magyarázható az Y szórásnégyzetével (következik az Y szórásnégyzetéből)
- Ugyanígy:
⇒ az Y szórásnégyzetének 62%-a magyarázható az X szórásnégyzetével (következik az X szórásnégyzetéből)

Varga J.

Statisztika: gyakori módszerek és hibák

89

Egyenes-fektetés („lineáris regresszió”)

- Tényleg egyenes rajzolható az adatokra?
Ha nem: nemlineáris görbeillesztés!
- Kb. normális eloszlás szerint szórnak az adatok az egyenes körül?
- Mindenütt azonos a szórás?
Ha nem: súlyozott illesztés!
- Pontosan ismert az X (nem valószínűségi változó) ?
- Függetlenek az adatpontok?

Varga J.

Statisztika: gyakori módszerek és hibák

90

r² értelmezése egyenes-illesztésnél

- Az Y szórásnégyzetének az a hányada, amit az X változása okoz
- A pontok szóródása az illesztett egyenes körül az Y teljes varianciájának (1-r²) hányadát teszi ki

Varga J.

Statisztika: gyakori módszerek és hibák

91

(Hyper)sík fektetése („multilin. regr.”)

- Minden független változóval lineáris az Y kapcsolata?
- Kb. normális eloszlás szerint szórnak az adatok a sík körül?
- Mindenhol azonos a szórás?
Ha nem: súlyozott illesztés!
- Minden független változó pontosan ismert (nem valószínűségi változó) ?
- Függetlenek az adatpontok?

Varga J.

Statisztika: gyakori módszerek és hibák

92

„Rank correlation”

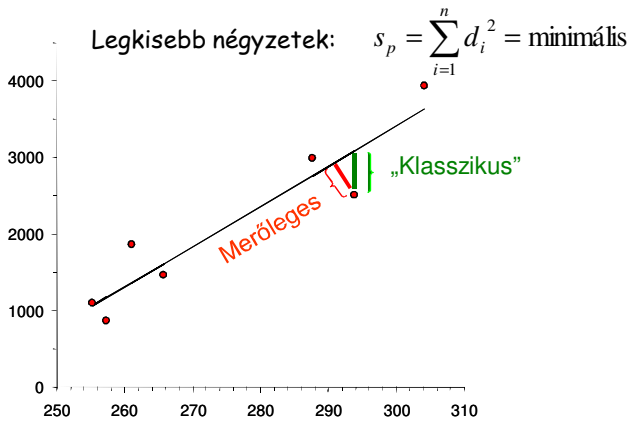
- Nem érzékeny a kieső pontokra
- Ha nem lineáris a kapcsolat
- Egyetlen eloszlású adatoknál
- Ha mérési határon túli adatok is vannak (<alsó határ)
- Ha kategóriákba sorolt változók összefüggését nézzük

Varga J.

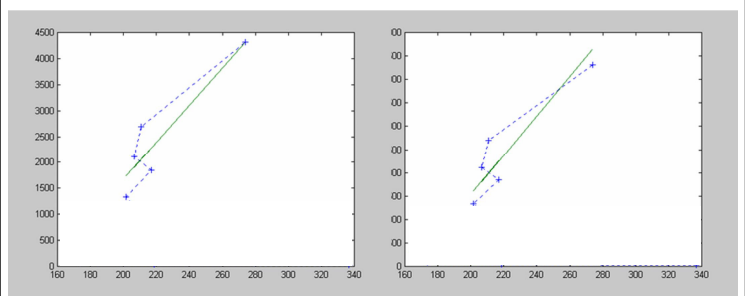
Statisztika: gyakori módszerek és hibák

93

Egyenes-fektetés: modellek



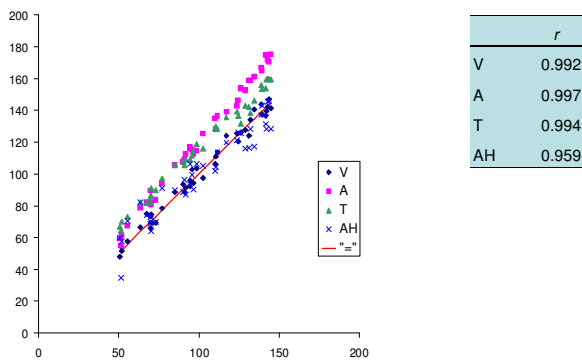
PI.: Az illesztett egyenesek különböznek



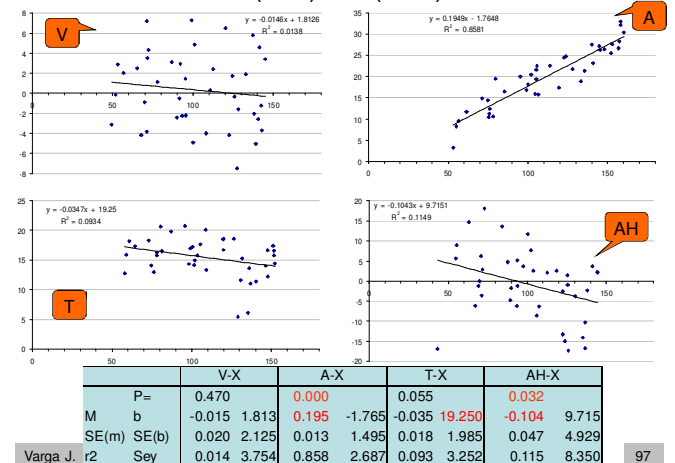
„Klasszikus”

Merőleges

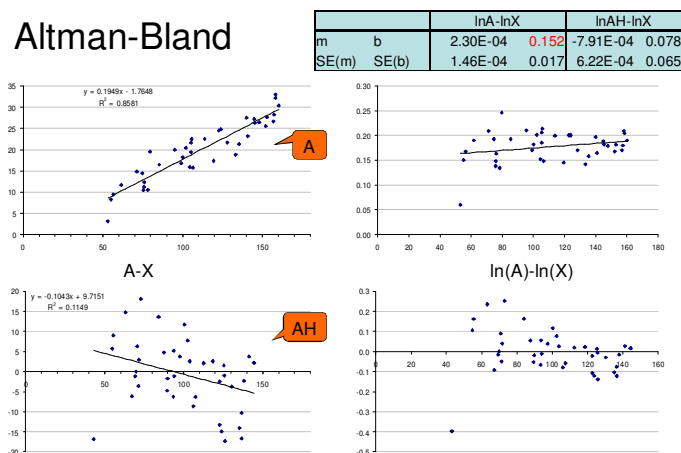
Két módszer összehasonlítása



Altman-Bland: (Y-X) vs. (Y+X)/2



Altman-Bland



Módszerek összevetése: Tanulság

- A (Pearson-féle) korrelációs együttható NEM alkalmas ugyanazt a paramétert mérő különböző módszerek összehasonlítására
- Az eltérés két módszer között 3 féle lehet:
 - Véletlen hiba
 - Állandó torzítás („bias”)
 - Arányos (értéktől függő) hiba
- Az Altman-Bland diagram segít ezek elkülönítésében:
 - A tengelymetszet különbözik 0-tól: állandó torzítás
 - A meredekség különbözik 0-tól: arányos hiba

Varga J.

Statisztika: gyakori módszerek és hibák

99

Hiba-típusok

- **Torzítás** („inaccuracy“): a mért érték eltérése a valóditól mérőszám: $\bar{x} - \mu$
- **Pontosság** („precision“): mennyire szórnak a mért értékek mérőszám: relatív hiba (variation coefficient):

$$VC = \frac{SD}{\bar{x}} \cdot 100\%$$

Varga J.

Statisztika: gyakori módszerek és hibák

100

Mérési hiba fajtái:

- Durva hiba
 - elnézés,
 - csere
 - ...
- Rendszeres (szisztematikus) hiba
 - a módszer vagy az eszközök nem megfelelő kalibrálása
- Véletlen hiba
 - a minta változékonysága
 - bemérési pontatlanság
 - a mérés bizonytalanága

Varga J.

Statisztika: gyakori módszerek és hibák

101

Hiba-terjedés

- Ha az f paramétert az x_1, x_2, \dots, x_n mért értékekből számoljuk:

$$f = f(x_1, x_2, \dots, x_n)$$

- Akkor a mérési hibája kb:

$$\Delta f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot \Delta x_i^2$$

Varga J.

Statisztika: gyakori módszerek és hibák

102

Példák: Hibaterjedés

$$\Delta f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot \Delta x_i^2$$

Példák egyszerű képletekre:

f	$\partial f / \partial x$	$\partial f / \partial y$
$a \cdot x + b \cdot y$	a	b
x / y	$1/y$	$-x/y^2$
$\ln(x)$	$1/x$	

- Beütésszámok: minta (b) és háttér (h) hibájuk: \sqrt{b} és \sqrt{h}
- Akkor a különbségük hibája:

$$\Delta(b-h) = \sqrt{\sqrt{b}^2 + \sqrt{h}^2} = \sqrt{b+h}$$

Varga J.

Statisztika: gyakori módszerek és hibák

103

Tanulság

- A valószínűségi változókból (véletlen hibával terhelt mennyiségekből) számolt mintafüggvények is hibával terheltek
- A hiba terjedése nem lineáris
- A számolt paraméter hibája sokkal magasabb is lehet, mint az egyes összetevőké.

Varga J.

Statisztika: gyakori módszerek és hibák

104

Néhány nem-paraméteres próba

	Paraméteres	Nem-paraméteres
1 minta középértéke	1 mintás t	Wilcoxon előjeles sorrend
2 középérték, párosított	párosított t	Wilcoxon, párosított
2 középérték, párosítatlan	2 mintás t, d	Mann-Whitney
Több minta, csoportosított	Ismétléses ANOVA	Friedman
Több minta	ANOVA	Kruskal-Wallis
Részarány, párosított		McNemar
Részarány, párosítatlan		Fisher-féle exakt χ^2

Varga J.

Statisztika: gyakori módszerek és hibák

105

Párosított Wilcoxon-próba

- Függetlenek a párok?
- Indokolt a párba állítás?
- Pontosan 2 csoportot hasonlítunk össze?
- A véletlen hatás egyértelműen nem normális eloszlású?
Ha igen: párosított t-próba!
- A különbségek szimmetrikus eloszlásúak?

Varga J.

Statisztika: gyakori módszerek és hibák

106

Mann-Whitney próba

- Független hibák?
- Nem párosítható adatok?
- Pontosan 2 csoport?
- Azonos eloszlásúak?
- Biztosan a mediánt akarod összehasonlítani?
- Ha egyoldalú próba: valóban lehet az?
- Nem normális eloszlású populáció?

Varga J.

Statisztika: gyakori módszerek és hibák

107

Friedman-próba

- Az összetartozó adatok csak egy szempontban különböznek? (A lényeg az egyéni variabilitás kizárása!)
- Függetlenek az egyedek?
- A véletlen hatás egyértelműen nem normális eloszlású?
Ha igen: (ismétléses) ANOVA!

Varga J.

Statisztika: gyakori módszerek és hibák

108

Kruskal-Wallis próba

- Független „hibák”?
Egy adat eltolódása nem befolyásolja a többit?
- Nem összetartozó adatok?
Ha igen: Friedman!
- A csoportok egyértelműen nem normális eloszlásúak?
Ha igen: ANOVA!
- Tényleg csak a mediánok különbsége érdekel?
Nem a csoportok átfedése?
- Hasonló eloszlásúak a csoportok?

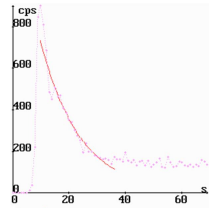
Varga J.

Statisztika: gyakori módszerek és hibák

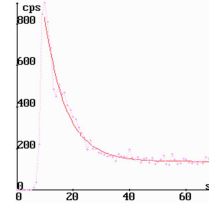
109

Paraméter-illesztés: (Bi)exponenciális

Exponenciális ($a \cdot \exp(b \cdot x)$)
 $a = 1499 \pm 114.8$
 $b = -7.144E-05 \pm 3.813E-06$



BIEXP. ILL. ($a \cdot \exp(b \cdot x) + c \cdot \exp(d \cdot x)$)
 $a = 2603 \pm 315.6$
 $b = -0.000132 \pm 1.111E-05$
 $c = 138.2 \pm 24.27$
 $d = -6.989E-14 \pm 3.052E-06$



Túl sok paraméteres modell egyes komponensei egymás rovására bizonytalanok lehetnek!

Varga J.

Statisztika: gyakori módszerek és hibák

110

Százalékok

- Mikor számolható?
- 100 minta fölött?!
- Részarány kifejezése kis számoknál:
pl. 7/16 írandó 63.6% helyett!

Varga J.

Statisztika: gyakori módszerek és hibák

111

Adatformatum: Tizedes jegyek, kerekítés

- A nyers adatokat azon tizedesjeggyel bezárólag írjuk ki, amely a szórás első értékes jegye.
- Ha a szórás első értékes jegye „1”, még egy jegyet kiírunk
- Az átlagot, szórást általában ennél 1 jeggyel pontosabban
- Pl.: ha $SD=0.04356$
 Adat: 12.07
 Átlag: 12.008, szórás: 0.044

Varga J.

Statisztika: gyakori módszerek és hibák

112

Szórás megadása

- Általában nem szeretik a „ \pm ” -t

	Helytelen	Helyes
Átlag és szórás	4.05 \pm 0.24	4.05 (SD 0.24)
Átlag és szórása	4.052 \pm 0.038	4.052 (SE 0.038)
Konfidencia-intervallum	4.052 \pm 0.076	3.98 to 4.13

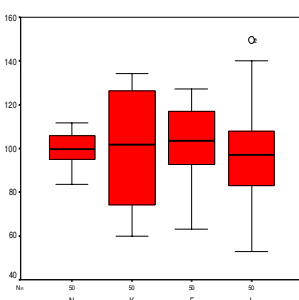
Varga J.

Statisztika: gyakori módszerek és hibák

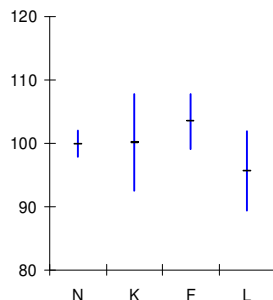
113

Hibavonalak

Adatoknál:
Terjedelemből vagy referencia-tartomány



Átlagoknál:
Konfidencia-intervallum



Varga J.

Statisztika: gyakori módszerek és hibák

114