

Korrelációs ellipszis alkalmazása és szerkesztése

SVÁB JÁNOS

MTA Mezőgazdasági Kutatóintézete, Martonvásár

Kutatómunka során gyakran vizsgáljuk két változó, X és Y együttes alakulását. Kétféle kérdésünk lehet: 1. egy általunk választott X -értékhez mekkora Y -értékek tartoznak, vagy mekkora átlagos Y -érték várható; 2. X és Y értékeinek lehetséges kombinációi milyen X , Y kombinációhatárok között mekkora gyakorisággal fordulnak elő.

Az első kérdésre a regressziós egyenlet, legegyszerűbb esetben a jól ismert $Y' = a + bX$ lineáris regressziós egyenlet ad választ. A különböző X -értékekhez tartozó Y -értékek szóródását és átlagos értékét ekkor az $1/A$. ábra szerinti *tolerancia-* és *konfidencia-sáv* fejezi ki. A regressziós egyenletben az X -változó értékei lehetnek előre tervezhetők, sőt ismételhetők is. Ekkor X nem véletlen változó. Ennek jellemző példái a dóziskísérletek, ahol $X = a$ dózis, $Y = a$ célváltozó, véletlen változó. De lehet X véletlen változó is, pl. $X =$ talaj humusz % és $Y =$ termés összefüggés-vizsgálatakor. Előbbit elsőfajú, utóbbit másodfajú regresszióknak nevezi a szakirodalom [1].

Az elsőfajú lineáris regresszió csak $Y = a_Y + b_{YX}X$ alakú lehet, még akkor is, ha Y -ből következtetünk X -re. A másodfajú lineáris regresszió esetén, ha Y -ből következtetünk X -re, $X = a_X + b_{XY}Y$ új egyenletet kell meghatározni, mert itt b_{YX} nem egyenlő b_{XY} -nal.

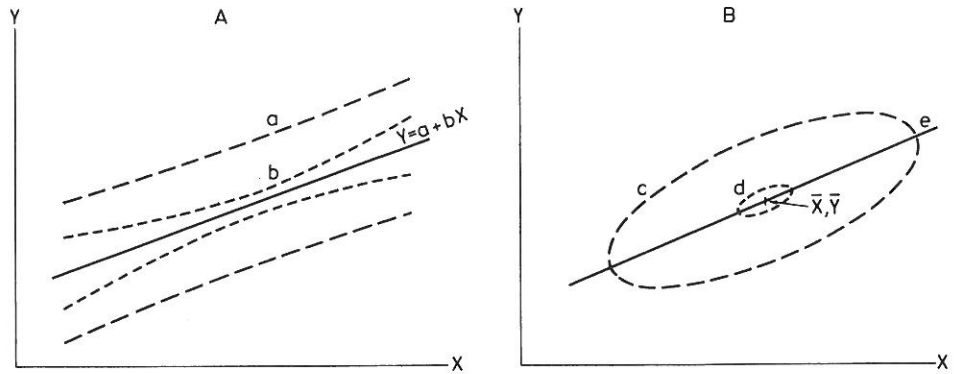
A második kérdés a másodfajú regresszió azon igen gyakori esetére vonatkozik, amikor nem kívánunk X -ről Y -ra, sem viszont következtetni, hanem az X , Y kombinációk együttes előfordulásának szóródását és várható gyakoriságát keressük. Ekkor az összefüggést regressziós egyenes (b_{YX} és/vagy b_{XY}) helyett korrelációs ellipszissel ábrázoljuk. A korrelációs ellipszis is kétféle lehet, az $1/B$. ábra szerint. Ha az egyedi X , Y kombinációkra vonatkozik, akkor *tolerancia-*, ha az \bar{X} , \bar{Y} főátlagra vonatkozik, akkor *konfidencia-ellipszis*. E két ellipszis analóg a szórásból és a középérték-szórásból számított *tolerancia-*, illetve *konfidencia-határokkal*, amikor csak egy változót vizsgálunk.

Három véletlen változó kombinációi háromdimenziós derékszögű koordináta-rendszerben ábrázolhatók. A kombinációkat ábrázoló pontfelhő sűrűsége akkor *korrelációs ellipszoiddal* érzékeltethető.

Ebben a dolgozatban a korrelációs ellipszis szerkesztését ismertetem, mivel — tudomásom szerint — ez a magyar szakirodalomban ez idáig nem történt meg, és a külföldi szakirodalomban is nehezen található meg, biometriai könyvekben nincs leírva. Bemutatok továbbá egy szerkesztési megoldást arra az általános esetre, amikor

a két változó mértékegységének *ábrázolt mérete* különböző. A szakirodalom ugyanis csak azokat az eseteket ismerteti, amikor a két változó mértékegységének az ábrázolt hossza azonos (pl. mindkettő 1 cm), függetlenül attól, hogy mi a változók mértékegysége.

Ezúton mondok köszönetet dr. JÓZSA SÁNDOR matematikusnak matematikai tanácsaiért és útbaigazításaiért. Köszönöm SIMITS KATALIN tanszéki mérnöknek a számítások elvégzését, ellenőrzését és az ábrák elkészítését, WELLISCH PÉTERnek pedig a kézirat elolvasásakor tett értékes megjegyzéseit.



1. ábra

Regresszió (A) és korreláció (B). a) tolerancia-sáv; b) konfidencia-sáv; c) tolerancia-ellipszis; d) konfidencia-ellipszis; e) nagytengely

Alkalmazás

A korrelációs ellipszis tulajdonképpen „harang alakú” kétdimenziós normális eloszlás vetülete a síkra. Ezért ugyanúgy használjuk modellnek, szemléltető támpontnak két véletlen változó együttes vizsgálatokor, mint az ismert haranggörbe alakú egydimenziós normális eloszlás vetületét a számegyenesre ($X \pm s$ stb.) egyetlen véletlen változó eloszlásának elemzésekor. Kétdimenziós normális eloszlás esetén a változók egyenként is normális eloszlásúak és nem kell hogy összefüggjenek. De ha összefüggnek egymással, regressziójuk lineáris, ami ábrázolva azt jelenti, hogy a köralapú „harang alak” az összefüggés mértékétől függően elkeskenyedik, alapja, azaz vetülete a síkra ellipszissé válik, egyenes nagy- és kistengellyel (lásd 5. ábrát).

A két változó szórásának és mértékegységének nem kell azonosnak lennie. E két esetben azonban ábrázolásakor óvatosságnak kell lenni, hogy ne kapjunk torzítottan elnyújtott ellipszist, ami téves értelmezéshez vezethet.

A korrelációs ellipszist háromféleképpen alkalmazhatjuk.

a) Ha tudjuk, vagy jóhiszeműen elvárható a vázolt feltételek teljesülése, akkor a $P\%$ -os szintű tolerancia-ellipszist ábrázoljuk a pontfelhő vagy az egy vagy két lineáris regressziós egyenes helyett.

b) Ábrázoljuk a pontfelhőt és a $P\%$ szintű tolerancia-ellipszist. Ekkor az a fő célunk, hogy összehasonlítsuk az X , Y kombinációk tényleges szóródását a kétdimenziós normális eloszlás esetén várható szóródással.

c) Ha kérdésünk az \bar{X} , \bar{Y} főátlag kombinációra vonatkozik, konfidencia-ellipszissel ábrázoljuk \bar{X} , \bar{Y} konfidencia-tartományát.

Példák:

1. Összefüggést keresünk a talaj P_2O_5 és K_2O ppm tartalma között. Vizsgálhatjuk, hogy milyen érték-kombináció-határokon belül fordul elő a minták 50%-a, 80%-a, 95%-a stb. Ekkor tolerancia-ellipszist szerkesztünk $P=0,5; 0,2; 0,05$ valószínűségi szintre. De vizsgálhatjuk, hogy a reprezentált talajokban milyen kombináció-határok között várható a kálium és a foszfor ppm közös középértéke, $P=0,01; 0,05$ stb. valószínűségi szinten. Ekkor konfidencia-ellipszissel adjuk meg a közös középérték-kombináció határértékeit.

2. Gyakran előfordul, hogy két változó értékei önmagukban nem extrémek, de együttes előfordulásuk szakmailag alig értelmezhető, feltételezhetően hibás a meghatározás, vagy különleges helyzetet takar. Ilyen lehet pl. a 10% $CaCO_3$ -ot tartalmazó talajban mért 6,2-es pH-érték. A tolerancia-ellipszis objektív eszközt ad az ilyen szélsőséges kombinációk automatikus felismeréséhez, elkülönítéséhez. Ez nagy adathalmazok elemzése esetén különösen jelentős, amikor az adatkombinációk szubjektív szakmai megítélése a nagy adattömeg nehéz áttekinthetősége miatt szinte lehetetlen.

3. Ugyanazt a tulajdonságot n számú objektumon két módszerrel mértük. Vizsgáljuk, hogy a két módszerrel kapott értékek mennyire korrelálnak egymással, azaz milyen pontos az átszámíthatóság, továbbá, hogy mely tartományban pontosabb vagy bizonytalanabb az átszámíthatóság. Ekkor az X - és az Y -tengely a két módszer értéktartománya, és a pontfelhő sűrűsödése, gócai jelzik az átszámíthatóság pontosságát, tartományait. Módszertani megfigyeléseinket a tolerancia-ellipszis segítségével így nagymértékben pontosabbá tehetjük.

4. A tolerancia-ellipszis, minthogy kétdimenziós normális eloszlás, vagyis egy „harang” vízszintes metszetének a vetülete az X , Y -síkra, az azonos sűrűségi, valószínűségi pontokat köti össze. Ezért egyetlen ábrán akár több ellipszist is szerkeszthetünk különböző valószínűségi, sűrűségi szintekre. Ezeket a megfigyelt pontfelhő sűrűségi gócaival vagy eloszlásával összevetve igen hasznos segítséget kapunk a két változó véletlen kombinálódását erősen befolyásoló tényezők hatásának felismerésére.

5. Ha a pontfelhő és a számított ellipszis között valamilyen szabályszerű eltérés figyelhető meg, pl. a pontfelhő tojás alakú, csúccsal az origó felé, akkor gyakran az egyik vagy mindkét változó megfelelő transzformációja után újból megszerkesztjük az ellipszist, most már a transzformálásnak megfelelő beosztású tengelyekkel. A transzformáció különösen indokolt, ha az X - és az Y -változók egyike lineáris, a másik négyzetes, köbös mértékegységű, pl. gyümölcs átmérője és súlya közötti összefüggés esetén.

Irodalmi áttekintés

A korrelációs ellipszis elemi geometriai ismeretekkel nem szerkeszthető meg. Az ellipszis alakja függ ugyanis a két változó összefüggésének szorosságától, azaz az r korrelációs koefficiensről, a két változó s_x és s_y szórásától, az ábra két tengelyének a beosztásától, végül az ellipszis nagysága a választott $P\%$ tolerancia-, illetve konfidencia-szinttől.

A magyar szakirodalomban tudomásom szerint csak JUVAN CZ—PAKSY [3] könyve ad néhány képletet a korrelációs ellipszis szerkesztésére. De a képletek nem elegendők a szerkesztés megoldásához. Képleteket közöl a Geigy gyár tudományos táblázatokat tartalmazó könyve [1]. A témakör felől tájékoztatást ad HEALY [2]. Újabban az ellipszis hajlásszögét és különböző szimmetria-tengelyeinek az összefüggését SCHNEIDER [4] taglalta részletesen. Háromdimenziós ellipszoiddal THÖNI [5] foglalkozott. Egyébként szakcikkek, biometriai, matematikai statisztikai könyvek is bemutatnak néha korrelációs ellipszist, de szerkesztési képleteket ezek nem ismertetnek.

Az egyébként is szűkös ez irányú szakirodalomban közölt képletek legfőbb hiányossága, mint már említettem, az, hogy az ellipszis főtengelyének hosszára és hajlásszögére vonatkozó képletek csak arra az esetre érvényesek, ha az X - és az Y -változó mértékegysége az ábrán azonos hosszúságú, függetlenül attól, hogy a mértékegységek valójában mit jelentenek. Így pl. 1 cm-nek véve az X -tengelyen a pH-t, és ugyancsak 1 cm-nek véve 1 ppm P_2O_5 talajvizsgálati értéket, nyilván gyakorlatilag lehetetlenné válna az ábrázolása. Kevésbé szélsőséges esetekben is igen elnyújtott ellipszist kapnánk. Ezt a számítás-ábrázolási ellentmondást egy γ -val jelölt korrekciós együtthatóval kiküszöböltem.

A korrelációs ellipszis szerkesztése

Példa:

43 napraforgó-tábla humusz % és Arany-féle kötöttségi szám (K_A) korrelációját, érték-kombinációik előfordulásának gyakorisági eloszlását vizsgáltuk. A korrelációs ellipszis szerkesztéséhez a következő mutatókra van szükség:

$$\begin{aligned} n &= 43; \text{ az adatpárok, objektumok száma} \\ \bar{X} &= 2,58; \text{ a humusz \%}, \text{ az } X\text{-változó középértéke;} \\ \bar{Y} &= 43,6; \text{ a kötöttségi szám, az } Y\text{-változó átlaga;} \\ s_x &= 0,7174; \text{ az } X\text{-változó szórása;} \\ s_y &= 9,993; \text{ az } Y\text{-változó szórása;} \\ r &= 0,6559; \text{ az } X\text{- és az } Y\text{-változó lineáris korrelációs koefficiense.} \end{aligned}$$

A tengelyek beosztása az ábrán

Az ábrát általában úgy célszerű szerkeszteni, hogy az ellipszis hajlásszöge közel 45° -os, vagy annál valamivel kisebb legyen. Ezt azzal érjük el, hogy a két változó adatainak a szóródástartományát (ez durván a szórás hatszorosa) közel egyforma hosszúságúnak, cm-ben kifejezve, q cm-nek ábrázoljuk.

Először megválasztjuk az X -tengely, az abszcissa beosztását; 1 cm-re eső X -szakaszt i_x -szel jelölve

$$i_x \approx \frac{6s_x}{q}. \quad (1)$$

Ha $n < 30$, akkor $6s_x$ helyett $4s_x$ -et tegyünk a számlálóba.

Az i_X -ből számítjuk ki az Y -tengely beosztását:

$$i_Y \approx \frac{s_Y}{s_X} i_X. \quad (2)$$

Akkor kapunk igen szemléletes ábrát, ha

$$0,8 < \frac{i_Y s_X}{i_X s_Y} < 2,0. \quad (3)$$

Az ábra elkészítését megkönnyíti, ha i_X és i_Y értéke 1, 2 vagy 5, illetve ezen számok tizede, százada, tízszerese, százszorosa stb. A legelőnyösebb, ha a (3) képlet szerinti hányados 1-et ad.

Példánkban tervezzük azt, hogy az ábra 15 cm-es vízszintes szakaszán férjenek el a megfigyelt X -értékek. Vagyis $q = 15$ cm. Ez természetesen önkényes mérték választás. Ha fele méretű ábrát tervezünk, akkor pl. $q = 8$ cm-t választanánk. Tehát $q = 15$ cm esetén

$$i_X \approx \frac{6 \cdot 0,7174}{15} = 0,287 \approx \frac{0,2 \text{ humusz } \%}{1 \text{ cm}}$$

az X -tengely beosztása. Az i_X -et lehetőleg lefelé kerekítsük.

$$i_Y \approx \frac{9,993}{0,7174} \cdot 0,2 = 2,79 \approx \frac{5,0 K_A}{1 \text{ cm}}$$

az Y -tengely beosztása. Az i_Y -t lehetőleg felfelé kerekítsük.

Végül is

$$\frac{i_Y s_X}{i_X s_Y} = \frac{5 \cdot 0,7174}{0,2 \cdot 9,993} = 1,795.$$

Ha a számított i_Y -t 5,0 helyett 2-re kerekítenénk, a hányadosunk értéke 0,7179 lenne. Ez kevés. Köztes kerekítés, pl. 3 vagy 4-re, ábrázolástechnikai okok miatt hátrányos.

Az ellipszist határoló téglalap

Első lépésként a korrelációs ellipszist kívülről határoló és a 2. ábra szerinti négy pontban érintő téglalapot szerkesztjük meg.

A téglalap, egyben a korrelációs ellipszis középpontja az X , Y derékszögű koordináta-rendszerben:

$$\text{középpont} = \bar{X}, \bar{Y}. \quad (4)$$

A $P\%$ -szintű tolerancia-ellipszist határoló téglalapnak a középponttól mért

$$\text{félhossza: } h_X = \pm k_{P\%} s_X \quad (5)$$

$$\text{félmagassága: } h_Y = \pm k_{P\%} s_Y, \quad (6)$$

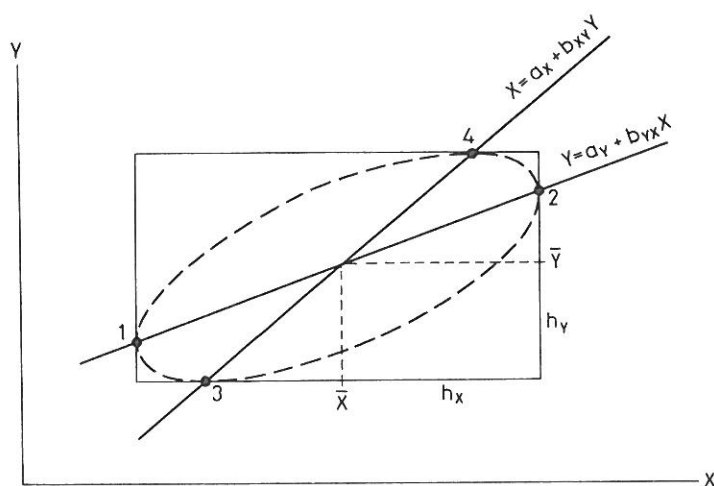
ahol

$$k_{P\%} = \sqrt{\frac{2F_{P\%}(n+1)}{n-2} \left(\frac{n-1}{n}\right)}. \quad (7)$$

Az F -érték szabadságfokainak száma 2 és $n-2$ a számlálóra, illetve a nevezőre.

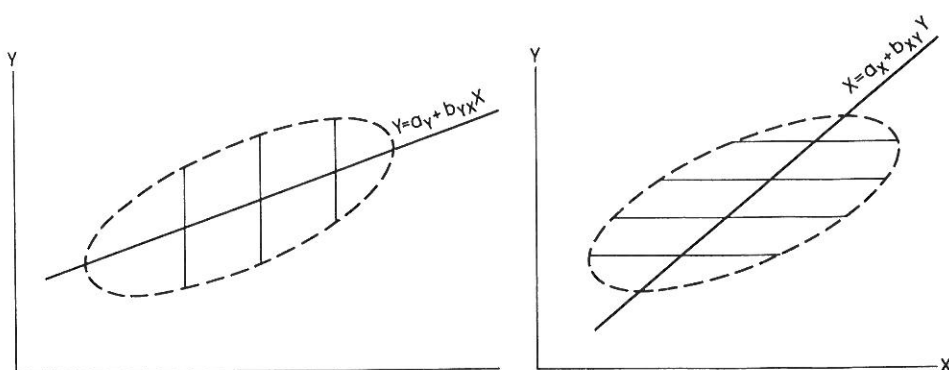
A $k_{p\%}$ szorzó az egydimenziós $P\%$ normális eloszláshoz szignifikancia-vizsgálatokhoz alkalmazott $t_{p\%}$ -kal analóg. Ha $k_{p\%}=1$ -et választunk, akkor egy szórási méretű korrelációs ellipszist kapunk. Az egy szórási tolerancia-ellipszis $n \rightarrow \infty$ esetén az objektumok 39,3%-át fedi le, szemben az egydimenziós egy szórási tolerancia-határok 68,3%-ot fedő intervallumával.

A szemléletesség érdekében $n < 100$ esetén $P = 10\%$ szintű tolerancia-ellipszis szerkesztése javasolható, azaz $F_{10\%}$ -értékeket helyettesítünk a (7) képletbe. Ez az ellipszis a pontok 90%-át fedi. Ha $n < 30$, akkor egy szórási, azaz $k_{p\%} = 1$ -gyel számított tolerancia-ellipszis szerkesztése a szemléletes megoldás. Ha konfidencia-



2. ábra

A határoló téglalap és a korrelációs ellipszis négy érintkezési pontja



3. ábra

A korrelációs ellipszis szimmetriakapcsolata a kétféle regressziós egyenessel

ellipszist szerkesztünk, akkor a (7) képletet még $\sqrt{1/n}$ -nel megszorozzuk, és n -től függetlenül választjuk meg a $P\%$ -szintet, pl. a szokásos 5% -ra.

Példánkban $P = 10\%$ -os szintre

$$k_{10\%} = \sqrt{\frac{2 \cdot 2,44(43+1)}{43-2} \left(\frac{43-1}{43}\right)} = 2,262$$

A $P = 10\%$ -os tolerancia-ellipszist határoló téglalap félhossza

$$h_x = \pm 2,262 \cdot 0,7174 = \pm 1,62,$$

félmagassága

$$h_y = \pm 2,262 \cdot 9,993 = \pm 22,60.$$

A határoló téglalap és a korrelációs ellipszis 2. ábra szerinti négy érintkezési pontjának koordinátái:

az 1. és 2. pont koordinátái

$$\begin{aligned} X_{1,2} &= \bar{X} \pm h_x = 2,58 \pm 1,62 && = 0,96 \text{ és } 4,20 \\ Y_{1,2} &= \bar{Y} \pm b_{YX} h_x = 43,6 \pm 9,136 \cdot 1,62 && = 28,8 \text{ és } 58,4 \end{aligned} \quad (8)$$

a 3. és 4. pont koordinátái

$$\begin{aligned} X_{3,4} &= \bar{X} \pm b_{XY} h_y = 2,58 \pm 0,04709 \cdot 22,60 && = 1,52 \text{ és } 3,64 \\ Y_{3,4} &= \bar{Y} \pm h_y = 43,6 \pm 22,60 && = 21,0 \text{ és } 66,2. \end{aligned} \quad (9)$$

A (8) és a (9) képletben a b_{YX} és a b_{XY} a két változó összefüggésének lineáris regressziós koeficiensei az

$$Y = a_Y + b_{YX} X, \text{ vagy } Y = \bar{Y} + b_{YX}(X - \bar{X}), \quad (10)$$

illetve az

$$X = a_X + b_{XY} Y, \text{ vagy } X = \bar{X} + b_{XY}(Y - \bar{Y}) \quad (11)$$

lineáris regressziós egyenletekből. A két regressziós koeficiens kiszámítható a megadott mutatókból:

$$b_{YX} = r \frac{s_Y}{s_X} = 0,6559 \cdot \frac{9,993}{0,7174} = 9,136 \quad (12)$$

és

$$b_{XY} = r \frac{s_X}{s_Y} = 0,6559 \cdot \frac{0,7174}{9,993} = 0,04709. \quad (13)$$

Az 1. és a 2. pontot a (10) egyenlet, a 3. és a 4. pontot a (11) egyenlet szerinti regressziós egyenes köti össze, ahogy az a 2. ábrán látható. A két regressziós egyenes a határoló téglalap érintkezési pontjaiban metszi a korrelációs ellipszist. A két regressziós egyenlet az ellipszis két szimmetriatengelye olyképpen, hogy az ellipszis kerületének pontjai a (10) egyenlet szerinti egyenesre az Y -tengely irányában

szimmetrikusak, míg a (11) egyenlet szerinti egyenesre az X -tengely irányában szimmetrikusak, ahogy azt a 3. ábra szemlélteti.

A korrelációs ellipszis egyenlete:

$$Y = \bar{Y} + b_{YX}(X - \bar{X}) \pm v\sqrt{h_X^2 - (X - \bar{X})^2}, \quad (14)$$

ahol

$$v = \frac{s_Y}{s_X} \sqrt{1 - r^2} \quad (15)$$

konstans. Példánkban

$$v = \frac{9,993}{0,7174} \sqrt{1 - 0,6559^2} = 10,515.$$

Példánk adataiból a $P = 10\%$ -os tolerancia-ellipszis egyenlete:

$$Y = 43,6 + 9,136(X - 2,58) \pm 10,515\sqrt{1,62^2 - (X - 2,58)^2}.$$

A (14) egyenlet első két tagja a (10) képlet szerinti regressziós egyenlettel azonos. A harmadik, a \pm előjelű tag adja meg az ellipszis szimmetrikus pontjait a regressziós egyeneshez. Ha a (14) és a (15) egyenletben X -t és Y -t megcseréljük, ugyanezt az ellipszist kapjuk, de úgy, hogy a (11) regressziós egyenesre az X -tengely irányában lesznek az ellipszis \pm pontjai szimmetrikusak.

Példánk ellipsziséét a (14) képlet alapján az 1. táblázat szerkezete szerint számítjuk ki. A $P = 10\%$ -os tolerancia-ellipszis pontjait az Y^- - és az Y^+ -oszlop adja meg. Előbbi a regressziós egyenes alatti, utábbi a regressziós egyenes feletti pontokat jelenti. A $P = 10\%$ -os tolerancia-ellipszist és a megfigyelési pontfelhőt példánk adataiból a 4. ábra szemlélteti. Ugyanezen az ábrán mutatjuk be az egy szórásnyi tolerancia-ellipszist is.

Az 1. táblázat első négy sorába a határoló téglalap előzőleg már kiszámított érintkezési pontjainak X és $X - \bar{X}$ koordinátáit írjuk, majd az itteni képletekkel (algoritmussal) újra kiszámítjuk az Y^- - és az Y^+ -koordinátákat. Ezzel ellenőrizzük, hogy számításainkba nem csúszott-e hiba. Ezután a legkisebb és a legnagyobb X érintkezési pontok között könnyen szerkeszthető X -értékeket választunk, melyekhez majd kiszámítjuk az Y^- - és Y^+ -koordinátákat. A legkisebb és a legnagyobb X -értékek mellett erősen sűrűsítjük az X -értékeket, mert ott erős az ellipszis görbülése, azaz közeli X -értékekhez viszonylag távoli Y -értékek tartoznak. Az Y^- -oszlopban egyetlen érték sem lehet kisebb, mint a négy érintkezési pont legkisebb Y^- -értéke, az Y^+ -oszlopban nem lehet nagyobb, mint a négy érintkezési pont legnagyobb Y^+ -értéke.

Az utolsó négy sor kitöltésére később térünk vissza.

Az 1. táblázat fejlécéből és a (14) egyenletből látható, hogy az ellipszis méretét a h_X -érték határozza meg. A négyzetgyök alatti $X - \bar{X}$ abszolút értéke ugyanis nem lehet nagyobb h_X -nél, mert negatív értéket kapnánk a gyökjel alatt. Ha tehát megváltoztatjuk a $P\%$ -szintet, vagy konfidencia-ellipszist számítunk, a képletben csak a h_X változik. Az egy szórásnyi tolerancia-ellipszis esetén $h_X = s_X$.

1. táblázat
Korrelációs ellipszis szerkesztéséhez a kerület pontjainak
X, Y koordinátái a (14) egyenlet alapján*

X	$X - \bar{X}$	$\bar{Y} + b(X - \bar{X})$	$\pm v\sqrt{h_x^2 - (X - \bar{X})^2}$	Y ⁻	Y ⁺
0,96	-1,62	28,80	±0,00	28,8	28,8
4,20	1,62	58,40	±0,00	58,4	58,4
1,52	-1,06	33,92	±12,88	21,0	46,8
3,64	1,06	53,28	±12,88	40,4	66,2
0,98	-1,60	28,98	±2,67	26,3	31,6
1,00	-1,58	29,17	±3,76	25,4	32,9
1,10	-1,48	30,08	±6,93	23,2	37,0
1,20	-1,38	30,99	±8,92	22,1	39,9
1,40	-1,18	32,82	±11,67	21,2	44,5
2,00	-0,58	38,30	±15,91	22,4	54,2
2,58	0,00	43,60	±17,03	26,6	60,6
3,00	0,42	47,44	±16,45	31,0	63,9
3,50	0,92	52,01	±14,02	38,0	66,0
3,80	1,22	54,75	±11,21	43,5	66,0
4,00	1,42	56,57	±8,20	48,4	64,8
4,10	1,52	57,49	±5,89	51,6	63,4
4,15	1,57	57,94	±4,20	53,7	62,1
4,18	1,60	58,20	±2,67	55,5	60,9
0,98	-1,60	28,98	-2,67	26,3	
4,18	1,60	58,20	+2,67		60,9
2,33	-0,25	41,32	+16,83	29,0	
2,83	0,25	45,88	-16,83		58,2

* A példa adatai: $\bar{X}=2,58$; $\bar{Y}=43,6$; $b_{YX}=9,136$; $v=10,515$; $h_x=k_{10\%} \cdot s_x=1,62$; $s_x=0,7174$

A nagytengely hajlásszöge és egyenlete

Az ellipszis nagytengelye átmegy az \bar{X} , \bar{Y} középponton. Szimmetriatengely, melyre az ellipszis \pm pontjai merőlegesen, tehát a kistengellyel párhuzamosan szimmetrikusak. Az ellipszis kistengelye ugyancsak átmegy az \bar{X} , \bar{Y} középponton és merőleges a nagytengelyre.

A nagy- és a kistengely megszerkesztése nem szükséges. De előnyös, mert csak így határozható meg az ellipszis négy „szélső” pontja, a négy tengelyvég koordinátái. Növeli az összefüggés szemléletességét is, ha a nagytengelyt berajzoljuk a korrelációs ellipszisbe. A regressziós egyenest, illetve egyeneseket ne ábrázoljuk, mert azok, bár ugyancsak szimmetriatengelyek, a regressziós modellt szemléltetik, ezért megtévesztők.

A nagytengely iránytangense Y/X mértékegységben kifejezve

$$b_0 = \frac{1}{\gamma} [A + \sqrt{1 + A^2}], \tag{16}$$

ahol
$$\gamma = \frac{i_X}{i_Y}, \quad (17)$$

és
$$A = \frac{\gamma^2 s_Y^2 - s_X^2}{2\gamma r s_X s_Y} \left(= \frac{\gamma^2 S Q_Y - S Q_X}{2\gamma S P} \right). \quad (18)$$

Ha az X és az Y változó mértékegységének ábrázolt hossza (cm-ben) azonos, akkor $i_X = i_Y$, és így $\gamma = 1$. A szakirodalomban — tapasztalatom szerint — csak ezt az esetet szokták ismertetni. A közölt képletekben ugyanis nem szerepel a γ szorzó. Ez ábrázolási nehézségeket okoz, ha a két változó mértékegysége különböző, vagy azonos ugyan a mértékegységük, de szórásuk igen különböző.

Példánkban $i_X = 0,2$ és $i_Y = 5,0$, ezért

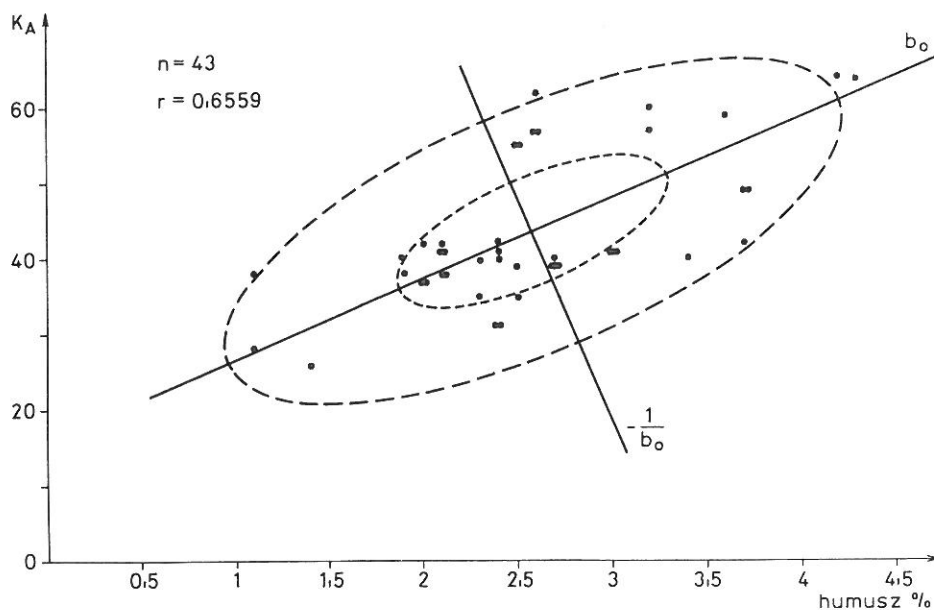
$$\gamma = \frac{0,2}{5,0} = 0,04,$$

továbbá

$$A = \frac{0,04^2 \cdot 9,993^2 - 0,7174^2}{2 \cdot 0,04 \cdot 0,6559 \cdot 0,7174 \cdot 9,993} = \frac{-0,35489}{0,37617} = -0,9434.$$

Így

$$b_0 = \frac{1}{0,04} [-0,9434 + \sqrt{1 + (-0,9434)^2}] = 10,784.$$



4. ábra

A példa adataiból szerkesztett $P = 10\%$ -os és egyszórásnyi tolerancia-ellipszis

A korrelációs ellipszis nagytengelyének egyenlete tehát

$$\begin{aligned} Y &= \bar{Y} + b_0(X - \bar{X}), \\ \text{azaz } Y &= \bar{Y} + 10,784(X - \bar{X}). \end{aligned} \quad (19)$$

Egyébként az Y/X mértékegységben kifejezett b_0 -iránytangens értéke mindig a két regressziós koeficiens között fekszik, és pedig

$$b_{YX} < b_0 < 1/b_{XY}. \quad (20)$$

Az ellipszis nagytengelyének geometriai ábra szerinti (cm/cm) iránytangense

$$\operatorname{tg} \varphi = \gamma b_0 = 0,04 \cdot 10,784 = 0,4314, \quad (21)$$

amiből példánkban

$$\varphi = 23,33^\circ.$$

Az ideális szerkesztés az lenne, ha $\varphi = 45^\circ$. Ekkor a határoló téglalap éppen négyzet alakú, és az ellipszis nagytengelye azonos a határoló négyzet átlójával. Ha φ nem 45° , akkor a nagytengely és a határoló téglala átlója nem esik egybe. Egyébként a határoló téglala átlójának az iránytangensét a szakirodalom b_1 -gyel szokta jelölni,

$$b_1 = \frac{s_Y}{s_X} = \frac{h_Y}{h_X}. \quad (22)$$

Az átló ugyancsak szimmetriatengelye a korrelációs ellipszisnek, oly módon, ahogy azt az 5/E. ábra szemlélteti, ahol forgatás útján szerkesztettük meg az ellipszist.

A korrelációs ellipszis kistengelyének geometriai hajlásszöge $\varphi - 90^\circ$, példánkban $-66,67^\circ$, Y/X -ben kifejezett hajlásszöge $-\frac{1}{b_0}$, példánkban $-0,0927$. A kétféleképpen kifejezett hajlásszög kapcsolata:

$$\operatorname{tg}(\varphi - 90^\circ) = -\frac{1}{\gamma b_0}. \quad (23)$$

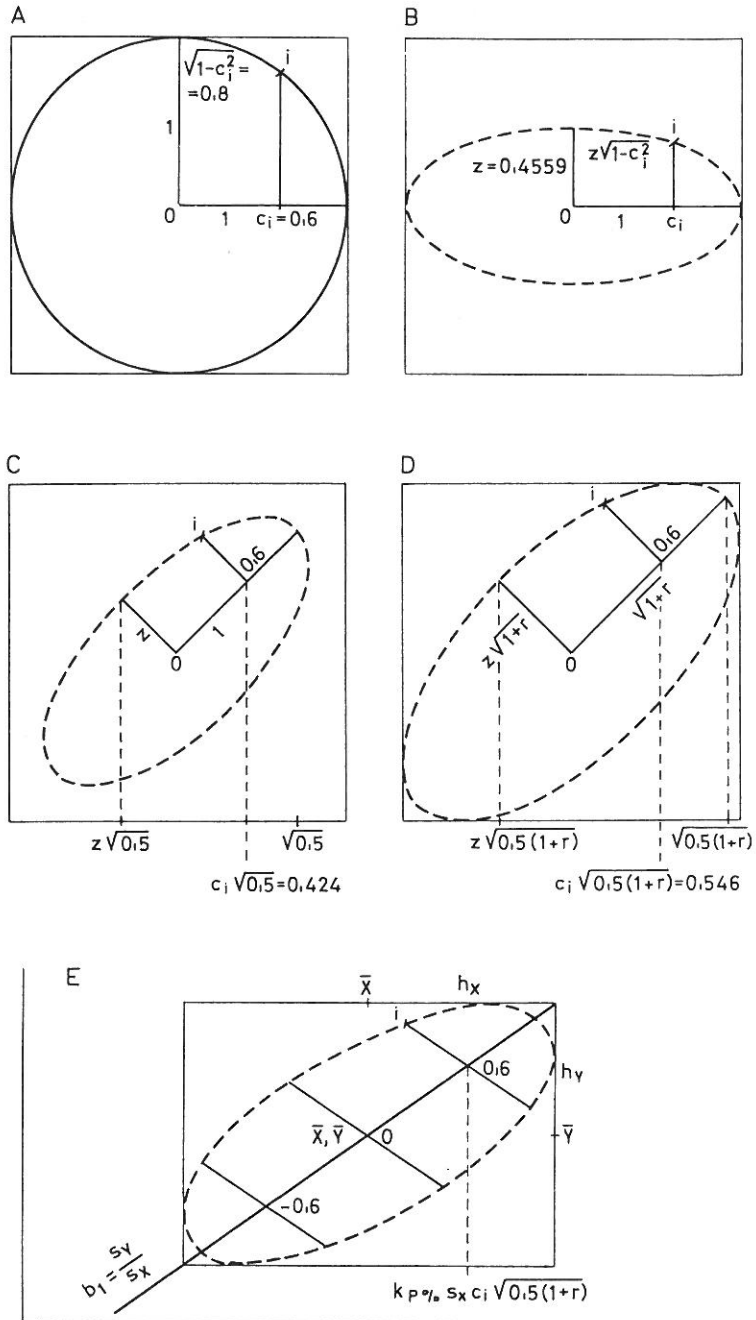
A nagy- és a kistengely végeinek koordinátái

A nagytengely végeinek koordinátái:

$$\begin{aligned} X &= \bar{X} \pm t_1 \cos \varphi \\ Y &= \bar{Y} \pm \frac{1}{\gamma} t_1 \sin \varphi. \end{aligned} \quad (24)$$

A kistengely végeinek koordinátái:

$$\begin{aligned} X &= \bar{X} \pm t_2 \sin \varphi \\ Y &= \bar{Y} \pm \frac{1}{\gamma} t_2 \cos \varphi, \end{aligned} \quad (25)$$



5. ábra

Korrelációs ellipszis szerkesztése forgatással. A körkerület i -vel definiált pontjának helye az ellipszisen. A. A kör; B. Lapítás; C. Forgatás; D. Tágítás; E. Átskálázás

ahol t_1 a nagytengely, t_2 a kistengely hossza az X -változó mértékegységében kifejezve.

$$t_1, t_2 = k_{p\%} \sqrt{0,5[B \pm \sqrt{B^2 - 4s_X^2 \gamma^2 s_Y^2 (1-r)^2}]}, \quad (26)$$

ahol

$$B = s_X^2 + \gamma^2 s_Y^2. \quad (27)$$

A példában

$$\begin{aligned} B &= 0,7174^2 + 0,04^2 \cdot 9,993^2 = 0,6744 \\ t_1, t_2 &= 2,262 \sqrt{0,5[0,6744 \pm \sqrt{0,6744^2 - 4 \cdot 0,7174^2 \cdot 0,04^2 \cdot 9,993^2 (1 - 0,6559)^2}]} \\ &= 2,262 \sqrt{0,5[0,6744 \pm 0,5171]}, \end{aligned}$$

amiből

$$t_1 = 1,746 \text{ és } t_2 = 0,6344.$$

A nagytengely végeinek koordinátái a (23) képlettel:

$$X = 2,58 \pm 1,746 \cdot \cos 23,33^\circ = 0,98 \text{ és } 4,18$$

$$Y = 43,6 \pm \frac{1}{0,04} \cdot 1,746 \cdot \sin 23,33^\circ = 26,3 \text{ és } 60,9.$$

A kistengely végeinek koordinátái a (24) képlettel:

$$X = 2,58 \pm 0,6344 \cdot \sin 23,33^\circ = 2,33 \text{ és } 2,83$$

$$Y = 43,6 \mp \frac{1}{0,04} \cdot 0,6344 \cdot \cos 23,33^\circ = 58,2 \text{ és } 29,0.$$

A pontosabb ábrázolás érdekében e négy pont koordinátáit célszerű az 1. táblázatba, utolsó négy sorként beírni.

Szerkesztés forgatással

A korrelációs ellipszis előállítható az egységnyi sugarú kör ellipszissé lapításával, forgatásával, nyújtásával, végül átskálázásával. Az egyes lépéseket az 5. ábra szemlélteti.

a) Egységnyi sugarú kört és határoló négyzetet szerkesztünk. Jelöljük i -vel a kör bármely pontját, a vízszintes tengelyt pedig c -vel. Ekkor az i -pont koordinátái az abszcisszán c_i , az ordinátán pedig $\pm \sqrt{1 - c_i^2}$.

b) A kört ellipszissé lapítjuk oly módon, hogy a kerület $\pm \sqrt{1 - c^2}$ ordinátáit megszorozzuk a korrelációs koefficiensből számított

$$z = \sqrt{\frac{1-r}{1+r}} \left(= \sqrt{\frac{1-0,6559}{1+0,6559}} = 0,4559 \right) \quad (28)$$

lapítottsági koefficienssel.

c) Az ellipszist az óramutató járásával ellentétesen 45° -kal elforgatjuk; innen a $\sqrt{0,5}$ -ös szorzó a (31) képletben;

d) A kerület koordinátáit $\sqrt{(1+r)}$ -rel megszorozzuk, hogy az ellipszis érintse a határoló négyzetet.

e) A tengelyeket átskálázzuk X , Y beosztásra. Ezért átszorozzuk az ellipszis kerületének koordinátáit X és Y szórásával és a $k_{P\%}$ szorzóval.

A vázolt lépések a számítások során részben összevonhatók. Az egységnyi sugarú kör kerületének koordinátái az X -, Y -tengelyű tolerancia-ellipszis kerületén:

$$X = \bar{X} \pm s_X g [c - (\pm z \sqrt{1 - c^2})] \quad (29)$$

$$Y = \bar{Y} \pm s_Y g [c + (\pm z \sqrt{1 - c^2})] \quad (30)$$

ahol

$$g = k_{P\%} \sqrt{0,5(1+r)}. \quad (31)$$

E két egyenletből levezethető a (14) egyenlet, ugyanis a (29) és (30) egyenletek összegéből

$$c = \frac{1}{k_{P\%} \sqrt{2(1+r)}} \left[\frac{X - \bar{X}}{s_X} + \frac{Y - \bar{Y}}{s_Y} \right]. \quad (32)$$

A (29) egyenletben c -t a (32) egyenlet jobb oldalával helyettesítve és Y -t X -szel kifejezve a (14) egyenlethez jutunk.

Előjel-szabályok: a (29) és a (30) egyenletben kétszer fordul elő \pm előjel. A két képlet alkalmazását megkönnyíti és a számításokat leegyszerűsíti a következő előjel-szabályok figyelembevétele:

1. Ha \bar{X} és \bar{Y} után a „+” előjellel számolunk, akkor az ellipszis melléktengelyétől jobbra eső koordinátákat kapjuk meg. Ha a „-” előjellel számolunk, akkor a bal oldali azonos abszolút értékű, de mínusz előjelű c -értékek X -, Y -koordinátáit kapjuk. Ezért mínusz előjelű c -értékkel felesleges számolni.

2. Ha a $z\sqrt{1-c^2}$ kifejezés előjele „+”, akkor az ellipszis felső, ha előjele „-”, az ellipszis alsó felét kapjuk. Így tehát a (29) és a (30) képletek előjelei közül a

- + + az ellipszis jobb felső negyedét = I
- - az ellipszis bal alsó negyedét = III
- + - az ellipszis jobb alsó negyedét = IV
- + az ellipszis bal felső negyedét = II

határozzák meg.

Példaképpen számítsuk ki példánk adataiból a kör kerületének $i=0,6$; $0,8$ pontjának koordinátáit az X -, Y -tengelyű, $P=10\%$ tolerancia-ellipszis kerületén:

$$\begin{aligned} g &= 2,262 \sqrt{0,5(1+0,6559)} = 2,0582 \\ X &= 2,58 + 0,7174 \cdot 2,0582 [0,6 - 0,4559 \sqrt{1-0,6^2}] \\ &= 2,58 + 1,4766 [0,2353] = 2,93 \\ Y &= 43,6 + 9,993 \cdot 2,0582 [0,6 + 0,4559 \sqrt{1-0,6^2}] \\ &= 43,6 + 20,568 [0,9647] = 63,4. \end{aligned}$$

Ellenőrzésként az $X = 2,93$ -at helyettesítsük a (14) egyenletbe:

$$Y = 43,6 + 9,136(2,93 - 2,58) + 10,515\sqrt{1,62^2 - (2,93 - 2,58)^2} = 63,45.$$

A 2. táblázat a forgatásos szerkesztése eljárás számításainak célszerű szerkezetét, technikáját mutatja. Itt számszerűen is követhetők a koordináták előjel szerinti szimmetriaviszonyai. Látható, hogy pozitív előjelű c koordinátákra kell csak számítást végezni, a bal oldali koordinátákat az $X - \bar{X}$ és az $Y - \bar{Y}$ előjelének megváltoztatásával kapjuk meg.

2. táblázat
Korrelációs ellipszis kerületének koordinátái forgatással*

c (1)	$\sqrt{1-c^2}$ (2)	$z(2)$ (3)	$X - \bar{X}$ $s_x g(1-3)$ (4)	$Y - \bar{Y}$ $s_y g(1+3)$ (5)	$X_{I,III}$ $\bar{X} + (4)$	$Y_{I,III}$ $\bar{Y} + (5)$	$X_{II,IV}$ $\bar{X} - (4)$	$Y_{II,IV}$ $\bar{Y} - (5)$
1,0		0,0	1,48	20,6	4,06	64,2	1,10	23,0
0,96	0,280	0,128	1,24	22,3	3,82	65,9	1,34	21,3
	-0,280	-0,128	1,61	17,1	4,19	60,7	0,97	26,5
0,90	0,436	0,199	1,04	22,6	3,62	66,2	1,54	21,0
	-0,436	-0,199	1,62	14,4	4,20	58,0	0,96	29,2
0,80	0,600	0,274	0,78	22,1	3,36	65,7	1,80	21,5
	-0,600	-0,274	1,59	10,8	4,17	54,4	0,99	32,8
0,60	0,800	0,365	0,35	19,8	2,93	63,4	2,23	23,8
	-0,800	-0,365	1,42	4,8	4,00	48,4	1,16	38,8
0,30	0,954	0,435	-0,20	15,1	2,38	58,7	2,78	28,5
	-0,954	-0,435	1,09	-2,8	3,67	40,8	1,49	46,4
0,0	1,0	0,456	-0,67	9,4	1,91	53,0	3,25	34,2
	-1,0	-0,456	0,67	-9,4	3,25	34,2	1,91	53,0

* A példa adatai: $\bar{X} = 2,58$; $\bar{Y} = 43,6$; $s_x = 0,7174$; $s_y = 9,993$; $r = 0,6559$; $k_{10\%} = 2,262$; $z = 0,4559$; $g = 2,0582$

A pontos szerkesztés érdekében a $c = 0,0$; $0,3$; $0,6$; $0,8$; $0,9$; $0,96$ és $1,0$ -hez tartozó X -, Y -koordinátákat érdemes kiszámítani. Ez 13 sort jelent a 2. táblázatban és 28 pontot határozunk meg az ellipszis kerületén.

Az 5/E. ábrán a $c = -0,6$, $0,0$ és $+0,6$ abszcissza értékekhez tartozó X , Y koordinátaértékeket ábrázoltuk. Az ábrából látható, hogy az eredeti kör abszcisszájának az elforgatása a

$$b_1 = s_y/s_x = h_y/h_x \tag{33}$$

hajlásszögű tengelyt adja, mely a határoló téglalap átlója. Az ellipszis kerületének koordinátái erre a tengelyre is szimmetrikusak, ahogy az 5/E ábra szemlélteti, de nem merőlegesen. Az ellipszis nagytengelye és az átlója csak akkor esik egybe, ha $\varphi = 45^\circ$. Ekkor $b_0 = b_1$.

Ilyenformán a korrelációs ellipszisnek tehát négy szimmetriatengelye van, b_{YX} , b_{XY} , b_0 és b_1 hajlásszögű. Ezek közül a b_0 hajlásszögű adja a nagytengelyt. Az ellipszis nagytengelyét a forgatásos számolási eljárásnál is (16)–(27) képletekkel határozzuk meg.

Az 1. vagy a 2. táblázat szerint számoljunk?

A 2. táblázat szerinti számolás gyorsabb, és pontosabban meghatározható, hogy hol sűrítjük az X , Y pontokat. Ezenkívül a 2. táblázat első két oszlopa sablonosan elő is készíthető. Hátránya azonban, hogy ábrázolásakor az X - és Y -értékek nem kerek számok, így az ábra készítése nagyobb körültekintést igényel, mint az 1. táblázat szerinti módszerrel, ahol az X -értékeket, minthogy egész számok, előre berajzolt vonalakhoz illesztjük, pl. milliméterpapír használatakor.

Egyelőre nincs elég tapasztalat, hogy melyik az előnyösebb eljárás. A kevés fellelhető szakirodalom az 1. táblázat szerinti megoldást ismerteti. A magam részéről is úgy látom, hogy ha a kutató szerkeszti az ellipszist, egyszerűbb és pontosabb az 1. táblázat szerinti módszer. Számítógépes programozáshoz azonban a 2. táblázat szerinti forgatóasztalos megoldás talán előnyösebb, mert itt előre rögzíthető, hogy a terület mely pontjait határozzuk meg. Az 1. táblázat szerinti megoldásnál ugyanis a h_x méretétől függ, hogy mely X -értékeket válasszuk meg előre.

Összefoglalás

Két véletlen változó, tulajdonság korrelációjának ábrázolásakor derékszögű koordinátarendszerben a korrelációs ellipszis ábrázolása nagyon megkönnyíti az összefüggést kifejező pontfelhő megítélését, a szélsőséges kombinációk, vagy a pontsűrűsödések felismerését. Módszert mutatunk be a X - és Y -tengely skálabeosztásának célszerű megválasztására, bemutatunk továbbá két módszert a korrelációs ellipszis kerületének és tengelyeinek megszerkesztésére. Az egyik módszer a lineáris regresszió-analízisen, a másik módszer az egységnyi sugarú kör ellipszissé lapításán, elforgatásán és tengelyeinek átskálázásán alapszik. A két módszer ugyanazt az ellipszist adja. Kézzel történő ábrázoláshoz az első módszer előnyösebb. A szerkesztés részleteit, a négy szimmetriatengely (b_0 , b_1 , b_{yx} , b_{xy}) értelmezését ábrák szemléltetik.

Bemutatunk egy γ -val jelölt szorzófaktort, amelyet akkor kell használni, ha az X - és az Y -tengely mértékegységének az ábrázolt hossza különbözik. A dolgozat bevezetésében megtárgyaljuk a korrelációs ellipszis két változatának, a tolerancia- és a konfidencia-ellipszisnek értelmezését és alkalmazási körét az agrokémiai kutatás területén.

Irodalom

- [1] Documenta Geigy: Wissenschaftliche Tabellen. 183—184. G. Thieme Verl. Stuttgart. 1975.
- [2] HEALY, M. J. R.: Drawing a probability ellipse. 202—204. Applied Statistics 21. 1972.
- [3] JUVAN CZ, I. & PAKSY A.: Orvosbiometria. Medicina Könyvkiadó, Budapest. 1982.
- [4] SCHNEIDER, B.: Structural relationship analysis of observational data in clinical chemistry. Proceeding 3rd Hung. Biometric Conference, Budapest, 13—15 April 1981.
- [5] THÖNI, H.: Darstellung eines Toleranz und Vertrauensellipsoids. 57—61. EDV.12.2. 1981.

Érkezett: 1983. április 12.

Applying and Constructing Correlation Ellipses

J. SVÁB

Agricultural Research Institute of the Hungarian Academy of Sciences, Martonvásár (Hungary)

Summary

When representing the correlation between two random variables in an orthogonal coordinate system, plotting a correlation ellipse may make it much easier to interpret the scatter of dots, to find out extreme combinations or the clustering of dots.

The well-known formula of correlation ellipses [1] is

$$Y = \bar{Y} + b_{YX}(X - \bar{X}) \pm v \sqrt{h_x^2 - (X - \bar{X})^2},$$

where

$$v = \frac{s_y}{s_x} \sqrt{1 - r^2} \quad \text{and} \quad h_x = s_x \sqrt{\frac{2F_{P\%}(n+1)}{n-1} \left(\frac{n-1}{n} \right)}$$

(see Fig. 2).

The construction of the axes of the ellipse based on the known formulae supposes that the unit of measurement of variables X and Y should be plotted on the graph in equal unit of length, regardless of their actual dimension. In many cases this restraint results in distorted graphs, depending on the unit of measurement of the two variables, or the difference between their standard deviations.

To overcome this difficulty a coefficient $\gamma = i_x/i_y$ has been introduced, where i_x means the region of variable X on 1 cm length of the X axis, and i_y means the region of variable Y on 1 cm length of the Y axis. The formulae completed by a coefficient γ as described in literature are shown in this paper as formulae No. (16)–(27).

The paper discusses also an other solution of the construction of correlation ellipses, where one is starting from a unit circle: corresponding to the value of the correlation coefficients (r) the circle is flattened down to ellipse shape, then rotated and enlarged to the wanted size, constructing finally the ellipse with arbitrary scaling of the X and Y axes. The whole procedure is illustrated in Fig. 5. Actual calculations, determination of the dots of the ellipse are done according to formulae (28)–(31) and the scheme shown in Table 2.

The terms of tolerance and confidence belt and ellipse are discussed, too (Fig. 1), the symmetry condition of the correlation ellipse are outlined in Figs. 2, 3 and 5. Fig. 4 shows the tolerance ellipse of the example used in the paper, at the $P=0.1$ level of significance and $s=1$ (one standard deviation) scale.

Table 1. X, Y coordinates of the dots of the ellipse for the construction of a correlation ellipse on the basis of equation (14). *Statistics of the example used in the paper.

Table 2. Coordinates of the perimeter of a correlation ellipse by rotation. *Statistics of the example used in the paper.

Fig. 1. Regression (A) and correlation (B). a) tolerance belt; b) confidence belt; c) tolerance ellipse; d) confidence ellipse; e) principal axis.

Fig. 2. Four points of contact of the surrounding rectangle and the correlation ellipse.

Fig. 3. The symmetry conditions of the correlation ellipse to the two kinds of regression line.

Fig. 4. Tolerance ellipse of the example used in the paper, at the $P=0.1$ level of significance and $s=1$ (one standard deviation) scale.

Fig. 5. Construction of a correlation ellipse by rotation. The place of point "r" of the circle's perimeter on the ellipse. A. Unit circle; B. Flattening; C. Rotation; D. Enlargement; E. Rescaling.

Anwendung und Konstruktion von Korrelationsellipsen

J. SVÁB

Landwirtschaftliches Forschungsinstitut der Ungarischen Akademie der Wissenschaften, Martonvásár (Ungarn)

Zusammenfassung

Soll die Korrelation zwischen zwei Zufallsvariablen in einem rechtwinkligen Koordinatensystem dargestellt werden, so kann die Korrelationsellipse die Deutung des Punkteschwarms, das Erkennen extremer Kombinationen oder die Clusterbildung der Punkte erheblich erleichtern.

Die bekannte Formel der Korrelationsellipse ist [1]

$$Y = \bar{Y} + b_{YX}(X - \bar{X}) \pm v \sqrt{h_X^2 - (X - \bar{X})^2},$$

wo

$$v = \frac{s_Y}{s_X} \sqrt{1 - r^2} \quad \text{und} \quad h_X = s_X \sqrt{\frac{F_{P\%}(n+1)}{n-2} \left(\frac{n-1}{n} \right)}$$

ist (s. Abb. 2).

Für die Konstruktion der Ellipsenachsen nach den bekannten Formeln ist vorgegeben, dass die Masseinheit der Variablen X und Y im Diagramm mit gleichen Längenmass dargestellt wird. Diese Bedingung führt aber oft zu verzerrten Abbildungen je nach dem, in welchen Masseinheiten die zwei Variablen ausgedrückt sind, oder wie stark ihre Standardabweichungen voneinander verschieden sind.

Zur Überbrückung dieses Problems wurde ein Koeffizient $\gamma = i_X/i_Y$ eingeführt, wo i_X das Bereich der Variablen X auf 1 cm Längenabschnitt der X -Achse, i_Y das Bereich der Variablen Y auf 1 cm Längenabschnitt der Y -Achse ausdrückt. Die aus der Fachliteratur übernommenen, mit dem Koeffizienten γ ergänzten Formeln sind hier als Formeln (16)–(27) angeführt.

Für die Konstruktion einer Korrelationsellipse wird in der vorliegenden Arbeit auch eine andere Lösung beschrieben, bei der von einem Einheitskreis ausgegangen wird, der dann auf Grund der r -Werte (Korrelationskoeffizienten) zu einer Ellipse abgeflacht, verdreht, auf das erwünschte Mass vergrößert und schliesslich in einem Koordinatensystem mit beliebig erwählter Masseinteilung der X - und Y -Achse die gewünschte Ellipse aufgezeichnet wird. Dieser Vorgang ist in Abb. 5. veranschaulicht. Die Rechenarbeit, die Bestimmung der Ellipsenpunkte erfolgt den Formeln (28)–(31) gemäss und nach dem in Tab. 2. gezeigten Schema.

Die Begriffe des Toleranz- und Konfidenzgürtels, bzw. der Ellipse (Abb. 1.), sowie die Symmetrieverhältnisse der Korrelationsellipse werden auch erörtert und in Abb. 2., 3. und 5. graphisch dargestellt. Abb. 4. veranschaulicht die Toleranzellipse des besprochenen Beispiels bei $P = 10\%$ und $s = 1$, d. h. im Grössenmass von einer Einheit der Standardabweichung.

Tab. 1. X -, sowie Y -Koordinaten der zur Konstruierung der Korrelationsellipse benötigten Punkte auf Grund der Gleichung (14). *Masszahlen des angeführten Beispiels.

Tab. 2. Koordinaten des Perimeters der Korrelationsellipse nach Anwendung einer Verdrehung. *Masszahlen des angeführten Beispiels.

Abb. 1. Regression (A) und Korrelation (B). a) Toleranzgürtel; b) Konfidenzgürtel; c) Toleranzellipse; d) Konfidenzellipse; e) Hauptachse.

Abb. 2. Die vier Berührungspunkte der Korrelationsellipse und ihres Tangentenrechtecks.

Abb. 3. Symmetrieverhältnis der Korrelationsellipse mit den zwei, auf verschiedene Weise berechneten Regressionsgeraden.

Abb. 4. Die mit Hilfe der Masszahlen des besprochenen Beispiels konstruierte Toleranzellipse bei $P = 10\%$ und $s = 1$.

Abb. 5. Konstruktion der Korrelationsellipse mit Hilfe von Verdrehen. Lage des Punktes i des Einheitskreises auf dem Perimeter der Ellipse. A. Der Kreis; B. Abflachen; C. Verdrehen; D. Vergrössern; E. Erwählen der neuen Masseinteilung der X - und Y -Achse.

Построение и использование эллипса корреляции

Я. ШВАБ

Научно-исследовательский институт сельского хозяйства, Венгерской Академии Наук, Мартовашар (Венгрия)

Резюме

При изображении корреляции двух случайных величин в ортогональной системе координат, построение эллипса корреляции в значительной мере облегчает оценку рассеивания точек, различение выдающихся комбинаций, определение плотности распределения точек, точечных кластеров.

Для эллипса корреляции известно уравнение [1]

$$Y = Y + b_{YX}(X - \bar{X}) \pm v \sqrt{h_X^2 - (X - \bar{X})^2},$$

где

$$v = \frac{s_Y}{s_X} \sqrt{1 - r^2} \quad \text{и} \quad h_X = s_X \sqrt{\frac{2F_{P\%}(n+1)}{n-2} \left(\frac{n-1}{n} \right)}$$

(смотри рисунок 2.).

Построение осей эллипса по указанным уравнениям предполагает, что единицы длины переменных X и Y какими бы они не были, на рисунке должны иметь одинаковую длину. Это условие часто ведет к искажению рисунка зависимо от того, каковы единицы измерения двух переменных величин или от степени их рассеивания.

Для преодоления этой трудности ввели коэффициент $\gamma = i_X/i_Y$, где i_X область переменной X в 1 см оси X , i_Y область переменной Y в 1 см оси Y .

Встречающиеся в специальной литературе уравнения, дополненные коэффициентом γ показывают уравнения, приведенные в данной работе ([16]—[27]).

В работе приведена и такая возможность построения эллипса корреляции, когда оно начинается из окружности определенного радиуса, затем окружность формируют в эллипс согласно коэффициентам корреляции (r), поворачивают, увеличивают до требуемых размеров, затем строят эллипс по свободно выбранному подразделению осей X и Y . Все это показано на рисунке 5. Сам расчет, определение точек эллипса ведется по уравнениям [28]—[31] с помощью схематической таблицы 2.

Работа обсуждает также понятия полоса толерантности, полоса доверия и эллипс толерантности и доверия. (Рис. 1), а также условия симметрии эллипса корреляции. Это показано на рисунках 2., 3. и 5. На рисунке 4. приведен пример эллипса толерантности $P = 10\%$ и $s = 1$, т. е. в одном стандартном размере.

Табл. 1. Координаты X , Y точек окружности для построения эллипса корреляции, на основании уравнения (14). *Статистические показатели.

Табл. 2. Координаты периметра эллипса корреляции при повороте.
*Статистические показатели.

Рис. 1. Регрессия (А) и корреляция (В). а) полоса толерантности. б) полоса доверия. с) эллипс толерантности. d) эллипс доверия. е) главная ось.

Рис. 2. Ограничивающий прямоугольник и четыре точки соприкосновения эллипса корреляции.

Рис. 3. Симметрическая связь эллипса корреляции с двумя видами регрессионных прямых.

Рис. 4. Пример построения эллипса. $P = 10\%$, $s = 1$.

Рис. 5. Построение эллипса корреляции поворотом. Место точки окружности обозначенной i в эллипсе. А. Окружность. В. Доведение окружности до формы эллипса. С. Поворот. Д. Увеличение размеров. Е. Переградуировка.