10. Lineáris regresszió

Valószínűségi változók kovarianciáját eddig csak diszkrét esetben vizsgáltuk, annak ellenére, hogy ugyanaz a definíció alkalmas folytonos valószínűségi változók kovarianciájának definiálására is. Amiért ezt a témát mégis eddig halogattuk, az az együttes sűrűségfüggvény fogalmának hiánya volt, amely fogalom lehetővé teszi a kovariancia kiszámolását folytonos esetben is.

A kovariancia és szórás fogalmak alkalmazásaként a lineáris regressziót is itt tárgyaljuk. Lineáris regresszió alatt elsősorban egy statisztikai modellt értünk, ami a változók közötti lineáris kapcsolatra alapozva vezet le összefüggéseket a változók viselkedésére. A modellt használják prediktív, illetve magyarázó célzattal is. Az előbbi alkalmazás a becsléselmélet, míg utóbbi a hipotézisvizsgálat témaköréhez sorolható, amik a statisztika részterületei.

Ugyanakkor a lineáris regressziónak van egy tisztán valószínűségszámítási vonatkozása is, amihez nincs szükség a statisztika alapfogalmaira, mint a minta vagy a becslés. Ez annak a kérdésnek a környékre, hogy hogyan lehet adott X és Y valószínűségi változók esetén olyan α, β számokat választani, hogy $\beta X + \alpha$ a lehető legközelebb legyen Y-hoz.

10.1. Szórás és kovariancia folytonos esetben

Legyen X folytonos valószínűségi változó, és jelölje f_X a sűrűségfüggvényét. Hogy tudjuk meghatározni X szórását?

Korábban vizsgáltuk már az X várható értékét, sőt $g(X)$ transzformált várható értékét is, ahol $g : \mathbb{R} \to \mathbb{R}$ tetszőleges folytonos függvény. Emiatt az X szórásnégyzetét is ki tudjuk számolni (ahogy azt a normális eloszlás esetében már számoltuk is):

$$\mathbb{D}^2(X) \stackrel{\text{def}}{=} \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x - \Big(\int_{.\infty}^{\infty} x f_X(x)\mathrm{d}x\Big)^2.$$

Ebből pedig X szórása $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$.

A szórásnégyzet (illetve szórás) jelentése ilyen esetben is átlagtól való négyzetes eltérés (és annak gyöke). Szemléletesen azt méri, mennyire „terül szét” a sűrűségfüggvény a várható érték körül.[46]

10.1.1. Példa. Legyen $Z \sim \mathrm{Exp}(\lambda)$ valamilyen λ pozitív valósra. Ekkor két parciális integrálással adódik, hogy

$$\mathbb{E}(Z^2) = \int_0^{\infty} z^2 \lambda e^{-\lambda z}\mathrm{d}z = \Big[-e^{-\lambda z} z^2\Big]_0^{\infty} - \int_0^{\infty} -e^{-\lambda z} 2z\mathrm{d}z = \int_0^{\infty} 2z e^{-\lambda z}\mathrm{d}z =$$
$$= \Big[2z\Big(-\frac{1}{\lambda}\Big)e^{-\lambda z}\Big]_0^{\infty} - \int_0^{\infty} 2\Big(-\frac{1}{\lambda}\Big)e^{-\lambda z}\mathrm{d}z = \frac{2}{\lambda}\int_0^{\infty} e^{-\lambda z}\mathrm{d}z = \frac{2}{\lambda^2}.$$

Tehát

$$\mathbb{D}^2(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \frac{2}{\lambda^2} - \Big(\frac{1}{\lambda}\Big)^2 = \frac{1}{\lambda^2} \qquad \Longrightarrow \qquad \mathbb{D}(Z) = \frac{1}{\lambda}.$$

Ezen gondolatmeneten továbbhaladva észrevehetjük, hogy folytonos valószínűségi változók kovarianciája is értelmes a kovariancia eredeti definíciójával, feltéve, hogy az ott szereplő várható értékek léteznek. Sőt, az alábbi állítás is érvényben marad:

$$\mathrm{cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Konkrét esetben számolási nehézséget tipikusan az $\mathbb{E}(XY)$ tag jelent, hiszen az XY valószínűségi változó eloszlása az (X, Y) valószínűségi vektorváltozó együttes eloszlásától függ, nem csak X és Y peremeloszlásaitól. A következő állítás segítségével XY eloszlásának kiszámolása nélkül is meghatározható $\mathbb{E}(XY)$.

[46] A sűrűségfüggvény alakjáról számos további származtatott mennyiség nyilatkozik, mint a valószínűségi változó átlagos abszolút eltérése, a csúcsossága (más néven lapultsága), vagy a ferdesége.

10.1.2. Állítás. *Legyen $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g : \mathbb{R}^n \to \mathbb{R}$ olyan függvény, amire $\mathbb{E}(g(X_1, \dots, X_n))$ létezik. Ekkor*

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) \cdot f_{\underline{X}}(x_1, \dots x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

Ha g folytonos és nemnegatív, akkor $\mathbb{E}(g(X_1, \dots, X_n))$ létezik, beleértve, hogy értéke esetleg $+\infty$.

Az állításnak speciális esete, hogy ha (X, Y) folytonos valószínűségi vektorváltozó, akkor

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y,$$

feltéve, hogy a várható érték létezik (ugyebár a $g : (x, y) \mapsto x \cdot y$ függvény nem nemnegatív).

10.1.3. Példa. Jelölje X az éves összes csapadékmennyiséget (1000 mm-ben számolva), Y pedig az évben eladott esernyők számát (1000 db-ban számolva). Tegyük fel, hogy az együttes sűrűségfüggvényük:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{5}(4 - 2x^2 + xy - y^2) & \text{ha } 0 < x < 1 \text{ és } 0 < y < 2, \\ 0 & \text{egyébként.} \end{cases}$$

Számoljuk ki X és Y kovarianciáját. Az előző állítás szerint

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y = \int_0^2 \int_0^1 xy \cdot \frac{1}{5}(4 - 2x^2 + xy - y^2) \mathrm{d}x \mathrm{d}y =$$

$$= \frac{1}{5} \int_0^2 \int_0^1 (4xy - 2x^3 y + x^2 y^2 - xy^3) \mathrm{d}x \mathrm{d}y = \frac{1}{5} \int_0^2 \left[2x^2 y - \frac{1}{2} x^4 y + \frac{1}{3} x^3 y^2 - \frac{1}{2} x^2 y^3 \right]_{x=0}^1 \mathrm{d}y =$$

$$= \frac{1}{5} \int_0^2 \Big(\frac{3}{2} y + \frac{1}{3} y^2 - \frac{1}{2} y^3 \Big) \mathrm{d}y = \frac{1}{5} \Big[\frac{3}{4} y^2 + \frac{1}{9} y^3 - \frac{1}{8} y^4 \Big]_0^2 = \frac{1}{5} \Big(3 + \frac{8}{9} - 2 \Big) = \frac{17}{45}.$$

A kovarianciához szükségünk van még a várható értékekre. Annyi csak a probléma, hogy ehhez az f_X sűrűségfüggvény még nem áll rendelkezésünkre. Szerencsére azt tudjuk, hogy a peremeloszlás sűrűségfüggvénye hogyan számolható az együttes sűrűségfüggvényből:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) \mathrm{d}x = \int_{-\infty}^{\infty} x \cdot \int_{-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}y \mathrm{d}x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y) \mathrm{d}y \mathrm{d}x.$$

Ezen a ponton észre is vehetjük, hogy $\mathbb{E}(X)$ igazából a $g(x, y) = x$ függvény szerinti transzformált várható értéke, így hamarabb eljutunk ugyanehhez a formulához. Némi integrálással kapjuk, hogy

$$\mathbb{E}(X) = \int_0^2 \int_0^1 x \cdot \frac{1}{5}(4 - 2x^2 + xy - y^2) \mathrm{d}x \mathrm{d}y = \frac{7}{15}$$

$$\mathbb{E}(Y) = \int_0^2 \int_0^1 y \cdot \frac{1}{5}(4 - 2x^2 + xy - y^2) \mathrm{d}x \mathrm{d}y = \frac{4}{5}$$

$$\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{17}{45} - \frac{7}{15} \cdot \frac{4}{5} = \frac{1}{225} \approx 0{,}0044.$$

A kovariancia illetve szórás korábban tárgyalt tulajdonságai szintén teljesülnek, függetlenül attól, hogy folytonos esetről beszélünk-e vagy sem.

10.1.4. Lemma. *Legyen (X, Y, Z) valószínűségi vektorváltozó. Ekkor teljesülnek az alábbiak, feltéve, hogy a bennük szereplő mennyiségek értelmezettek:*

(1) Ha $c \in \mathbb{R}$, akkor $\mathbb{D}(X + c) = \mathbb{D}(X)$ és $\mathbb{D}(cX) = |c|\mathbb{D}(X)$.
(2) $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\mathrm{cov}(X, Y)$.
(3) $\mathbb{D}^2(X) = 0$ pontosan akkor teljesül, ha $\mathbb{P}(X = c) = 1$ valamilyen $c \in \mathbb{R}$-re.
(4) Ha X és Y függetlenek, akkor $\mathrm{cov}(X, Y) = 0$, speciálisan $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
(5) (bilineáris) Ha $b, c \in \mathbb{R}$ akkor $\mathrm{cov}(X, bY + cZ) = b \cdot \mathrm{cov}(X, Y) + c \cdot \mathrm{cov}(X, Z)$.

Megjegyzés. A lemma 4. pontja általánosabban alkalmazható, ha felhasználjuk az alábbi lemmát.

10.1.5. Lemma. *Ha X és Y független valószínűségi változók, g és h folytonos, valós függvények, akkor $g(X)$ és $h(Y)$ is függetlenek.*

A lemma nem nyilvánvaló, de itt nem bizonyítjuk.

Valószínűségi vektorváltozó esetén a szórásnégyzeteket és kovarianciákat mátrixba rendezve szokás kezelni. Ennek a motivációja nem a kompakt leírhatóság, hanem az, hogy a valószínűségi vektorváltozókkal való számolásokban természetes módon előkerül a kovarianciamátrix vektorokkal vett szorzata, a mátrix determinánsa, illetve nyoma is, lásd például a többváltozós normális eloszlást a 12. előadáson.

⚠ **10.1.6. Definíció.** Az $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó **kovarianciamátrix**a az alábbi $n \times n$-es valós mátrix:

$$\operatorname{cov}(\underline{X}) = \begin{pmatrix} \operatorname{cov}(X_1,X_1) & \operatorname{cov}(X_1,X_2) & \dots & \operatorname{cov}(X_1,X_n) \\ \operatorname{cov}(X_2,X_1) & \operatorname{cov}(X_2,X_2) & & \vdots \\ \vdots & & \ddots & \\ \operatorname{cov}(X_n,X_1) & \dots & & \operatorname{cov}(X_n,X_n) \end{pmatrix},$$

azaz $\operatorname{cov}(\underline{X})_{i,j} = \operatorname{cov}(X_i, X_j)$ minden $1 \le i, j \le n$ esetén.

De hol van ebben szórásnégyzet? Mivel $\mathbb{D}^2(X_1) = \operatorname{cov}(X_1, X_1)$, így a mátrix főátlójában lévő elemek a vektorváltozó koordinátáinak szórásnégyzetei.

10.1.7. Állítás. *Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Ekkor*

(1) $\operatorname{cov}(\underline{X})$ *szimmetrikus, azaz* $\operatorname{cov}(X_i, X_j) = \operatorname{cov}(X_j, X_i)$ *minden* $1 \le i, j \le n$ *esetén.*

(2) $\operatorname{cov}(\underline{X})$ *pozitív szemidefinit mátrix, azaz* $\sum_{i=1}^n \sum_{j=1}^n a_i \operatorname{cov}(X_i, X_j) a_j \ge 0$ *minden* $(a_1, \dots, a_n) \in \mathbb{R}^n$ *esetén, és pontosan akkor* 0*, ha* $\sum_{i=1}^n a_i X_i$ *1-valószínűséggel konstans.*

Bizonyítás. A kovariancia szimmetrikussága a definíciója szimmetrikusságából adódik, ezt nem ragozzuk. A pozitív szemidefinitség belátását kezdjük az extrém esettel: tegyük fel, hogy $\sum_{i=1}^n a_i X_i$ 1-valószínűséggel konstans valószínűségi változó, azaz $\mathbb{P}\big(\sum_{i=1}^n a_i X_i = c\big) = 1$ valamilyen $c \in \mathbb{R}$ esetén. Az előző lemma 3-as pontja szerint ez ekvivalens azzal, hogy a valószínűségi változó szórásnégyzete 0. Továbbá, a lemma 5-ös pontja miatt

$$\mathbb{D}^2\Big(\sum_{i=1}^n a_i X_i\Big) = \operatorname{cov}\Big(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\Big) = \sum_{i=1}^n \sum_{j=1}^n a_i \operatorname{cov}(X_i, X_j) a_j. \tag{8}$$

Tehát ha a valószínűségi változó 1-valószínűséggel konstans, akkor a jobb oldalon lévő összeg is 0. Az érvelés fordított irányba ugyanígy elmondható, így az állítás „pontosan akkor” része teljesül. Az egyenlőtlenség belátásához már csak azt kell észrevennünk, hogy a szórásnégyzet nemnegatív, ezért a (8) egyenlet jobb oldala is mindig nemnegatív. □

10.1.8. Példa. Írjuk fel az előző példában szereplő (X, Y) valószínűségi vektorváltozó kovarianciamátrixát. Ehhez szükségünk van $\mathbb{D}^2(X)$-re és $\mathbb{D}^2(Y)$-ra is. A korábbiakkal analóg átalakításokkal, illetve polinomok integrálásával kapjuk, hogy

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f_X(x,y) \mathrm{d}x\mathrm{d}y - \Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_X(x,y) \mathrm{d}x\mathrm{d}y\Big)^2 =$$

$$= \int_0^2 \int_0^1 x^2 \cdot \frac{1}{5}(4 - 2x^2 + xy - y^2) \mathrm{d}x\mathrm{d}y - \Big(\int_0^2 \int_0^1 x \cdot \frac{1}{5}(4 - 2x^2 + xy - y^2) \mathrm{d}x\mathrm{d}y\Big)^2 = \frac{7}{90}.$$

És hasonlóan $\mathbb{D}^2(Y) = \frac{58}{225}$. Tehát ha $\underline{Z}$ jelöli az (X, Y) valószínűségi vektorváltozót, akkor

$$\operatorname{cov}(\underline{Z}) = \begin{pmatrix} \frac{7}{90} & \frac{1}{225} \\ \frac{1}{225} & \frac{58}{225} \end{pmatrix}.$$

10.2. Lineáris regresszió

Tegyük fel, hogy egy esernyőket áruló bolt tulajdonosai vagyunk, és kapunk egy hosszútávú előrejelzést a jövő évi csapadékmennyiségről. Jobb híján ezen előrejelzés alapján próbáljuk megtippelni, mekkora készletet rendeljünk, azaz körülbelül hány esernyőt fogunk eladni. Hogyan kellene tippeljünk, ha a korábbi évek alapján van némi elképzelésünk a csapadékmennyiség és az eladott esernyők száma közti összefüggésről? Ilyen becslésre az egyik lehetséges módszerünk a lineáris regresszió.

Jelölje X az éves csapadékmennyiséget, Y pedig az eladott esernyők számát, ahogy a második példában. Tegyük fel, hogy (X, Y) együttes sűrűségfüggvénye a példában szereplő $f_{X,Y}$. A lineáris regresszió alapötlete, hogy próbáljuk meg Y-t az X-nek egy lineáris függvényével, azaz $\beta \cdot X + \alpha$ alakban, a lehető legjobban közelíteni.

Vegyük észre, hogy a „legjobb közelítés” nem egy jóldefiniált fogalom: azt még meg kéne mondanunk, mi alapján tekintünk egy közelítést jónak vagy rossznak. Erre többféle megközelítés is bevethető,[47] de a legalapvetőbb, az ún. **legkisebb négyzetek módszere**.

⚠ **10.2.1. Definíció.** Legyenek X és Y valószínűségi változók. Ekkor Y-nak az X-re vett **lineáris regresszió**ján azt a $\beta X + \alpha$ valószínűségi változót értjük, ahol $\alpha, \beta \in \mathbb{R}$, és az

$$\mathbb{E}\Big(\big(Y - (\beta X + \alpha)\big)^2\Big) \tag{9}$$

mennyiség minimális.

Ennek az optimalizálási problémának a megoldása lényegében mindig létezik és egyértelmű:

⚠ **10.2.2. Állítás.** *Legyenek X és Y olyan valószínűségi változók, amire $\mathbb{D}^2(X)$, $\mathbb{D}^2(Y)$ és $\mathrm{cov}(X, Y)$ véges, továbbá $\mathbb{D}^2(X) \neq 0$. Ekkor a (9) egyenletben szereplő várható érték pontosan akkor minimális, ha*

$$\beta = \frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)} \qquad \textit{és} \qquad \alpha = \mathbb{E}(Y) - \frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)}\mathbb{E}(X).$$

10.2.3. Definíció. Az Y valószínűségi változó X-re vett **regressziós egyenes**e az

$$\{(x, y) \in \mathbb{R}^2 \mid y = \beta x + \alpha\}$$

egyenes a síkon, ahol β és α értéke a fenti állításban szerepel.

Vizuálisabban, az (X, Y) valószínűségi vektorváltozó lehetséges értékeinek a síkján az eloszlást „legjobban közelítő” egyenes a regressziós egyenes. A lineáris regresszió akkor lesz jól használható modell, ha az (X, Y) együttes eloszlása ezen egyenes környékén koncentrálódik.

Megjegyzés. A β-ra és az α-ra vonatkozó egyenleteket nem feltétlenül egyszerű sem megjegyezni, sem megindokolni. Egy heurisztika (de nem bizonyítás) a helyes α és β megtalálására, hogy olyannak válasszuk őket, amire Y-nak és $\beta X + \alpha$-nak ugyanaz a várható értéke és az X-el vett kovarianciája. Emiatt

$$\mathbb{E}(Y) = \mathbb{E}(\beta X + \alpha) = \beta\mathbb{E}(X) + \alpha \qquad \text{és} \qquad \mathrm{cov}(X,Y) = \mathrm{cov}(X, \beta X + \alpha) = \beta\mathbb{D}^2(X) + 0,$$

amely egyenletekből adódik is β és α értéke.

Egy hasonló, kompaktabb megközelítés a korreláció fogalmán keresztül vezet. Idézzük fel, X és Y korrelációja:

$$\mathrm{corr}(X,Y) \stackrel{\mathrm{def}}{=} \frac{\mathrm{cov}(X,Y)}{\mathbb{D}(X)\mathbb{D}(Y)}$$

[47] A lineáris regresszió alternatív változatai, amelyek máshogy definiálják a „legjobb közelítés” fogalmát: súlyozott lineáris regresszió, ridge regresszió, avagy az ℓ_1 regresszió.

egy -1 és 1 közti valós szám, ami X és Y lineáris összefüggését méri. Azt állítjuk, hogy ha $\beta X+\alpha$ az Y lineáris regressziója X-re, akkor teljesül, hogy

$$\frac{(\beta X+\alpha)-\mathbb{E}(Y)}{\mathbb{D}(Y)}=\frac{X-\mathbb{E}(X)}{\mathbb{D}(X)}\cdot\operatorname{corr}(X,Y).$$

Más szavakkal, ha Y standardizáltjába az első Y helyére a $\beta X+\alpha$ regressziót helyettesítjük, akkor az eredmény X standardizáltjának korreláció-szorosa. Ez az azonosság egyszerű átrendezéssel belátható.

Bizonyítás. A következő függvényt kellene minimalizálnunk:

$$h(\alpha,\beta)=\mathbb{E}\Big(\big(Y-(\beta X+\alpha)\big)^2\Big)=\mathbb{E}\Big(Y^2+\beta^2X^2+\alpha^2-2\beta XY-2\alpha Y+2\alpha\beta X\Big)=$$

$$=\mathbb{E}(Y^2)+\beta^2\mathbb{E}(X^2)+\alpha^2-2\beta\mathbb{E}(XY)-2\alpha\mathbb{E}(Y)+2\alpha\beta\mathbb{E}(X).$$

Az eredeti formából látszik, hogy h nemnegatív (hiszen valószínűségi változó négyzetének várható értéke), továbbá az átalakított formából világos, hogy α-ban és β-ban h másodfokú polinom. Egy ilyen polinomnak csak ott lehet globális minimuma, ahol mind az α-ban, mind a β-ban vett parciális derivált eltűnik.

Bár egy (α_0,β_0) pontban a parciális deriváltak eltűnése nem elégséges feltétele annak, hogy ez a pont a h függvény globális minimuma legyen, jelen esetben a nemnegativitás miatt mégis ez a helyzet. Valóban, indirekt tegyük fel, hogy az (α_0,β_0) pontban eltűnik mindkét parciális derivált, de $h(\alpha_1,\beta_1)<h(\alpha_0,\beta_0)$. Nézzük a függvényt a két pontot összekötő egyenesen, vagyis tekintsük az $f(t)=h(t\alpha_0+(1-t)\alpha_1,t\beta_0+(1-t)\beta_1)$ egyváltozós függvényt. Mivel ezt h-ból lineáris behelyettesítéssel kaptuk, így polinom kell legyen t-ben, ami legfeljebb másodfokú. Sőt, a 0-ban vett deriváltját is ki tudjuk fejezni h parciális deriváltjaival az (α_0,β_0) pontban, ezért $f'(0)=0$, hiszen a parciális deriváltak eltűnnek. Összefoglalva, f egy olyan legfeljebb másodfokú polinom, amire $f'(0)=0$, és f mindenhol nemnegatív (ebből már látjuk, hogy f vagy egy felfelé álló parabola, vagy konstans), de mégis $h(\alpha_1,\beta_1)=f(1)<f(0)=h(\alpha_0,\beta_0)$. Ez ellentmondás, ilyen polinom nincs.

Visszatérve a globális minimum pontos értékére, h parciális deriváltjai a következők:

$$\beta\text{ szerint: }2\beta\mathbb{E}(X^2)-2\mathbb{E}(XY)+2\alpha\mathbb{E}(X)\qquad\text{és}\qquad\alpha\text{ szerint: }2\alpha-2\mathbb{E}(Y)+2\beta\mathbb{E}(X).$$

Vagyis a parciális deriváltak közös nullhelyeit megadó egyenletek:

$$\alpha\mathbb{E}(X)+\beta\mathbb{E}(X^2)=\mathbb{E}(XY)\qquad\text{és}\qquad\alpha+\beta\mathbb{E}(X)=\mathbb{E}(Y).$$

Ez egy 2×2-es lineáris egyenletrendszer α-ban és β-ban. Megoldása:

$$\beta=\frac{\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)}{\mathbb{E}(X^2)-\mathbb{E}(X)^2}=\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\qquad\text{és}\qquad\alpha=\mathbb{E}(Y)-\beta\mathbb{E}(X)=\mathbb{E}(Y)-\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\mathbb{E}(X),$$

amik éppen a kívánt egyenletek. □

10.2.4. Példa. Mit kapunk a felvezető példa esetében, ahol X a csapadékmennyiség, Y az eladott esernyők száma? A már kiszámolt kovarianciamátrix koordinátáiból rögtön felírhatók az Y-nak az X-re vett lineáris regressziójának együtthatói:

$$\beta=\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}=\frac{1/225}{7/90}=\frac{2}{35},\qquad\alpha=\mathbb{E}(Y)-\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\mathbb{E}(X)=\frac{4}{5}-\frac{2}{35}\frac{7}{15}=\frac{58}{75}.$$

Tehát ha X-re kapunk egy előrejelzést, akkor ezen együtthatókkal közelíthetjük Y-t. Némi értelmezést hozzáadva: az eső mennyiségének emelkedése csak kismértékben fogja növelni a már alapból magas készletszükségletet.

Mivel a lineáris regresszió csak közelítés, így fontos információ lehet, hogy mekkora hibával találja el Y értékét. (Hiba alatt itt átlagos négyzetes hibát, vagyis szórásnégyzetet értünk.)

10.2.5. Állítás. *Legyen az Y valószínűségi változó X-re vett lineáris regressziója $\beta X + \alpha$. Ekkor az eltérés szórásnégyzete:*

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) - \frac{\operatorname{cov}(X,Y)^2}{\mathbb{D}^2(X)}.$$

Bizonyítás. A szórásnégyzet fentebb felsorolt tulajdonságai és $\beta = \frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}$ miatt:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y - \beta X) = \mathbb{D}^2(Y) + \beta^2\mathbb{D}^2(X) - 2\operatorname{cov}(Y, \beta X) =$$

$$= \mathbb{D}^2(Y) + \frac{\operatorname{cov}(X,Y)^2}{\big(\mathbb{D}^2(X)\big)^2}\mathbb{D}^2(X) - 2\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\operatorname{cov}(Y,X) = \mathbb{D}^2(Y) - \frac{\operatorname{cov}(X,Y)^2}{\mathbb{D}^2(X)}.$$

Éppen ez volt az állítás. □

Megjegyzés. Másképpen felírva:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) \cdot \big(1 - \operatorname{corr}(X,Y)^2\big).$$

Speciálisan, minél nagyobb a korreláció X és Y közt, annál kisebb rész járul hozzá a hiba szórásnégyzetéhez $\mathbb{D}^2(Y)$-ból. Továbbá, ez az átfogalmazás azt is mutatja, hogy a fenti állításból következik a 6.5 alfejezet állítása.

10.2.6. Példa. Az előző példa esetében

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \frac{58}{225} - \frac{(1/225)^2}{(7/90)^2} \approx 0{,}2545.$$

Vagyis az eladások jócskán eltérhetnek a lineáris regresszió által becsült értéktől.

Hasznos észben tartani, hogy statisztikai témakörben nem ugyanezt értik lineáris regresszió alatt. A különbség, hogy ott nem feltételezik, hogy a valószínűségi változók eloszlása ismert, de általában azt sem, hogy (az esetlegesen ebből levezethető) kovariancia és szórásnégyzet értékeit ismernénk. Így a statisztikai értelemben vett lineáris regresszióba beleértik azt is, hogy a β és α értékek maguk is becsült mennyiségek, egy véges nagyságú minta alapján. Ez lényegesen eltérő egyenleteket és értelmezést jelent, de ettől még a lineáris regresszió ötlete ugyanaz marad: keressünk közelítőleg lineáris összefüggést a vizsgált változók között.