A várható érték tulajdonságai (6. előadás)

- (összeg várható értéke) Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$
- (konstans kiemelése) Ha az X valószínűségi változó várható értéke létezik, és c tetszőleges valós szám, akkor
$$\mathbb{E}(c \cdot X) = c \cdot \mathbb{E}(X).$$

A várható érték tulajdonságai (6. előadás)

- (összeg várható értéke) Ha X, Y valószínűségi változók, és X, Y, $X+Y$ várható értéke létezik, akkor
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$
- (konstans kiemelése) Ha az X valószínűségi változó várható értéke létezik, és c tetszőleges valós szám, akkor
$$\mathbb{E}(c \cdot X) = c \cdot \mathbb{E}(X).$$
- (szorzat várható értéke független esetben) Ha az X és Y valószínűségi változók függetlenek, és X, Y, $X \cdot Y$ várható értéke létezik, akkor
$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

A várható érték tulajdonságai (6. előadás)

- (összeg várható értéke) Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$
- (konstans kiemelése) Ha az X valószínűségi változó várható értéke létezik, és c tetszőleges valós szám, akkor
$$\mathbb{E}(c \cdot X) = c \cdot \mathbb{E}(X).$$
- (szorzat várható értéke független esetben) Ha az X és Y valószínűségi változók függetlenek, és $X, Y, X \cdot Y$ várható értéke létezik, akkor
$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$
- (függvény várható értéke) Ha $g : \mathbb{R} \to \mathbb{R}$ olyan függvény, hogy $\mathbb{E}(X)$ létezik, és az X lehetséges értékei $x_1, x_2, \ldots$, akkor
$$\mathbb{E}(g(X)) = \sum_{i=1}^{\infty} g(x_i)\mathbb{P}(X = x_i).$$

Összeg várható értéke

Állítás

Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{k,m} (x_k + y_m)\mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k,m} x_k \mathbb{P}(X = x_k, Y = y_m) + \sum_{k,m} y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k} x_k \mathbb{P}(X = x_k) + \sum_{m} y_m \mathbb{P}(Y = y_m) = \mathbb{E}(X) + \mathbb{E}(Y),
\end{aligned}$$

ahol azt használtuk, hogy az $\{X = x_k, Y = y_m\}$ események kizáróak, uniójuk $\{X = x_k\}$, és hasonlóképpen a másik tagban az Y esetén.

Összeg várható értéke

Állítás

Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{k,m} (x_k + y_m)\mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k,m} x_k \mathbb{P}(X = x_k, Y = y_m) + \sum_{k,m} y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k} x_k \mathbb{P}(X = x_k) + \sum_{m} y_m \mathbb{P}(Y = y_m) = \mathbb{E}(X) + \mathbb{E}(Y),
\end{aligned}$$

ahol azt használtuk, hogy az $\{X = x_k, Y = y_m\}$ események kizáróak, uniójuk $\{X = x_k\}$, és hasonlóképpen a másik tagban az Y esetén.

Következmény: $\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)$.

Összeg szórása független esetben

Állítás

Ha X, Y **független** *valószínűségi változók, és X, Y várható értéke létezik, akkor*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X + Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{k,m} x_k \cdot y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&\overset{(*)}{=} \sum_{k,m} x_k y_m \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_m) \\
&= \left(\sum_k x_k \mathbb{P}(X = x_k) \right) \cdot \left(\sum_m y_m \mathbb{P}(Y = y_m) \right) = \mathbb{E}(X) \cdot \mathbb{E}(Y),
\end{aligned}$$

ahol a $(*)$ lépésben használtuk a függetlenségnek a diszkrét valószínűségi változókra vonatkozó alakját.

A szórásnégyzet tulajdonságai

- (nemnegativitás) $D^2(X) \geq 0$ és $D(X) \geq 0$ mindig teljesül
- (szorzás és eltolás) ha a, b valós számok, X véges szórású valószínűségi változó, akkor

$$D^2(aX + b) = a^2 D^2(X) \quad \Rightarrow \quad D(aX + b) = |a| D(X).$$

A szórásnégyzet tulajdonságai

- (nemnegativitás) $D^2(X) \geq 0$ és $D(X) \geq 0$ mindig teljesül
- (szorzás és eltolás) ha a, b valós számok, X véges szórású valószínűségi változó, akkor

$$D^2(aX + b) = a^2 D^2(X) \quad \Rightarrow \quad D(aX + b) = |a| D(X).$$

- (összeg szórása független esetben) ha az X, Y valószínűségi változók függetlenek és szórásuk létezik, akkor

$$D^2(X+Y) = D^2(X)+D^2(Y) \quad \Rightarrow \quad D(X+Y) = \sqrt{D^2(X) + D^2(Y)}.$$

- van olyan valószínűségi változó, melynek várható értéke véges, de a szórása nem létezik (például: $\mathbb{P}(X = k) = c/k^3$ megfelelő c-vel)

A binomiális eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó binomiális eloszlású n renddel és p paraméterrel, azaz

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n),$$

akkor

$$\mathbb{E}(X) = np; \qquad D(X) = \sqrt{np(1-p)}.$$

Példa. Egy kérdőív egy kérdésére $n = 1000$ megkérdezett közül mindenki a többiektől függetlenül $p = 0,65$ valószínűséggel válaszol. Ekkor a válaszadók számának (melyet jelöljünk X-szel) a várható értéke:

$$\mathbb{E}(X) = np = 1000 \cdot 0,65 = 650,$$

míg a szórása:

$$D(X) = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0,65 \cdot 0,35} = 15,08.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Legyen X binomiális eloszlású n renddel és p paraméterrel. Azaz: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeresek száma. Vezessük be az alábbi indikátorokat $j = 1, 2, \ldots, n$ esetén:

$$X_j = \begin{cases} 1 & \text{ha a } j. \text{ kísérlet sikeres;} \\ 0 & \text{ha a } j. \text{ kísérlet nem sikeres.} \end{cases}$$

Ekkor X éppen az indikátorok összege (az egyesek száma):

$$X = X_1 + X_2 + \ldots + X_n = \sum_{j=1}^{n} X_j.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Legyen X binomiális eloszlású n renddel és p paraméterrel. Azaz: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeresek száma. Vezessük be az alábbi indikátorokat $j = 1, 2, \ldots, n$ esetén:

$$X_j = \begin{cases} 1 & \text{ha a } j. \text{ kísérlet sikeres;} \\ 0 & \text{ha a } j. \text{ kísérlet nem sikeres.} \end{cases}$$

Ekkor X éppen az indikátorok összege (az egyesek száma):

$$X = X_1 + X_2 + \ldots + X_n = \sum_{j=1}^{n} X_j.$$

Mivel minden j-re

$$\mathbb{E}(X_j) = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p,$$

ezért

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n) = p + \ldots + p = np.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Ugyanazokat az indikátorokat használjuk:

$$X_j = \begin{cases} 1 & \text{ha a } j\text{.kísérlet sikeres;} \\ 0 & \text{ha a } j\text{.kísérlet nem sikeres.} \end{cases}$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Ugyanazokat az indikátorokat használjuk:

$$X_j = \begin{cases} 1 & \text{ha a } j\text{.kísérlet sikeres;} \\ 0 & \text{ha a } j\text{.kísérlet nem sikeres.} \end{cases}$$

Most $X_j = X_j^2$, hiszen $0^2 = 0$ és $1^2 = 1$, és már láttuk, hogy $\mathbb{E}(X_j) = p$. Ezért

$$D^2(X_j) = \mathbb{E}(X_j^2) - \mathbb{E}(X_j)^2 = \mathbb{E}(X_j) - \mathbb{E}(X_j)^2 = p - p^2 = p(1-p),$$

Mivel az X_j indikátorok **függetlenek**, és az összegük X:

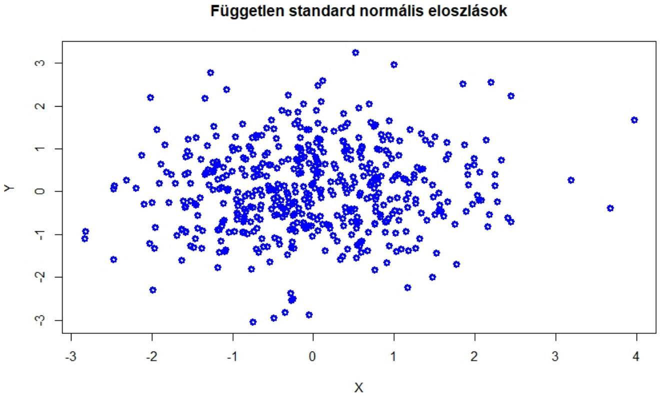
$$D(X) = \sqrt{D^2(X_1 + X_2 + \ldots + X_n)} = \sqrt{D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)} =$$
$$= \sqrt{p(1-p) + p(1-p) + \ldots + p(1-p)} = \sqrt{np(1-p)}.$$

A hipergeometrikus eloszlás várható értékére vonatkozó összefüggés szintén indikátorokkal igazolható.

Kovariancia és korrelációs együttható

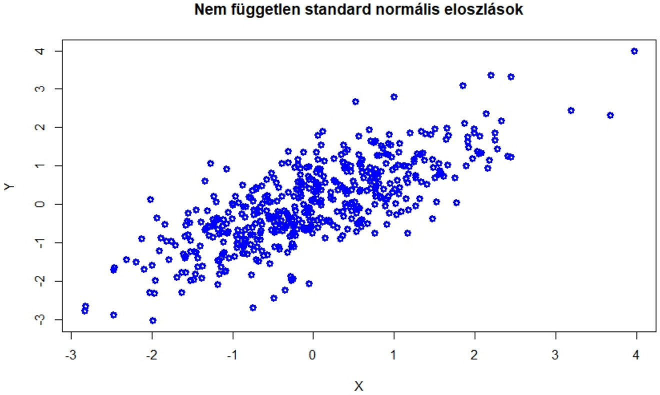
- két valószínűségi változó lehet
  - **független**: például két találomra választott ember jövedelme, vagy,
  - **nem független**: például egy találomra választott ember jövedelme most, illetve fél év múlva
- az **összefüggőség mértéke** különböző lehet:
  - egy találomra választott felnőtt életkora és jövedelme „erősen összefüggő”, a fiataloké és időseké általában alacsonyabb;
  - egy találomra választott felnőtt életkora és testmagassága „gyengén összefüggő”, hiszen egy fiatal felnőtt nőhet, az idősek pedig valamennyit veszítenek a testmagasságukból, de egyik változás sem nagyon jelentős.
- a kapcsolat erősségének jellemzésére többféle mérőszám használható, ezek között van a **kovarianca** és a **korrelációs együttható**. Ez utóbbinak a „nagy” értékei „erős, lineáris jellegű” összefüggésre utalnak.

Független normális eloszlások



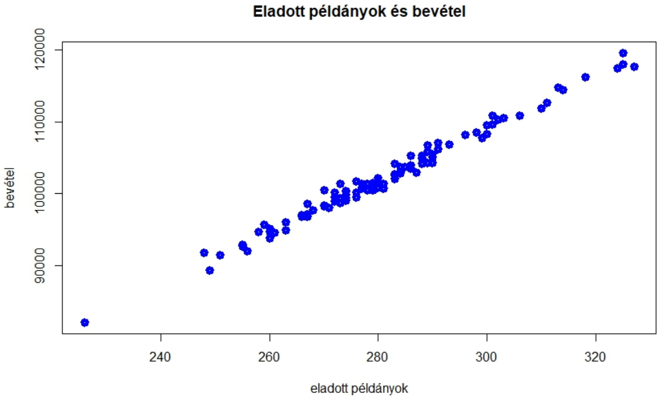
500 darab véletlen pont a síkon, melyek koordinátái **független** standard nomális eloszlásúak. A koordináták között nincs kapcsolat: a kovariancia és a korrelációs együttható is 0 lesz.

Pozitív korreláció



500 elemű minta a következő többdimenziós normális eloszlásból: $(X, \frac{X+Z}{\sqrt{2}})$, ahol $X, Z \sim N(0,1)$ függetlenek.
Minél nagyobb X, „valószínűleg” annál nagyobb $(X+Z)/\sqrt{2}$ is $\rightarrow$ ennek megfelelően a két koordináta közötti **kovariancia** és **korrelációs együttható** is **pozitív** lesz.

Erős pozitív korreláció



100 elemű minta az $(X+Y, 300X+400Y)$ eloszlásból, ahol $X \sim$ Poisson(100) és $Y \sim$ Poisson(180) függetlenek. A megfigyelések szinte teljesen egy pozitív meredekségű egyenesre illeszkednek $\rightarrow$ a **korrelációs együttható pozitív** és **majdnem 1**, ami „nagy”, mert ennek 1 lesz a lehetséges legnagyobb értéke.

A kovariancia

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor X és Y **kovarianciája**:

$$\mathrm{cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\big).$$

A kovariancia

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor X és Y **kovarianciája**:

$$\mathrm{cov}(X,Y) = \mathbb{E}\big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\big).$$

- **A kovariancia kiszámítása:**

$$\mathrm{cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y).$$

- **Szimmetria.** $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$.
- **Kapcsolat a szórásnégyzettel.** $\mathrm{cov}(X,X) = D^2(X)$.

A kovariancia

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor X és Y **kovarianciája**:

$$\mathrm{cov}(X,Y) = \mathbb{E}\big((X-\mathbb{E}(X))\cdot(Y-\mathbb{E}(Y))\big).$$

- **A kovariancia kiszámítása:**

$$\mathrm{cov}(X,Y) = \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\mathbb{E}(Y).$$

- **Szimmetria.** $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$.
- **Kapcsolat a szórásnégyzettel.** $\mathrm{cov}(X,X) = D^2(X)$.
- **Függetlenséggel való kapcsolat.** Ha az X és Y valószínűségi változók **függetlenek**, akkor $\mathrm{cov}(X,Y) = 0$.

  Fordítva nem igaz: $\mathrm{cov}(X,Y) = 0$ esetén nem biztos, hogy X és Y függetlenek.

A kovariancia tulajdonságai

- Konstanssal való kovariancia. $\mathrm{cov}(X, c) = 0$, ha c valós szám.
- **Linearitás.** Egyrészt

$$\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z),$$

  másrészt tetszőleges $c \in \mathbb{R}$ valós számra

$$\mathrm{cov}(c \cdot X, Y) = c \cdot \mathrm{cov}(X, Y).$$

- **Összeg szórásnégyzete.** $D^2(X+Y) = D^2(X)+D^2(Y)+2\mathrm{cov}(X, Y)$. Továbbá

$$D^2\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} D^2(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j).$$

- Különbség szórásnégyzete. $D^2(X-Y) = D^2(X)+D^2(Y)-2\mathrm{cov}(X, Y)$.

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X;
- a B újságból eladott példányok száma Y.
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y-é 180.
- Az A újság ára 300 forint, a B-é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája?

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X;
- a B újságból eladott példányok száma Y.
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y-é 180.
- Az A újság ára 300 forint, a B-é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája? Azaz mennyi $\mathrm{cov}(X+Y, 300X+400Y)$?

A számolás előtt: **pozitív**, **negatív** vagy 0 kovarianciára tippelnénk?

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X;
- a B újságból eladott példányok száma Y.
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y-é 180.
- Az A újság ára 300 forint, a B-é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája? Azaz mennyi $\mathrm{cov}(X+Y, 300X+400Y)$?

A számolás előtt: **pozitív**, **negatív** vagy 0 kovarianciára tippelnénk?

Mivel **minél nagyobb** a példányszám, **„valószínűleg” annál nagyobb a bevétel**, **pozitív** kovarianciára számíthatunk.

Kovariancia: példa

X és Y **független**ek, **Poisson-eloszlású**ak, X paramétere 100, az Y-é 180. Ekkor az **eladott példányok számának** és a **bevételnek** a kovarianciája:

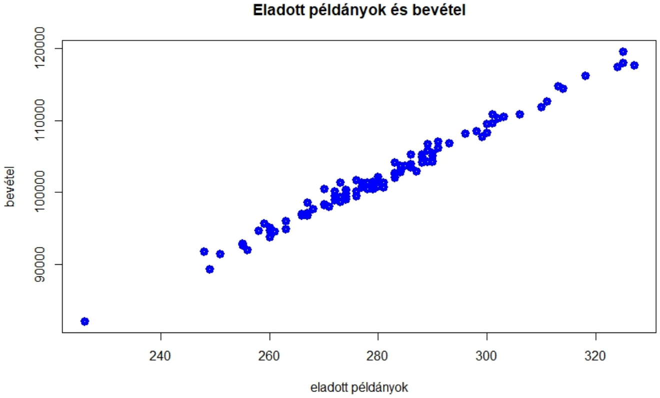
$$\begin{aligned}
\operatorname{cov}(X+Y, 300X+400Y) &\overset{(a)}{=} \operatorname{cov}(X, 300X) + \operatorname{cov}(X, 400Y) + \\
&\quad + \operatorname{cov}(Y, 300X) + \operatorname{cov}(Y, 400Y) = \\
&\overset{(a,b)}{=} 300 \cdot \operatorname{cov}(X, X) + 400 \cdot \operatorname{cov}(Y, Y) = \\
&\overset{(b)}{=} 300D^2(X) + 400D^2(Y) = \\
&\overset{(c)}{=} 300 \cdot 100 + 400 \cdot 180 = 102000,
\end{aligned}$$

ahol felhasználtuk, hogy (a) a kovariancia **lineáris**;

(b) **független** valószínűségi változók kovarianciája 0, illetve egy valószínűségi változó saját magával vett kovarianciája a szórásnégyzete;

(c) egy λ paraméterű **Poisson-eloszlású** valószínűségi változó **szórásnégyzete** λ.

Kovariancia: példa



A bevétel $(300X + 400Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ független megfigyelésből. Kovariancia: $\mathrm{cov}(X + Y, 300X + 400Y) = 102000$.

Korrelációs együttható: bevezetés

- A **kovariancia** bevezetésének célja, hogy két valószínűségi változó közötti **összefüggés erősségét** tudjuk mérni.
- A korábbi példában: a példányszám és a bevétel kovarianciája **102000** volt.
- Viszont ha a bevételt nem forintban, hanem ezer forintos egységben mérjük:

  $$X : \text{példányszám} \quad Y : \text{bevétel forintban} \quad Z : \text{bevétel ezer forintban},$$

  akkor

  $$\operatorname{cov}(X, Z) = \operatorname{cov}\left(X, \frac{Y}{1000}\right) = \frac{\operatorname{cov}(X, Y)}{1000} = 102.$$

  Vagyis a kovariancia a **mértékegységtől függ** $\Rightarrow$ hasznos egy olyan mennyiség, ami szintén az összefüggés erősségét méri, de a mértékegység választásától függetlenül.
- Ilyen lesz a **korrelációs együttható**.

Korrelációs együttható: definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója**:

$$R(X,Y) = \begin{cases} \frac{\mathrm{cov}(X,Y)}{D(X)D(Y)}, & \text{ha } D(X) > 0, D(Y) > 0; \\ 0, & \text{ha } D(X) = 0 \text{ vagy } D(Y) = 0. \end{cases}$$

Korrelációs együttható: definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója**:

$$R(X,Y) = \begin{cases} \frac{\mathrm{cov}(X,Y)}{D(X)D(Y)}, & \text{ha } D(X) > 0, D(Y) > 0; \\ 0, & \text{ha } D(X) = 0 \text{ vagy } D(Y) = 0. \end{cases}$$

- **Lehetséges értékek**. A korrelációs együttható értéke mindig -1 és 1 közé esik:
$$|R(X,Y)| \leq 1.$$
- **Lineáris összefüggés**. Legyen $a > 0$ valós szám, b tetszőleges valós szám. Ekkor
$$R(X, aX + b) = 1 \qquad \text{és} \qquad R(X, -aX + b) = -1.$$
- Tegyük fel, hogy $|R(X,Y)| = 1$. Ekkor léteznek olyan a és b valós számok, hogy az $Y = aX + b$ egyenlet 1 valószínűséggel teljesül. Vagyis a korrelációs együttható lehetséges legnagyobb értékei lineáris összefüggés esetén érhetők el.

Korrelációs együttható: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X;
- a B újságból eladott példányok száma Y.
- Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180.
- Az A újság ára 300 forint, a B-é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a korrelációs együtthatója?

Korrelációs együttható: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X;
- a B újságból eladott példányok száma Y.
- Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180.
- Az A újság ára 300 forint, a B-é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a korrelációs együtthatója?

$$R(X+Y, 300X+400Y) = \frac{\operatorname{cov}(X+Y, 300X+400Y)}{D(X+Y)D(300X+400Y)} = \\ = \frac{102000}{D(X+Y)D(300X+400Y)}$$

a korábbi számolás alapján, így a szórásokat kell meghatároznunk.

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y-é 180.
Ekkor az **eladott példányok számának** szórása:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y-é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X+Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100+180} = 16,73.$$

A bevétel szórása:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y-é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X+Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100+180} = 16,73.$$

A bevétel szórása:

$$D(300X + 400Y) = \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} =$$
$$= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = 6148,17.$$

Ezek alapján a korrelációs együttható:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y-é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X+Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100+180} = 16,73.$$

A bevétel szórása:

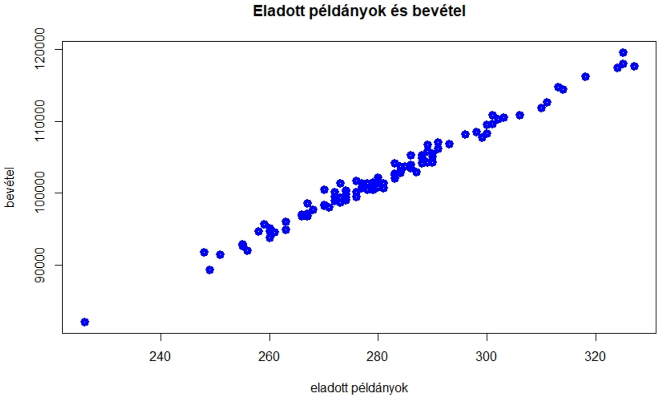
$$D(300X+400Y) = \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} = $$
$$= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = 6148,17.$$

Ezek alapján a korrelációs együttható:

$$R(X+Y, 300X+400Y) = \frac{\operatorname{cov}(X+Y, 300X+400Y)}{D(X+Y)D(300X+400Y)} =$$
$$= \frac{102000}{16,73 \cdot 6148,17} = 0,9915.$$

A korrelációs együttható lehetséges legnagyobb értéke **1**, így ez **erős pozitív korrelációt** jelent.

Korrelációs együttható: példa



A bevétel $(300X + 400Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ független megfigyelésből. Kovariancia: **102000**, korrelációs együttható: $0,9915$.

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é **4000**. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$\mathrm{cov}(X+Y, 300X+4000Y) = 300\cdot 100 + 4000\cdot 180 = 750000;$$

$$D(X+Y) = \sqrt{D^2(X)+D^2(Y)} = \sqrt{100+180} = 16,73;$$

$$D(300X+4000Y) = \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} = $$

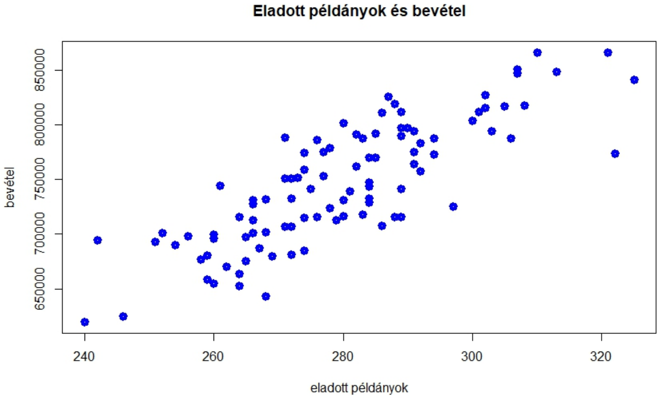
$$= \sqrt{300^2\cdot 100 + 4000^2\cdot 180} = 53749,42;$$

$$R(X+Y, 300X+4000Y) = \frac{cov(X+Y, 300X+4000Y)}{D(X+Y)D(300X+4000Y)} = $$

$$= \frac{750000}{16,73\cdot 53749,42} = 0,83.$$

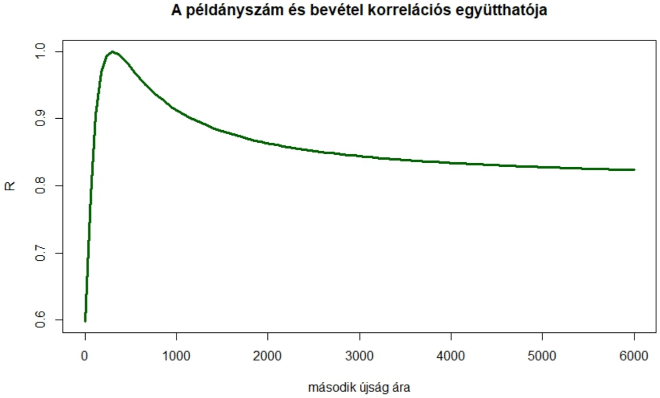
A korrelációs együttható értéke kisebb, mint hasonló ár esetén.

Korrelációs együttható: példa



A bevétel $(300X + 4000Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 750000, korrelációs együttható: $0,83$.

Korrelációs együttható: példa



A bevétel $(300X + cY)$ és az eladott példányszám $(X + Y)$ korrelációs együtthatója c függvényében, ahol az első újság ára 300, a másodiké c

Korrelációs együttható: példa

Tegyük fel, hogy X és Y **független**, 4 szórású Poisson-eloszlású valószínűségi változók. Számítsuk ki X és $-2X + Y$ korrelációs együtthatóját.

Korrelációs együttható: példa

Tegyük fel, hogy X és Y **független**, 4 szórású Poisson-eloszlású valószínűségi változók. Számítsuk ki X és $-2X+Y$ korrelációs együtthatóját.

$$R(X,-2X+Y)=\frac{\operatorname{cov}(X,-2X+Y)}{D(X)D(-2X+Y)}=\frac{(-2)\cdot\operatorname{cov}(X,X)+\operatorname{cov}(X,Y)}{D(X)D(X+Y)}=$$

$$=\frac{-2D^2(X)}{D(X)D(-2X+Y)}=\frac{-2D^2(X)}{D(X)\cdot\sqrt{D^2(-2X)+D^2(Y)}}=$$

$$=\frac{-2D^2(X)}{D(X)\cdot\sqrt{((-2)^2+1)\cdot D^2(X)}}=\frac{(-2)\cdot 4^2}{4\cdot\sqrt{5}\cdot 4}=-\frac{2}{\sqrt{5}}=-0,89.$$

Korrelációs együttható: példa

Tegyük fel, hogy X és Y **független**, 4 szórású Poisson-eloszlású valószínűségi változók. Számítsuk ki X és $-2X+Y$ korrelációs együtthatóját.

$$R(X,-2X+Y)=\frac{\operatorname{cov}(X,-2X+Y)}{D(X)D(-2X+Y)}=\frac{(-2)\cdot\operatorname{cov}(X,X)+\operatorname{cov}(X,Y)}{D(X)D(X+Y)}=$$

$$=\frac{-2D^2(X)}{D(X)D(-2X+Y)}=\frac{-2D^2(X)}{D(X)\cdot\sqrt{D^2(-2X)+D^2(Y)}}=$$

$$=\frac{-2D^2(X)}{D(X)\cdot\sqrt{((-2)^2+1)\cdot D^2(X)}}=\frac{(-2)\cdot 4^2}{4\cdot\sqrt{5}\cdot 4}=-\frac{2}{\sqrt{5}}=-0,89.$$

Erős negatív korrelációt kaptunk.

Itt $\operatorname{cov}(X,Y)=0$ és a szórásnégyzetek összeadódnak, mert X és Y **függetlenek**. A függetlenségen kívül valójában elég lett volna annyit feltenni, hogy **azonos a szórásuk**, azaz $D(X)=D(Y)$ – sem a Poisson-eloszlástól, sem a 4-től nem függ a végeredmény. Azt is használtuk, hogy

$$D^2(cX)=c^2D^2(X)\quad\Leftrightarrow\quad D(cX)=|c|D(X).$$

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- tengerparton töltött idő és egészség:

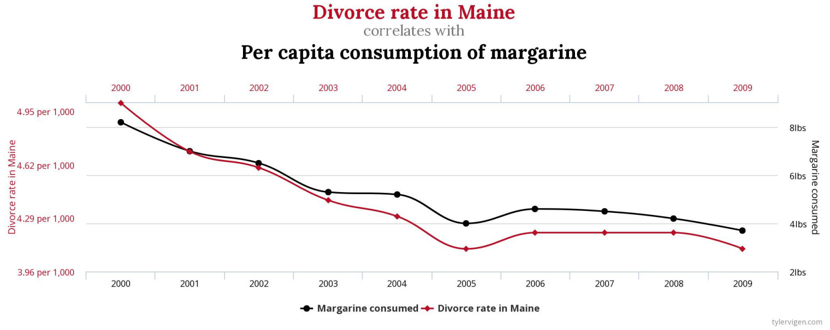
Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- tengerparton töltött idő és egészség:
  ha van is pozitív korreláció, **nem biztos, hogy van ok-okozati összefüggés**, a tengerparton töltött idő összefügg az anyagi helyzettel, ami az egészséggel, de csak a tengerparttól nem biztos, hogy egészséges lesz valaki, illetve aki beteg, kevésbé megy a tengerpartra
- a válások aránya Maine államban és a fejenkénti margarinfogyasztás az USA-ban:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- tengerparton töltött idő és egészség:
  ha van is pozitív korreláció, **nem biztos, hogy van ok-okozati összefüggés**, a tengerparton töltött idő összefügg az anyagi helyzettel, ami az egészséggel, de csak a tengerparttól nem biztos, hogy egészséges lesz valaki, illetve aki beteg, kevésbé megy a tengerpartra
- a válások aránya Maine államban és a fejenkénti margarinfogyasztás az USA-ban:van pozitív korreláció ($R = 0,9926$), de **feltehetően nincs ok-okozati összefüggés** (forrás és további példák:
  `http://tylervigen.com/spurious-correlations`)

Korreláció és ok-okozat



A válások aránya Maine államban és a fejenkénti margarinfogyasztás az USA-ban, korrelációs együttható: $0,9926$

`http://tylervigen.com/spurious-correlations`

„Big data” analízis: 200-300 mennyiség között könnyen található néhány olyan pár, amik ok-okozati összefüggés nélkül is nagy pozitív korrelációval rendelkeznek, de olyanok is, amik között valós összefüggés van $\rightarrow$ mindez alaposabb vizsgálatot igényel.

Korrelálatlanság

Ha az X, Y valószínűségi változók **kovarianciája** 0, akkor azt mondjuk, hogy X és Y **korrelálatlanok**. Mi ennek a kapcsolata a **függetlenséggel**?

$\boxed{X \text{ és } Y \textbf{ függetlenek}}$ $\boxed{X \text{ és } Y \textbf{ korrelálatlanok}}$

Korrelálatlanság

Ha az X, Y valószínűségi változók **kovarianciája** 0, akkor azt mondjuk, hogy X és Y **korrelálatlanok**. Mi ennek a kapcsolata a **függetlenséggel**?

X és Y **függetlenek** $\Rightarrow$ $\not\Leftarrow$ X és Y **korrelálatlanok**

Legyen X és Y két független, szabályos kockadobás eredménye.

$U = X + Y$ **az összeg**

$V = X - Y$ **a különbség**

$\mathrm{cov}(U, V) = \mathrm{cov}(X + Y, X - Y) = D^2(X) - \mathrm{cov}(X, Y) +$
$+\mathrm{cov}(X, Y) - D^2(Y) = 0 \Rightarrow X$ és Y **korrelálatlanok**

Ugyanakkor U és V **nem függetlenek**, például mert

Korrelálatlanság

Ha az X, Y valószínűségi változók **kovarianciája** 0, akkor azt mondjuk, hogy X és Y **korrelálatlanok**. Mi ennek a kapcsolata a **függetlenséggel**?

X és Y **függetlenek** $\Rightarrow$ $\not\Leftarrow$ X és Y **korrelálatlanok**

Legyen X és Y két független, szabályos kockadobás eredménye.

$U = X + Y$ **az összeg**

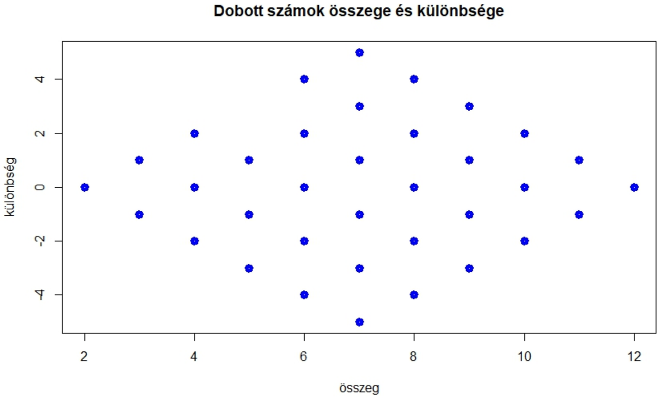
$V = X - Y$ **a különbség**

$\mathrm{cov}(U, V) = \mathrm{cov}(X + Y, X - Y) = D^2(X) - \mathrm{cov}(X, Y) +$
$+\mathrm{cov}(X, Y) - D^2(Y) = 0 \Rightarrow X$ és Y **korrelálatlanok**

Ugyanakkor U és V **nem függetlenek**, például mert

$$0 = \mathbb{P}(U = 11, V = 0) \neq \mathbb{P}(U = 11) \cdot \mathbb{P}(V = 0) = \frac{2}{36} \cdot \frac{1}{6}$$

Korrelálatlanság: példa



A dobott számok **különbségének** $(X - Y)$ és a dobott számok **összegének** $(X + Y)$ együttes előfordulása 100 megfigyelésből. Kovariancia: **0**, de $X + Y$ és $X - Y$ **nem függetlenek**.