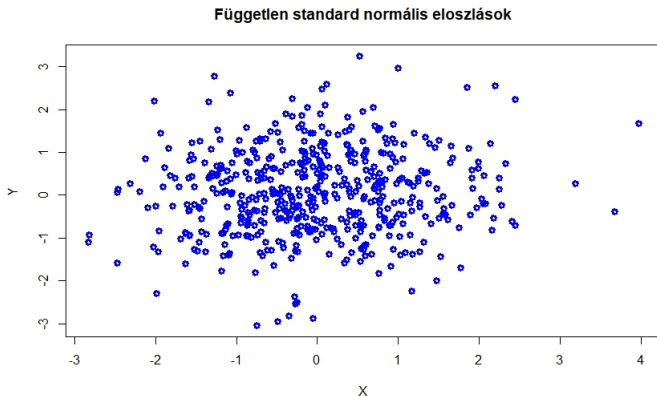


Kovariancia és korrelációs együttható (6. előadás)

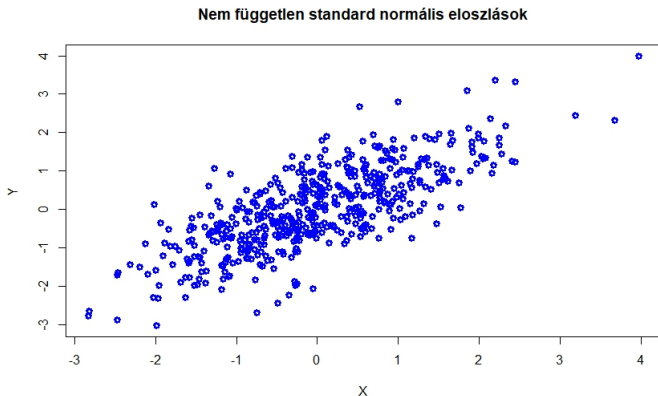
- két valószínűségi változó lehet
 - ▶ **független**: például két találmóra választott ember jövedelme, vagy,
 - ▶ **nem független**: például egy találmóra választott ember jövedelme most, illetve fél év múlva
- az **összefüggőség mértéke** különböző lehet:
 - ▶ egy találmóra választott felnőtt életkora és jövedelme „erősen összefüggő”, a fiataloké és időseké általában alacsonyabb;
 - ▶ egy találmóra 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 **kovariancia** 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 normális eloszlásúak. A koordináták között nincs kapcsolat: a kovariancia és a korrelációs együttható is **0** lesz.

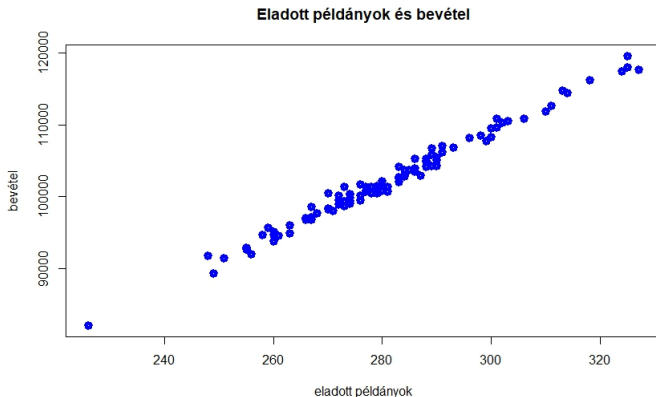
Pozitív korreláció



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 $X \sim \text{Poisson}(100)$ és $Y \sim \text{Poisson}(180)$ 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.

Házi feladat október 16., 10:15-ig

Annát minden nap a többitől függetlenül $p = 0,001$ valószínűséggel éri baleset. Legyen X az, hogy holnaptól hányadik nap az első, amikor baleset éri, Y pedig az, hogy az első balesete után hány nap telik el a másodikig. (azaz a második balesete holnaptól $X + Y$ nap múlva lesz.)

- Mennyi $X + Y$ várható értéke?
- Mennyi $X + Y$ szórása?
- Mennyi $X - Y$ várható értéke?
- Mennyi $X - Y$ szórása?

Házi feladat október 16., 10:15-ig, megoldás

X és Y is **geometriai eloszlású** $p = 0,001$ paraméterrel, továbbá tegyük fel, hogy **egymástól függetlenek**.

Ekkor $X + Y$ várható értéke:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \frac{1}{p} + \frac{1}{p} = 1000 + 1000 = 2000.$$

A függetlenséget felhasználva kapjuk $X + Y$ szórását:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{\frac{1-p}{p^2} + \frac{1-p}{p^2}} = \sqrt{\frac{2 \cdot 0,999}{0,001^2}} = 1413,5.$$

Házi feladat október 16., 10:15-ig, megoldás

X és Y is **geometriai eloszlású** $p = 0,001$ paraméterrel, továbbá tegyük fel, hogy **egymástól függetlenek**.

Ekkor $X + Y$ várható értéke:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \frac{1}{p} + \frac{1}{p} = 1000 + 1000 = 2000.$$

A függetlenséget felhasználva kapjuk $X + Y$ szórását:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{\frac{1-p}{p^2} + \frac{1-p}{p^2}} = \sqrt{\frac{2 \cdot 0,999}{0,001^2}} = 1413,5.$$

Hasonlóképpen $X - Y$ várható értéke:

$$\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = \frac{1}{p} - \frac{1}{p} = 1000 - 1000 = 0.$$

A függetlenséget felhasználva kapjuk $X - Y$ szórását:

$$D(X - Y) = \sqrt{D^2(X) + D^2(-Y)} = \sqrt{\frac{1-p}{p^2} + \frac{1-p}{p^2}} = \sqrt{\frac{2 \cdot 0,999}{0,001^2}} = 1413,5.$$

Házi feladat október 16., 10:15-ig, megoldás

```
> x<-rgeom(500, prob=0.001)    500 elemű minta geometriai eloszlásból  
> y<-rgeom(500, prob=0.001)    500 elemű minta geometriai eloszlásból  
> hist(x, col="orange", xlab="értékek", ylab="gyakoriságok",  
main="Első baleset hisztogramja")  
> lines(abline(v=1000, lwd="5", col="blue"))
```

Házi feladat október 16., 10:15-ig, megoldás

```
> hist(x+y, col="#6db7f7", xlab="értékek", ylab="gyakoriságok",  
main="Az összeg hisztogramja")
```

```
> lines(abline(v=2000, lwd="5", col="blue"))
```

```
> mean(x+y)       $\mathbb{E}(X + Y) = 2000$ 
```

```
[1] 2045.012
```

```
> sd(x+y)
```

```
[1] 1395.061       $D(X + Y) = 1413,5$ 
```

```
> hist(x-y, col="#6db7f7", xlab="értékek", ylab="gyakoriságok",  
main="A különbség hisztogramja")
```

```
> lines(abline(v=0, lwd="5", col="blue"))
```

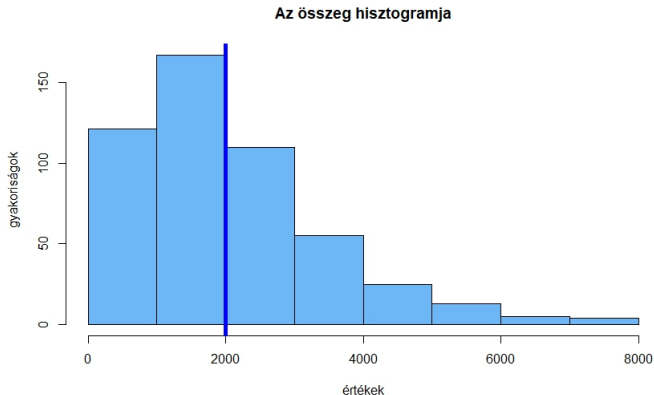
```
> mean(x-y)
```

```
[1] 9.452       $\mathbb{E}(X - Y) = 0$ 
```

```
> sd(x-y)
```

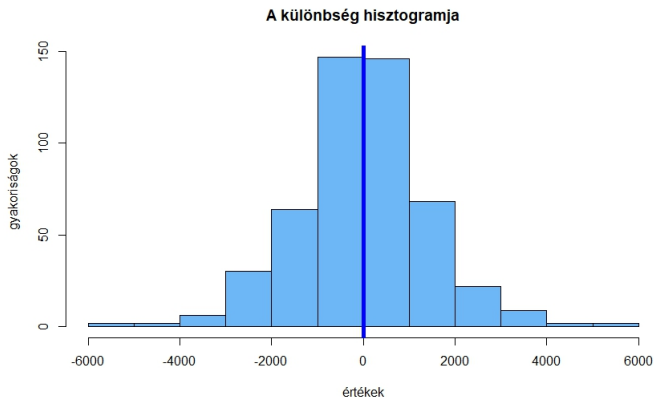
```
[1] 1450.261       $D(X - Y) = 1413,5$ 
```

Házi feladat október 16., 10:15-ig, megoldás



Histogram 500 elemű mintából, ahol az eloszlás $X + Y$ eloszlása úgy, hogy X és Y függetlenek, geometriai eloszlásúak $p = 0,001$ paraméterrel ($\mathbb{E}(X + Y) = 2000, D(X + Y) = 1413,5$)

Házi feladat október 16., 10:15-ig, megoldás



Histogram 500 elemű mintából, ahol az eloszlás $X - Y$ eloszlása úgy, hogy X és Y függetlenek, geometriai eloszlásúak $p = 0,001$ paraméterrel ($\mathbb{E}(X - Y) = 0$, $D(X - Y) = 1413,5$)

A kovariancia

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

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

A kovariancia

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

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

- **A kovariancia kiszámítása:**

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

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

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

A kovariancia tulajdonságai

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

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

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

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

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

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

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

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 $\text{cov}(X + Y, 300X + 400Y)$?

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 $\text{cov}(X + Y, 300X + 400Y)$?

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

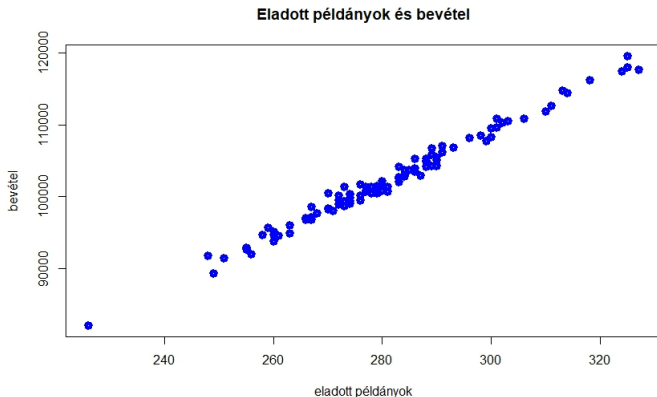
$$\begin{aligned}\text{cov}(X + Y, 300X + 400Y) &\stackrel{(a)}{=} \text{cov}(X, 300X) + \text{cov}(X, 400Y) + \\ &\quad + \text{cov}(Y, 300X) + \text{cov}(Y, 400Y) = \\ &\stackrel{(a,b)}{=} 300 \cdot \text{cov}(X, X) + 400 \cdot \text{cov}(Y, Y) = \\ &\stackrel{(b)}{=} 300D^2(X) + 400D^2(Y) = \\ &\stackrel{(c)}{=}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

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: $\text{cov}(X + Y, 300X + 400Y) = 102000$.

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 : példányszám Y : bevétel forintban Z : bevétel ezer forintban,

akkor

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

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**:

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

Korrelációs együttható: definíció

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

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

- **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

$$R(X, aX + b) = 1 \quad \text{és} \quad R(X, -aX + b) = -1.$$

- 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 érhetőek 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 **bevételek** a korrelációs együtthatója?

$$\begin{aligned} R(X + Y, 300X + 400Y) &= \frac{\text{cov}(X + Y, 300X + 400Y)}{D(X + Y)D(300X + 400Y)} = \\ &= \frac{102000}{D(X + Y)D(300X + 400Y)} \end{aligned}$$

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:

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

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:

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

A bevétel szórása:

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

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:

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

A bevétel szórása:

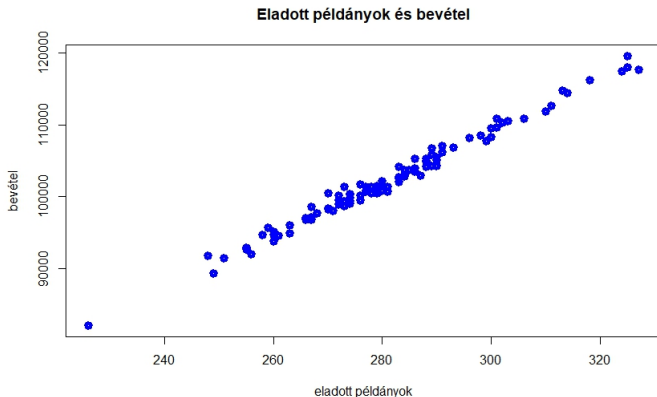
$$\begin{aligned} D(300X + 400Y) &= \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = \mathbf{6148,17}. \end{aligned}$$

Ezek alapján a korrelációs együttható:

$$\begin{aligned} R(X + Y, 300X + 400Y) &= \frac{\text{cov}(X + Y, 300X + 400Y)}{D(X + Y)D(300X + 400Y)} = \\ &= \frac{\mathbf{102000}}{16,73 \cdot 6148,17} = \mathbf{0,9915}. \end{aligned}$$

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?

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

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

$$\begin{aligned} D(300X + 4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 4000^2 \cdot 180} = 53749,42; \end{aligned}$$

$$\begin{aligned} R(X + Y, 300X + 4000Y) &= \frac{\text{cov}(X + Y, 300X + 4000Y)}{D(X + Y)D(300X + 4000Y)} = \frac{750000}{16,73 \cdot 53749,42} \\ &= 0,083. \end{aligned}$$

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?

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

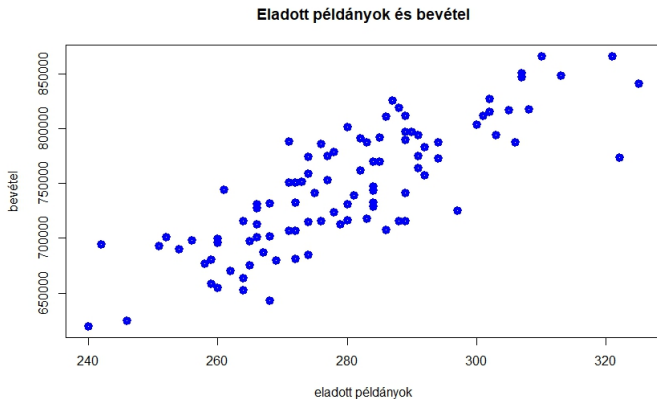
$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = 16,73;$$

$$\begin{aligned} D(300X + 4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 4000^2 \cdot 180} = 53749,42; \end{aligned}$$

$$\begin{aligned} R(X + Y, 300X + 4000Y) &= \frac{\text{cov}(X + Y, 300X + 4000Y)}{D(X + Y)D(300X + 4000Y)} = \frac{750000}{16,73 \cdot 53749,42} \\ &= 0,083. \end{aligned}$$

A korrelációs együttható értéke majdnem 0, azaz nincs jelentős korreláció az eladott példányok száma és a bevétel között, ha az újságok ára nagyon eltérő.

Korrelációs együttható: példa



A bevétel ($300X + 4000Y$) és az eladott példányszám ($X + Y$) együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 750000, korrelációs együttható: 0,083.

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.

$$\begin{aligned}R(X, -2X + Y) &= \frac{\text{cov}(X, -2X + Y)}{D(X)D(-2X + Y)} = \frac{(-2) \cdot \text{cov}(X, X) + \text{cov}(X, Y)}{D(X)D(X + Y)} = \\&= \frac{-2D^2(X)}{D(X)D(-2X + Y)} = \frac{-2D^2(X)}{D(X) \cdot \sqrt{D^2(-2X) + D^2(Y)}} = \\&= \frac{-2D^2(X)}{D(X) \cdot \sqrt{((-2)^2 + 1) \cdot D^2(X)}} = \frac{(-2) \cdot 4^2}{4 \cdot \sqrt{5} \cdot 4} = -\frac{2}{\sqrt{5}} = -0,89.\end{aligned}$$

Erős negatív korrelációt kaptunk.

Itt $\text{cov}(X, Y) = 0$, mert X és Y **függetlenek**. A függetlenséget az összeg szórásának kiszámításakor is használtuk, viszont X -ről és Y -ről ezen 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

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

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:

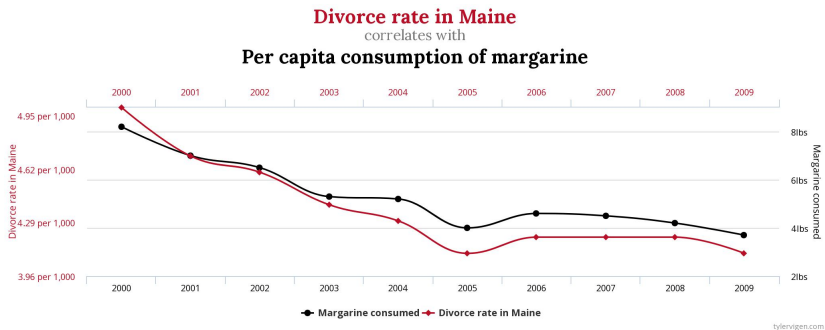
Korreláció és ok-okozat

- 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 margarin fogyasztás 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 margarin fogyasztás 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 → mindez alaposabb vizsgálatot igényel.

Korrelátlanság

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

X és Y **függetlenek**

X és Y **korrelátlanak**

Korrelátlanság

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

X és Y **függetlenek**



X és Y **korrelátlanak**

Legyen X és Y két független, szabályos kockadobás eredménye.

$U = X + Y$ az **összeg**

$V = X - Y$ a **különbség**

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

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

Korrelátlanság

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

X és Y **függetlenek**



X és Y **korrelátlanak**

Legyen X és Y két független, szabályos kockadobás eredménye.

$U = X + Y$ az **összeg**

$V = X - Y$ a **különbség**

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

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**.

Házi feladat november 6., 10:15-ig

Kétszer dobunk szabályos dobókockával. Legyen X az első dobás, Y a második.

Számítsuk ki X és $rX + Y$ korrelációs együtthatóját tetszőleges r esetén, és ábrázoljuk ezt r függvényében (kézzel készített ábra is megfelelő).

Milyen r -ek esetén lesz pozitív, negatív, illetve 0 a korrelációs együttható?