

Statisztika

Politológus képzés

Daróczi Gergely

Politológia Tanszék

2012. február 20.



PÁZMÁNY PÉTER
KATOLIKUS EGYETEM
Bölcsészettudományi Kar

1 A mérési hiba

- Megbízhatóság és érvényesség

2 A kutatás megtervezése

- A kutatás lehetséges céljai
- A kutatás egységei
- Ökológiai tévkövetkeztetés
- Változók és mérési szintek

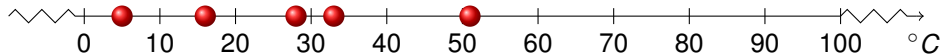
3 Mérési szintek

4 A változók közötti kapcsolatról

- Grafikus példák
- A relációk típusai
- További példák

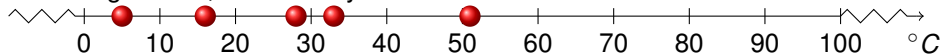
5 A kutatás főbb fázisai

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:



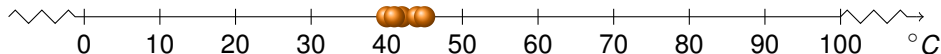
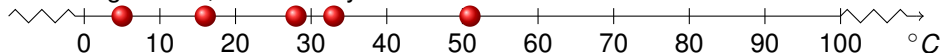
Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

Nem megbízható, nem érvényes mérés:



Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

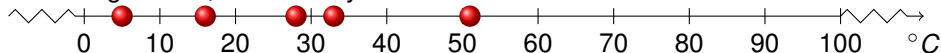
Nem megbízható, nem érvényes mérés:



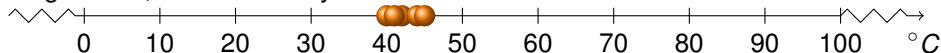
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

Nem megbízható, nem érvényes mérés:



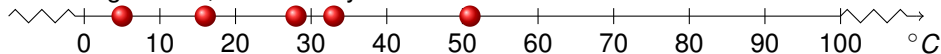
Megbízható, de nem érvényes mérés:



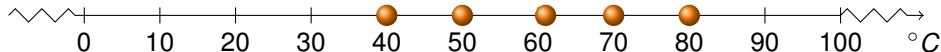
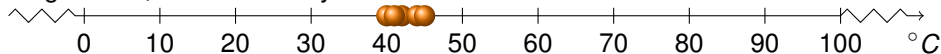
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

Nem megbízható, nem érvényes mérés:



Megbízható, de nem érvényes mérés:



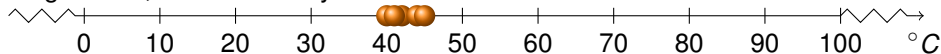
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

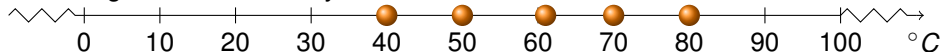
Nem megbízható, nem érvényes mérés:



Megbízható, de nem érvényes mérés:



Nem megbízható, de érvényes mérés:



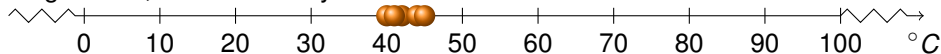
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

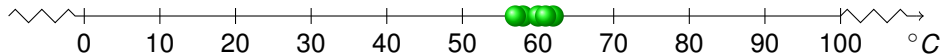
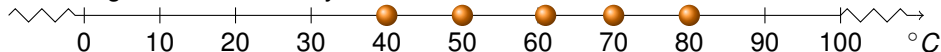
Nem megbízható, nem érvényes mérés:



Megbízható, de nem érvényes mérés:



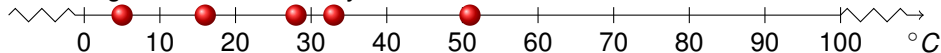
Nem megbízható, de érvényes mérés:



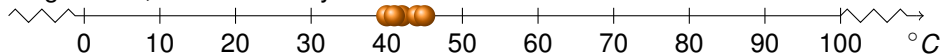
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

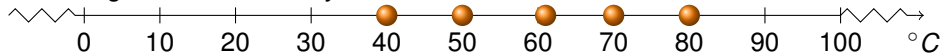
Nem megbízható, nem érvényes mérés:



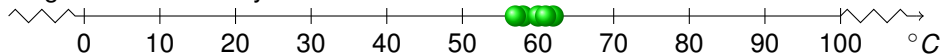
Megbízható, de nem érvényes mérés:



Nem megbízható, de érvényes mérés:



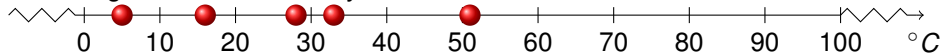
Megbízható és érvényes mérés:



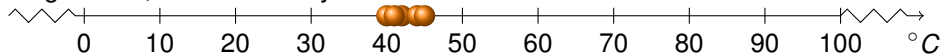
Megbízhatóság és érvényesség

Méréseket végeztünk egy vödör 60°C hőmérsékletű vízben:

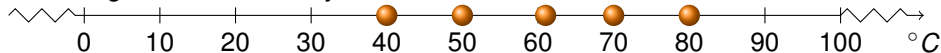
Nem megbízható, nem érvényes mérés:



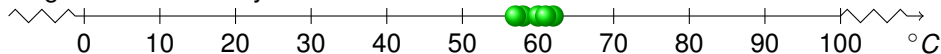
Megbízható, de nem érvényes mérés:



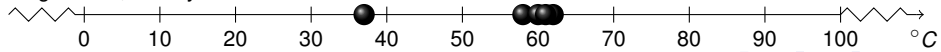
Nem megbízható, de érvényes mérés:



Megbízható és érvényes mérés:

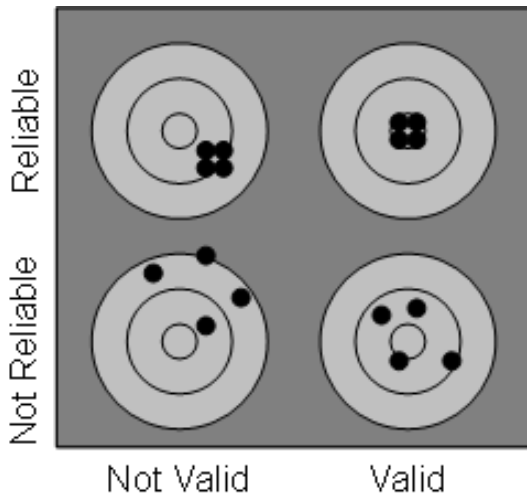


Megbízható, érvényes?



Megbízhatóság és érvényesség

Újratöltve

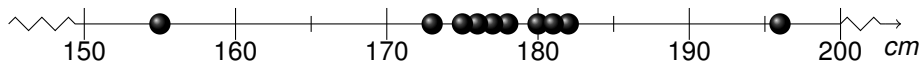


Forrás: ccustomerthink.com

Megbízhatóság és érvényesség

A társadalomtudományban

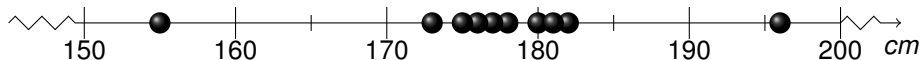
10 hallgat magasságát mértük:



Megbízhatóság és érvényesség

A társadalomtudományban

10 hallgat magasságát mértük:



Forrás: <http://www.wallpaperhere.com/>

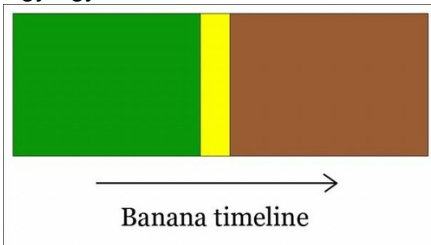
A következő kérdések közül melyek lehetnek megfelelő kutatási témák?

- Hány hajléktalan ember él a fővárosban?
- Melyek a leginkább elterjedt öngyilkossági eljárások?
- Milyen színű a banán?

A kutatás lehetséges céljai

Értelmetlennek tűnő kutatási témák

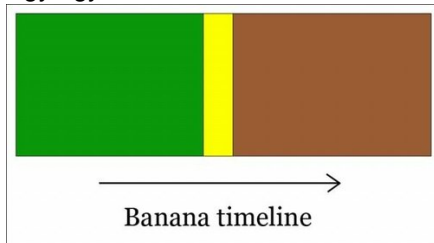
Egy egyszerű idősor:



A kutatás lehetséges céljai

Értelmetlennek tűnő kutatási témák

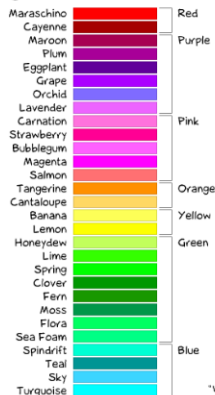
Egy egyszerű idősor:



Egy valódi adatfelvétel (N=225.000)
adatai alapján:

Color names if
you're a girl...

Color names if
you're a guy...



Doghouse Diaries
"We take no as an answer."

A következő kérdések közül melyek lehetnek megfelelő kutatási témák?

- Hány hajléktalan ember él a fővárosban?
- Melyek a leginkább elterjedt öngyilkossági eljárások?
- Milyen színű a banán?
- Mennyibe kerül egy vödör alma?
- Milyen kapcsolat állhat fenn az emberek hajszíne és testsúlya között?
- Melyik a legolvasottab könyv a világon?

Mik lehetnek a témaválasztás fő faktorai?

Definíció: feltáró, leíró és magyarázó kutatások

A kutatás egységei

Mintavételi egység, megfigyelési egység, eset

Milyen alanyokat vizsgálunk az alábbiakban?

- Az idősebb emberek jobban félnek a bűnözőktől, mint a fiatalabbak.
- A gazdasági fejlődés valóban a várható életkor csökkenésével jár együtt?
- Melyik a legnagyobb magyar focidrukker csapat?
- Minél hosszabb az eljegyzés, annál hosszab a házasság.
- Mely foglalkozás bír a legnagyobb presztízzsel?

A megfigyelési egység minden esetben egybeesik a mintavételi egységgel?

A Turisztikai Kutató Központ adatai alapján Budapesten 2010-ben 2 millió turista fordult meg. A jelentés azt is tartalmazta, miszerint a Balatonnál ugyanezen évben 2,5 millió ember fordult meg.

Állítható-e a fentiek alapján, hogy...

- Budapest lakó kevésbé barátságos, mint a Balaton körül élő emberek?
- Budapesten kevesebb a látványosság, mint a Balaton környékén?
- a tó kedveltebb turista-célpont?
- a turisták szerint a Balatonnál jobb időzni, mint Budapesten?
- a Balaton környékén több turisztikai adót fizettek, mint a fővárosban?

Egyáltalán: mi mondható el a fenti adatsor alapján?

Emile Durkheim (1897): *Le Suicide*

- A férfiak hajlamosabbak öngyilkosságot elkövetni, mint a nők.
- Az öngyilkosság jellemzőbb az özvegy, elvált vagy házasságon kívül élőkre, mint a házasokra.
- Gyerekes emberek körében kisebb az öngyilkossági arány.
- Az öngyilkossági ráta a protestánsoknál magasabb, mint a katolikusoknál vagy a zsidóknál. Ez annak köszönhető, hogy a katolikus országokban sokkal nagyobb a társadalmi integráltság, mint a protestánsoknál.
- Az öngyilkossági arány kimagasló a katonák körében.

Tétel

Annak a veszélye, hogy egyes emberekről mint elemzési egységekről fogalmazunk meg állításokat csoportokon végzett megfigyelések alapján.



Változók és mérési szintek

Egy konkrét példa

Hány éves Ön?

- >18
- 19-24
- 25-30
- 31-40
- 41-100
- 100<

Mutasd be a fenti példában a változót és annak attribútumait!

A változók típusai

- Belső változók
 - Függő változó
 - kvalitatív változók
 - kvantitatív változók
 - Független változó
 - kvalitatív változók
 - kvantitatív változók

Küldő változók

- Kontroll-változó
- Egyéb változók

Hozz 1-1 példát a fentebbi típusokra!

Kvalitatív változók:

- Nominális: egymást kizáró, a teljes válasz-univerzumot lefedő kategóriák
- Ordinális: u.a., mint fentebb + a kategóriák sorbarendeázhetők

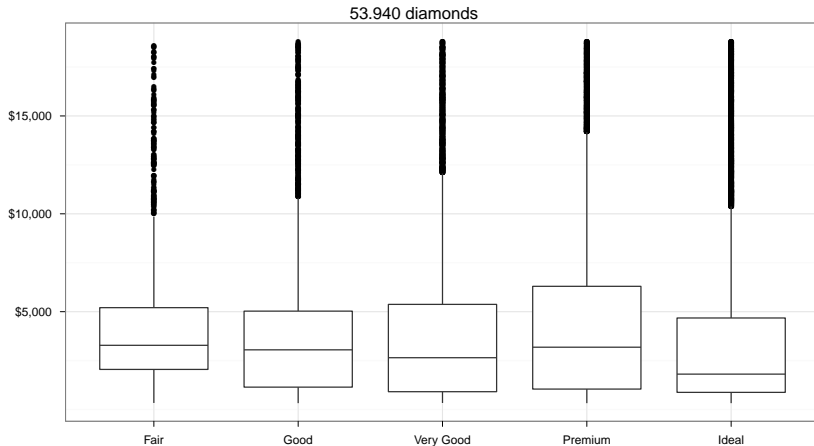
	Nominális	Ordinális	Intervallum	Arány
Klasszifikáció	X	X	X	X
Sorrend		X	X	X
Egyenlő távolságok			X	X
Nullpont				X

Kvantitatív változók:

- Intervallum: u.a., mint fentebb + a kategóriák egyenlő távolságra találhatók
- Arány: u.a., mint fentebb + nullpont

A változók közötti kapcsolatról

Egy grafikus ábrázolás



```
ggplot(diamonds, aes(cut, price)) + geom_boxplot() + xlab('') + ylab('') +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

Az adatbázis szerkezete

ggplot2/diamonds

Prices of 50,000 round cut diamonds

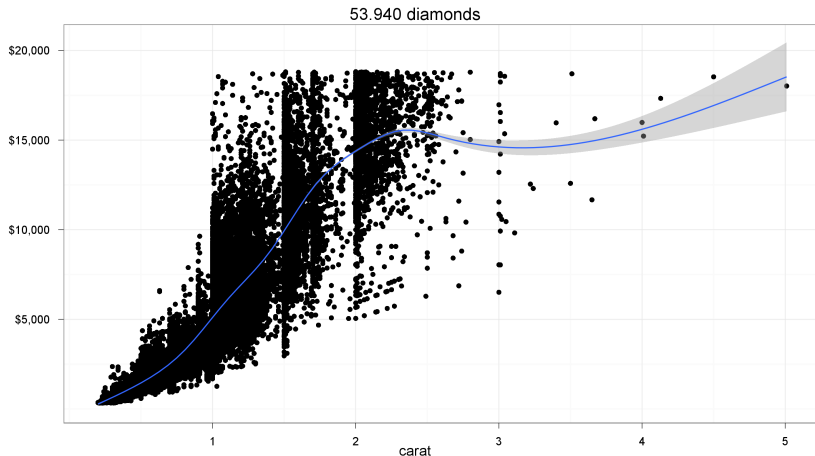
Description:

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

- price. price in US dollars (\\$326--\\$18,823)
- carat. weight of the diamond (0.2--5.01)
- cut. quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- colour. diamond colour, from J (worst) to D (best)
- clarity. a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
- x. length in mm (0--10.74)
- y. width in mm (0--58.9)
- z. depth in mm (0--31.8)
- depth. total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43--79)
- table. width of top of diamond relative to widest point (43--95)

A változók közötti kapcsolatról

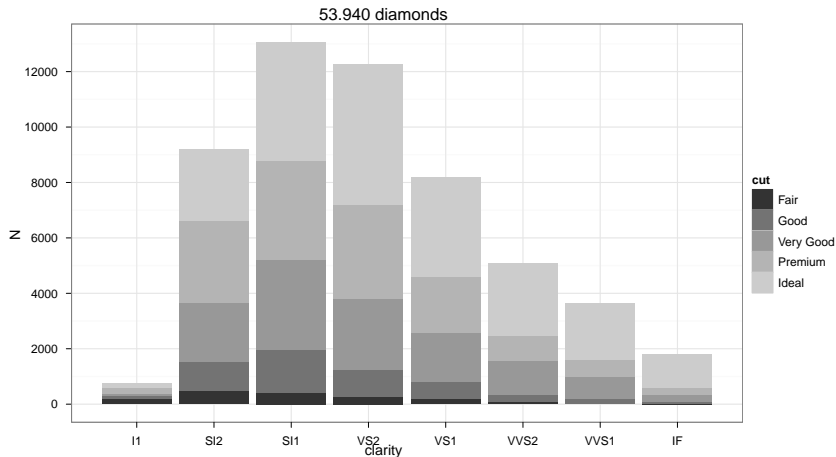
Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price)) + geom_point() + geom_smooth() + ylab('') +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

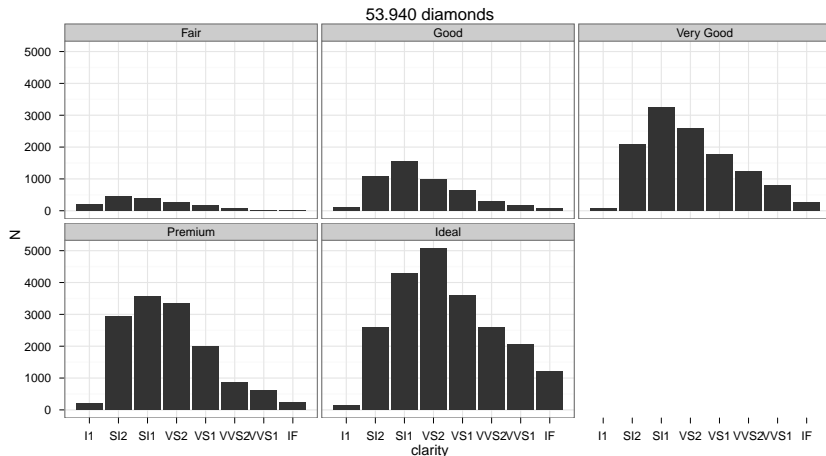
Egy grafikus ábrázolás



```
ggplot(diamonds, aes(clarity, fill=cut)) + geom_bar() + ylab("N") +  
theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

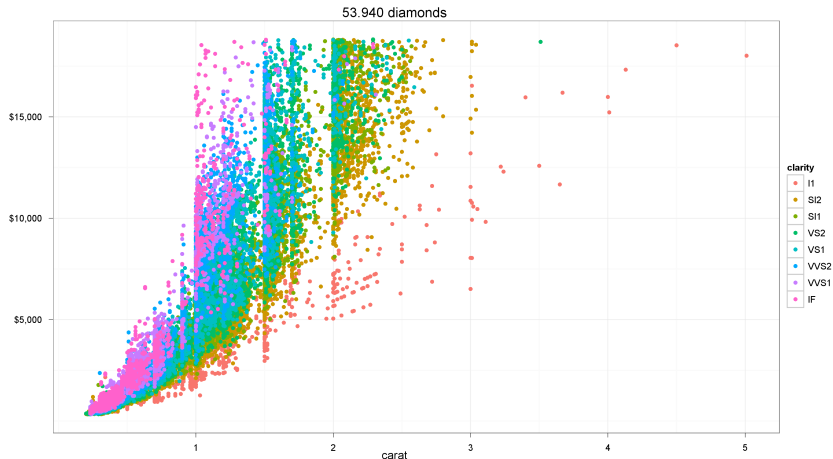
Egy grafikus ábrázolás



```
ggplot(diamonds, aes(clarity)) + geom_bar() + ylab("N") + facet_wrap(~ cut) +  
theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

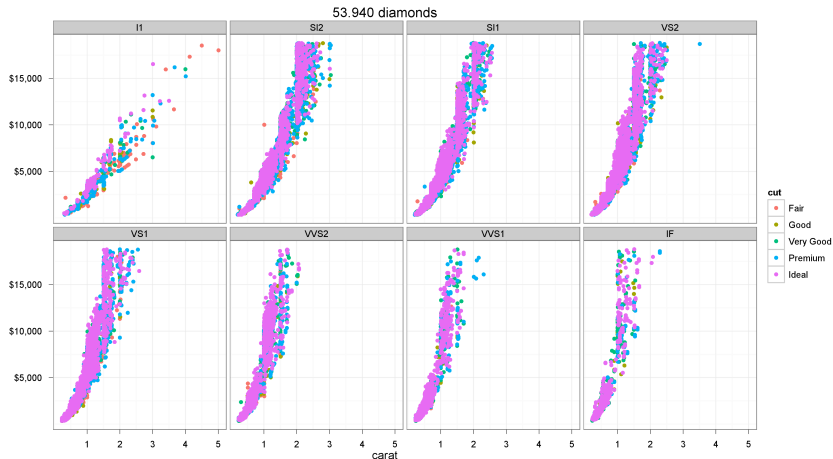
Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price, color=clarity)) + geom_point() + ylab("") +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

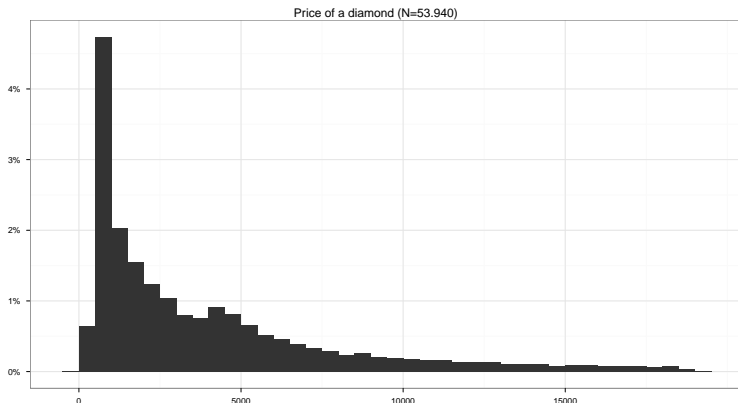
Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price, color=cut)) + geom_point() + ylab("") + facet_wrap(~ clarity, nr=2) + scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

Egy grafikus ábrázolás



Online kutatást végeztünk az Interneten gyémántot árusító emberek között. A megkérdezettek a fentebbi válaszokat adták.

Mit gondolhatunk a kutatás érvényességéről és megbízhatóságáról?

A változók közötti kapcsolat lehetséges típusai

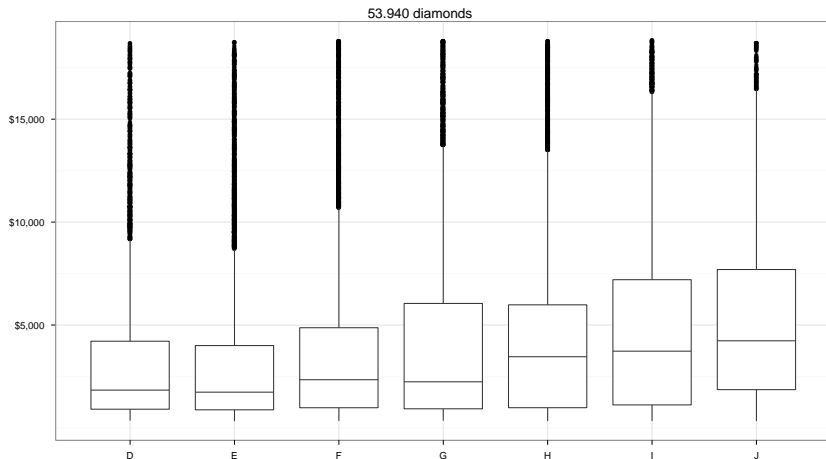
Lehetséges kapcsolat két változó között:

- együttjárás,
- korreláció ,
- hamis/látszólagos összefüggés,
- hatás,
- a hatás iránya,
- **okság.**

Statisztikai szignifikáns...

A változók közötti kapcsolatról

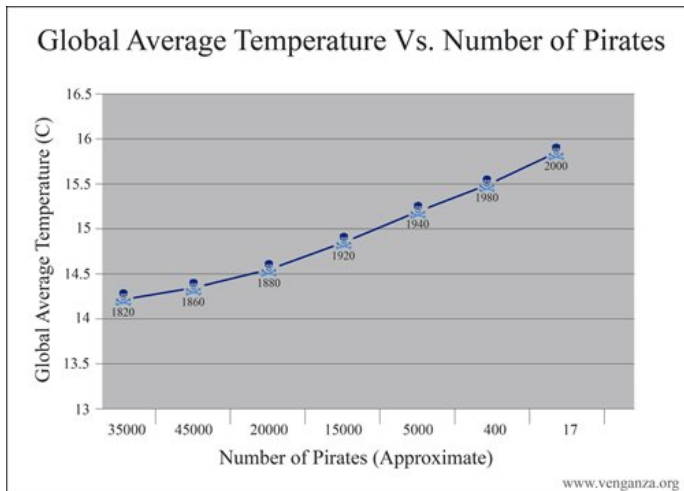
A hatás iránya



```
ggplot(diamonds, aes(color, price)) + geom_boxplot() + xlab("") + ylab("") +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

A változók közötti kapcsolatról

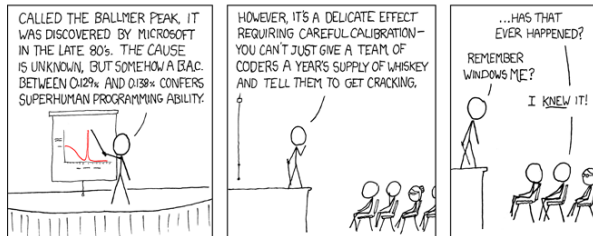
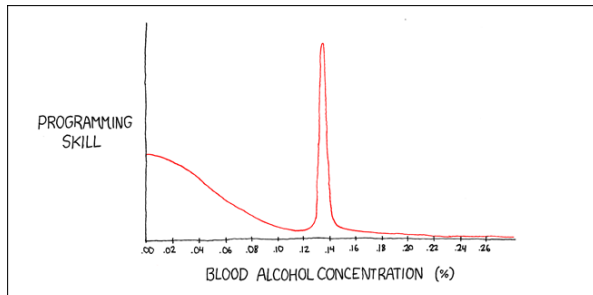
Magas korreláció



A korrelációs együttható magas értéke nem cáfolható. És?

A változók közötti kapcsolatról

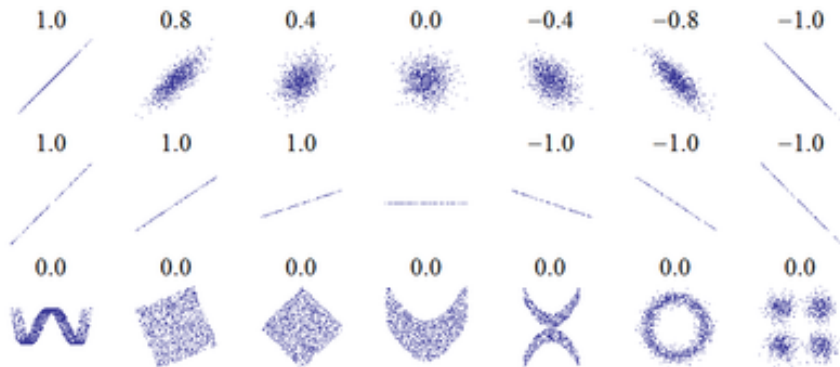
Korrelálatlan változók



Forrás: <http://xkcd.com/323/>

A változók közötti kapcsolatról

Korreláció



Pozitív („egyeses arányosság”: $R = 1$), negatív („fordított arányosság”: $R = -1$), lineáris, négyzetes ... és korrelálatlan ($R = 0$) kapcsolat
R: korrelációs együttható

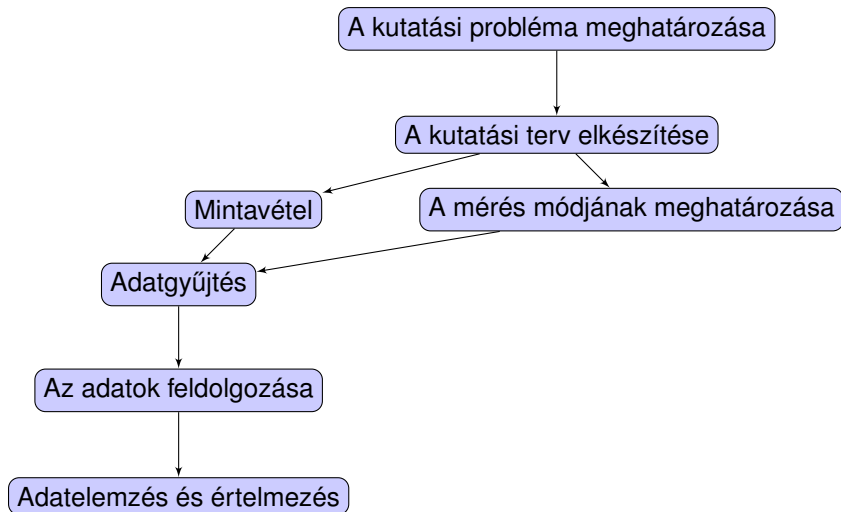
Milyen változók, milyen mérési szintek feleltethetők meg az alábbi kérdéseknek/kijelentéseknek?

- Az idősebb emberek jobban félnek a bűnözőktől, mint a fiatalabbak.
- A gazdasági fejlődés valóban a várható életkor csökkenésével jár együtt?
- Melyik a legnagyobb magyar focidrukker csapat?
- Minél hosszabb az eljegyzés, annál hosszab a házasság.
- Mely foglalkozás bír a legnagyobb presztízzsel?

Attribútumok?

A kutatás főbb fázisai

Egy folyamatábra



Köszönöm a figyelmet!

Daróczi Gergely
daroczi.gergely@btk.ppke.hu