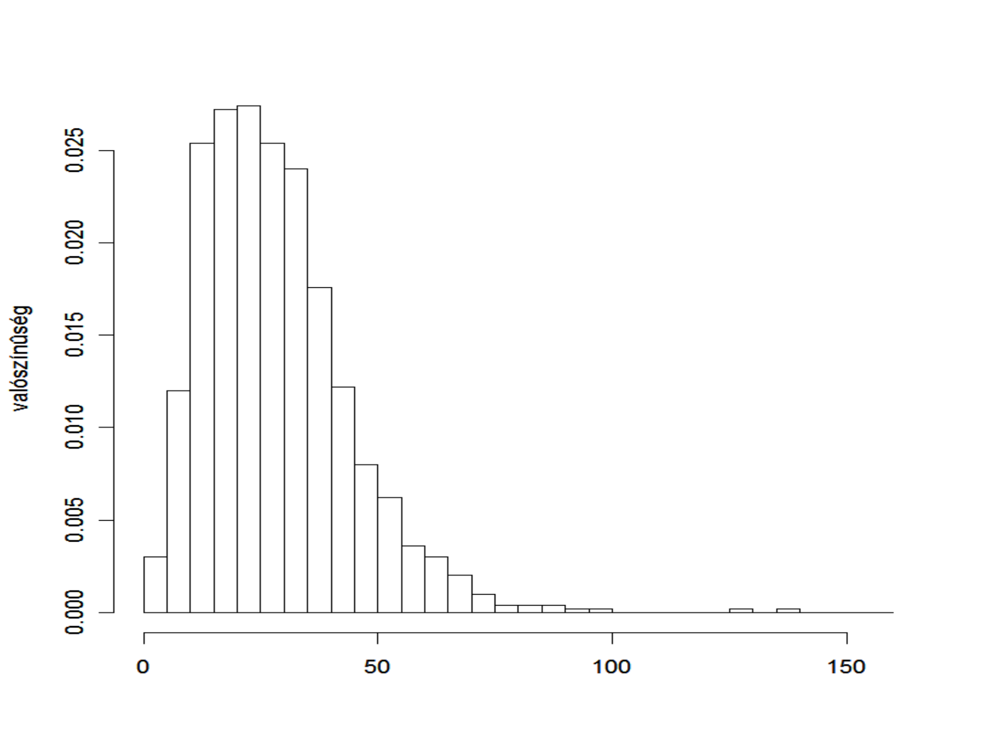
Khi-négyzet eloszlás

Statisztika II., 7. alkalom

A khi-négyzet eloszlás

A khi négyzet eloszlást (Pearson) leggyakrabban kategorikus adatok elemzésére használjuk.
N darab standard normális eloszlású változó négyzetes összegeként kapjuk a khi-négyzet eloszlást.
Jelölése: χ^2.



A khi-négyzet eloszlás

Általános képlete:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

A nominális változó K darab lehetséges értéke esetén a szabadságfok f=K-1

Illeszkedés vizsgálat (egy minta):

H0: változó tapasztalati eloszlása megfelel egy feltételezett eloszlásnak.

Példa (Varga, 2000)

	Koronás címer	Kádár címer	Rákosi címer	Kossuth címer	Összesen
Oi	280	48	12	36	376
Ei	94	94	94	94	376

H0: A négy címerfajtát az emberek egyenlő mértékben kedvelik.

$$\chi^2 = \sum_i \frac{(O-E)^2}{E} = \frac{(280-94)^2 + (48-94)^2 + (12-94)^2 + (36-94)^2}{94} = 497.87$$

f=4-1=3

A kritikus érték még 0.01-nél is csak 11.3, így 99%-os valószínűséggel, a négy címerfajtát nem egyelő mértékben kedvelik.

A khi-négyzet eloszlás

Két változó kapcsolata:

Kategorikus esetben a függetlenség és homogenitásvizsgálat a khi-négyzet eloszlás segítségével történik.

A próba szabadságfoka f=(K-1)(G-1), ahol K és G a két változó lehetséges értékeinek száma.

Függetlenségvizsgálat:

H0: Az oszlopoknak megfelelő gyakoriságok függetlenek a sorok gyakoriságaitól

Homogenitásvizsgálat:

H0: Az oszlopoknak és soroknak megfelelő gyakoriságok függnek egymástól.

Ehhez először a várt gyakoriságok kiszámítása szükséges.

(Megfigyelt gyakoriságok)			
	Y1	Y2	Össz.
X1	O11	O12	O1+
X2	O21	O22	O2+
Össz.	O+1	O+2	N

$$P(A \cap B) = P(A)P(B)$$

(Várt gyakoriságok)		
	Y1	Y1
X1	E11=O1+*O+1/N	E12= O1+*O+2/N
X2	E21= O2+*O+1/N	E22= O2+*O+2/N

A khi-négyzet eloszlás (függetlenségvizsgálat)

Nyugtató hatása enyhe depresszióra (Megfigyelt gyakoriságok)			
	Depressziós	Nem depressziós	Össz.
Nyugtató	335	76	411
Placebó	302	105	407
Össz.	637	181	818

H0: A gyógyulás szempontjából mindegy, hogy placebót kap a depressziós vagy nyugtatót.

Nyugtató hatása enyhe depresszióra (Várt gyakoriságok)		
	Depressziós	Nem depressziós
Nyugtató	637*411/818=320.1	181*411/818=90.9
Placebó	637*407/818=316.9	181*407/818=90.1

$$\chi^2 = \frac{(335-320.1)^2}{320.1} + \frac{(302-316.9)^2}{316.9} + \frac{(76-90.9)^2}{90.9} + \frac{(105-90.1)^2}{90.1} = 6.3$$

df=(2-1)(2-1)=1

p=0.0118

A khi-négyzet eloszlás (homogenitásvizsgálat)

A homogentásvizsgálat formailag ugyanúgy történik, mint a függetlenségvizsgálat.

Fontos különbség, hogy a kérdés nem az, hogy az egyik változó hatással van-e a másikra, hanem az, hogy a két változó hatással van-e egymásra.

Példák:

- Az, hogy mennyire ért egyet a bálnák megmentéséért folytatott küzdelemmel (abszolút, inkább igen, közepesen, inkább nem, egyáltalán nem) összefügg-e azzal, hogy mennyire ért egyet az esőerők megmentésével (abszolút, inkább igen, közepesen, inkább nem, egyáltalán nem)?
- A tréning hangulatával való elégedettség összefügg-e a tréning hatékonyságával való elégedettséggel?
- A kedvenc sport és a kedvenc nyaralási helyszín összefügg-e?

Homogenitásvizsgálat versus függetlenségvizsgálat

A következő mintavételi eljárásokat különböztethetjük meg:

Poisson
Multinomiális
Prospektive
Retrospektive
Random kísérlet

Függetlenségvizsgálat csak a Poisson és a multinomiális esetben elképzelhető, homogenitásvizsgálat minden esetben.

A khi-négyzet eloszlásból származtatható asszociációs mérőszámok

A khi-négyzet próba teszteli két változó függetlenségét (vagy egyik függését a másiktól).

Ha az érték nulla, függetlenek, de a kapcsolat erősségének a statisztika értéke nem jó mutatója, mert függvénye a mintanagyságnak, a szabadsági foknak, így egyéb mérőszámokat használunk az asszociáció erősségének kifejezésére. Az ideális mutató a korrelációhoz hasonlóan nulla és egy közötti értékeket vehetne fel.

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

A ϕ (phi) együttható értéke függetlenség esetén nulla.

2x2-es kontingencia táblázat esetén maximális értéke egy, ennél nagyobb táblázat esetén értéke túllépheti az egyet.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

A Pearson féle kontingencia (C) együttható értéke szintén nulla függetlenség esetén. Nulla és egy közötti balról zárt intervallumbeli értékeket vehet fel (egynél kisebb).

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

A Cramer féle V együttható függetlenség esetén a nulla értéket veszi fel, értékei 0 és 1 közt mozognak, mindkét szélsőértéket felvehetik. k a képletben a kisebb a két változó lehetséges értékek száma közül.