Nemparaméteres eljárások, binomiális-próba

Statisztika II., 6. alkalom

Binomiális eloszlás

A binomiális eloszlású változónak két értéke van (a kódolás 1 és 0 leggyakrabban)

Pl. beteg, nem beteg
férfi, nő
kezelt, nem kezelt

Az elosztást egyértelműen meghatározza az egyik érték aránya, valószínűségének meghatározása. Valószínűség hipotetikus populáció vagy kísérlet esetén.

Pl. $p_{\text{beteg}}=0.33$
$p_{\text{férfi}}=0.42$
$p_{\text{kezelt}}=0.57$

A változó átlagát a 1-essel kódolt alternatíva aránya adja. Jelölése: π . Másik alternatíva: $(1-\pi)$

Pl. 1,1,0,1,0,1 $$\pi = \frac{1+1+0+1+0+1}{6} = \frac{4}{6} = 0.66$$

Egy minta, valószínűség tesztelése

H0:Az minta átlaga alapján a populációban a valószínűségről való feltételezésünk tartható

H1: A feltevésünk a populációbeli valószínűségről nem tartható

Pl.43 egyetemistából 31 azt mondja első gyermekének fiút szeretne, 12 lányt.

H0: általában ugyanannyian akarnak fiút, mint lányt.

H1: nem ugyanannyian akarnak fiút, mint lányt

$$(\hat{\pi}) = 31/43 = 0.72$$

Kontingencia táblázat

A kontingencia táblázat az együttes gyakoriságokat foglalja össze.

Szamoai származású amerikaiak (1976-1981)			
	Érrendszeri halálok	Nem érrendszeri halálok	Össz.
Túlsúlyos	16	2045	2061
Nem túlsúlyos	7	1044	1051
Össz.	23	3089	3112

A sorok és az oszlopok különböző változók lehetséges értékeinek gyakoriságait foglalják össze.

Az egyes cellák az együttes gyakoriságokat tartalmazzák.

A kontingencia táblázat binomiális változók esetében 2x2-es (a sorösszegeket és az oszlopösszegeket nem számítva).

Kontingencia táblázat

Szamoai származású amerikaiak (1976-1981)			
	Érrendszeri halálok	Nem érrendszeri halálok	Össz.
Túlsúlyos	16	2045	2061
Nem túlsúlyos	7	1044	1051
Össz.	23	3089	3112

$(\textit{Túlsúlyos})\hat{\pi}_2 = 16/2061 = 0.0077$

$(\textit{Nem})\hat{\pi}_1 = 7/1051 = 0.0066$

$(\hat{\pi}_2 - \hat{\pi}_1) = 0.0011$

Esélyek (odds) az arányok helyett

Sokszor sokkal inkább célszerű az egyik alternatíva esélyét meghatározni a másikhoz képest a populáció eloszlásában az arány helyett az összes válasz tekintetében.

Pl. 0.05 az arányok közötti eltérés ($\hat{\pi}$)

0.45 és 0.5 esetében ez sokkal kevésbé jelentős, mint 0.05 és 0.1 esetében.

$$\omega = \pi / (1 - \pi)$$

A mintában:

$$\hat{\omega} = \hat{\pi} / (1 - \hat{\pi})$$

Depressziós esetek a nyugtató szedésekor:0.742.

Az esélye a depressziónak: 0.742/(1-0.742)=2.876 az 1-hez.

Az esély egy szám 0 és plusz végtelen között.

A 0.5-ös arány megegyezik az 1:1-hez eséllyel.

Ha az igen esélye $\hat{\omega}$, a nemé $1/\hat{\omega}$ és az igen aránya $\hat{\omega}/(\hat{\omega}+1)$

Esélyek arányai (Odds ratio)

Az esélyek hányadai releváns alternatívát nyújtanak két minta összehasonlítására.

Ha az esélyek a két populációban $\hat{\omega}_1$ és $\hat{\omega}_2$, az odds ratio az $\phi = \hat{\omega}_2 / \hat{\omega}_1$

Az esélyek hányada azt adja meg, hogy hányszor valószínűbb egy jelenség az egyik csoportban, mint a másikban.

Pl.

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

$$\phi = \hat{\omega}_2 / \hat{\omega}_1 = (302/105)/(335/76) = 105*335/(76*302) = 1.53$$

Azaz placebót szedve a 1.53-szor nagyobb az esélye a depressziónak, mint nyugtatót szedve.

Miért jó?

- A gyakorlatban ez sokkal stabilabb mutató, mint az arány.
- Retrospektív tanulmányoknál csak ez használható

Prospektív versus retrospektív vizsgálat

A függő és független változó az esetek nagy részében meghatározható.

Ha a független változó két értékénél egy-egy mintát veszünk vagy random módon a két csoportba elosztjuk a mintánkat, a vizsgálat prospektív.

Pl. Nyugtató és placebó.

Ha azonban a függő változó két értékére nézve veszünk egy-egy mintát, a vizsgálat retrospektív.

Ekkor számunkra releváns módon csak az odds ratiot lehet számolni.

Pl. Dohányzás és tüdőrák.

Az arányok vizsgálatánál fontos az, hogy honnan közelítünk (függő vagy független változó).

Az esély bárhonnan nézve 2 az 1-hez.

	Tüdőrák	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]