

# Neparametarski testovi: Hi-kvadrat test



**PREDAVANJE BR.6**

# Uvod



- P6: Testiranje pretpostavke o raspodjeli vjerovatnoća kategoričke varijable
- P7: Analiza dvodimenzionalnih kontigencijskih tabela

# Uvod



- Kategorička varijabla – koja sadrži određeni broj iscrpnih i uzajamno isključivih kategorija
- Iscrpnost: mora biti moguće da svakog člana svrstamo u neku od kategorija
- Isključivost: određeni član populacije može pripadati samo jednoj kategoriji
- Primjeri kategoričkih varijabli: bračni status, pol, obrazovanje itd.

# Uvod



- Empirijska raspodjela kategorične varijable - učestalost pojedinih kategorija varijable na uzorku.
- Raspodjela vjerovatnoća kategorične varijable - vjerovatnoće pojedinih kategorija varijable u populaciji.
- Vjerovatnoće kategorija u populaciji ( $\pi$ ) - ocijenjene proporcijom ( $p$ ) date kategorije na slučajnom uzorku.

# Uvod



- Razlike između aritmetičkih sredina – kvantitativni brojni podaci koji su normalno raspoređeni
- Ako su varijable kategoričke i podaci izraženi u frekvencijama, ili ako distribucija odstupa od normalne – hi kvadrat test
- Upotreba: ako želimo da utvrdimo da li neke dobijene (opažene) frekvencije odstupaju od očekivanih frekvencija pod određenom hipotezom

# $\chi^2$ test



$$\chi^2 = \sum \frac{(f_0 - f_t)^2}{f_t}$$

- $f_0$  – opažene frekvencije
- $f_t$  – očekivane (teorijske) frekvencije
- Testovi zasnovani na hi-kvadrat rasporedu obuhvataju niz problema koji se mogu odnositi na modalitete jednog ili više obilježja.
- Najčešće korišćeni: testovi oblika rasporeda i testovi nezavisnosti obilježja

# TEST RASPODELE – $\chi^2$



- Nulta hipoteza: Raspodjela vjerovatnoća kategoričke varijable u populaciji ima određenu strukturu.
  - Struktura vjerovatnoća može biti različita: uniformna (podjednaka vjerovatnoća svih kategorija) ili neka druga.
- Nulta hipoteza, ako se pretpostavlja da je raspodjela vjerovatnoća uniformna:  $\pi_1 = \pi_2 = \dots \pi_g$
- Nulta hipoteza za varijablu sa tri kategorije bi mogla glasiti i:  $\pi_1 = 0.2$ ;  $\pi_2 = 0.6$ ;  $\pi_3 = 0.2$ .

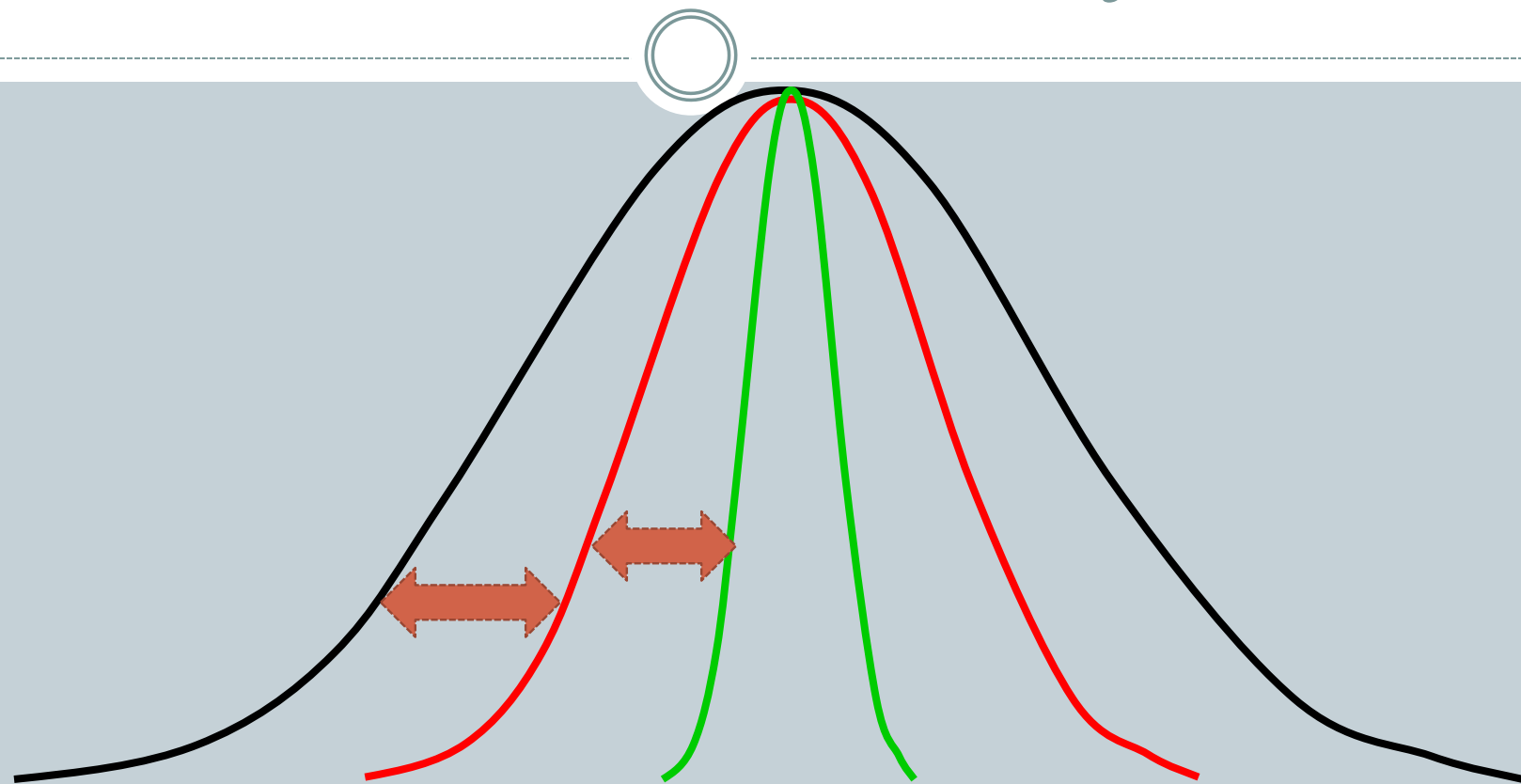
$$\chi^2$$



- Ho: nema razlika između empirijskih i teorijskih frekvenci
- Ako teorijske frekvence = uniformna distribucija
  - Ho: distribucija empirijskih frekvenci je uniformna
- Ako  $p > 0.05$  – ne odbacujemo Ho
  - Distribucija empirijskih frekvenci jeste uniformna
  - Svi odgovori ljekara podjednako zastupljeni!
- **NE MORA DISTRIBUCIJA BITI UNIFORMNA!**
- $\chi^2$  – test bilo koje distribucije
  - Npr. normalne!



# Normalnost distribucije



- Hi kvadrat test ( $\chi^2$ )
- Kolmogorov Smirnov test – ne moramo da poznajemo teorijsku distribuciju

# Primjer hi-kvadrat testa na jednom uzorku



1. Skup od 48 ljekara iznio je mišljenje o tome da li treba davati analgetike ženama na porođaju. Dobijeni su odgovori dati u sljedećoj tabeli. Da li odgovori pokazuju neko značajno odstupanje od onoga što bismo očekivali kad bi odgovori bili dati “nasumice”, tj. sasvim slučajno?

Nulta hipoteza: Nema razlike između dobijenih odgovora.

Kad bi odgovori bili dati slučajno, svaki bi imao jednaku vjerovatnoću, pa bi očekivali da se svaki odgovor pojavi isti broj puta.

	<b>Da</b>	<b>Ne znam</b>	<b>Ne</b>	<b>Ukupno</b>
<b>Opažene frekvencije</b>	26	12	10	48
<b>Teorijske frekvencije</b>	16	16	16	48

# Primjer hi-kvadrat testa na jednom uzorku



fo	ft	fo-ft	(fo-ft)^2	(fo-ft)^2/ft
26	16	10	100	6.25
12	16	-4	16	1
10	16	-6	36	2.25

hi-kvadrat

9.5

	Observed N	Expected N	Residual
1	26	16.0	10.0
2	12	16.0	-4.0
3	10	16.0	-6.0
Total	48		

	VAR00001
Chi-Square	9.500 <sup>a</sup>
df	2
Asymp. Sig.	.009

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 16.0.

- Primijetimo: kad ne bi postojala nikakva razlika između opaženih i teorijskih frekvencija, hi-kvadrat bi bio jedna nuli
- Tablica graničnih vrijednosti hi-kvadrat raspodjele data je u tabeli H u Dodatku knjige.
- granična vrijednosti  $\chi^2$  uz 2 stepena slobode (N-1) uz rizik greške 5% - 5.991. Dakle odbacujemo nultu hipotezu.
- SPSS: Analyze/Nonparametric Tests/Legacy Dialogs/Chi-square test

# Primjer hi-kvadrat testa na jednom uzorku



2. Pretpostavimo da je uzet slučajan uzorak od 2400 lica iz skupa od 960.000 nezaposlenih. Na nivou značajnosti 5%, da li se raspored nezaposlenih lica po školskoj spremi razlikuje od očekivanog (uniformnog) samo u granicama slučajnosti, odnosno, mogu li se ovih 2400 nezaposlenih lica smatrati kao uzorak iz skupa u kome su sva 4 modaliteta podjednako zastupljena.

Struktura nezaposlenih prema školskoj spremi data je u sljedećoj tabeli:

Školska sprema	Broj nezaposlenih
Visoka	560
Viša	570
Srednja	660
Niža	610
	<b>2400</b>

# Primjer hi-kvadrat testa na jednom uzorku



Školska sprema	Broj nezaposlenih		$f_o - f_t$	$(f_o - f_t)^2$	$(f_o - f_t)^2 / f_t$
	$f_o$	$f_t$			
Visoka	560	600	-40	1600	2.67
Viša	570	600	-30	900	1.50
Srednja	660	600	60	3600	6.00
Niža	610	600	10	100	0.17
	<b>2400</b>	<b>2400</b>			<b>10.33</b>

	Observed N	Expected N	Residual
1.00	560	600.0	-40.0
2.00	570	600.0	-30.0
3.00	660	600.0	60.0
4.00	610	600.0	10.0
Total	2400		

	Sprema
Chi-Square	10.333 <sup>a</sup>
df	3
Asymp. Sig.	.016

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 600.0.

# Za vježbu



1. Uz pomoć testa psihomotorike testirano je 200 ljudi. Test je takve prirode da daje samo tri kategorije rezultata: A – slab, B – prosječan i C – dobar. Rezultati su pokazali da 40 ispitanih pokazuje slabe psihomotoričke sposobnosti, 110 prosječne, a 50 dobre. Odstupa li taj rezultat značajno od rezultata koje bismo očekivali da svojstvo normalno raspoređeno među ispitanicima?

	Observed N	Expected N	Residual
1	40	50.0	-10.0
2	110	100.0	10.0
3	50	50.0	.0
Total	200		

### Test Statistics

	Psihomotorik a
Chi-Square	3.000 <sup>a</sup>
df	2
Asymp. Sig.	.223

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 50.0.

# Za vježbu



2. Jedan profesor Statistike tvrdio je da među njegovih 50 studenata tridesetorica moraju pasti, da je 15 prosječnih, a 5 vrlo dobrih. Odstupa li takva raspodjela statistički značajno od onoga što bismo mogli očekivati pod vidom normalne raspodjele, tj. da li je 50% učenika u srednjoj kategoriji?

uspješnost_studenata			
	Observed N	Expected N	Residual
pali	30	12.5	17.5
prosječni	15	25.0	-10.0
vrlo dobri	5	12.5	-7.5
Total	50		

Test Statistics	
	uspješnost_s tudenata
Chi-Square	33.000 <sup>a</sup>
df	2
Asymp. Sig.	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 12.5.