

KROSTABULACIJE



University of Montenegro

Krostabulacije – Chi square

- Chi square statistika je neparametrički test koji koristimo za testiranje veza među *nominalnim i ordinalnim varijablama*
- Neparametrički test – ne pretpostavlja normalnu distribuciju podataka
- Umjesto za numeričke vrijednosti, koristi se ocjenu distribucije na kategorijama (demokrata, republikanac, visok prihod, nizak prihod itd.)

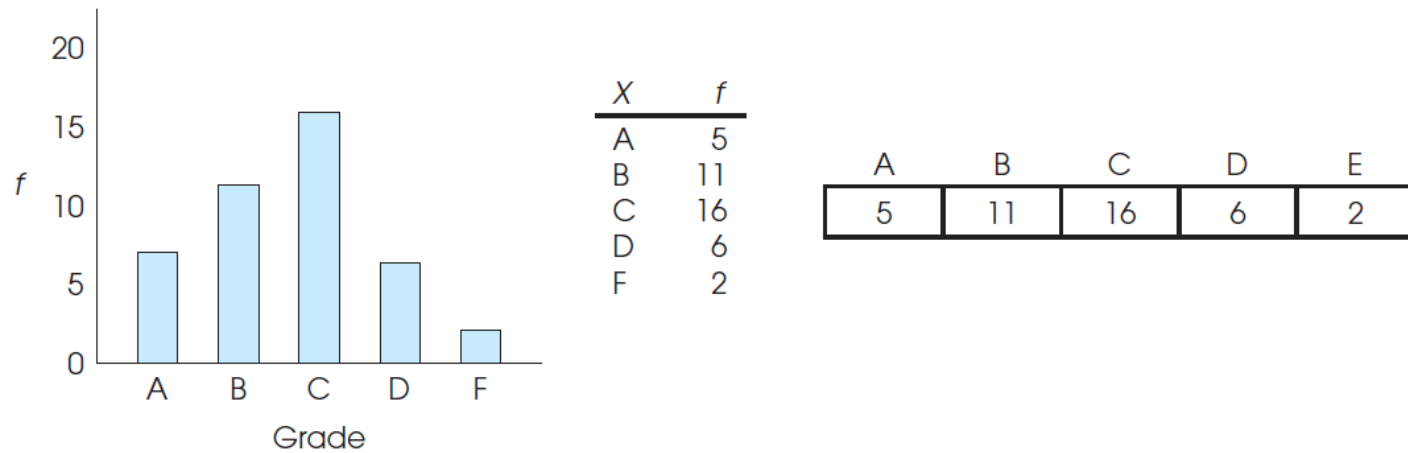


FIGURE 15.1

A distribution of grades for a sample of $n = 40$ students. The same frequency distribution is shown as a bar graph, as a table, and with the frequencies written in a series of boxes.

- Dvije verzije Chi square testa
- *Goodness of fit* – za procjenu distribucije jedne varijable u populaciji
- *Independence* (nezavisnost) – za ocjenu veze između dvije ili više varijabli (krostabulacija)

Goodnes of fit test

- Razlika između posmatrane i pretpostavljene distribucije
- *Posmatrana* – distribucija odgovora koju smo dobili u uzorku
- *Pretpostavljena* – distribucija koju smo formulisali nultom hipotezom

Chi square

$$\text{chi-square} = \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

- f_o = *posmatrana frekvencija u polju*
- f_e = *očekivana frekvencija u polju*



University of Montenegro

Test nezavisnosti

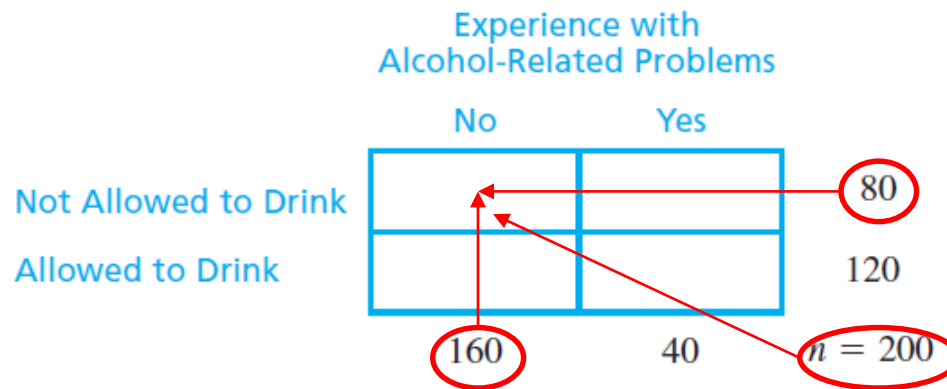
- Odnos između dvije varijable
- Pretpostavka - nema odnosa među varijablama
- Uticaj modela vaspitanja na problem sa alkoholizmom

	Experience with Alcohol-Related Problems		
	No	Yes	
Not Allowed to Drink	71	9	80
Allowed to Drink	89	31	120
	160	40	$n = 200$

- Pretpostavljene vrijednosti:

Experience with Alcohol-Related Problems

	No	Yes	
Not Allowed to Drink			80
Allowed to Drink			120
	160	40	$n = 200$





University of Montenegro

Pretpostavljene vrijednosti

- Određivanje pretpostavljenih vrijednosti

$$f_e = \frac{f_c f_r}{n}$$

- f_c – broj obzervacija u koloni
- f_r – broj obzervacija u redu
- n – ukupan broj obzervacija u uzorku

Pretpostavljene vrijednosti

Experience with Alcohol-Related Problems

	No	Yes
Not Allowed to Drink	64	80
Allowed to Drink	160	40

$n = 200$

120

$$\frac{160 * 80}{200} = 64$$

- Chi square test – *6.381*

		Experience with Alcohol-Related Problems		
		No	Yes	
Not Allowed to Drink	71	9	80	
Allowed to Drink	89	31	120	
	160	40	$n = 200$	

- *Stepeni slobode df – broj kolona minus 1, pomnoženo sa brojem redova minus 1.*

		Experience with Alcohol-Related Problems		
		No	Yes	
Allowed to Drink	64	16	80	
Not Allowed to Drink	96	24	120	
	160	40	$n = 200$	

- Tablica značajnosti:

<i>df</i>	<i>Proportion in Critical Region</i>				
	<i>0.10</i>	<i>0.05</i>	<i>0.025</i>	<i>0.01</i>	<i>0.005</i>
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59

Independence test

1. Za svako polje u krostabulaciji odrediti pretpostavljenu (očekivanu) vrijednost.
2. Od posmatrane vrijednosti oduzeti očekivanu
3. Kvadrirati razlike
4. Podijeliti rezultat sa očekivanom vrijednošću
5. Sabrati vrijednosti za sva polja
6. Provjeriti u tablici kritičnu vrijednost χ^2 testa za stepene slobode
7. Donijeti odluku o hipotezi i interpretirati podatke

Primjer

- Uticaj pola na zainteresovanost za politiku?
- EVS 2017

	Veoma zainteresovan/a	Donekle zainteresovan/a	Donekle nezainteresovan/a	Uopšte nisam zainteresovan/a
Muškarci	18.1%	39.6%	25.7%	16.5%
Žene	9.7%	34.5%	31.6%	23.8%

Chi square 1292.92*** značajan na $p < 0.001$

Primjer

- Pol i religioznost

	Religiozan/a	Nisam religiozan/a
Muškarci	65%	35%
Žene	72%	28%
Chi square test 312.00*** značajan na $p < 0.001$		