

KROSTABULACIJE

- 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.)

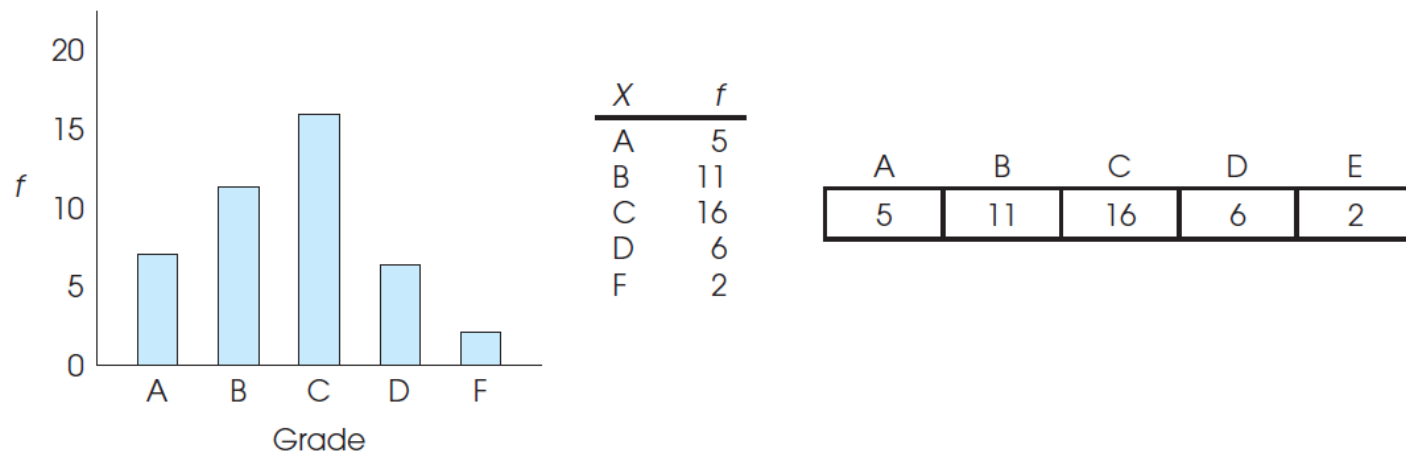


FIGURE 15.1

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

Primjer

- Studentima umjetnosti data je abstraktna slika bez očiglednog vrha i dna, koju su u vježbi trebali da okače pravilno.
- Istraživačka pretpostavka (**H0**) je da nijedna od četiri strane ne bi trebala da bude preferirana.
- Dakle, da je frekvenicija u svakoj kategoriji podjednaka (25%)

Top up (correct)	Bottom up	Left side up	Right side up
25%	25%	25%	25%

Primjer

- Za uzorak veličine $n=50$, nulta hipoteza pretpostavlja da će u svakoj kategoriji biti po 12.5 studenata.

Observed Frequencies

Expected Frequencies

	Top up (Correct)	Bottom up	Left Side up	Right Side up
Observed Frequencies	18	17	7	8
Expected Frequencies	12.5	12.5	12.5	12.5

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*

Računanje

- Gornja strana: $(18-12.5)^2/12.5 = 2.42$
- Donja strana: $(17-12.5)^2/12.5 = 1.62$
- Lijeva strana: $(7-12.5)^2/12.5 = 2.42$
- Desna strana: $(8-12.5)^2/12.5 = 1.62$

- Chi square = *8.08*

- Možemo li da odbacimo nultu hipotezu?

Stepeni slobode

- Potrebna nam je dodatna informacija, koja uzima u obzir broj kategorija koje testiramo.
- Stepeni slobode – degrees of freedom *df* (broj kategorija mogućih odgovora -1)

- 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



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	$n = 200$	

Razlike?

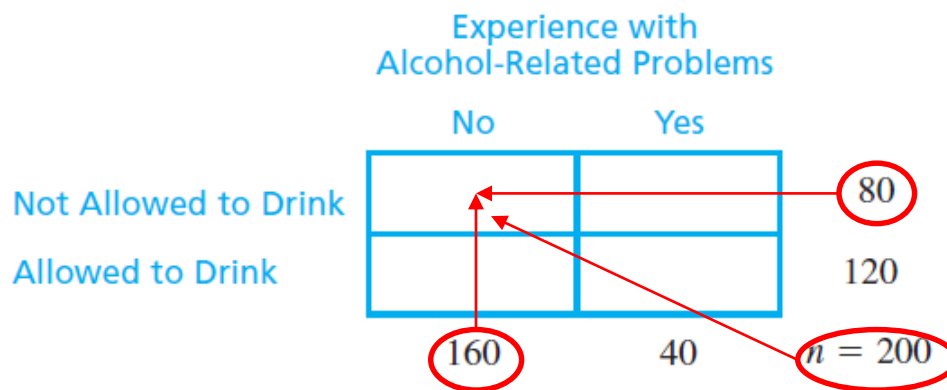
- Stepeni slobode (*df*) – broj redova minus 1, pomnoženo sa brojem kolona minus 1.

$$df = (R-1)(C-1) = (2-1)(2-1) = 1$$

- 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

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



University of Montenegro

Independence test

- Od ovog koraka, chi square independence test je identičan kao i goodness of fit 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 vrijednosću
 5. Sabrati vrijednosti za sva polja
 6. Provjeriti u tablici kritičnu vrijednost chi square testa za stepene slobode
 7. Donijeti odluku o hipotezi i interpretirati podatke

Zadatak

- Istraživači žele da ispitaju uticaj visokog obrazovanja na visinu mjesečnih primanja, pretpostavljajući da visoko obrazovani ljudi zarađuju više nego oni koji to nisu. Prikupili su podatke na uzorku od **n=15**. Testirajte nultu hipotezu da su ove dvije varijable nezavisna jedan u odnosu na drugu.

Zadatak

	Visoko obrazovanje	Prihodi
1.	Da	Visoki
2.	Da	Srednji
3.	Ne	Srednji
4.	Ne	Niski
5.	Ne	Niski
6.	Da	Visoki
7.	Ne	Srednji
8.	Ne	Visoki
9.	Da	Srednji
10.	Da	Visoki
11.	Ne	Srednji
12.	Da	Srednji
13.	Da	Niski
14.	Ne	Visoki
15.	Da	Srednji