

STATISTIKA ZA PSIHologe

VAJA 8 - REŠITVE 3. NALOGE

1. ¹ Zanima nas, ali v populaciji obstaja povezanost med lastništvom živali in izbrano fakulteto (VF=Veterinarska fakulteta, MF=Medicinska fakulteta). S spletno anketo smo pridobili naslednje podatke.

		Fakulteta		Vsota
		VF	MF	
Živali	Da	42	63	105
	Ne	2	33	35
Vsota		44	96	140

Pri stopnji značilnosti $\alpha = 0,05$ testirajte ali sta dani spremenljivki v populaciji asociirani.

Rešitev:

¹L. Lusa: Naloge iz Biostatistike

<i>Populacija</i>	<i>Študentje medicinske in veterinarske fakultete</i>
<i>Vzorec</i>	<i>140 študentov medicinske in veterinarske fakultete</i>
<i>Raziskovalno vprašanje</i>	<i>Ali sta izbrana fakulteta in lastništvo živali v populaciji povezani?</i>
<i>Statistični test</i>	<i>Test χ^2 (hi-kvadrat). Uporabimo ga, ko želimo preučevati povezanost med dvema imenskima spremenljivkama.</i>
<i>Ničelna domneva</i>	<i>V populaciji sta izbrana fakulteta in lastništvo domačih živali neodvisni (ni povezanosti). Ali: v populaciji sta deleža študentov, ki imajo domačo žival, enaka na medicinski in na veterinarski fakulteti.</i>
<i>Alternativna domneva</i>	<i>V populaciji sta fakulteta in lastništvo domačih živali povezani. Ali: v populaciji je delež študentov, ki imajo domačo žival, različen.</i>
<i>Testna statistika</i>	<p><i>Pearsonova statistika:</i></p> $\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(f_{ij} - \tilde{f}_{ij})^2}{\tilde{f}_{ij}}$ <p><i>k: število vrstic, l: število stolpcev, f_{ij}: opažena frekvenca v celici, ki se nahaja v vrstici i in v stolpcu j, \tilde{f}_{ij}: pričakovana frekvenca v celici, ki se nahaja v vrstici i in v stolpcu j.</i></p>
<i>Porazdelitev testne statistike ob veljavni ničelni domnevi</i>	<i>$\chi^2(1)$: χ^2 z eno stopinjo prostosti (df: degree of freedom). Stopinje prostosti so parameter, ki določa obliko porazdelitve χ^2 in je odvisen od števila možnih vrednosti spremenljivk: $df = (k-1)(l-1)$.</i>
<i>Predpostavke za uporabo testa</i>	<i>statistične enote (študentje) so neodvisne vzorec je velik, $\tilde{f}_{ij} \geq 5$</i>

Koliko študentov bi pričakovali v vsaki celici, če bi bili fakulteta in lastništvo živali neodvisni (če bi držala ničelna domneva). Izpolnite tabelo pričakovanih frekvenc \tilde{f}_{ij} .

	Fakulteta		Vsota
	VF	MF	
Živali	Da	72	105
	Ne	24	35
Vsota	44	96	140

Izračun testne (Pearsonove) statistike:

$$\chi^2 = \frac{(42-33)^2}{33} + \frac{(63-72)^2}{72} + \frac{(2-11)^2}{11} + \frac{(33-24)^2}{24} = 14,32$$

H_0 zavrնemo in sprejmemo H_1 natanko tedaj, ko velja

$$\chi^2 > \chi_{1-\alpha}^2((k-1)(l-1)),$$

oziroma v našem primeru ko

$$\chi^2 > \chi_{0,95}^2(1).$$

Pogledamo v tabelo za χ^2 in ugotovimo $\chi_{0,95}^2(1) = 3,84$. Vrednost Pearsonove statistike je večja od kritične vrednosti, torej zavrնemo ničelno hipotezo v prid alternativni. Spremenljivki izbira fakultete in lastništvo živali sta statistično značilno (celo statistično zelo značilno) asocirani.

2. ² Denimo, da so se še trije raziskovalci (A, B in C) ukvarjali z istim raziskovalnim vprašanjem in so dobili naslednje rezultate (rezultati v prvi tabeli so tisti, ki smo jih dobili z našo raziskavo).

Naši rezultati		Fakulteta		Vsota
		VF	MF	
Živali	Da	42	63	105
	Ne	2	33	35
Vsota		44	96	140

Raziskovalec A		Fakulteta		Vsota
		VF	MF	
Živali	Da	22	48	70
	Ne	22	48	70
Vsota		44	96	140

Raziskovalec B		Fakulteta		Vsota
		VF	MF	
Živali	Da	42·2	63·2	105·2
	Ne	2·2	33·2	35·2
Vsota		44·2	96·2	140·2

Raziskovalec C		Fakulteta		Vsota
		VF	MF	
Živali	Da	42·3	63·3	105·3
	Ne	2·3	33·3	35·3
Vsota		44·3	96·3	140·3

Brez dodatnih izračunov uredite rezultate od najmanjše do največje vrednosti testne statistike.

Rešitev: Pravilni vrstni red: A, naši rezultati, B, C.

Raziskovalec A ima najmanjšo vrednost testne statistike, ker je bil njegov vzorec velik kot naš, ampak je opazil manjšo razliko med skupinama. V njegovem primeru so opaženi deleži lastništva domačih živali natanko enaki na obeh fakultetah, torej se opažene f_{ij} in pričakovane \tilde{f}_{ij} vrednosti ujemajo in je vrednost testne statistike natanko 0.

Raziskovalca B in C sta opazila isto razliko v deležih kot mi, ampak sta imela večji vzorec. Ista odstopanja v deležih pri večjem vzorcu generirajo večje vrednosti testne statistike.

²L. Lusa: Naloge iz Biostatistike

Intuitivno: pri večjem vzorcu postane manj verjetno, da se bodo pojavile velike razlike med skupinama, če sta v populaciji deleži enaki, torej je smiselno, da so ob isti opaženi razliki vrednosti testne statistike večje, če je vzorec večji.

3. Heck in soavtorji so v članku *Offsetting Risks: High School Gay-Straight Alliances and Lesbian, Gay, Bisexual, and Transgender (LGBT) Youth* raziskovali povezanost spremenljivke Vrsta srednješole (GSA+ - participants who attended a high school with a gay-straight student alliance; GSA- - participants who attended a high school without a gay-straight student alliance) z različnimi drugimi spremenljivkami. Mi bi radi raziskali ali sta spremenljivki Vrsta srednje šole in status (v zvezi, poročen, samski) v preučevani populaciji pri stopnji značilnosti $\alpha = 0,05$ povezani. Razberite ustrezne podatke iz tabele in naredite statistično analizo.

Table 1
Demographic Characteristics of the Entire Sample and a Comparison of GSA+ and GSA- Groups to Determine Potential Covariates

Variable	Entire sample (n = 145)	GSA+ (n = 79)	GSA- (n = 66)	GSA+ and GSA- comparison	
	n (%)	n (%)	n (%)	t/ χ^2	p-value
Gender				3.33	.189
Female	87 (60.0)	52 (65.8)	35 (53.0)		
Male	48 (33.1)	21 (26.6)	27 (40.9)		
Transgender/other	10 (6.9)	6 (7.6)	4 (6.1)		
Ethnicity				6.09	.297
African American	14 (9.7)	6 (7.6)	8 (12.1)		
Asian American	8 (5.6)	5 (6.4)	3 (4.5)		
White	102 (70.8)	56 (71.8)	46 (69.7)		
Hispanic/Chicano	7 (4.9)	2 (2.6)	5 (7.6)		
Indian/Native American	1 (0.7)	0 (0.0)	1 (1.5)		
Other	12 (8.3)	9 (11.5)	3 (4.5)		
Sexual orientation				8.42	.015
Bisexual	54 (37.2)	37 (46.8)	17 (25.8)		
Gay/lesbian	83 (57.2)	40 (50.6)	43 (65.1)		
Other	8 (5.6)	2 (2.5)	6 (9.1)		
Relationship status				3.85	.146
Dating	42 (29.0)	24 (30.4)	18 (27.3)		
Married/committed	31 (21.3)	21 (26.6)	10 (15.2)		
Single	72 (49.7)	34 (43.0)	38 (57.6)		
Population				18.11	.003
Less than 2,500	9 (6.2)	0 (0.0)	9 (13.6)		
2,500–4,999	9 (6.2)	4 (5.1)	5 (7.6)		
5,000–9,999	19 (13.1)	8 (10.3)	11 (16.7)		
10,000–49,999	43 (29.7)	31 (39.7)	12 (18.2)		
50,000–250,000	36 (24.8)	21 (26.9)	15 (22.7)		
More than 250,000	28 (19.3)	14 (17.9)	14 (21.2)		
Education				1.60	.660
12 years	25 (17.2)	13 (16.5)	12 (18.2)		
13 years	53 (36.6)	26 (32.9)	27 (40.9)		
14 years	49 (33.8)	30 (38.0)	19 (28.8)		
15 years	18 (12.4)	10 (12.7)	8 (12.1)		
	Mean (SD)	Mean (SD)	Mean (SD)		
Age	19.17 (0.76)	19.06 (0.75)	19.30 (0.74)	-1.91	.058
Community climate	2.90 (1.19)	2.67 (1.13)	3.18 (1.21)	-2.63	.009
CTQ-SF total score	21.28 (8.35)	21.02 (8.83)	21.58 (7.80)	-0.39	.694
CTQ-SF sexual	6.25 (3.43)	6.16 (3.87)	6.35 (2.83)	-0.32	.749
CTQ-SF physical	6.46 (2.86)	6.39 (3.00)	6.53 (2.71)	-0.29	.774
CTQ-SF emotional	8.57 (4.02)	8.47 (4.08)	8.70 (3.98)	-0.34	.735
Outness inventory	26.41 (8.71)	26.22 (9.51)	26.65 (7.72)	-0.29	.775
Sexual orientation	5.52 (1.39)	5.28 (1.37)	5.82 (1.35)	-2.38	.019

Note. GSA+ = participants who attended a high school with a gay-straight student alliance; GSA- = participants who attended a high school without a gay-straight student alliance. Population refers to the population of the city or town where the participant attended high school for the longest period of time. CTQ-SF = Childhood Trauma Questionnaire-Short Form that provides a total score and three subscales for sexual, physical, and emotional abuse.

by other students”) and so that participants could indicate whether they felt the victimization was related to their sexual orientation. Researchers studying bullying have often relied on retrospective self-reports, and the accurate re-

porting of bullying experiences by victims has been demonstrated (Olweus, 1993).

For each item, participants indicated how often they experienced each form of at-school victimization using a Likert scale ranging from 0 (this

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Rešitev: Obe spremenljivki sta imenski, torej uporabimo **hi kvadrat test**. Zapišemo H_0 : Status (X) in Vrsta srednje šole ($GSA+$ ali $GSA-$) (Y) v populaciji nista asociirani.

H_1 : Status (X) in Vrsta srednje šole ($GSA+$ ali $GSA-$) (Y) v populaciji sta asociirani. Preverimo predpostavke za uporabo testa: enostavni slučajni vzorec, enote neodvisne in $\tilde{f}_{ij} \geq 5$ (pogledamo spodaj v tabelo pričakovanih frekvenc in vidimo, da je pogoj izpolnjen).

Status/Vrsta srednje šole	GSA+	GSA-	Skupaj
V zvezi	24	18	42
Poročen	21	10	31
Samski	34	38	72
Skupaj	79	66	145

Izračunamo še tabelo pričakovanih frekvenc

Status/Vrsta srednje šole	GSA+	GSA-	Skupaj
V zvezi	22,9	19,1	42
Poročen	16,9	14,1	31
Samski	39,2	32,8	72
Skupaj	79	66	145

Izračunamo vrednost Pearsonove statistike:

$$\chi^2 = \frac{(24 - 22,9)^2}{22,9} + \frac{(18 - 19,1)^2}{19,1} + \frac{(21 - 16,9)^2}{16,9} + \frac{(10 - 14,1)^2}{14,1} + \frac{(34 - 39,2)^2}{39,2} + \frac{(38 - 32,8)^2}{32,8}$$

in dobimo $\chi^2 = 3,82$.

H_0 zavrnamo in sprejmemo H_1 natanko tedaj, ko velja

$$\chi^2 > \chi_{1-\alpha}^2((k-1)(l-1)),$$

kjer je k število vrstic in l število stolpcev v tabeli. V našem primeru ko

$$\chi^2 > \chi_{0,95}^2(2).$$

Pogledamo v tabelo za χ^2 in ugotovimo $\chi_{0,95}^2(2) = 5,99$. Vrednost Pearsonove statistike je manjša od kritične vrednosti, torej ne moremo zavrniti ničelne hipoteze v prid alternativni. Na podlagi podatkov ne moremo zaključiti, da obstaja v populaciji povezanost med vrsto srednje šole in statusom.