

# Припрема за писмени из предмета Статистика у библиотекама: задатак са нормалном расподелом

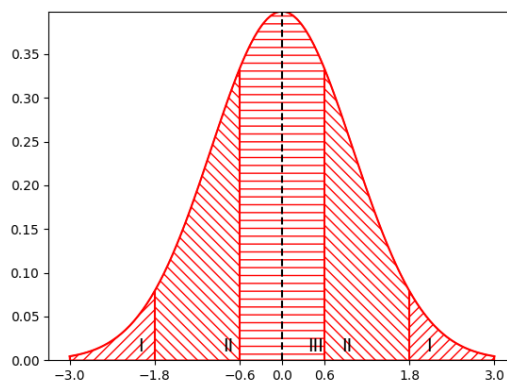
Милош Утвић

11. септембар 2024.

## 1 Задаци из збирке [1], одељак Normalna kriva (стр. 12)

43. Под претпоставком да се оцене знања из једног школског предмета нормално расподељују, како ће бити подељено 180 ученика у пет категорија, на основу оцена 1, 2, 3, 4 и 5 (израчунати број ученика у свакој категорији)?

**Решење:** Иако се у литератури обично користи биномна расподела као апроксимација нормалне, овде ће бити приказан поступак који је предложен у [1]. Површину испод стандардне нормалне (Гаусове) криве и изнад  $x$ -осе треба поделити на 5 делова које одговарају траженим категоријама. Пошто је  $x$ -оса бесконачна права, а део површине испод криве у интервалу  $[-3, 3]$  заузима 99,74% укупне површине испод криве, поделићемо коначан интервал  $[-3, 3]$  на 5 делова исте дужине. Пошто је дужина интервала  $[-3, 3]$  једнака  $6 = (3 - (-3))$ , дужина петине целог интервала, односно сваког појединачног подинтервала, је  $6/5 = 1,2$ . Крајње тачке подинтервала рачунамо тако што на леви крај целог интервала  $(-3)$  додајемо дужину подинтервала  $(1,2)$ , а исти резултат би се добио и кад бисмо од десног краја целог интервала  $(3)$  одузимали дужину подинтервала  $(1,2)$ . Према томе, крајњи подинтервала су  $-3$ ;  $-3 + 1,2 = -(3 - 1,2) = -1,8$ ;  $-1,8 + 1,2 = -(-1,8 - 1,2) = -0,6$ ;  $-0,6 + 1,2 = 0,6$ ;  $0,6 + 1,2 = 1,8$  и  $1,8 + 1,2 = 3$  (Слика 1).



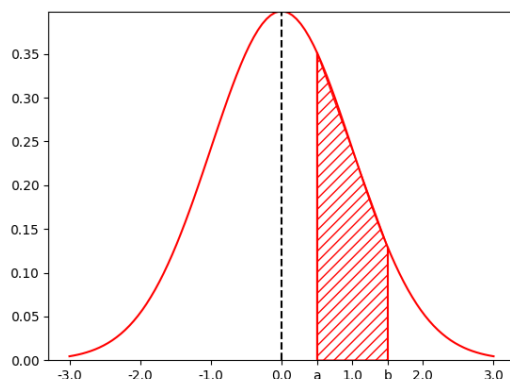
Слика 1: Подела интервала  $[-3, 3]$  на 5 подинтервала исте дужине.

Како бисмо избегли рачунске грешке и надокнадили неурачунатих 0,26% површине, крајње подинтервале ћемо проширити до одговарајуће бесконачности ( $-3$  и  $3$  више нису крајње тачке подинтервала), тако да је  $x$ -оса коначно подељена на следећих 5 интервала:  $[-\infty, -1,8]$ ,  $[-1,8, -0,6]$ ,  $[-0,6, 0,6]$ ,  $[0,6, 1,8]$  и  $[1,8, +\infty]$ . Према [1], површине испод нормалне криве и изнад ових подинтервала одговарају релативним фреквенцијама тражених категорија (израженим у процентима).

Ако је  $[a, b]$  неки интервал на  $x$ -оси, у даљем тексту ознака  $\Pi(a; b)$  представљаће површину испод стандардне нормалне (Гаусове) криве и изнад интервала  $[a, b]$ . Према томе, тражимо површине  $\Pi(-\infty; -1,8)$ ,  $\Pi(-1,8; -0,6)$ ,  $\Pi(-0,6; 0,6)$ ,  $\Pi(0,6; 1,8)$  и  $\Pi(1,8; +\infty)$  које одговарају релативним фреквенцијама тражених категорија (израженим у процентима).

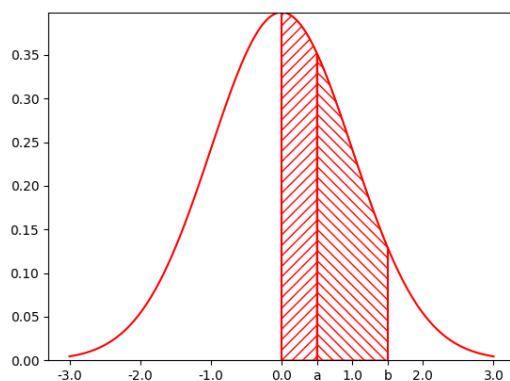
Тражене површине рачунамо користећи Таблицу 1 (стандардна нормална расподела) и следећа правила:

ПР<sub>1</sub> Ако је  $x \geq 0$ , таблична вредност  $T(x)$  представља површину испод стандардне нормалне (Гаусове) криве и изнад интервала  $[0, x]$ , тј.  $\Pi(0; x) = T(x)$  (Слика 2).



Слика 2: Површина испод нормалне криве која одговара табличној вредности за  $x$ .

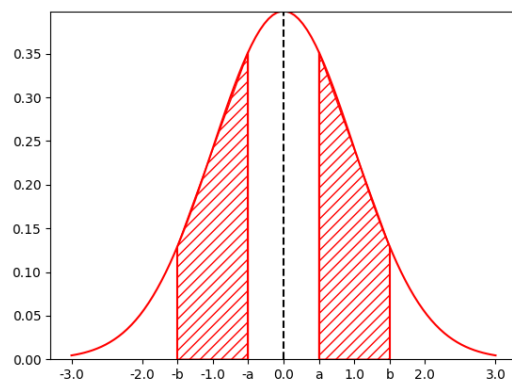
ПР<sub>2</sub> Ако  $0 \leq a < b$ , тј.  $[a, b]$  је неки интервал на позитивном делу  $x$ -осе, тада се са слике јасно уочава да је  $\Pi(a; b) = \Pi(0; b) - \Pi(0; a) = T(b) - T(a)$  (Слика 3).



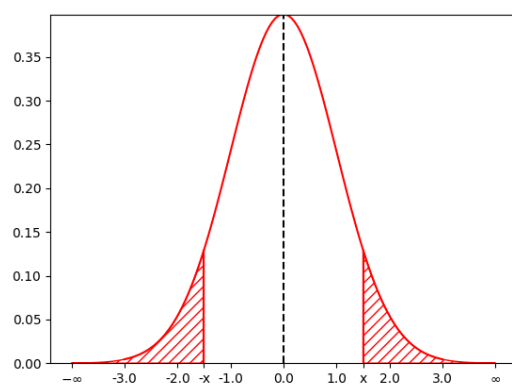
Слика 3: Површина испод нормалне криве која одговара позитивном интервалу  $[a, b]$ .

ПР<sub>3</sub> Због симетрије стандардне нормалне криве важи да је  $\Pi(-x; 0) = \Pi(0; x) = T(x)$ ,  $x \geq 0$ . Такође, ако  $0 \leq a < b$ , тада важи  $\Pi(-b; -a) = \Pi(a; b) = T(b) - T(a)$ . Специјално,  $\Pi(-x; x) = \Pi(-x; 0) + \Pi(0; x) = 2 \cdot \Pi(0; x) = 2 \cdot T(x)$  (Слика 4).

ПР<sub>4</sub> Целокупна површина испод графика стандардне нормалне криве једнака је  $1 = 100\%$ , тј.  $\Pi(-\infty; +\infty) = 100\%$ . Самим тим, због поменуте симетрије важи да је  $\Pi(-\infty; 0) = \Pi(0; +\infty) = 50\%$ . Одатле одмах произилази да је за  $x \geq 0$  испуњено:  $\Pi(-\infty; -x) = \Pi(x; +\infty) = \Pi(0; +\infty) - \Pi(0; x) = 50\% - T(x)$  (Слика 5).



Слика 4: Површине испод нормалне криве које одговарају интервалима  $[a, b]$  и  $[-b, -a]$  су једнаке.



Слика 5: Површине испод нормалне криве које одговарају интервалима  $[x, +\infty]$  и  $[-\infty, -x]$  су једнаке.

Одатле одмах добијамо да је

- $\Pi(-\infty; -1,8) = \Pi(1,8; +\infty) = 50\% - \Pi(0; 1,8) = 50\% - T(1,8) = 50\% - 46,41\% = 3,59\%$  и то је релативна фреквенција која одговара првој и последњој (5.) категорији (правила  $\Pi_3$  и  $\Pi_4$ ).
- $\Pi(-1,8; -0,6) = \Pi(0,6; 1,8) = T(1,8) - T(0,6) = 46,41\% - 22,57\% = 23,84\%$  и то је релативна фреквенција која одговара другој и четвртој категорији (правила  $\Pi_3$  и  $\Pi_2$ ).
- $\Pi(-0,6; 0,6) = \Pi(-0,6; 0) + \Pi(0; 0,6) = 2 \cdot \Pi(0; 0,6) = 2 \cdot T(0,6) = 2 \cdot 22,57\% = 45,14\%$  и то је релативна фреквенција која одговара средњој (трећој) категорији (правило  $\Pi_3$ ).
- Провера ( $\sum f_r = 100\%$ ):  $3,59\% + 23,84\% + 45,14\% + 23,84\% + 3,59\% = 100,00\%$ . Дакле, све је у реду, збир релативних фреквенција је заиста  $100,00\%$ .

На крају остаје да израчунамо оно што се тражи у задатку, а то су апсолутне фреквенције 5 могућих оцена на скупу од 180 ученика, под претпоставком да су оцене расподељују нормално. Ми смо, израчунавајући површине испод Гаусове криве, одредили релативне фреквенције, те остаје само да искористимо образац који повезује релативне ( $f_r$ ) и апсолутне фреквенције ( $f$ ) за узорак величине  $N$ :

$$f_r = \frac{f}{N} \Leftrightarrow f = N \cdot f_r. \quad (1)$$

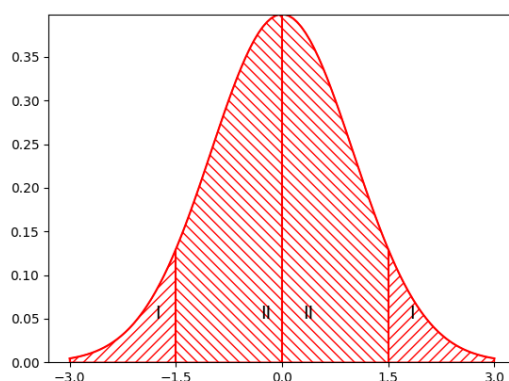
Према томе, израчунате површине (у процентима) треба помножити величином узорка (у овом примеру  $N = 180$  ученика) и заокружити на најближи цео број:

- $180 \cdot 3,59\% = 180 \cdot 3,59/100 = 6,462 \approx 6$ , фреквенција која одговара првој и последњој (5.) категорији, тј. оценама 1 и 5.
- $180 \cdot 23,84\% = 180 \cdot 23,84/100 = 42,912 \approx 43$ , фреквенција која одговара другој и четвртој категорији, тј. оценама 2 и 4.
- $180 \cdot 45,14\% = 180 \cdot 45,14/100 = 81,252 \approx 81$ , фреквенција која одговара средњој (трећој) категорији, односно оцени 3.
- Провера ( $\sum f = N$ ):  $6 + 43 + 81 + 43 + 6 = 179 < 180 = N$ , зато поправљамо средњу (највећу) фреквенцију тако да укупан збир фреквенција буде једнак величини узорка. Према томе, коначне фреквенције оцена 1–5 су редом 6, 43, 82, 43 и 6.  $\square$

44. Како ће на нормалној кривој бити расподељено 160 случајева у четири категорије?

**Решење:** Као у претходном задатку, делимо коначан интервал  $[-3, 3]$  на 4 подинтервала исте дужине (Слика 1), а онда крајње подинтервале проширујемо до крајева  $x$ -осе, тако да је  $x$ -оса коначно подељена на следећа 4 интервала:  $[-\infty, -1,5]$ ,  $[-1,5, 0]$ ,  $[0, 1,5]$  и  $[1,5, +\infty]$ . Према [1], површине испод нормалне криве и изнад ових подинтервала одговарају релативним фреквенцијама тражених категорија (израженим у процентима).

- $\Pi(-\infty; -1,5) = \Pi(1,5; +\infty) = 50\% - \Pi(0; 1,5) = 50\% - T(1,5) = 50\% - 43,32\% = 6,68\%$  и то је релативна фреквенција која одговара првој и четвртој категорији (правила  $\Pi_3$  и  $\Pi_4$ ).
- $\Pi(-1,5; 0) = \Pi(0; 1,5) = T(1,5) = 43,32\%$  и то је релативна фреквенција која одговара другој и трећој категорији (правила  $\Pi_3$  и  $\Pi_2$ ).



Слика 6: Подела интервала  $[-3, 3]$  на 4 подинтервала исте дужине.

- Провера ( $\sum f_r = 100\%$ ):  $6,68\% + 43,32\% + 43,32\% + 6,68\% = 100,00\%$ . Дакле, све је у реду, збир релативних фреквенција је заиста  $100,00\%$ .

Применом формуле (1) за  $N = 160$  добијамо:

- $160 \cdot 6,68\% = 160 \cdot 6,68/100 = 10,688 \approx 11$ , фреквенција која одговара првој и четвртој категорији.
- $160 \cdot 43,32\% = 160 \cdot 43,32/100 = 69,312 \approx 69$ , фреквенција која одговара другој и трећој категорији.
- Провера ( $\sum f = N$ ):  $11 + 69 + 69 + 11 = 160 = N$ , нема потребе за поправком фреквенција јер је њихов укупан збир једнак величини узорка. Према томе, коначне фреквенције су редом 11, 69, 69 и 11.  $\square$

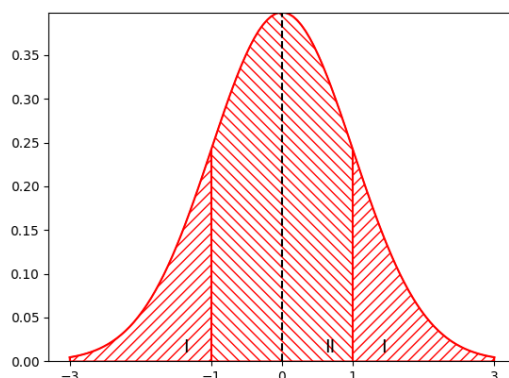
## 2 $\chi^2$ (хи-квадрат) тест

Преузето из [1], стр. 228: *Претпоставимо да је жири сиручњака проценио прилагоденост групе од 120 радника на потпуно нове услове производње. Том приликом 10 радника представљено је у категорију „сасвим прилагодени” (I); 85 радника у категорију „углавном прилагодени” (II); а 25 радника у категорију „неприлагодени” (III). Да ли је расподела фреквенција у три категорије варијабле прилагодености — нормална?*

Овде је питање непрецизно, а донекле и бесмислено јер дискретна расподела не може да буде нормална (то својство могу имати само континуалне или непрекидне расподеле). Питање се више односи на то да ли ова расподела значајно одступа од нормалне расподеле или не. Применићемо  $\chi^2$  (хи-квадрат) тест да бисмо одговорили на то питање.

Типична **нулта хипотеза** ( $H_0$ )  $\chi^2$  теста се заснива на претпоставци да између добијених (опажених, енгл. observed frequencies, у ознаци  $f_o$ ) и теоријских (очекиваних, енгл. expected frequencies, у ознаци  $f_e$ ) фреквенција нема разлике (односно да је разлика случајно настала), при чему су теоријске фреквенције израчунате на основу претпоставке о одређеној расподели. Пошто се овде питање односи на нормалну расподелу, нулта хипотеза ће претпоставити да фреквенције имају нормалну расподелу и теоријске фреквенције ће рачунати на начин описан у задацима 43 и 44, при чему, у овом случају, имамо 3 категорије, па делимо коначан интервал  $[-3, 3]$  на 3 подинтервала исте дужине (Слика 2). Потом крајње подинтервале проширујемо до крајева  $x$ -осе, тако да је  $x$ -оса коначно подељена на следећа 3 интервала:  $[-\infty, -1]$ ,  $[-1, 1]$  и  $[1, +\infty]$ . Према [1], површине испод нормалне

криве и изнад ових подинтервала одговарају релативним теоријским фреквенцијама датих категорија (израженим у процентима).



Слика 7: Подела интервала  $[-3, 3]$  на 3 подинтервала исте дужине.

- $\Pi(-\infty; -1) = \Pi(1; +\infty) = 50\% - \Pi(0; 1) = 50\% - T(1) = 50\% - 34,13\% = 15,87\%$  и то је релативна фреквенција која одговара првој и трећој категорији (правила ПР<sub>3</sub> и ПР<sub>4</sub>).
- $\Pi(-1; 1) = \Pi(-1; 1) + \Pi(0; 1) = 2 \cdot \Pi(0; 1) = 2 \cdot T(1) = 2 \cdot 34,13\% = 68,26\%$  и то је релативна фреквенција која одговара другој (средњој) категорији (правила ПР<sub>3</sub> и ПР<sub>2</sub>).
- Провера ( $\sum f_r = 100\%$ ):  $15,87\% + 68,26\% + 15,87\% = 100,00\%$ . Дакле, све је у реду, збир релативних фреквенција је заиста  $100,00\%$ .

Применом формуле (1) за  $N = 120$  добијамо:

- $120 \cdot 15,87\% = 120 \cdot 15,87/100 = 19,044 \approx 19$ , фреквенција која одговара првој и трећој категорији.
- $120 \cdot 68,26\% = 120 \cdot 68,26/100 = 81,912 \approx 82$ , фреквенција која одговара другој (средњој) категорији.
- Провера ( $\sum f = N$ ):  $19 + 82 + 19 = 120 = N$ , нема потребе за поправком фреквенција јер је њихов укупан збир једнак величини узорка. Према томе, коначне теоријске (очекиване) фреквенције  $f_e$  су редом 19, 82 и 19.

Опажене фреквенције  $f_o$  (10, 85, 25) и теоријске фреквенције  $f_e$  (19, 82 и 19) се очигледно разликују, питање је само да ли је та разлика статистички значајна. У ту сврху применићемо  $\chi^2$  тест.

|                             | I      | II     | III    | N      |                             |       |       |          |
|-----------------------------|--------|--------|--------|--------|-----------------------------|-------|-------|----------|
| $f_o$                       | 10     | 85     | 25     | 120    |                             |       |       |          |
| $f_e$                       | 19     | 82     | 19     | 120    |                             |       |       |          |
| $f_o - f_e$                 | -9     | 3      | 6      |        |                             |       |       |          |
| $(f_o - f_e)^2$             | 81     | 9      | 36     |        |                             |       |       |          |
| $\frac{(f_o - f_e)^2}{f_e}$ | 4,2632 | 0,1098 | 1,8947 | 6,2677 | $d_f$                       | 0,05  | 0,01  | $\chi^2$ |
|                             |        |        |        |        | $(2 - 1) \cdot (3 - 1) = 2$ | 5,991 | 9,210 | 6,268    |

Пошто је  $5,991 < \chi^2 = 6,268 < 9,210$ , закључујемо да је вредност  $\chi^2$  статистички значајна на нивоу 0,05 и на том нивоу одбацујемо нулту хипотезу, што значи да на нивоу грешке од 0,05 (или са 95% вероватноће) дата расподела значајно одступа од нормалне расподеле.  $\square$

### Задатак из збирке [1], одељак Hi - kvadrat (стр. 33)

94. На једно питање из инвентара личности 300 испитаника је дало следеће одговоре:

|           |               |              |
|-----------|---------------|--------------|
| Слажем се | Неодлучан сам | Не слажем се |
| 87        | 111           | 102          |

Утврдити да ли се овај распоред (добитених) фреквенција значајно разликује од распореда који би се очекивао:

- а) на основу претпоставке о једнакој вероватноћи јављања одговора;  
 б) на основу претпоставке о нормалној расподели одговора.

**Решење:** Овде су заправо дата два задатка који се решавају над истим подацима и истим тестом ( $\chi^2$ ), али са различитом нултом хипотезом и, самим тим, са различито израчунатим вредностима теоријских (очекиваних) фреквенција.

**Решење а):** Нулта хипотеза: између добијених фреквенција  $f_o$  одговора и теоријских фреквенција  $f_e$  одговора нема разлике (односно, разлика је случајно настала), при чему теоријске фреквенције одговора имају **униформну** расподелу (односно, одговори имају подједнаку вероватноћу).

Подједнака вероватноћа одговора значи да би теоријски сваки одговор требало да се јави подједнак број пута. Укупан број испитаника је 300, а различитих одговора је 3, па би сваки одговор требало да се јави  $300/3 = 100$  пута, односно теоријске фреквенције су 100, 100 и 100.

Опажене фреквенције  $f_o$  (87, 111, 102) и теоријске фреквенције  $f_e$  (100, 100 и 100) се очигледно разликују, питање је само да ли је та разлика статистички значајна. У ту сврху применићемо  $\chi^2$  тест.

|                             | I    | II   | III  | N    |                             |       |       |          |
|-----------------------------|------|------|------|------|-----------------------------|-------|-------|----------|
| $f_o$                       | 87   | 111  | 102  | 300  |                             |       |       |          |
| $f_e$                       | 100  | 100  | 100  | 300  |                             |       |       |          |
| $f_o - f_e$                 | -13  | 11   | 2    |      |                             |       |       |          |
| $(f_o - f_e)^2$             | 169  | 121  | 4    |      |                             |       |       |          |
| $\frac{(f_o - f_e)^2}{f_e}$ | 1,69 | 1,21 | 0,04 | 2,94 | $d_f$                       | 0,05  | 0,01  | $\chi^2$ |
|                             |      |      |      |      | $(2 - 1) \cdot (3 - 1) = 2$ | 5,991 | 9,210 | 2,94     |

Пошто је  $\chi^2 = 2,94 < 5,991 < 9,210$ , закључујемо да вредност  $\chi^2$  није статистички значајна, те не можемо одбацити нулту хипотезу, што значи да дата расподела не одступа значајно од униформне расподеле (али то не значи да смемо да тврдимо да је расподела одговора униформна).  $\square$

**Решење б):** Нулта хипотеза: између добијених фреквенција  $f_o$  одговора и теоријских фреквенција  $f_e$  одговора нема разлике (односно, разлика је случајно настала), при чему теоријске фреквенције одговора имају **нормалну** расподелу.

Претпоставка о нормалној расподели захтева да релативне теоријске фреквенције израчунамо на исти начин као у примеру са почетка одељка 2, у коме су такође постојале три категорије. У случају да је број различитих одговора био 5 или 4, применили бисмо поступак из задатака 43 и 44 и поново бисмо добили исте вредности релативних теоријских фреквенција. Према ономе што смо већ видели у поменутом примеру, релативне теоријске фреквенције су 15,87%, 68,26% и 15,87%. Међутим, сада укупна величина узорка није 120, већ 300, па ће и апсолутне фреквенције имати другачије вредности. При томе користимо исти образац (1) са стране 4:

- $300 \cdot 15,87\% = 300 \cdot 15,87/100 = 47,61 \approx 48$ , фреквенција која одговара првом и трећем одговору.
- $300 \cdot 68,26\% = 300 \cdot 68,26/100 = 204,78 \approx 205$ , фреквенција која одговара другом (најфреквентнијем) одговору.
- Провера ( $\sum f = N$ ):  $48 + 205 + 48 = 301 > 300 = N$ , , зато поправљамо средњу (највећу) фреквенцију тако да укупан збир фреквенција буде једнак величини узорка. Према томе, коначне теоријске (оčekиване) фреквенције  $f_e$  су редом 48, 204 и 48.

Опажене фреквенције  $f_o$  (87, 111, 102) и теоријске фреквенције  $f_e$  (48, 204 и 48) се очигледно разликују, питање је само да ли је та разлика статистички значајна. У ту сврху применићемо  $\chi^2$  тест.

|                             | I       | II      | III   | N       |                             |       |       |          |
|-----------------------------|---------|---------|-------|---------|-----------------------------|-------|-------|----------|
| $f_o$                       | 87      | 111     | 102   | 300     |                             |       |       |          |
| $f_e$                       | 48      | 204     | 48    | 300     |                             |       |       |          |
| $f_o - f_e$                 | 39      | -93     | 54    |         |                             |       |       |          |
| $(f_o - f_e)^2$             | 1521    | 8649    | 2916  |         |                             |       |       |          |
| $\frac{(f_o - f_e)^2}{f_e}$ | 31,6875 | 42,3971 | 60,75 | 134,835 | $d_f$                       | 0,05  | 0,01  | $\chi^2$ |
|                             |         |         |       |         | $(2 - 1) \cdot (3 - 1) = 2$ | 5,991 | 9,210 | 134,835  |

Пошто је  $5,991 < 9,210 < \chi^2 = 134,835$ , закључујемо да је вредност  $\chi^2$  статистички значајна на нивоу 0,01 и на том нивоу одбацујемо нулту хипотезу, што значи да на нивоу грешке од 0,01 (или са 99% вероватноће) дата расподела значајно одступа од нормалне расподеле.  $\square$

## Литература

- [1] Statistika za psihologe : sa zbirkom zadataka / Čedomir Dragičević (zadaci za vežbanja: Čedomir Dragičević i Lazar Tenjović). - 3. izd. - Beograd : Centar za primenjenu psihologiju, 2011 (Beograd : Centar za primenjenu psihologiju). - 255, 87 str. : graf. prikazi ; 24 cm. - (Biblioteka Psihološki priručnici / [Centar za primenjenu psihologiju]).