

Kontingenční tabulky – řešení.

Datový soubor BARVA VLASŮ.

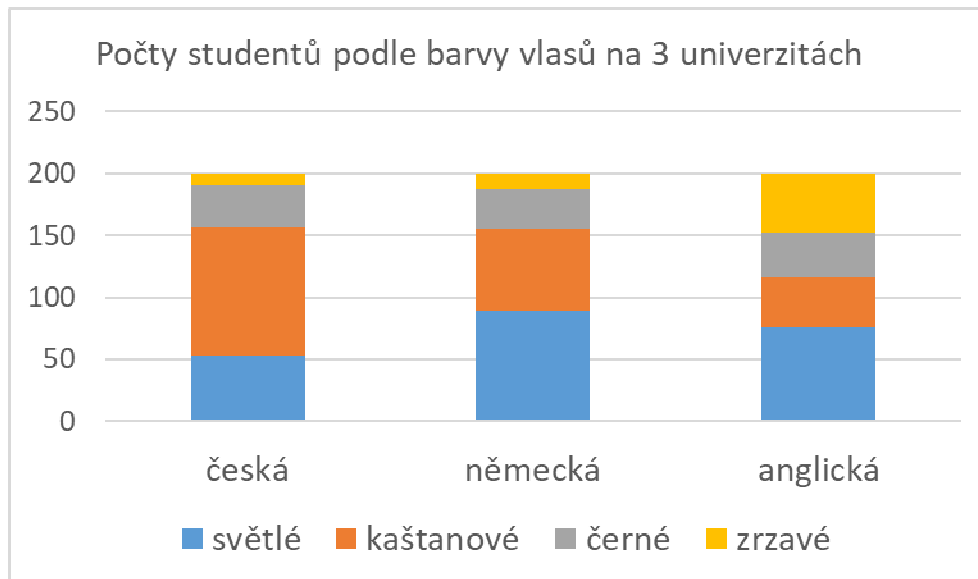
Na třech evropských univerzitách byl proveden pilotní průzkum barvy vlasů na 200 náhodně vybraných studentech. Zajímá nás, zda se pravděpodobnosti výskytu čtyř barev vlasů mezi univerzitami liší.

Představení datového souboru:

jako kontingenční tabulku.

| univerzita | vlasý světlé | vlasý kaštanové | vlasý černé | vlasý zrzavé | Řádk. součty |
|------------|--------------|-----------------|-------------|--------------|--------------|
| česká | 53 | 103 | 35 | 9 | 200 |
| německá | 89 | 66 | 32 | 13 | 200 |
| anglická | 76 | 40 | 36 | 48 | 200 |
| Vš. skup. | 218 | 209 | 103 | 70 | 600 |

Mohli bychom využít také například sloupcové diagramy:



Tento graf je z Excelu.

Statistica umí taky, ale datovou tabulku musíte přepsat do nové tabulky, která má vzhled jako kontingenční tabulka s názvy sloupců a řádků jako názvy proměnných a případů (řádků), a potom zvolit Grafy → 2D grafy → Sloupcové grafy.

Nulová hypotéza: čtveřice pravděpodobnosti výskytu barev vlasů jsou stejné pro všechny tři univerzity.

Test homogeneity v kontingenční tabulce.

Podmínka:

nejmenší očekávaná četnost je ≥ 5 .

To je splněno, nejmenší očekávaná četnost je 23,333.

Souhrnná tab.: Očekávané četnosti (02 barva vlasů)

Četnost označených buněk > 10

Pearsonův chí-kv. : 77,6315, sv=6, p=,000000

| univerzita | vlasý světlé | vlasý kaštanové | vlasý černé | vlasý zrzavé | Řádk. součty |
|------------|--------------|-----------------|-------------|--------------|--------------|
| česká | 72,6667 | 69,6667 | 34,3333 | 23,3333 | 200,0000 |
| německá | 72,6667 | 69,6667 | 34,3333 | 23,3333 | 200,0000 |
| anglická | 72,6667 | 69,6667 | 34,3333 | 23,3333 | 200,0000 |
| Vš. skup. | 218,0000 | 209,0000 | 103,0000 | 70,0000 | 600,0000 |

Výsledek testu: Na základě (Pearsonova) chí-kvadrát testu zamítáme nulovou hypotézu o shodnosti pravděpodobnostní struktury barevnosti vlasů na třech univerzitách. Hodnota testové statistiky $\chi^2 = 77.63$, srovnáváme s chí-kvadrát rozdělením o 6 stupních volnosti. P-hodnota < 0.001 .

Struktura pravděpodobností barevnosti vlasů za platnosti nulové hypotézy, tedy kdyby četnosti výběrů byly vzájemně srovnatelné: jsou to sloupcové marginální pravděpodobnosti, spočtu jako

Statistica spočte následující tabulku, když v záložce MOŽNOSTI zvolím v levém sloupci „Procenta celkového počtu“.

| Kontingenční tabulka (02 barva vlasů) Četnost označených buněk > 10 (Marginální součty nejsou označeny) | | | | | | |
|---|------------|--------------|-----------------|-------------|--------------|--------------|
| | univerzita | vlasý světlé | vlasý kaštanové | vlasý černé | vlasý zrzavé | Řádk. součty |
| Četnost | česká | 53 | 103 | 35 | 9 | 200 |
| Celková četn. | | 8,83% | 17,17% | 5,83% | 1,50% | 33,33% |
| Četnost | německá | 89 | 66 | 32 | 13 | 200 |
| Celková četn. | | 14,83% | 11,00% | 5,33% | 2,17% | 33,33% |
| Četnost | anglická | 76 | 40 | 36 | 48 | 200 |
| Celková četn. | | 12,67% | 6,67% | 6,00% | 8,00% | 33,33% |
| Četnost | Vš. skup. | 218 | 209 | 103 | 70 | 600 |
| Celková četn. | | 36,33% | 34,83% | 17,17% | 11,67% | |

Datový soubor BARVA OČÍ.

Na české univerzitě byla navíc zaznamenána barva očí. Z datové tabulky vybírám údaje pro „vzácné“ kombinace: Ptáme se, zda je mezi zbarvením vlasů a očí nějaká závislost.

| Barva očí | Barva vlasů | |
|-----------|-------------|--------|
| | Černé | Zrzavé |
| Modré | 3 | 5 |
| Hnědé | 24 | 1 |

Nulová hypotéza: výskyt barevných variací očí a vlasů jsou na sobě nezávislé.

Test nezávislosti v kontingenční tabulce.

Podmínka: nejmenší očekávaná četnost je ≥ 5 .

Není splněno. Můžeme použít Yatesovu korekci na spojitost nebo Fisherův exaktní test. Oboje Statistica počítá pro čtyřpolní tabulky.

Výsledek testu:

| Statist. | Statist. : oči(2) x vlasý(2) (02 barva) | | |
|--------------------------------|---|------|-----------------|
| | Chí-kvadr. | sv | p |
| Pearsonův chí-kv. | 13,94250 | df=1 | p=,00019 |
| M-V chí-kvadr. | 12,31098 | df=1 | p=,00045 |
| Yatesův chí-kv. | 10,28729 | df=1 | p=,00134 |
| Fisherův přesný, 1-str. | | | p=,00129 |
| Fisherův přesný, 2-str. | | | p=,00129 |
| McNemarův chí-kv. (A/D) | ,2500000 | df=1 | p=,61708 |
| McNemarův chí-kv. (B/C) | 11,17241 | df=1 | p=,00083 |
| Fí pro tabulky 2 x 2 | -,650000 | | |
| Tetrachorická korelace | -,885028 | | |
| Kontingenční koeficient | ,5449883 | | |

| Souhrnná tab.: Očekávané četnosti (02 barva očí 2x2) Četnost označených buněk > 10 Pearsonův chí-kv. : 13,9425, sv=1, p=,000188 | | | |
|---|-------------|--------------|--------------|
| oči | vlasý černé | vlasý zrzavé | Řádk. součty |
| modré | 6,55 | 1,45 | 8,00 |
| hnědé | 20,45 | 4,55 | 25,00 |
| Vš. skup. | 27,00 | 6,00 | 33,00 |

| Kontingenční tabulka (02 barva očí 2x2) Četnost označených buněk > 10 (Marginální součty nejsou označeny) | | | | |
|---|-----------|-------------|--------------|--------------|
| | oči | vlasý černé | vlasý zrzavé | Řádk. součty |
| Četnost | modré | 3 | 5 | 8 |
| Celková četn. | | 9,09% | 15,15% | 24,24% |
| Četnost | hnědé | 24 | 1 | 25 |
| Celková četn. | | 72,73% | 3,03% | 75,76% |
| Četnost | Vš. skup. | 27 | 6 | 33 |
| Celková četn. | | 81,82% | 18,18% | |

Zamítám nulovou hypotézu o nezávislosti barevnosti vlasů a očí. Kvůli malým četnostem musím použít Yatesův chí-kvadrát = 10,29, p-hodnota = 0,00134. Nebo Fisherův přesný test s p-hodnotou 0,00129.

Odhad pravděpodobnosti výskytu barev očí za předpokladu nezávislosti na barvě vlasů v tabulce.