**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.

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 homogenity** v kontingenční tabulce.

**Podmínka**:

nejmenší očekávaná četnost je ≥ 5.

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

**Výsledek testu**: Na základě (Pearsonova) chí-kvadrát testu zamítáme nulovou hypotézu o shodnosti pravděpodobnostní struktury barevnosti vlasů na třech univerzitách. Hodnota testové statistiky χ2 = 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

$$\hat{p}\_{světlé}=\frac{218}{600}=36.33 \%, \hat{p}\_{kaštanové}=\frac{209}{600}=34.86 \% , \hat{p}\_{černé}=\frac{103}{600}=17.17 \% , \hat{p}\_{zrzavé}=\frac{70}{600}=11.67 \%$$

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



------------------------------------------------------------------------------------------------------------------------------------------

**Datový soubor BARVA OČÍ.**

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

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.

**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:**

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.