

Téma 1: Bodové zpracování četností

Vzorový příklad: U 20 studentů 1. ročníku byly zjišťovány známky z matematiky, angličtiny a údaje o pohlaví (viz skripta Popisná statistika, příklad 2.4). Příslušný datový soubor se jmenuje **znamky.sta**. Proveďte bodové zpracování četností.

Postup ve STATISTICE:

1. Do programu STATISTICA načtete datový soubor **znamky.sta**.
2. Znaků nazvěte X, Y, Z, vytvořte jim návěští (X - známka z matematiky, Y - známka z angličtiny, Z - pohlaví studenta) a popište, co znamenají jednotlivé varianty (u znaků X a Y: 1 - výborně, 2 - velmi dobře, 3 - dobře, 4 - neprospěl, u znaku Z: 0 - žena, 1 - muž). Soubor uložte.
Návod: Kurzor nastavíme na Var1 – 2x klikneme myší – Name X – Long Name známka z matematiky, Text label – výborně, Numeric – 1, velmi dobře, Numeric – 2, dobře, Numeric – 3, neprospěl, Numeric – 4, OK. U proměnné Y lze text label okopírovat z proměnné X – v Text Labels Editor zvolíme Copy from variable X.
(Přepínání mezi číselnými hodnotami a jejich textovým popisem se děje pomocí tlačítka s obrázkem štítku.)
3. U znaků X a Y vypočtete absolutní četnosti, relativní četnosti a relativní kumulativní četnosti.
Návod: Statistics - Basic Statistics/Tables – Frequency tables – OK – Variables X, Y, OK – Summary.
(Obě tabulky se uloží do workbooku a listovat v nich můžeme pomocí stromové struktury v levém okně.)

Řešení:

| Category | Frequency table: X: známka z matematiky (zr | | | |
|--------------|---|------------------|----------|--------------------|
| | Count | Cumulative Count | Percent | Cumulative Percent |
| výborně: | 7 | 7 | 35,00000 | 35,0000 |
| velmi dobře: | 3 | 10 | 15,00000 | 50,0000 |
| dobře | 2 | 12 | 10,00000 | 60,0000 |
| neprospěl: | 8 | 20 | 40,00000 | 100,0000 |
| Missing | 0 | 20 | 0,00000 | 100,0000 |

| Category | Frequency table: Y: známka z angličtiny (znar | | | |
|--------------|---|------------------|----------|--------------------|
| | Count | Cumulative Count | Percent | Cumulative Percent |
| výborně: | 4 | 4 | 20,00000 | 20,0000 |
| velmi dobře: | 4 | 8 | 20,00000 | 40,0000 |
| dobře | 7 | 15 | 35,00000 | 75,0000 |
| neprospěl: | 5 | 20 | 25,00000 | 100,0000 |
| Missing | 0 | 20 | 0,00000 | 100,0000 |

Komentář: Vidíme např., že jedničku z matematiky mělo 7 studentů, tj. 35%. Zkoušku z angličtiny složilo 15 studentů, tj. 75%.

4. Vytvořte sloupkový diagram absolutních četností znaků X a Y.
Návod: Graphs – Histograms – Variables X, Y – OK- vypneme Normal fit – Advanced – zaškrtneme Breaks between Columns, OK.

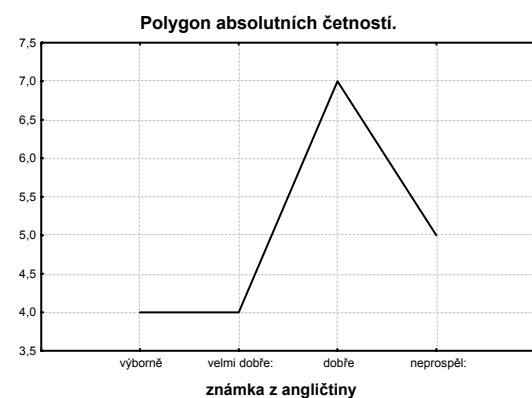
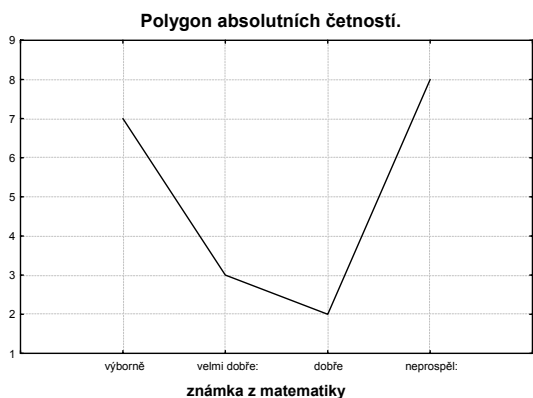
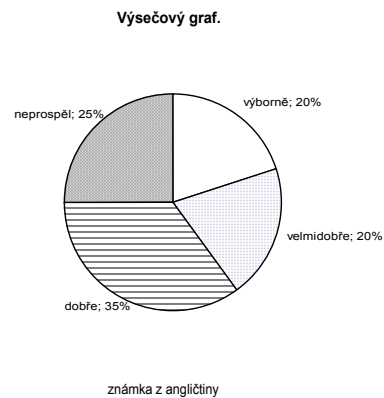
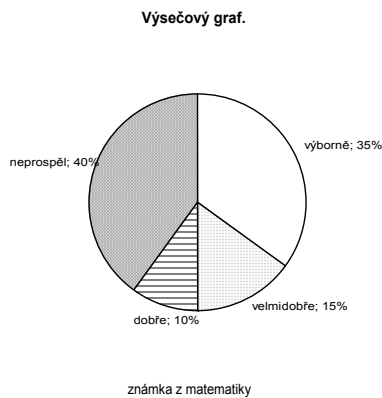
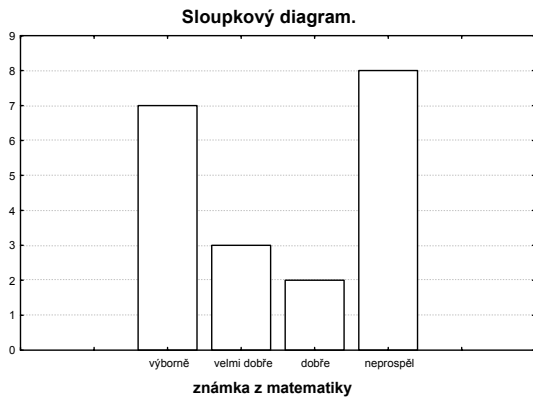
Vytvořte výšečový diagram absolutních četností znaků X a Y.

Návod: Graphs – 2D Graphs – Pie Charts – Variables X, Y – OK – Advanced – Pie legend Text and Percent (nebo Text and Value) – OK.

Vytvořte polygon absolutních četností znaků X a Y.

Návod: ve workbooku vstoupíme do tabulky rozložení četností proměnné X. Pomocí Edit – Delete - Cases vymažeme řádek označený Missing. Nastavíme se kurzorem na Count a kliknutím pravého tlačítka vstoupíme do menu Line Plot: Entire Columns. Vykreslí se polygon četností.

Řešení:



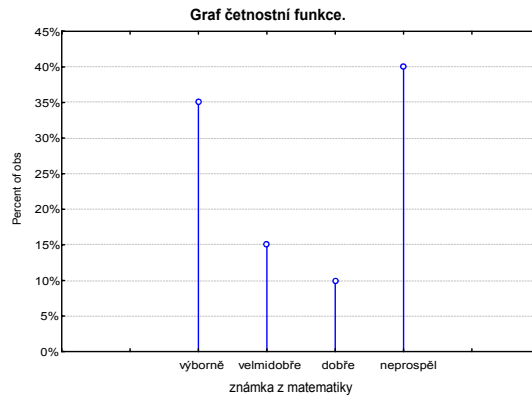
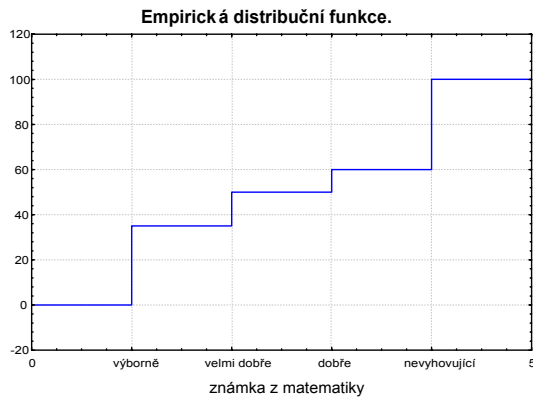
5. Vytvořte graf empirické distribuční funkce znaku X.

Návod: Při tvorbě histogramu zadáme v Advanced volbu Showing Type Cumulative, Y axis % - 2x klikneme myší na pozadí grafu – otevře se okno All Options – vybereme Plot: Bars – Type Rectangles.

Vytvořte graf četnostní funkce znaku X.

Návod: Při tvorbě histogramu zadáme v Advanced Y axis % - 2x klikneme myší na pozadí grafu – vybereme Plot General – zaškrtneme Markers – vybereme Plot: Bars – Type Lines.

Řešení:



Komentář: Z grafu empirické distribuční funkce bohužel nelze jednoduše odstranit svíslé čáry. Stojí za povšimnutí, že výška „schodu“ v grafu empirické distribuční funkce ve variantách „výborně“, „velmi dobře“, „dobře“ a „neprospěl“ odpovídá hodnotě četnostní funkce v příslušné variantě.

6. Z datového souboru vyberte pouze ženy (pouze muže) a úkol 3 proveďte pro ženy (pro muže).

Návod: Statistics - Basic Statistics/Tables – Frequency tables – OK – Variables X, Y, OK – Select Cases – zaškrtneme Selection Conditions – Include cases – zaškrtneme Specific, selected by Z = 0, OK.

Řešení:

Variační řady známek z matematiky a angličtiny pro ženy.

| Frequency table: X: známka z matematiky | | | | |
|---|-------|------------------|----------|--------------------|
| Category | Count | Cumulative Count | Percent | Cumulative Percent |
| výborně: | 5 | 5 | 50,00000 | 50,0000 |
| velmi dobře: | 2 | 7 | 20,00000 | 70,0000 |
| dobře | 1 | 8 | 10,00000 | 80,0000 |
| neprospěl: | 2 | 10 | 20,00000 | 100,0000 |
| Missing | 0 | 10 | 0,00000 | 100,0000 |

| Frequency table: Y: známka z angličtiny | | | | |
|---|-------|------------------|----------|--------------------|
| Category | Count | Cumulative Count | Percent | Cumulative Percent |
| výborně: | 4 | 4 | 40,00000 | 40,0000 |
| velmi dobře: | 2 | 6 | 20,00000 | 60,0000 |
| dobře | 1 | 7 | 10,00000 | 70,0000 |
| neprospěl: | 3 | 10 | 30,00000 | 100,0000 |
| Missing | 0 | 10 | 0,00000 | 100,0000 |

Variační řady známek z matematiky a z angličtiny pro muže.

| Category | Frequency table: X: známka z matematiky | | | |
|--------------|---|------------------|----------|--------------------|
| | Count | Cumulative Count | Percent | Cumulative Percent |
| výborně: | 2 | 2 | 20,00000 | 20,00000 |
| velmi dobře: | 1 | 3 | 10,00000 | 30,00000 |
| dobře | 1 | 4 | 10,00000 | 40,00000 |
| neprospěl: | 6 | 10 | 60,00000 | 100,00000 |
| Missing | 0 | 10 | 0,00000 | 100,00000 |

| Category | Frequency table: Y: známka z angličtiny | | | |
|--------------|---|------------------|----------|--------------------|
| | Count | Cumulative Count | Percent | Cumulative Percent |
| velmi dobře: | 2 | 2 | 20,00000 | 20,00000 |
| dobře | 6 | 8 | 60,00000 | 80,00000 |
| neprospěl: | 2 | 10 | 20,00000 | 100,00000 |
| Missing | 0 | 10 | 0,00000 | 100,00000 |

7. Nadále pracujte s celým datovým souborem. Vytvořte kontingenční tabulku absolutních četností znaků X a Y a graf simultánní četností funkce.

Návod: Statistics - Basic Statistics/Tables – Tables and banners – OK – Select cases – All – OK – Specify tables - List 1 X, List 2 Y, OK, Summary.

Vytvoření grafu simultánní četností funkce: Návrat do Crosstabulation Tables Result – 3D histograms – vybereme Axis Scaling – Mode Manual – Minimum 0 (a totéž provedeme pro Axis Y) – dále vybereme Graph Layout – Type – Spikes – OK. Graf lze natáčet pomocí Point of View.

Vytvořte kontingenční tabulku sloupcově a řádkově podmíněných relativních četností znaků X a Y.

Návod: Návrat do Crosstabulation Tables Result – Options - zaškrtneme ve sloupci Compute tables volbu Percentages of column counts (resp. Percentages of row counts).

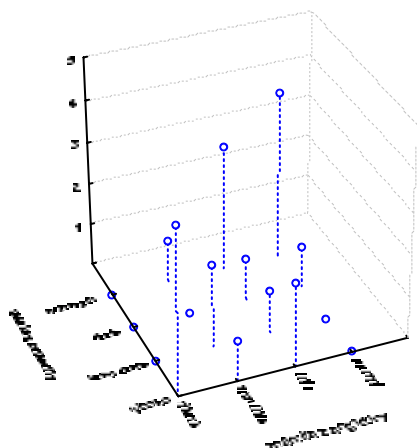
Řešení:

Kontingenční tabulka absolutních četností známek z matematiky a z angličtiny.

| Summary Frequency Table (znamky) | | | | | |
|-------------------------------------|---------|-------------|-------|-----------|------------|
| Marked cells have counts > 10 | | | | | |
| (Marginal summaries are not marked) | | | | | |
| X | Y | Y | Y | Y | Row Totals |
| | výborně | velmi dobře | dobře | neprospěl | |
| výborně | 4 | 1 | 2 | 0 | 7 |
| velmi dobře | 0 | 2 | 1 | 0 | 3 |
| dobře | 0 | 0 | 1 | 1 | 2 |
| neprospěl | 0 | 1 | 3 | 4 | 8 |
| All Grps | 4 | 4 | 7 | 5 | 20 |

Komentář: Z tabulky je patrné, že např. 4 studenti měli z obou předmětů známku „výborně“, 1 student měl z matematiky „výborně“ a současně z angličtiny „velmi dobře“ atd.

Simultánní četnostní funkce.



Kontingenční tabulka sloupcově a řádkově podmíněných relativních četností.

| Summary Frequency Table (znamky) | | | | | | |
|-------------------------------------|-------------|---------|-------------|--------|-----------|--------|
| Marked cells have counts > 10 | | | | | | |
| (Marginal summaries are not marked) | | | | | | |
| | X | Y | Y | Y | Y | Row |
| | | výborně | velmi dobře | dobře | neprospěl | Totals |
| Count | výborně | 4 | 1 | 2 | 0 | 7 |
| Column Percent | | 100,00% | 25,00% | 28,57% | 0,00% | |
| Row Percent | | 57,14% | 14,29% | 28,57% | 0,00% | |
| Count | velmi dobře | 0 | 2 | 1 | 0 | 3 |
| Column Percent | | 0,00% | 50,00% | 14,29% | 0,00% | |
| Row Percent | | 0,00% | 66,67% | 33,33% | 0,00% | |
| Count | dobře | 0 | 0 | 1 | 1 | 2 |
| Column Percent | | 0,00% | 0,00% | 14,29% | 20,00% | |
| Row Percent | | 0,00% | 0,00% | 50,00% | 50,00% | |
| Count | neprospěl | 0 | 1 | 3 | 4 | 8 |
| Column Percent | | 0,00% | 25,00% | 42,86% | 80,00% | |
| Row Percent | | 0,00% | 12,50% | 37,50% | 50,00% | |

Komentář: Podívejme se např. na druhý sloupec tabulky. V prvním řádku je číslo 25%. Znamená to, že 25% těch studentů, kteří měli „velmi dobře“ z angličtiny, dostali známku „výborně“ z matematiky. Nyní se podíváme např. na čtvrtý řádek tabulky. Ve třetím sloupci je číslo 37,5%. Znamená to, že 37,5% těch studentů, kteří neprospěli z matematiky, dostalo známku „dobře“ z angličtiny.