

Odpovědi

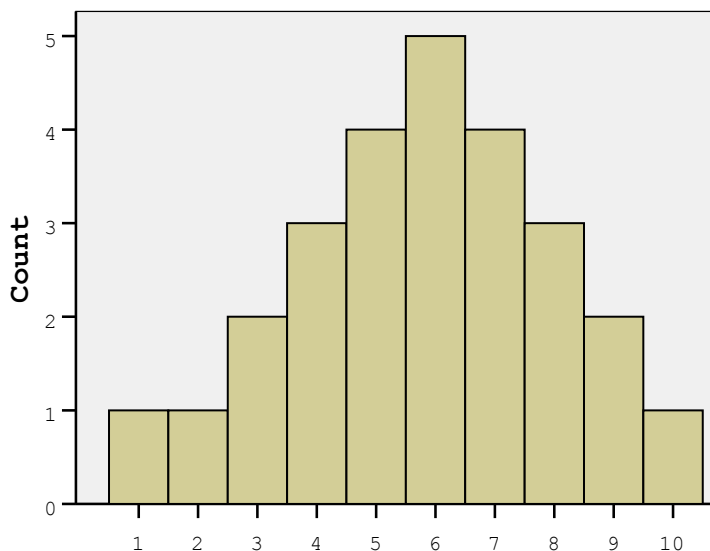
1. a

2. d

3. c

4. b

5. a



6. uniformní

7. a) protože body za písemky jsou poměrovou proměnnou, mohli bychom do tabulky četností uvést všechny možné hodnoty mezi 40 a 98 a u těch, kterých nikdo nedosáhl, uvést 0. Také bychom mohli o této proměnné uvažovat jako o spojitě proměnné (měřené hrubým měřítkem); potom bychom museli i u jednotlivých hodnot používat označení jako u intervalů (tj. 39,5 – 40,49 místo 40 atd.)

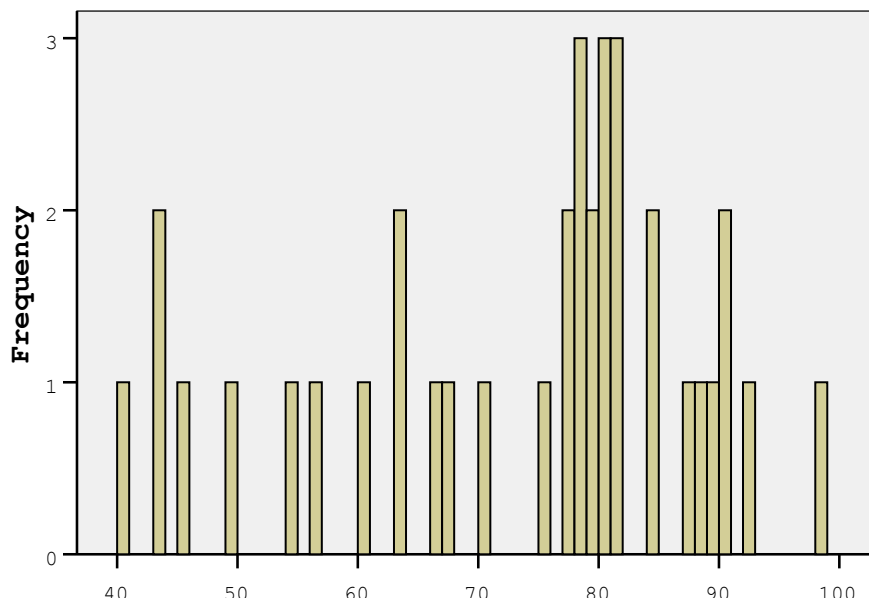
	četnost	relativní četnost	kumulativní relativní četnost
40	1	2,8	2,8
43	2	5,6	8,3
45	1	2,8	11,1
49	1	2,8	13,9
54	1	2,8	16,7
56	1	2,8	19,4
60	1	2,8	22,2
63	2	5,6	27,8
66	1	2,8	30,6
67	1	2,8	33,3
70	1	2,8	36,1
75	1	2,8	38,9
77	2	5,6	44,4
78	3	8,3	52,8
79	2	5,6	58,3
80	3	8,3	66,7
81	3	8,3	75,0
84	2	5,6	80,6
87	1	2,8	83,3

88	1	2,8	86,1
89	1	2,8	88,9
90	2	5,6	94,4
92	1	2,8	97,2
98	1	2,8	100,0
Celkem	36	100,0	

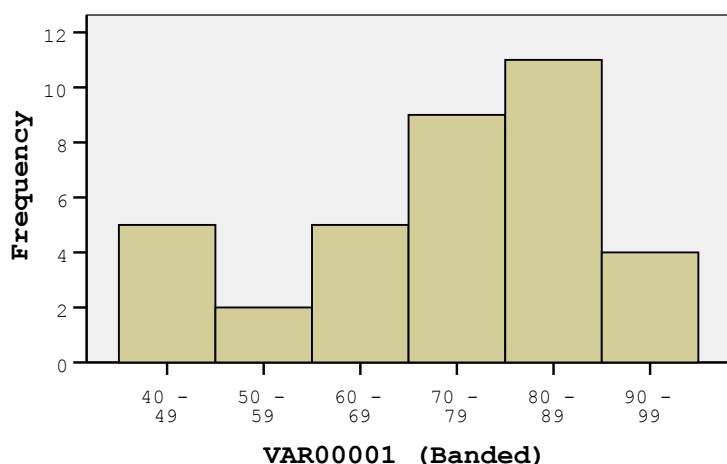
b)

	četnost	relativní četnost	kumulativní relativní četnost
40 - 49	5	13,9	13,9
50 - 59	2	5,6	19,4
60 - 69	5	13,9	33,3
70 - 79	9	25,0	58,3
80 - 89	11	30,6	88,9
90 - 99	4	11,1	100,0
Celkem	36	100,0	

8. a)



b)



9. sloupcový graf je grafickým znázorněním tabulky prostých četností (jedna hodnota – jeden sloupec), je vhodný pro diskrétní data na všech úrovních měření; histogram je grafickým znázorněním tabulky intervalových četností (jeden interval hodnot – jeden sloupec), je vhodný pro spojitá data na přinejmenším intervalové úrovni

10.1 tabulka četností skóre z tělesné výchovy:

Skór	Četnosti	Kumulativní četnost	Procenta	Kumulativní procenta
62	1	1	4,5	4,5
66	1	2	4,5	9,1
68	1	3	4,5	13,6
70	1	4	4,5	18,2
71	1	5	4,5	22,7
72	1	6	4,5	27,3
73	1	7	4,5	31,8
74	2	9	9,1	40,9
76	1	10	4,5	45,5
77	1	11	4,5	50
78	1	12	4,5	54,5
79	3	15	13,6	68,2
80	2	17	9,1	77,3
81	1	18	4,5	81,8
83	1	19	4,5	86,4
87	1	20	4,5	90,9
96	1	21	4,5	95,5
98	1	22	4,5	100
Celkem	22		100	

tabulka intervalových četností:

Interval	Četnost	Kumulativní četnost	Procenta	Kumulativní procenta
60 – 64	1	1	4,5	4,5
65 – 69	2	3	9,1	13,6
70 – 74	6	9	27,3	40,9
75 – 79	6	15	27,3	68,2
80 – 84	4	19	18,2	86,4
85 – 89	1	20	4,5	90,9
95 – 99	2	22	9,1	100

Celkem	22		100	
---------------	-----------	--	------------	--

10.2 a) matematika, b) tělesná výchova, c) anglický jazyk, d) český jazyk

10.3 matematika, histogram a)

10.4 anglický jazyk, histogram c)

Z příčin je nejpravděpodobnější c), rozdělení do dvou skupin. Ale i b) může způsobit tento efekt, pokud je počet doučujících se žáků velký a doučování je efektivní.

10.5 Tvrzení d) je správné.

10.6 c) Efekt stropu. Neboli, test byl velmi snadný, a protože má maximální možnou hranici získaných bodů, bude většina žáků u této hranice.

10.7 d). Přesněji jako outliers označujeme skór, který leží od horního či dolního kvartilu dále, než 3/2 interkvartilového rozpětí.

10.8 b) v tělesné výchově.

10.9 b) sloupcový diagram.

Boxplot není grafickou podobou tabulky četností, histogram se používá pro intervalové a poměrové úrovně měření s mnoha různými hodnotami a koláčový diagram se téměř nepoužívá, navíc ukazuje pouze relativní četnosti.

11.1 a) Poissonovo rozdělení ... popisuje rozložení řídkých, málo častých událostí

11.2 $\lambda = 2$ Lambda je frekvence za jednotku času.

11.3 b) normálním rozdělením ... pokud událost nastává častěji než desetkrát za sledovanou jednotku času, pak se dá aproximovat normálním rozdělením.

12.1 a) diskrétní uniformní. Diskrétní znamená, že rozložení nemůže nabývat všech hodnot, ale jen určitých vymezených – zde pouze jednotlivá čísla na kostce. Uniformní je rozložení, které nabývá všech hodnot se stejnou pravděpodobností.

12.2 d) normální. Součty nezávislých, stejně rozdělených veličin (a to dvě různé kostky jsou) se blíží k normálnímu rozdělení.

13. c) normální. Pokud je veličina daná velkým množstvím nezávislých vlivů, pak její rozdělení je přibližně normální – proto je normálně rozložena i například inteligence nebo výška.

14. Spojitost či diskrétnost se pouze na základě histogramů nedá posoudit. Ovšem rozdělení reprezentovaná sloupcovým grafem (poslední tři) nutně musí být diskrétní. U prvních devíti histogramů tedy nebude uvedeno zdali se jedná o diskrétní či spojitě rozdělení, k tomu bychom potřebovali znát měřenou veličinu a postup jejího měření.

14.1 unimodální, normální – podobně jsou rozděleny například vlastnosti jako inteligence nebo výška, toto rozdělení je velmi časté – proto se nazývá normální.

14.2 unimodální, zešikmené zleva, negativně zešikmené – poslední vlastnosti jsou shodné, zešikmení zleva a negativní zešikmení jsou synonyma.

14.3 bimodální – neboli má dva vrcholy.

14.4 unimodální, leptokurtické – leptokurtické rozdělení je docela podobné normálnímu, pouze je špičatější.

14.5 unimodální, platykurtické – platykurtické rozdělení je naopak méně špičaté, jakoby tlustší.

14.6 unimodální, zešikmené zprava, pozitivně zešikmené – zešikmené zprava je opět totožné s pozitivním zešikmením.

14.7 bimodální

14.8 multimodální – má více než dva vrcholy

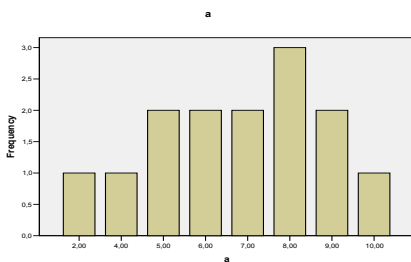
14.9 uniformní – vzorové uniformní rozdělení by mělo tvar obdélníku, náhodná data z uniformního rozdělení mohou vypadat i takto. Uniformní rozdělení nabývá každé možné hodnoty se stejnou pravděpodobností.

14.10 uniformní, diskretní – toto rozdělení představuje náhodná data při házení kostkou. Každá ze šesti hodnot může padnout se stejnou pravděpodobností, jiné hodnoty nejsou možné.

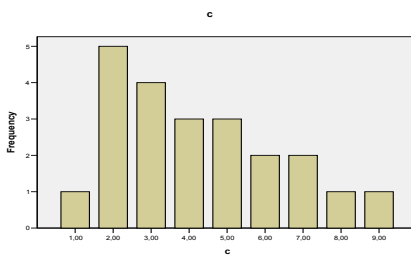
14.11 binomické, diskretní – vznikne například jako součet několika hodů mincí (hlava 1, lev 0; tedy součet alternativních rozdělení). V tomto případě ovšem byla pravděpodobnost jedné strany přibližně dvakrát tak velká, celkový výsledek je součtem osmi hodů.

14.12 alternativní, diskretní – rozdělení, které nabývá pouze dvou hodnot – jedné s pravděpodobností p a druhé s pravděpodobností $1 - p$. „Falešná“ mince z minulého rozdělení.

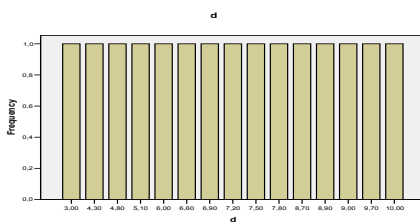
15. a. negativně zešikmené, unimodální



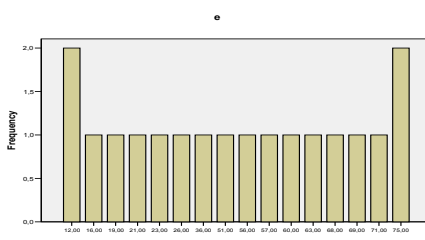
b. pozitivně zešikmené, unimodální



c. uniformní, symetrické

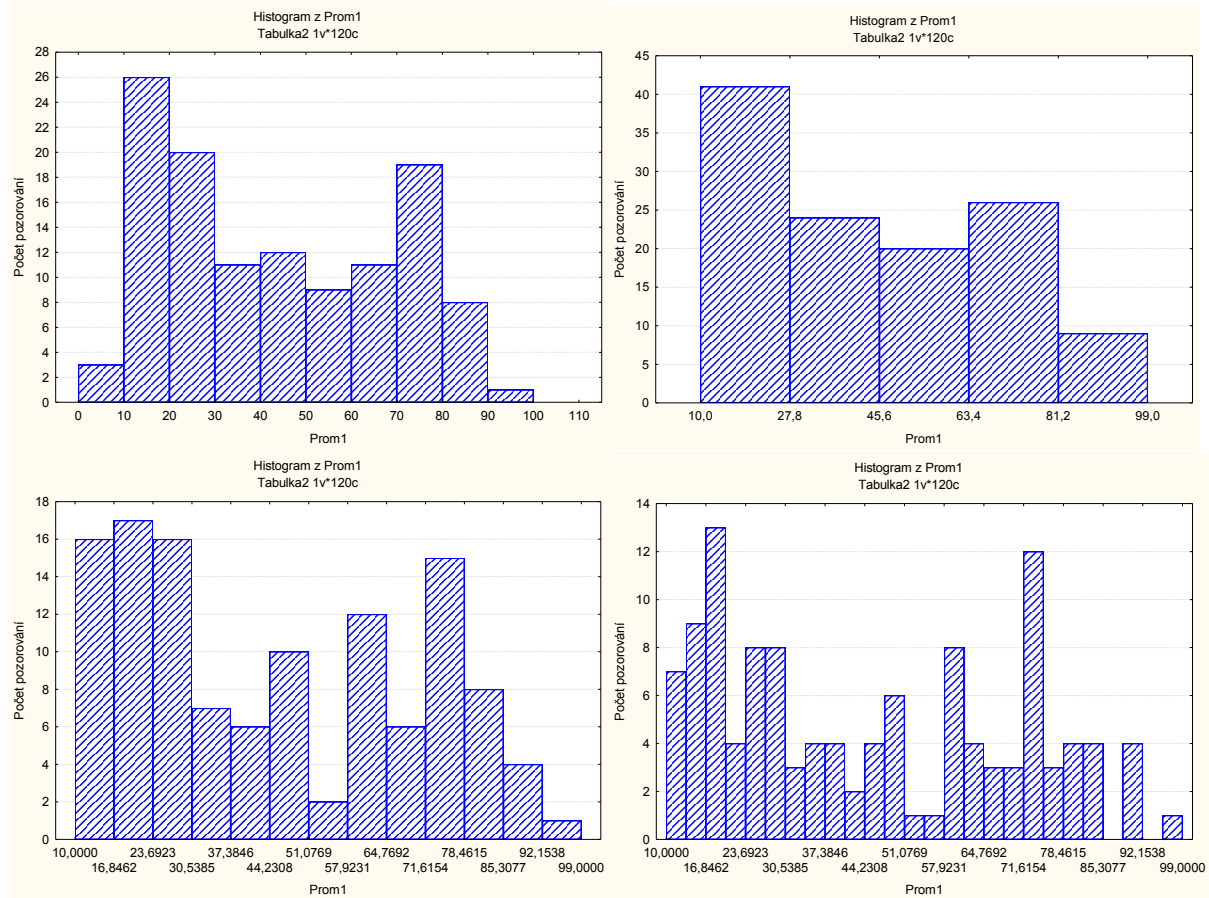


d. bimodální, symetrické

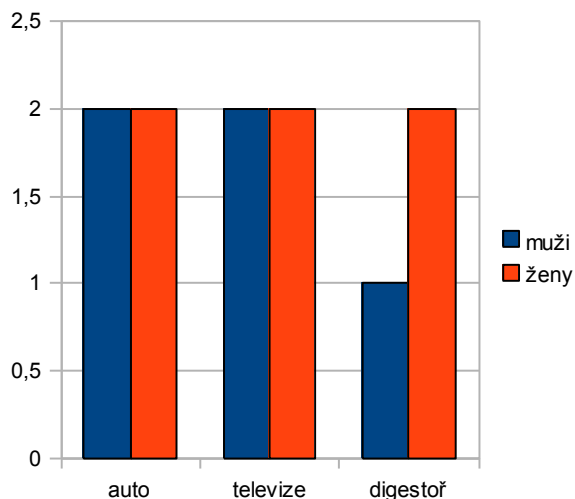


16. sloupcový diagram – jde o grafickou podobu tabulky četností, zachycuje diskrétní data; rozložení hodnot se zde zdá být blízké normálnímu. Zde vypadá velmi jako histogram, ale některé hodnoty na ose x chybí.

17. výsledný histogram se bude lišit podle zadané šířky intervalu



18. sloupcový diagram s tříděním



19.1 bimodální. Je to tím, že jde o data z dvoukohortového zrychleného longitudinálního výzkumu, tj. 2 skupiny – cca dvanáctiletí a šestnáctiletí.

19.2 Neplatí. Jedničky z matematiky se prudce vytrácí. Z češtiny ale také.

19.3 Zprava (pozitivně) zešikmené. Ve starší kohortě je méně zešikmené, u matematiky skoro symetrické. Já

sloupcový diagram.

19.4 Ordinální

19.5 pohlaví, kohorta, jedinec(ek)

19.6 Zleva (negativně) zešikmené. Hodně respondentů má s rodiči vřelý vztah. Čím je vztah horší, tím méně se vyskytuje (alespoň v našem vzorku).

19.7 Individualismus vznikl jako součtový skóre z několika málo položek dotazníku. Nenabývá mnoha hodnot, ale považujeme ho za spojitou škálu. Proto se histogram této proměnné při změně počtu intervalů dost výrazně vzhledově mění. U neg2 a depresivity to není tak výrazné. Usuzování na normalitu rozložení je dosti problematické. Snad jste si k tomu účelu nechali do histogramu zobrazit křivku normálního rozložení. Pro potřeby našeho kurzu bychom všechny proměnné v tomto úkolu mohli považovat za normálně rozložené.

20. Poissonovo

21. X-tý percentil je hodnota, pro kterou platí, že X % lidí (jevů) ve vzorku má/získalo tuto nebo menší hodnotu

22. v případě symetrického bimodálního rozložení

23. robustní (tj. např. pořadové)

24. b)

25.

Frequency	Stem & Leaf
3,00	0 . 334
14,00	0 . 55666667778899
2,00	1 . 02

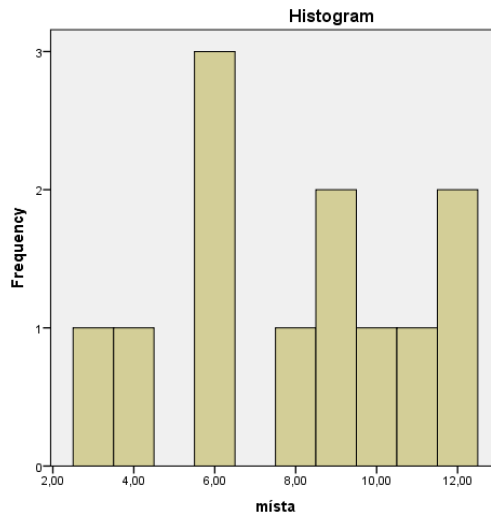
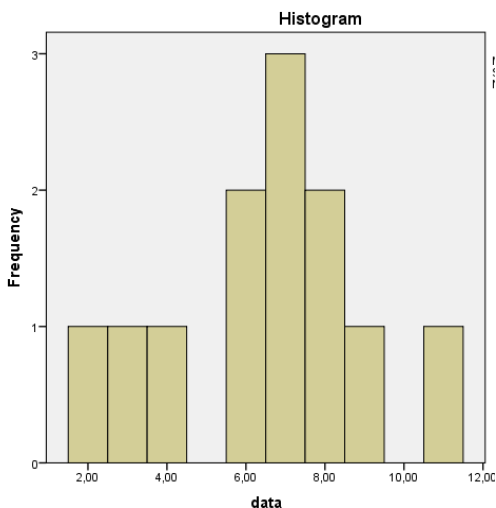
Stem width: 10,00
Each leaf 1 case(s)

nebo

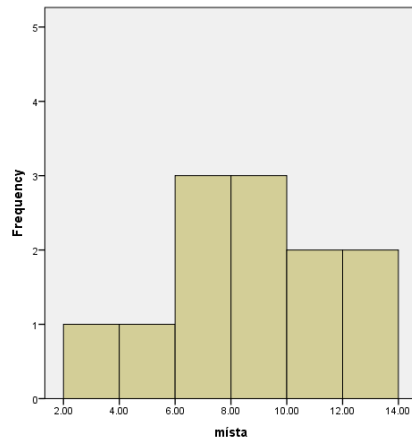
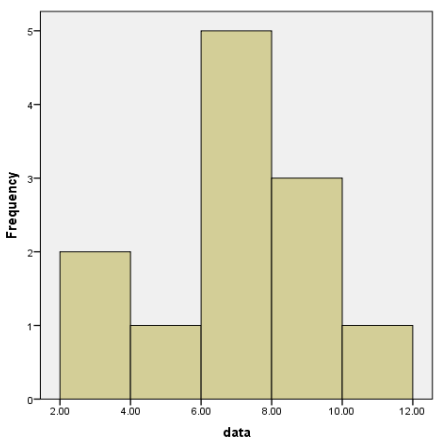
```
3 | 19
4 | 8
5 | 16
6 | 01699
7 | 258
8 | 7
9 | 07
10 | 02
```

O tvaru rozložení se zde uvažuje obtížně, protože dat je málo. Nicméně podle prvního diagramu lze konstatovat, že rozložení patrně bude unimodální a snad i celkem symetrické. Stojí za to si všimnout, že čím méně máme dat, tím výrazněji se proměňuje tvar rozložení při změně šířky intervalu.

26. Premennú „počet zapamatovaných míst“ a „počet zapamatovaných dat“ je najlepšie pre tento prípad zobrazit pomocou histogramu.



Pre zobrazenie dát je vhodné použiť histogram, nakoľko ide o pomerovú premennú. Histogramy vidíme na obrázkoch a všimame si aj vzniknuté „medzery“ (výskyt nulových frekvencií - četností pre viacero úrovní premennej. I keď aj z týchto histogramov je možné spraviť vizuálne a slovné porovnanie, pozrieme sa na dané histogramy s tým, že jednotlivé úrovne premennej zoskupíme do tried so šírkou 2.



Po určení šírky intervalu 2 sme znovu zostrojili histogramy. Ako vidíme z obrázkov, obe premenné sú unimodálne. Histogram „mista“ je symetrický naproti tomu histogram „data“ by sme ako symetrický označiť nemohli. V tomto histograme je zrejmé mierne zošikmenie zľava.

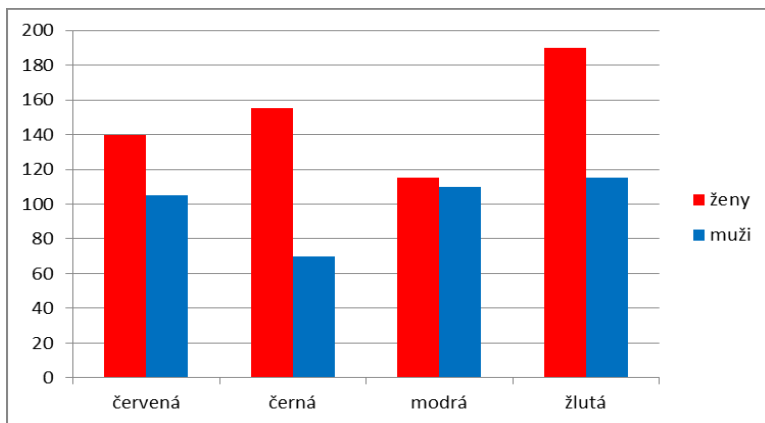
27. Os X v histograme znázorňuje rozličné úrovne skúmanej premennej (tieto hodnoty môžu reprezentovať hodnoty diskretných aj spojitých premenných; môžu byť reprezentované bodovými hodnotami ale aj intervalmi).

Os Y v histograme znázorňuje početnosť (četnosť) – frekvenciu výskytu jednotlivých hodnôt premennej.

28. Pre znázornenie rozloženia danej premennej je najvhodnejší graf s triedením podľa pohlavia. Pred samotným vyhotovením grafu je však nutné si upraviť dáta, a to zrátaním premenných vrátil/a a nechal/a si, nakoľko chceme zobraziť premennú „barva vybrané tužky“. Potom tabuľka bude mať tvar:

	barva vybrané tužky				Celkem
	červená	černá	modrá	žlutá	
ženy	140	155	115	190	600
muži	105	70	110	115	400
Celkem	245	225	225	305	1000

Z tejto tabuľky následne spravíme stĺpcový graf s triedením.

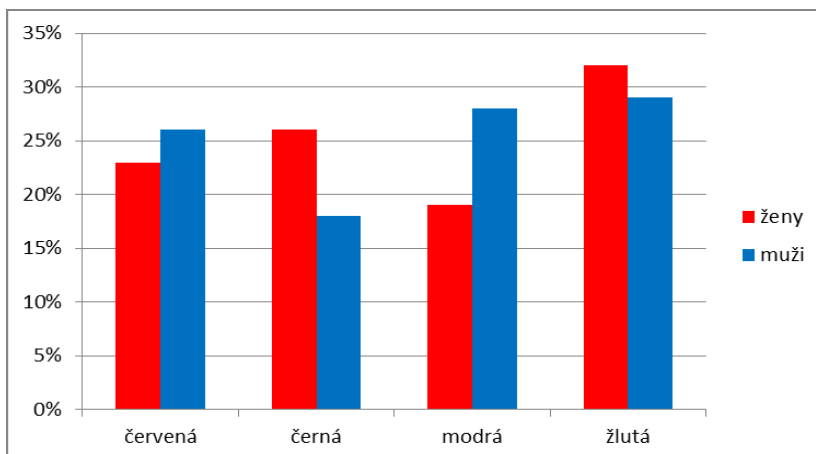


Ak sa pozrieme na tento stĺpcový graf, je vidno, že ženy výrazne prevyšujú mužov, čo spôsobuje to, že žien je viac vo výskumnom súbore ako mužov. Ako by sme teda zrovnocnili oba súbory pre zobrazenie grafom? Jednou z možností je zobraziť nie frekvencie (četnosti) ale percentá. Vyrátame percentá každej vybranej farby pre každý súbor.

Dostaneme tabuľku:

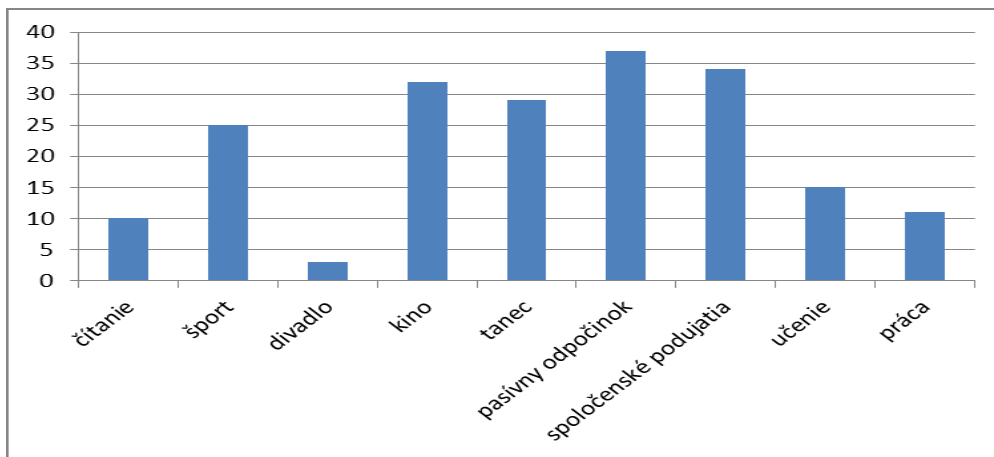
	barva vybrané tužky				Celkem
	červená	černá	modrá	žlutá	
ženy	140	155	115	190	600
%	23%	26%	19%	32%	100%
muži	105	70	110	115	400
%	26%	18%	28%	29%	100%
Celkem	245	225	225	305	1000

A graf z percentuálnej hodnoty bude vyzerat nasledovne:



Z daného grafu je porovnanie mužov a žien v rámci vybranej farby už naozaj zodpovedajúce realite. Vidíme, že muži si viac vyberajú červenú farbu a modrú farbu než ženy a ženy si viac vyberajú čiernu a žltú farbu než muži.

29. Pre zobrazenie dát je vhodný stĺpcový graf, nakoľko ide o nominálnu premennú. Jeho „vzhľad“ môže byť nasledovný: (iný spôsob zobrazenia stĺpcovým grafom by mohol byť napríklad ak by sme zoradili jednotlivé kategórie podľa frekvencie od najmenej po najväčšiu a naopak)



30. Napríklad by stem and leaf vyzeral takto:

```
.00 0 .
9.00 0 . 567888999
10.00 1 . 0111222244
12.00 1 . 555777778888
8.00 2 . 01123444
5.00 2 . 56779
3.00 3 . 013
.00 3 .
1.00 4 . 2
1.00 Extremes (>=48)
```

Stem width: 10.00

Each leaf: 1 case(s)

Rozloženie premennej sa z daného histogramu zdá unimodálne, zprava zošikmené a s dvomi odľahlými hodnotami.

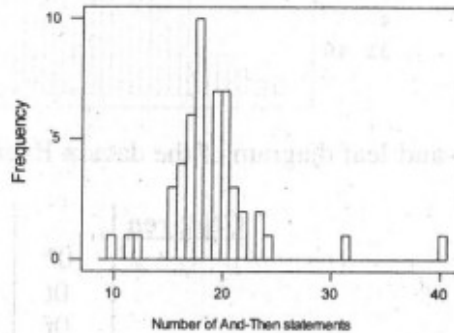
31. z Howella

2.1 Children's recall of stories:

a. Children's "and then...s" Frequency

Children's "and then...s"	Frequency
10	1
11	1
12	1
15	3
16	4
17	6
18	10
19	7
20	7
21	3
22	2
23	2
24	1
31	1
40	1

Distribution of And-Then statements for children



b. unimodal and positively skewed

2.3 The problem with making a stem-and-leaf display of the data in Exercise 2.1 is that almost all the values fall on only two leaves if we use the usual 10s' digits for stems.

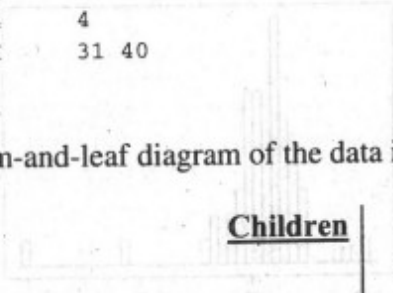
Stem	Leaf
1	0125556666777777888888888889999999
2	000000011122334
3	1
4	0

And things aren't much better even if we double the number of stems.

Stem	Leaf
1*	012
1.	555666677777788888888889999999
2*	000000011122334
2.	
3*	1
3.	
4*	0

Best might be to use the units digits for stems and add HI and LO for extreme values

Stem	Leaf
5	555
6	6666
7	7777777
8	88888888888
9	9999999
10	0000000
11	111
12	22
13	33
14	4
HI	31 40

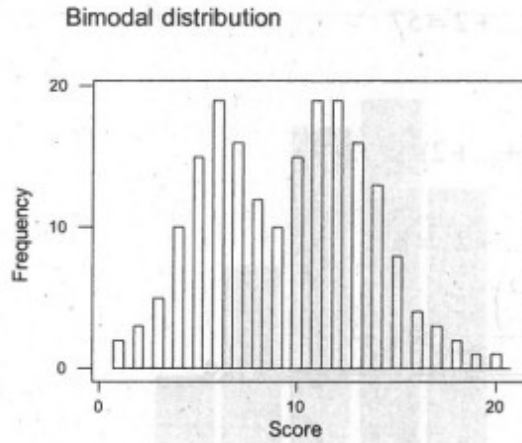


2.5 Stem-and-leaf diagram of the data in Exercises 2.1 and 2.4:

<u>Children</u>		<u>Adults</u>
	0*	1
	0t	34
	0f	55
	0s	7777
	0.	8888999999999
10	1*	000000001111111
2	1t	222223
555	1f	4444555
77777776666	1s	667
77777778888888888	1.	
1110000000	2*	
3322	2t	
4	2f	
	2s	
	2.	
40 31	Hi	

2.7 Invented bimodal data:

Score	Freq
1	2
2	3
3	5
4	10
5	15
6	19
7	16
8	12
9	10
10	15
11	19
12	19
13	16
14	13
15	8
16	4
17	3
18	2
19	1
20	1



2.11 The shape of the distribution of number of movies attended per month for the next 200 people you met would be positively skewed with a peak at 0 movies per month and a sharp drop-off to essentially the baseline by about 5 movies per month.

2.13 Stem-and-leaf for ADDSC

Stem	Leaf
2.	69
3*	0344
3.	56679
4*	00023344444
4.	55666778888999999
5*	00000000011223334
5.	55677889
6*	00012234
6.	55556899
7*	0024
7.	568
8*	
8.	55