Otázky k tématu 2 – četnosti, rozložení

1. Pokud je proměnná nominální (např. etnicita), preferovanou grafickou podobou tabulky četností je
   a) sloupcový diagram
   b) histogram

2. Která z následujících rozložení jsou pravděpodobně negativně zešikmená?
   a) rodinný příjem v Kč za měsíc
   b) věk při promoci na VŠ
   c) počet obyvatel měst v ČR
   d) skóry za lehký test

3. Histogram
   a) je podobný čárovému grafu
   b) obsahuje normální křivku
   c) je grafickou podobou tabulky četností
   d) znázorňuje vztah mezi dvěma proměnnými

4. Představte si, že určitý test z matematiky absolvuje 50 lidí – 25 matematiků a 25 lidí, kterým matematika nejde. Jaké rozložení nejspíše získáme?
   a) unimodální
   b) bimodální
   c) normální
   d) zešikmené

5. Rozložení následujících skórů je…
   1, 10, 6, 8, 7, 5, 5, 4, 9, 2, 9, 8, 6, 7, 8, 3, 4, 3, 5, 5, 7, 6, 4, 6, 6, 7
   a) unimodální a přibližně normální
   b) bimodální a negativně zešikmené
   c) normální a pozitivně zešikmené
   d) normální a negativně zešikmené

6. Jak nazýváme symetrické rozložení, v němž mají všechny hodnoty přibližně stejnou absolutní četnost?

7. Níže jsou uvedena data – výsledné skóry z jedné písemky. Vytvořte tabulku četností včetně kumulativních
a) tabulku prostých četností jednotlivých hodnot
b) tabulku intervalových četností při šířce intervalu 10

40	98	63	90	70	60	45	43	78
67	56	54	78	87	43	90	81	81
77	80	79	80	81	66	75	88	84
49	63	78	79	80	92	89	84	77

8. Zkonstruujte histogram ke každé tabulce četností vytvořené v úkolu 7.

9. Čím se liší sloupcový graf od histogramu?

10. Základní škola v Horních Kotěhůrkách se rozhodla zhodnotit kvalitu své výuky, a tak nechala všechny žáky otestovat ze čtyřech předmětů. V každém testu bylo možno získat 0 až 100 bodů, čím více, tím lepší výsledek. Bodové zisky žáků ze VII.B jsou shrnuty v tabulce.

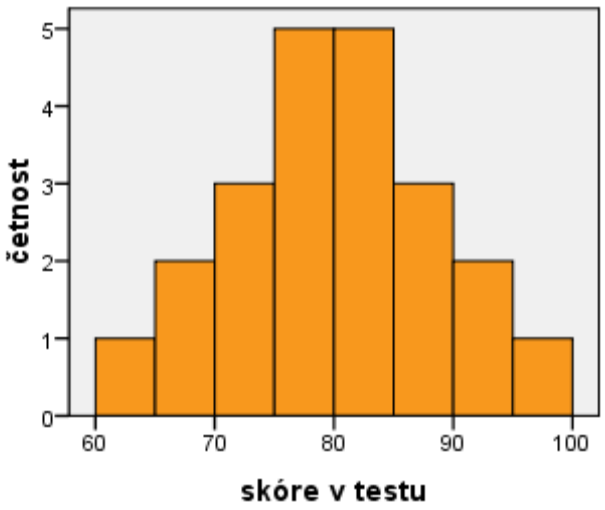
id žáka	matematika	český jazyk	anglický jazyk	tělesná výchova
1	92	94	92	83
2	77	98	97	74
3	86	89	67	68
4	70	96	89	79
5	81	91	73	96
6	68	99	76	80
7	77	94	66	87

8	61	83	91	66
9	71	90	94	72
10	82	97	86	77
11	86	95	74	70
12	79	99	81	76
13	65	82	66	73
14	80	97	62	62
15	90	77	69	98
16	81	89	88	81
17	97	93	71	79
18	76	99	90	71
19	75	90	85	78
20	84	95	69	74
21	72	92	73	80
22	89	88	91	79

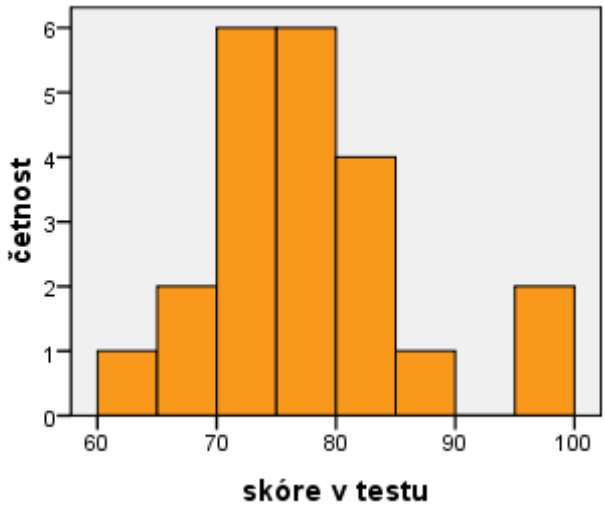
10.1 Vytvořte tabulku četností včetně kumulativních pro skóry z tělesné výchovy. Dále vytvořte tabulku intervalových četností při šířce intervalu 5.

10.2 Zde jsou histogramy jednotlivých předmětů. Který histogram patří ke kterému předmětu?

a)

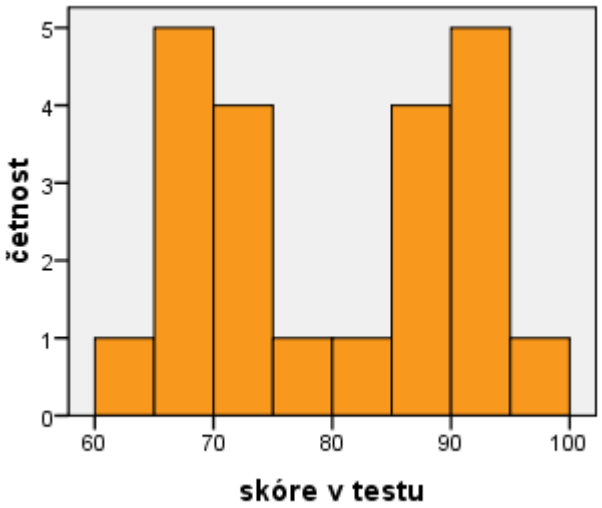


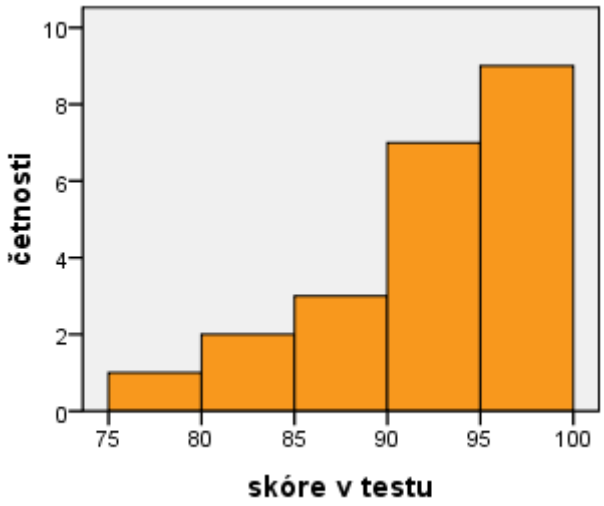
b)



c)

d)





10.3 Který z histogramů je nejpodobnější normálnímu rozdělení?

10.4 Který z histogramů **neodpovídá** unimodálnímu rozdělení? A která z následujících příčin to pravděpodobně způsobila?

a) Test z daného předmětu byl velmi snadný.
b) Pár žáků se navzájem doučuje a jsou v daném předmětu zdatnější.
c) Žáci jsou v daném předmětu rozděleni na dvě skupiny, začátečníky a pokročilé.
d) Předpoklady pro daný předmět jsou v populaci přibližně normálně rozděleny, žáci jen reprezentují populaci.

10.5 Rozdělení skórů z českého jazyka (histogram d) je zešikmené. Které z následujících tvrzení o něm je správně?

a) Je zešikmené zprava, neboli jde o kladné zešikmení.
b) Je zešikmené zprava, neboli jde o záporné zešikmení.
c) Je zešikmené zleva, neboli jde o kladné zešikmení.
d) Je zešikmené zleva, neboli jde o záporné zešikmení.

10.6 Efekt, který to zapříčinil, se nazývá:

a) Efekt podlahy.
b) Gaussův efekt.
c) Efekt stropu.
d) Efekt taháku.

10.7 Jako outliera bychom nejspíše nazvali:

a) Hodnotu, které nikdo nedosáhl.
b) Žáka, který měl v daném testu nejnižší skór.
c) Oblast mezi dvěma mody (vrcholy).
d) Skór, který je daleko od většiny ostatních skórů.

10.8 Ve kterém z předmětů lze najít outliery?

a) Ve všech předmětech.
b) V tělesné výchově.
c) V anglickém jazyce.
d) Ve všech kromě českého jazyka.

10.9 Učitel z matematiky se rozhodl testu využít a žáky oznámkoval:

100 – 91 ... známka 1
90 – 81 ... 2
80 – 71 ... 3
70 – 61 ... 4
a zbytek za 5.

Pokud by si napsal tabulku četností známek a chtěl ji lépe znázornit, jako její grafickou podobu by pravděpodobně zvolil:

a) boxplot
b) sloupcový diagram
c) histogram
d) koláčový diagram

11. Paní Votýpková provozuje v Horních Kotěhůrkách obchod se smíšeným zbožím. Vzhledem k tomu, že zákazníků chodí málo, má hodně času, a tak si aspoň začala zapisovat, kdo kdy přišel. Každý den má otevřeno 6 hodin a průměrně přijde každý den 12 zákazníků v náhodných časech.

11.1 Kdybychom u ní v obchodě dělali měsíc pozorování tím způsobem, že bychom si vzali vzorek 50 náhodně vybraných hodin z celkové otvírací doby a zaznamenávali počet zákazníků za každou sledovanou hodinu, pak by rozdělení proměnné „zákazníků za hodinu“ mělo přibližně tvar:

a) Poissonova rozdělení
b) Gaussova rozdělení
c) chí-kvadrátu
d) normálního rozdělení

11.2 Jaká bude hodnota λ tohoto rozdělení?

11.3 Pokud bychom namísto toho uvažovali rozdělení počtu zákazníků za celý den, dalo by se dost přesně nahradit:

a) chí-kvadrátem
b) normálním rozdělením
c) bimodálním rozdělením
d) uniformním rozdělením

12. Karel je velmi náruživým hazardním hráčem a v kostkách již prohrál značnou část svého majetku. Aby mohl hrát sofistikovaněji, rozhodl se si něco o pravděpodobnostech při hře v kostky zjistit. Po celou dobu budeme uvažovat klasické hrací kostky s šesti čísly, přitom pravděpodobnosti hodu jednotlivých čísel jsou stejné.

12.1 Jaké je rozdělení výsledku jednoho hodu hrací kostkou?

a) diskrétní uniformní
b) leptokurtické
c) spojité normální
d) poissonovo

12.2 V následujícím pokusu si Karel pokaždé hodí 6 kostkami a hozená čísla sečte. Jakému rozdělení bude nejvíce podobné rozdělení tohoto součtu?

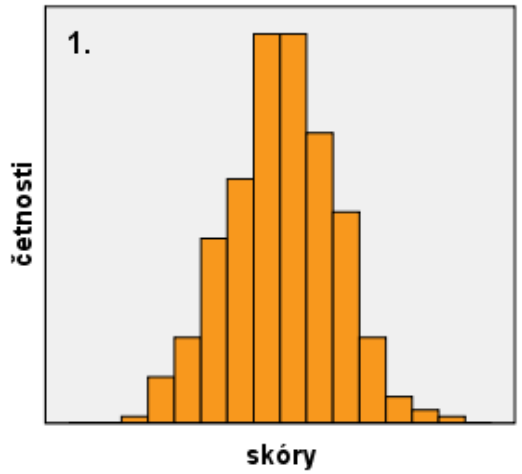
a) spojité uniformní
b) poissonovo
c) zešikmené zprava
d) normální

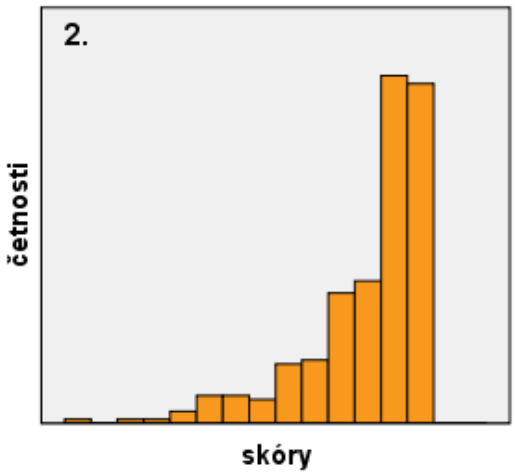
13. Pokud bych se ptal náhodných kolemjdoucích, jaký je můj věk, jaké by bylo rozdělení asi jejich odhadů?

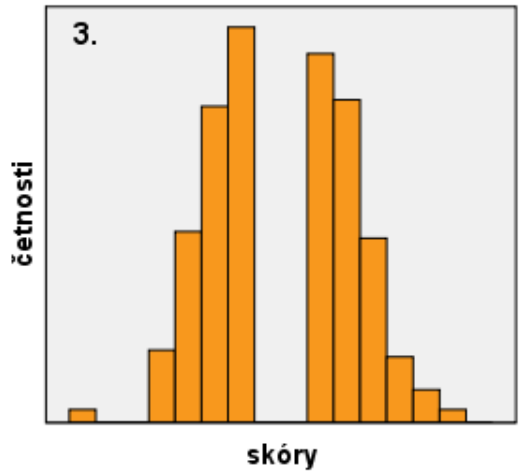
a) uniformní
b) bimodální
c) normální
d) poissonovo

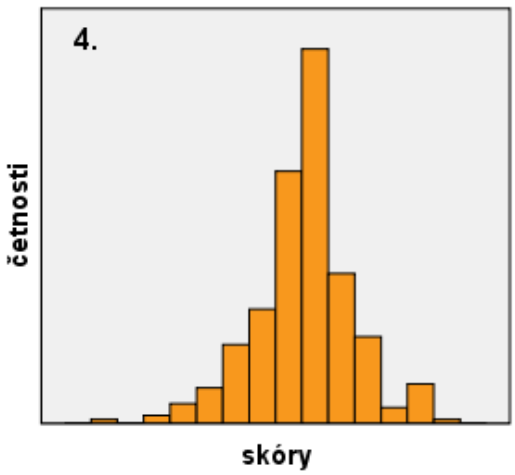
14. Následující histogramy, popřípadě sloupcové grafy, znázorňují různá rozložení. Každé z nich je možné popsat některými vlastnostmi z následujícího výběru, zpravidla několika z nich. Přiřaďte tedy ke každému grafu patřičné vlastnosti (jednu vlastnost je možno přiřadit k více grafům):

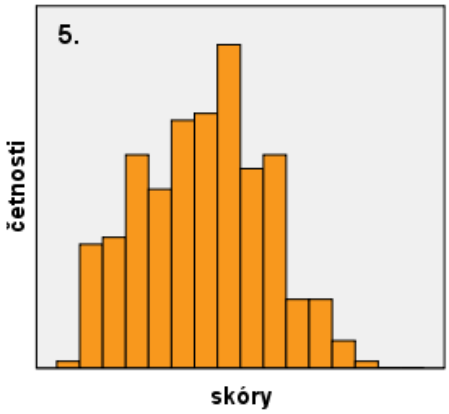
unimodální	bimodální	multimodální	
zešikmené zprava	zešikmené zleva	pozitivně zešikmené	negativně zešikmené
uniformní	normální	platykurtické	leptokurtické
alternativní	binomické		
spojité	diskrétní		

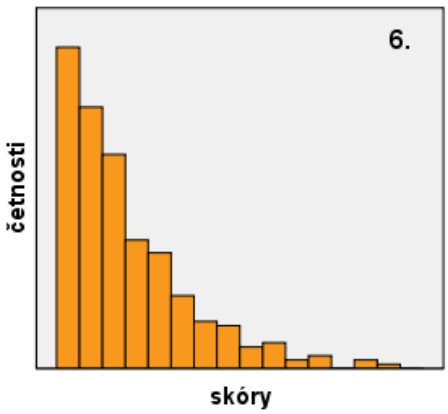

1.
četnosti
skóry

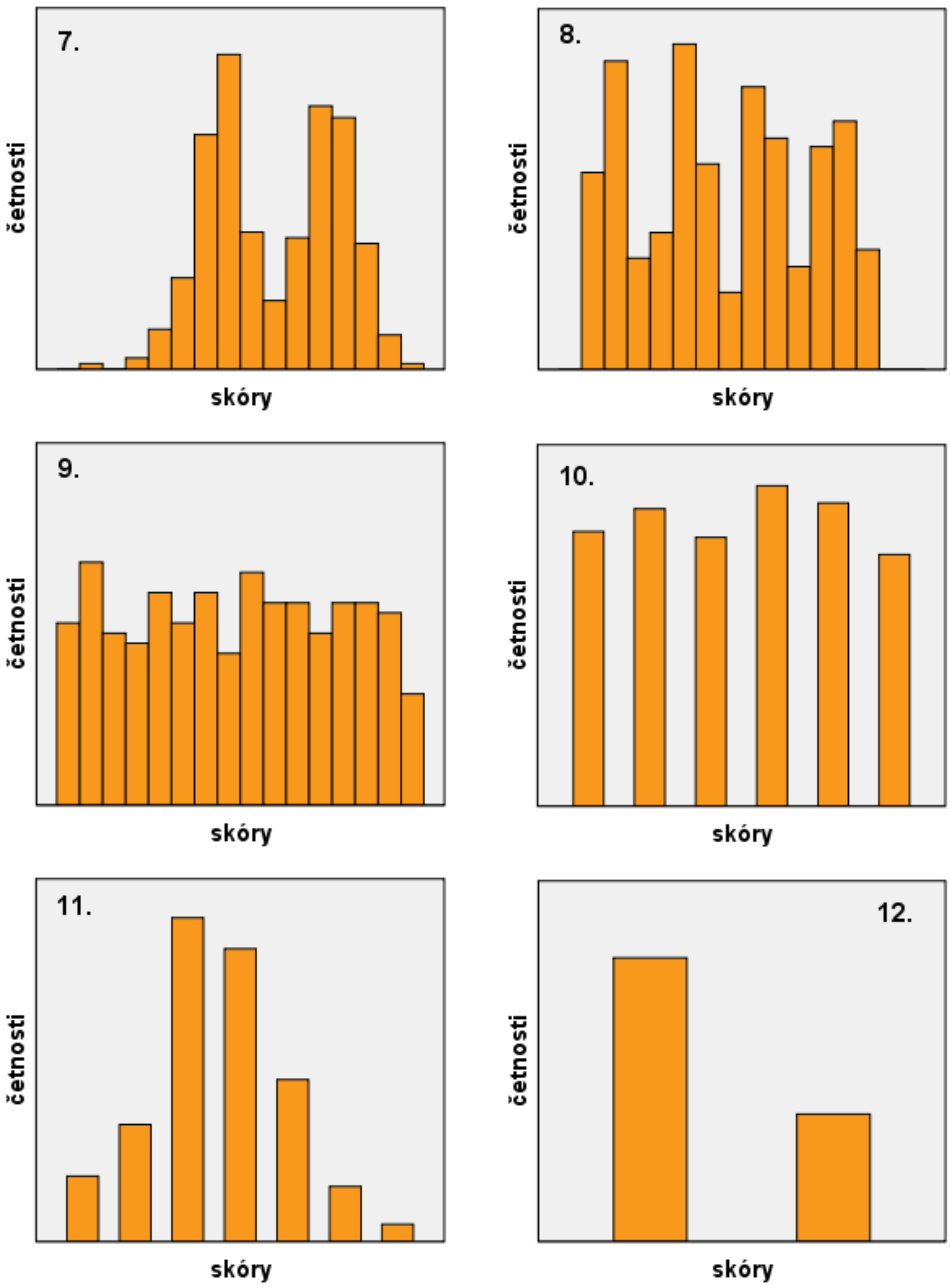

2.
četnosti
skóry


3.
četnosti
skóry


4.
četnosti
skóry


5.
četnosti
skóry

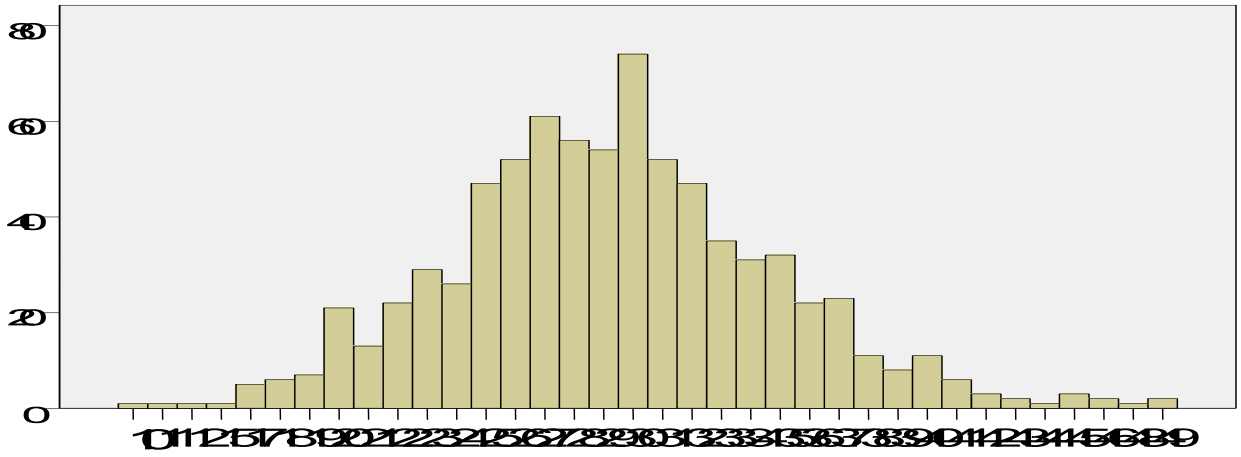

6.
četnosti
skóry



15. Následujících pět řad skórů představuje výsledky v různých testech. Graficky je znázorněte a popište tvar jejich rozložení:

a. 10; 6; 8; 7; 5; 5; 4; 9; 2; 9; 8; 6; 7; 8
b. 2; 9; 5; 1; 2; 4; 2; 6; 7; 2; 8; 5; 3; 4; 7; 2; 3; 5; 4; 3; 6; 3
c. 10,0; 9,7; 9,0; 8,9; 8,7; 7,8; 7,5; 7,2; 6,9; 6,6; 6,0; 5,1; 4,8; 4,3; 3,0
d. 12; 12; 16; 19; 21; 23; 26; 36; 51; 56; 57; 60; 63; 68; 69; 71; 75; 75

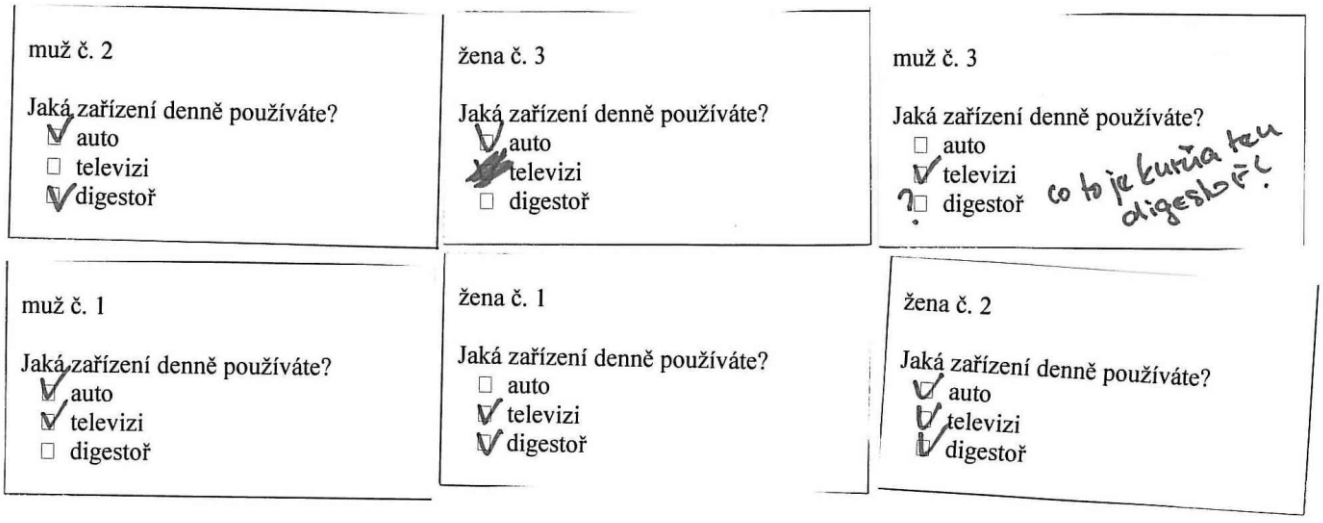
16. Pozorně si prohlédněte graf či diagram na následujícím obrázku. Jak se jmenuje? Podle čeho tak soudíte?



17. Ve statistickém programu své volby vytvořte tabulku četností a histogram následujících dat (řekněme opět, že jde o výsledné skóry z písemky). Zkuste si pohrát s různou šířkou intervalu v histogramu.

15	12	13	14	10	15	30	12	17	15	15	30
16	17	28	19	22	25	10	19	32	11	22	32
14	43	32	20	25	29	19	18	29	10	18	39
30	35	19	29	47	25	25	45	16	75	60	25
74	55	18	70	50	20	40	50	45	60	40	62
62	89	61	72	90	65	85	80	60	45	22	49
35	18	49	25	30	59	50	78	35	60	75	39
60	70	25	53	74	74	43	74	72	70	90	75
75	99	77	75	89	60	67	80	80	64	77	82
68	85	80	63	82	75	48	34	16	17	22	25

18. V rámci dotazníkového šetření jsme nasbírali mj. šest následujících dotazníků - od tří mužů a od tří žen. Pokud to potřebujete vědět, vězte, že cílem výzkumu je podívat se na genderové rozdíly v používání přístrojů. Jakou grafickou podobu tabulky četností byste použili pro prezentaci těchto dat? Načrtněte ji prosím.



19. Následující úlohy se týkají datového souboru „datička.xls“ (též datička.sav). [1]

19.1 Jaké je rozložení proměnné věk [vekr] v Datíčcích? (Daticka.xls) Čím by to mohlo být?

[1] Jde o cca 10 let stará, avšak reálná data z výzkumu Long2. Když v nich spočítáte něco zajímavého, rád o tom uslyším. Více viz Macek, P. , Mareš, J., Ježek, S., Valášková, M. Současný výzkum adolescentů: východiska, soubor, metody. In V. Smékal, P. Macek (Eds.), Utváření a vývoj osobnosti. Psychologické, sociální a pedagogické aspekty. Brno: Barrister & Principal, 2002. s. 129-146. ISBN 80-85947-83-8.

19.2 Ve mladší kohortě má více dětí jedničku z matematiky než jedničku z češtiny. Platí to i pro starší kohortu?

19.3 Jaké rozložení mají známky z matematiky i z češtiny v mladší kohortě? Liší se to nějak od starší kohorty? Jaké zobrazení jste k odpovědi použili?

19.4 Na jaké škále zde měříme proměnnou očekávané vzdělání? N O I P

19.5 V Datíčcích jsou 3 dichotomické proměnné, které to jsou?

19.6 Jaký tvar mají rozložení četností vřelostí rodičů? Jak by se dalo popsat běžnými slovy?

19.7 Pohrejte si s histogramem proměnné Individualismus v mladší kohortě. Vyzkoušejte různé počty (šíře) intervalů a klaďte si otázku, zda byste rozložení proměnné individualismus považovali za normální. Pak si zkuste totéž s proměnnou neg2 a depresivita ve starší kohortě.

20. Jaký typ rozložení mají obvykle řídké události?

21. Co je to percentil?

22. Je-li rozložení symetrické, obvykle se modus výrazně od průměru neliší. Za jaké podmínky to neplatí?

23. Liší-li se podstatně rozložení proměnné od normálního rozložení, klesá výpovědní hodnota popisných statistik, které nejsou …

24. Badatel pracuje s proměnnou „počet předmětů za celé studium, u nichž student musel opakovat zkoušku“. Tuto proměnnou prezentuje následující tabulkou s popisnými statistikami:

proměnná	N	rozpětí	M	SD
pocet_opak_predm	897	20	2,1	2,5

Rozložení proměnné je patrně
a) normální
b) pozitivně zešikmené
c) bimodální
d) uniformní
e) negativně zešikmené

25. Pro následující data proměnné „věk“ vytvořte diagram stonek a list (stem-and-leaf) a popište jejich rozložení.
10,2; 10,0; 9,7; 9,0; 8,9; 8,7; 7,8; 7,5; 7,2; 6,9; 6,9; 6,6; 6,1; 6,0; 5,6; 5,1; 4,8; 3,9; 3,1

26. Studentka Eva psala diplomku o autobiografické paměti manželů. Chtěla se dozvědět, co je pravdy na tom, že ženy si pamatují více z historie vztahu než jejich manželé. Ptala se jich tedy na to, kdy a kde se seznámili, kdy a kde si dali první pusu, kdy a kde měli poprvé sex apod. Celkem se každého zeptala na 12 takových událostí a zaznamenala si, u kolika z nich si pamatovali datum a u kolika místo. Také se ptala na délku současného vztahu v letech. Celkově to dalo dost práce, a tak máme data zatím ze 6 rodin:

Janů

Délka vztahu = 4	
Manželka	Manžel
data=9	data=3
místa=12	místa=6

Petrů

Délka vztahu = 5	
Manželka	Manžel
data=7	data=2
místa=9	místa=9

Vojtů

Délka vztahu = 8	
Manželka	Manžel
data=7	data=8
místa=3	místa=12

Jirků

Délka vztahu = 11	
Manželka	Manžel
data=6	data=7
místa=6	místa=11

Mirků

Délka vztahu = 14	
Manželka	Manžel
data=8	data=6
místa=4	místa=6

Nechoďdomů

Délka vztahu = 18	
Manželka	Manžel
data=11	data=4
místa=8	místa=10

Zajímá nás, jestli je rozložení (rozdělení) proměnných „počet zapamatovaných míst“ a „počet zapamatovaných dat“ podobné. Pokuste se obě rozložení graficky *zobrazit* tak, abychom je mohli vizuálně porovnat, a slovně je porovnejte.

27. Čo znázorňuje os X a os Y v histograme?

28. Profesor Šikula se zabývá atraktivitou barev. Uspořádal experiment, v němž mělo 1000 studentů vyplňovat tužkou škrtací test. Každý student dostal test a z velké nádoby si vybral tužku. Tužky byly nalakované různými barvami. Místo vyhodnocování škrtacích testů se profesor zaměřil na to, jakou barvu tužky si kdo vybral, a také na to, zda tužku vrátil (nebo si ji nechal). Výsledky celého experimentu jsou uvedeny v následující složené kontingenční tabulce:

		barva vybrané tužky				
		červená	černá	modrá	žlutá	Celkem
ženy	vrátila	120	130	90	170	510
	nechala si	20	25	25	20	90
muži	vrátil	85	45	90	100	320
	nechal si	20	25	20	15	80
	Celkem	245	225	225	305	1000

Graficky znázorněte rozložení proměnné *barva vybrané tužky* tak, abychom mohli srovnat rozložení této proměnné u mužů a u žen.

29. 196 študentov odpovedalo na otázku, jako najčastejšei trávia voľný čas v sobotu,. Výsledky sumarizuje tabulka:

kategória	čítanie	šport	divadlo	kino	tanec	pasívny odpočinok	spoločenské podujatia	učenie	práca
f	10	25	3	32	29	37	34	15	11

Premyslite si a následne vykonajte optimálnu reprezentáciu údajov. Zdôvodnite svoje rozhodnutie.

30. Nasledujúci súbor údajov vyjadruje počet bodov v teste dosiahnutý jednotlivými riešiteľmi.
Údaje zosumarizujte do podoby histogramu a popíšte rozloženie premennej.

12 23 11 7 27 12 9 9 11 33 17 14 5 12 22 24 24 17 10 8 14 26 21 24 31
18 20 17 48 15 8 17 30 29 8 11 17 18 18 6 9 21 25 18 12 15 15 27 42

31. Z Howella. Úkoly žádající stem-and-leaf diagram raději zkuste naplnit pomocí histogramu.

Many of the following exercises can be solved using either computer software or pencil and paper. The choice is up to you or your instructor. Any software package should be able to work these problems. Some of the exercises refer to a large data set named ADD.dat, which is available at www.uvm.edu/~dhowell/methods8/DataFiles/Add.dat. These data come from an actual research study (Howell and Huessy, 1985). The study is described in Appendix: Data Set on page (686).

2.1 Any of you who have listened to children tell stories will recognize that, unlike adults, they tend to recall stories as a sequence of actions rather than as an overall plot. Their descriptions of a movie are filled with the phrase “and then. . .” An experimenter with supreme patience asked 50 children to tell her about a given movie. Among other variables, she counted the number of “and then . . .” statements, which is the dependent variable. The data follow:

18 15 22 19 18 17 18 20 17 12 16 16 17 21 23 18 20 21 20 20 15 18 17 19 20
23 22 10 17 19 19 21 20 18 18 24 11 19 31 16 17 15 19 20 18 18 40 18 19 16

a. Plot an ungrouped frequency distribution for these data.

b. What is the general shape of the distribution?

2.2 Create a histogram for the data in Exercise 2.1 using a reasonable number of intervals.

2.3 What difficulty would you encounter in making a stem-and-leaf display of the data in Exercise 2.1?

2.4 As part of the study described in Exercise 2.1, the experimenter obtained the same kind of data for 50 adults. The data follow:

10 12 5 8 13 10 12 8 7 11 11 10 9 9 11 15 12 17 14 10 9 8 15 16
10 14 7 16 9 1 4 11 12 7 9 10 3 11 14 8 12 5 10 9 7 11 14 10 15 9

a. What can you tell just by looking at these numbers? Do children and adults seem to recall stories in the same way?

b. Plot an ungrouped frequency distribution for these data using the same scale on the axes as you used for the children's data in Exercise 2.1.

c. Overlay the frequency distribution from part (b) on the one from Exercise 2.1.

2.5 Use a back-to-back stem-and-leaf display (see Figure 2.5) to compare the data from Exercises 2.1 and 2.4.

2.6 Create a positively skewed set of data and plot it.

2.7 Create a bimodal set of data that represents some actual phenomenon and plot it.

2.11 What would you predict to be the shape of the distribution of the number of movies attended per month for the next 200 people you meet?

2.12 Draw a histogram for the GPA data in Appendix: Data Set referred to at the beginning of these exercises. (These data can also be obtained at http://www.uvm.edu/~dhowell/methods8/DataFiles/Add.dat.)

2.13 Create a stem-and-leaf display for the ADDSC score in Appendix: Data Set.