

Otázky k tématu 2 – četnosti, rozložení

- Pokud je proměnná nominální (např. etnicita), preferovanou grafickou podobou tabulky četností je
 - sloupcový diagram
 - histogram
- Která z následujících rozložení jsou pravděpodobně negativně zešikmená?
 - rodinný příjem v Kč za měsíc
 - věk při promoci na VŠ
 - počet obyvatel měst v ČR
 - skóry za lehký test
- Histogram
 - je podobný čárovému grafu
 - obsahuje normální křivku
 - je grafickou podobou tabulky četností
 - znázorňuje vztah mezi dvěma proměnnými
- Představte si, že určitý test z matematiky absolvuje 50 lidí – 25 matematiků a 25 lidí, kterým matematika nejde. Jaké rozložení nejspíše získáme?
 - unimodální
 - bimodální
 - normální
 - zešikmené
- Rozložení následujících skóru 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
 - unimodální a přibližně normální
 - bimodální a negativně zešikmené
 - normální a pozitivně zešikmené
 - normální a negativně zešikmené
- Jak nazýváme symetrické rozložení, v němž mají všechny hodnoty přibližně stejnou absolutní četnost?
- Níže jsou uvedena data – výsledné skóry z jedné písemky. Vytvořte tabulku četností včetně kumulativních
 - tabulku prostých četností jednotlivých hodnot
 - 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. Zkonstruuje histogram ke každé tabulce četností vytvořené v úkolu 7.

9. Čím se liší sloupcový graf od histogramu?

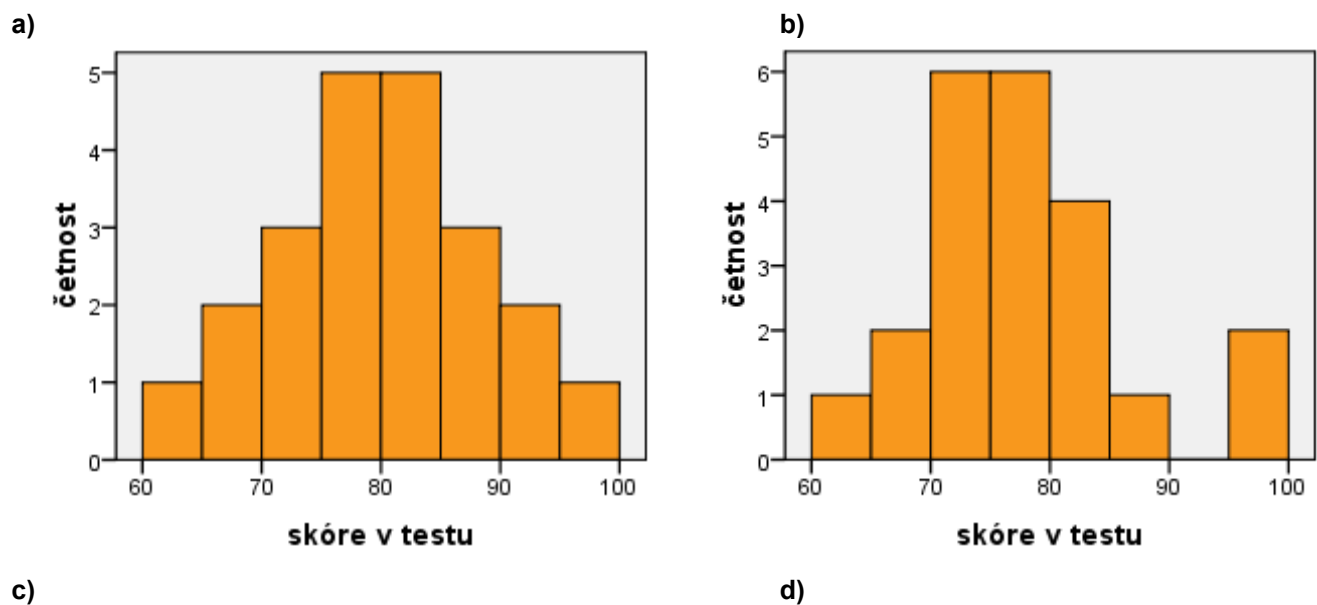
10. Základní škola v Horních Kotěhůrkách se rozhodla zhodnotit kvalitu své výuky, a tak nechala všechny žáky otestovat ze čtyřech předmětů. V každém testu bylo možno získat 0 až 100 bodů, čím více, tím lepší výsledek. Bodové zisky žáků ze VII.B jsou shrnuty v tabulce.

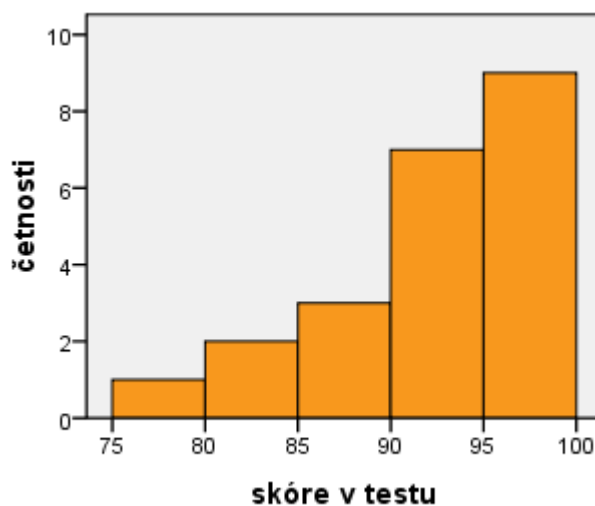
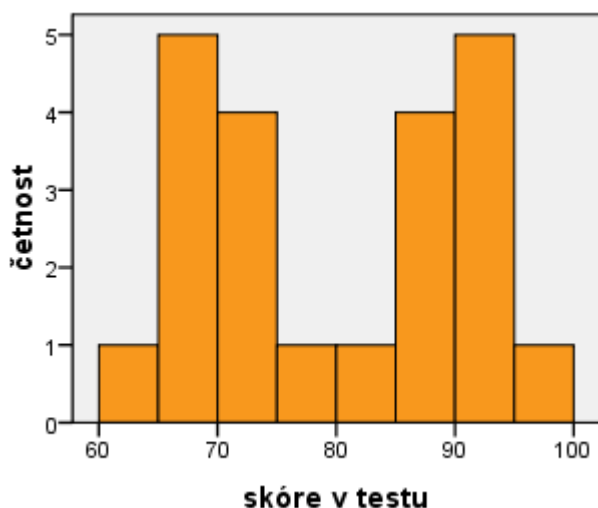
id žáka	matematika	český jazyk	anglický jazyk	tělesná výchova
1	92	94	92	83
2	77	98	97	74
3	86	89	67	68
4	70	96	89	79
5	81	91	73	96
6	68	99	76	80
7	77	94	66	87

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 skóre 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?





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?

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

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

- Je zešikmené zprava, neboli jde o kladné zešikmení.
- Je zešikmené zprava, neboli jde o záporné zešikmení.
- Je zešikmené zleva, neboli jde o kladné zešikmení.
- Je zešikmené zleva, neboli jde o záporné zešikmení.

10.6 Efekt, který to zapříčinil, se nazývá:

- Efekt podlahy.
- Gaussův efekt.
- Efekt stropu.
- Efekt taháku.

10.7 Jako outliera bychom nejspíše nazvali:

- Hodnotu, které nikdo nedosáhl.
- Žáka, který měl v daném testu nejnižší skóre.
- Oblast mezi dvěma mody (vrcholy).
- Skóre, který je daleko od většiny ostatních skóre.

10.8 Ve kterém z předmětů lze najít outliery?

- Ve všech předmětech.
- V tělesné výchově.
- V anglickém jazyce.
- 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 hodů jednotlivých čísel jsou stejné.

12.1 Jaké je rozdělení výsledku jednoho hodu hrací kostkou?

- a) diskrétní uniformní
- b) leptokurtické
- c) spojitě 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?

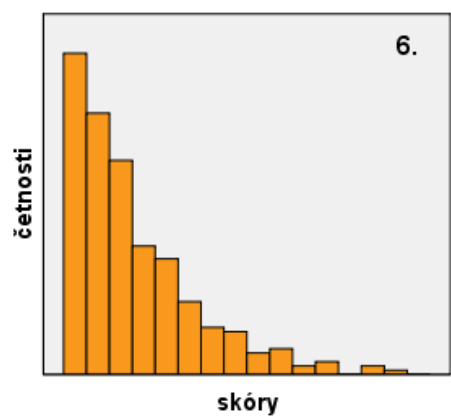
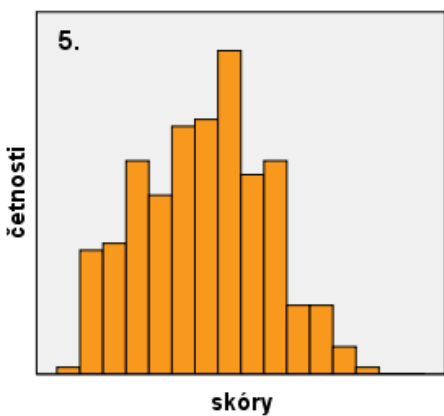
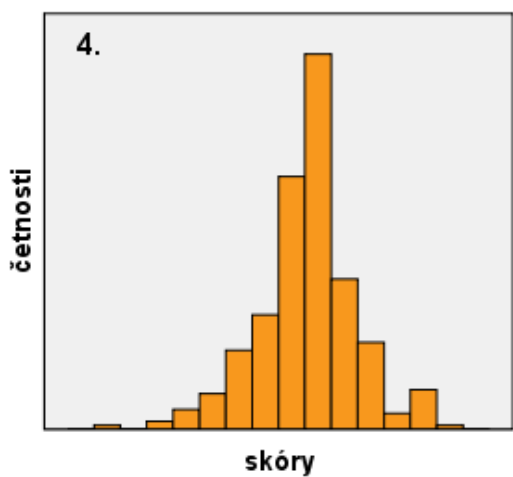
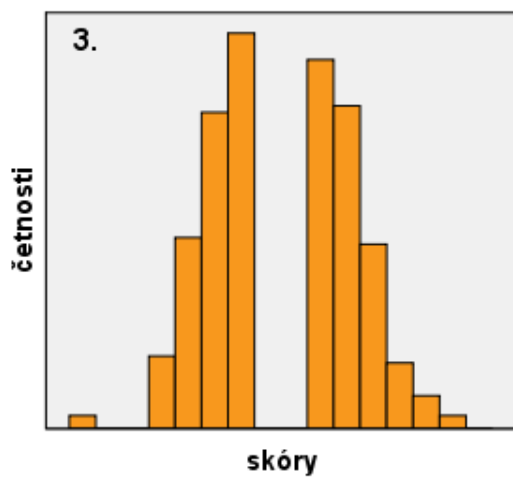
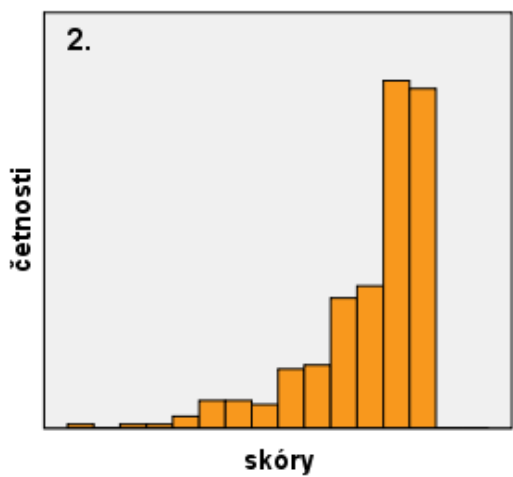
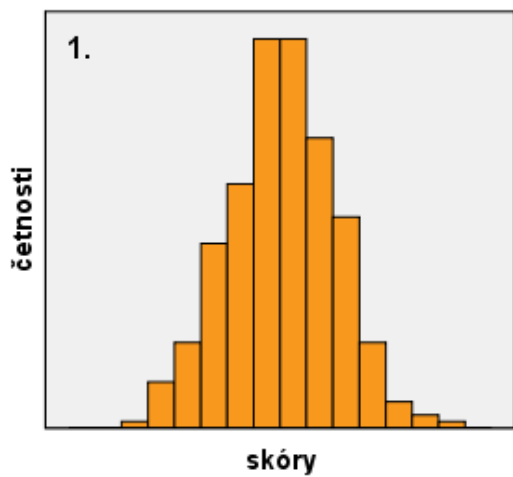
- a) spojitě uniformní
- b) poissonovo
- c) zešíkmené zprava
- d) normální

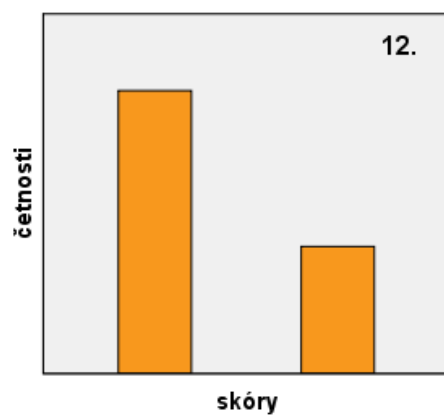
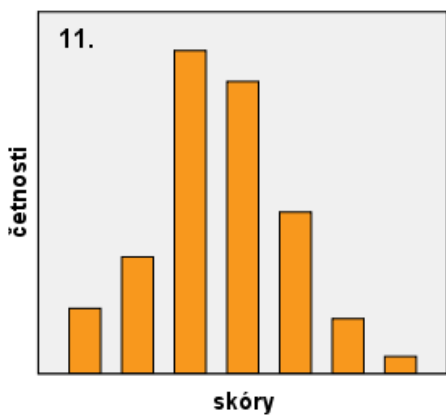
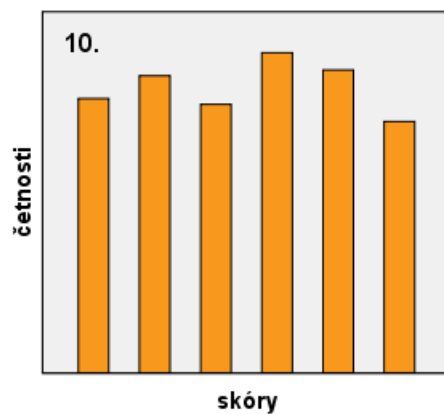
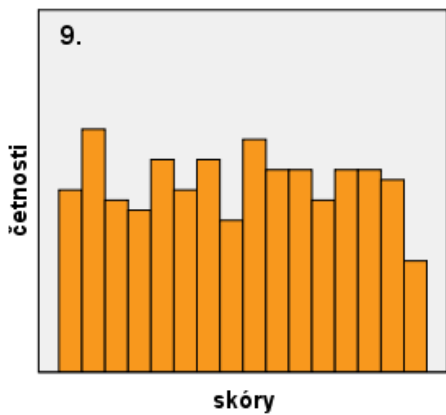
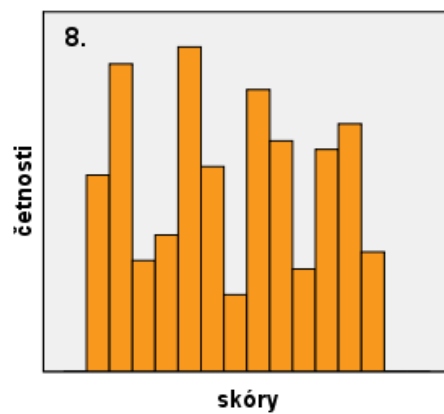
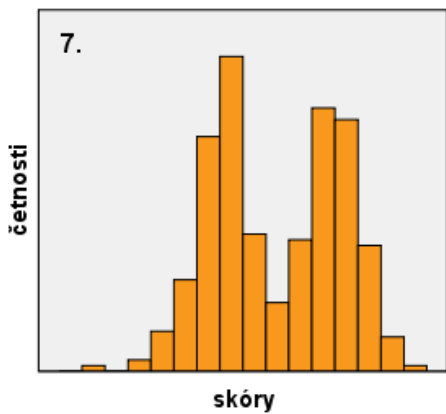
13. Pokud bych se ptal náhodných kolemjdoucích, jaký je můj věk, jaké by bylo rozdělení asi jejich odhadů?

- 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. Přiřadte tedy ke každému grafu patřičné vlastnosti (jednu vlastnost je možno přiřadit k více grafům):

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

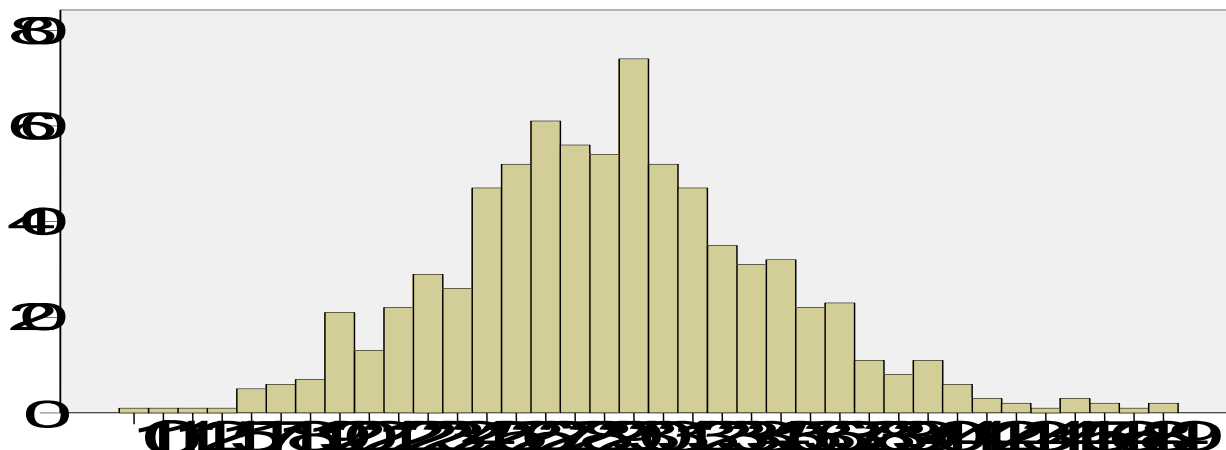




15. Následujících pět řád skóre představuje výsledky v různých testech. Graficky je znázorněte a popište tvar jejich rozložení:

- 10; 6; 8; 7; 5; 5; 4; 9; 2; 9; 8; 6; 7; 8
- 2; 9; 5; 1; 2; 4; 2; 6; 7; 2; 8; 5; 3; 4; 7; 2; 3; 5; 4; 3; 6; 3
- 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
- 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, vezte, 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.

<p>muž č. 2</p> <p>Jaká zařízení denně používáte?</p> <input checked="" type="checkbox"/> auto <input type="checkbox"/> televizi <input checked="" type="checkbox"/> digestoř	<p>žena č. 3</p> <p>Jaká zařízení denně používáte?</p> <input checked="" type="checkbox"/> auto <input checked="" type="checkbox"/> televizi <input type="checkbox"/> digestoř	<p>muž č. 3</p> <p>Jaká zařízení denně používáte?</p> <input type="checkbox"/> auto <input checked="" type="checkbox"/> televizi <input type="checkbox"/> digestoř <i>co to je kurča ten digestoř?</i>
<p>muž č. 1</p> <p>Jaká zařízení denně používáte?</p> <input checked="" type="checkbox"/> auto <input checked="" type="checkbox"/> televizi <input type="checkbox"/> digestoř	<p>žena č. 1</p> <p>Jaká zařízení denně používáte?</p> <input type="checkbox"/> auto <input checked="" type="checkbox"/> televizi <input checked="" type="checkbox"/> digestoř	<p>žena č. 2</p> <p>Jaká zařízení denně používáte?</p> <input checked="" type="checkbox"/> auto <input checked="" type="checkbox"/> televizi <input checked="" type="checkbox"/> digestoř

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 kladte 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 pusy, 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	Petrů	Vojtů
Délka vztahu = 4 Manželka Manžel data=9 data=3 místa=12 místa=6	Délka vztahu = 5 Manželka Manžel data=7 data=2 místa=9 místa=9	Délka vztahu = 8 Manželka Manžel data=7 data=8 místa=3 místa=12
Jirků	Mirků	Nechod'omů
Délka vztahu = 11 Manželka Manžel data=6 data=7 místa=6 místa=11	Délka vztahu = 14 Manželka Manžel data=8 data=6 místa=4 místa=6	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 škrtačí test. Každý student dostal test a z velké nádoby si vybral tužku. Tužky byly nalakované různými barvami. Místo vyhodnocování škrtačích 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				Celkem
		červená	černá	modrá	žlutá	
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 znázorníte rozložení proměnné *barva vybrané tužky* tak, abychom mohli srovnat rozložení této proměnné u mužů a u žen.

29. 196 študentov odpovedalo na otázku, ako najčastejšie trávia voľný čas v sobotu,. Výsledky sumarizuje tabuľka:

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 stonky a listu a popíšte rozloženie premennej.

Aký graf (okrem stonky a listu) by ste vybrali pre znázornenie daných údajov a prečo?

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