

# Příklady k procvičení látky běžných úvodů do teorie pravděpodobnosti a statistiky

Roman Biskup<sup>1</sup>

10. března 2012

<sup>1</sup>Mgr. Roman Biskup, Ph.D. – (zapálený) statistik ve výslužbě, aktuálně analytik v praxi ;-)



# Obsah

<b>1</b>	<b>Než začnete počítat</b>	<b>3</b>
1.1	Zadání příkladů . . . . .	3
1.2	Zadání samostatných prací . . . . .	3
<b>2</b>	<b>Pane pojdte si hrát!</b>	<b>5</b>
2.1	Vzorce, elementární calculus . . . . .	5
2.2	Kombinatorika . . . . .	8
<b>3</b>	<b>Základy teorie pravděpodobnosti</b>	<b>11</b>
3.1	Náhodný jev . . . . .	11
3.2	Náhodná veličina . . . . .	17
3.2.1	Diskrétní náhodná veličina . . . . .	17
3.2.2	Spojité náhodná veličina . . . . .	21
3.3	Zadání samostatné práce I . . . . .	25
<b>4</b>	<b>Zpracování datového souboru</b>	<b>27</b>
4.1	Třídění statistického souboru . . . . .	27
4.2	Charakteristiky úrovně, . . . . .	30
4.3	Zadání samostatné práce II . . . . .	33
<b>5</b>	<b>Základy teorie odhadu</b>	<b>35</b>
5.1	Bodový a intervalový odhad . . . . .	35
5.2	Zadání samostatné práce III . . . . .	39
<b>6</b>	<b>Testování hypotéz</b>	<b>41</b>
6.1	Testování hypotéz na základě jednoho výběru . . . . .	41
6.2	Testování hypotéz na základě dvou výběrů . . . . .	44
6.3	Testování hypotéz na základě více jak dvou výběrů . . . . .	48
6.4	Neparametrické testy . . . . .	51
6.5	Zadání samostatné práce IV . . . . .	53
<b>7</b>	<b>Regresní a korelační analýza</b>	<b>55</b>
7.1	Lineární regresní a korelační analýza . . . . .	55
7.2	Nelineární regresní a korelační analýza . . . . .	59
7.3	Vícenásobná regresní a korelační analýza . . . . .	64
7.4	Zadání samostatné práce V . . . . .	65
<b>8</b>	<b>Poděkování, reference a realizace</b>	<b>67</b>
8.1	Poděkování . . . . .	67
8.2	Reference . . . . .	67
8.3	Realizace . . . . .	68

<b>9</b>	<b>Seznam obrázků a tabulek</b>	<b>69</b>
9.1	Seznam tabulek . . . . .	69
9.1	Seznam obrázků . . . . .	69
<b>A</b>	<b>Kombinatorika letem světem – <i>Odvození vzorců</i></b>	<b>71</b>
A.1	Základní kombinatorické principy . . . . .	71
A.1.1	Princip bijekce . . . . .	71
A.1.2	Kombinatorické pravidlo o násobení . . . . .	71
A.1.3	Kombinatorické pravidlo o součtu . . . . .	72
A.1.4	Princip exkluze a inkluze . . . . .	72
A.1.5	Pravidlo logického stromu . . . . .	74
A.2	Permutace, variace a ty další . . . . .	75
A.2.1	Rozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují . . .	75
A.2.2	Rozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují * . .	75
A.2.3	Rozlišitelné přihrádky a rozlišitelné předměty, jež se mohou opakovat	76
A.2.4	Rozlišitelné přihrádky a rozlišitelné předměty, jež se opakují * . . .	76
A.2.5	Nerozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují . .	77
A.2.6	Nerozlišitelné přihrádky a rozlišitelné předměty, jež se mohou opakovat	78
A.3	Shrnutí . . . . .	80
<b>B</b>	<b>Jak počítá statistik ryby v rybníku</b>	<b>81</b>
<b>C</b>	<b>Vyhodnocování výsledků testování hypotéz na základě „<i>p</i>-value“</b>	<b>82</b>
C.1	Klasická metodologie vyhodnocování testování hypotéz . . . . .	82
C.2	Interpretace významu hodnoty „ <i>p</i> -value“ . . . . .	82
C.3	Není všechno zlato, co . . . . .	83
C.4	Modifikace oboustranného <i>p</i> -value . . . . .	84
<b>D</b>	<b>Průběžný test Statistika I</b>	<b>85</b>
<b>E</b>	<b>Průběžný test Statistika II</b>	<b>87</b>
E.1	Průběžný test Statistika II/a . . . . .	87
E.2	Průběžný test Statistika II/b . . . . .	88
<b>F</b>	<b>Průběžný test Statistika III</b>	<b>89</b>
F.1	Průběžný test Statistika III/a . . . . .	89
F.2	Průběžný test Statistika III/b . . . . .	89
F.3	Průběžný test Statistika III/c . . . . .	89
<b>G</b>	<b>Průběžný test Statistika IV</b>	<b>91</b>
<b>H</b>	<b>Zkouškový test Statistika</b>	<b>93</b>
H.1	Zkouškový test Statistika – 2005 . . . . .	93
H.2	Zkouškový test Statistika – 2009 . . . . .	95

# 1 Než začnete počítat

## 1.1 Zadání příkladů

- Příklady jsou vybírány tak, aby procvičili látku vykládanou na přednáškách ze „statisticky“.
- Číslo cvičení nemusí nutně odpovídat týdnu semestru. Informace jsou podávány na přednáškách.
- Obtížnost příkladů v rámci cvičení vzrůstá, navíc úkoly k řešení striktně v statistickém softwaru (STATISTICA komplet 6.1 Cz) mají vyšší číslo v rámci cvičení. Tyto příklady mají obvykle obsáhlejší datové zadání, které je ke stažení v adresáři `zadani/data/` toho kterého cvičení.
- Součástí cvičení mohou být malá teoretická cvičení – odvození vzorců, důkaz platnosti některých tvrzení, atd. Takové příklady budou označeny  $\odot T \odot$ . U takových cvičení bude obvykle upozorněno na stupeň obtížnosti – lehké, normální (neuváděno), těžké, pracné, ...

$\odot T \odot$   
Stupeň  
obtížnosti

## 1.2 Zadání samostatných prací

- Zadání samostatných prací v podstatě pokrývá probranou a procvičenou látku. Student by si na nich měl vyzkoušet statistické metody na datech z oblasti, která ho zajímá.
- Těžkopádnost zadání je způsobena nutností obecného zadání pro Vaše různá data, proto ji omluvte a pokud Vám některé formulace nebudou zřejmé konzultujte je s vyučujícím!
- Úlohy jsou zadávány tak, aby z nich byl zřejmá matematická interpretace problému a slouží především k procvičení postupů. Z toho důvodu je nutné procvičit převod slovní úlohy do její matematické formulace a naopak vyslovení závěru „zkousnutelného“ i matematicko-statistickým laikem, kterým byste Vy již být neměli.
- Komentář k řešení je tedy více než nutný. Výsledky Vaší práce by měly být srozumitelné a podat jak deduktivní tak induktivní informace o daném datovém souboru (viz jednotlivá zadání). Statistika je prostředkem kvantitativního hodnocení, jednotlivé úkoly sice procvičují jednotlivá témata, jež mají být osvojena v předmětu statistika, výsledkem však má být zpracování, vybraného souboru z několika pohledů, jež statistická metodologie nabízí.

## 1.2 Zadání samostatných prací



## 2 Pane pojdte si hrát!

### 2.1 Vzorce, elementární calculus

1. Opakování matematické symboliky např:

- (a) konstanty:  $e, \pi, i, \infty, \dots$
- (b) písmena řecké abecedy:  $\alpha, \beta, \delta, \Delta, \theta, \Theta, \lambda, \mu, \sigma, \Sigma, \phi, \Phi, \chi, \omega, \Omega, \dots$
- (c) funkce (proměnné  $x$ ), definiční obory, obory hodnot a jejich inverzní funkce:  $e^x, a^x, x^a, \log x, \ln x, \sin x, \cos x, \operatorname{tg} x, \operatorname{cotg} x, \max x, \min x, |x|, \dots$
- (d) číselné obory:  $\mathbb{N}, \mathbb{N}_0, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \dots$
- (e) spojky (logické a množinové), kvantifikátory:  $\neg, ', \text{non}, \wedge, \vee, \Rightarrow, \Leftrightarrow, \emptyset, \complement, \cap, \cup, \subset, \subseteq, =, \neq, \in, \forall, \exists, \dots$

2. Dosaďte do následujících vzorců (Pro informaci je v poznámkách pod čarou uvedeno, čeho se vlastně výpočty týkají, i když je to pro samotnou práci se vzorci nepodstatné):

- (a)  $V = a \cdot b \cdot c$ ,  
kde  $a = 2$ ;  $b = 1/2$  a  $c = 7$ ;<sup>1</sup>
- (b)  $k = 1 + 3,3 \cdot \log n$ ,  
kde  $n = 50$ ;<sup>2</sup>
- (c)  $KGZ = (R/Q) \cdot 100 \%$ ,  
kde  $R = 154$ ; a  $Q = 2928$ ;<sup>3</sup>
- (d)  $SGR = [(\ln W_t - \ln W_0) \cdot t^{-1}] \cdot 100 \%$   
kde  $W_t = 1200$ ;  $W_0 = 60$  a  $t = 153$ ;<sup>4</sup>
- (e)  $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$ ,  
kde  $n_1 = 10$ ;  $n_2 = 11$ ;  $\bar{x}_1 = 82,4$ ;  $\bar{x}_2 = 80,0$ ;  $s_1^2 = 12,1$ ;  $s_2^2 = 10,5$ .<sup>5</sup>
- (f) Vytvořte vzorec pro výpočet *kusové návratnosti ryb* ( $KNR$ ), což je procentuální podíl počtu ulovených ryb ( $N_u$ ) z počtu ryb vysazených  $N_n$ .<sup>6</sup>

<sup>1</sup>Vzorec pro výpočet objemu kvádrů, kde  $a, b$  a  $c$  jsou délky jeho hran.

<sup>2</sup>Tzv. *Sturgesův vzorec*, který slouží k výpočtu počtu intervalů  $k$ , do nichž je vhodné rozdělit  $n$  pozorování při intervalovém třídění statistického souboru. Více v oddílu 4.1.

<sup>3</sup>Vzorec pro výpočet *koeficientu zralosti gonád* ( $KZG$ ), je poměrem hmotnosti gonád ( $R$ ) k celkové hmotnosti těla ( $Q$ ) a udává se v procentech.

<sup>4</sup>Vzorec pro výpočet *specifické rychlosti růstu* (**S**pecific **G**rowth **R**ate), kde  $W_t$  je průměrná individuální hmotnost na konci období,  $W_0$  na začátku a  $t$  je délka období ve dnech.

<sup>5</sup>Výpočet testového kritéria u dvouvýběrového  $t$ -testu pro soubory se shodnými rozptyly (test na shodu středních hodnot), kde  $\bar{x}_i$  jsou výběrové průměry,  $s_i$  výběrové směrodatné odchylky a  $n_i$  rozsahy souborů, pro  $i = 1, 2$ . Více v oddílu 6.1.

<sup>6</sup>Kusová návratnost ryb se sleduje za jeden rok, nebo jako víceletý průměr a závisí: na typu vod, na druhu ryby, na stáří a velikosti vysazovaných ryb a na návštěvnosti revíru sportovními rybáři. Poskytuje informace o kvalitě zarybňování a o migraci ryb.

## 2.1 Vzorce, elementární calculus

- (g) Vytvořte vzorec pro výpočet *krmného koeficientu* ( $FCR$ ), což je podíl spotřeby krmiva v gramech za sledované období ( $F$ ) z přírůstku hmotnosti v gramech ( $FCE = W_t - W_0$ ); kde  $W_t$  je hmotnost obsádky na konci období,  $W_0$  na začátku období. V jakých jednotkách krmný koeficient vychází? Dále vypočtete hodnotu  $FCR$ , když při odchovu 300 kg násady kapra do tržní hmotnosti bylo použito 700 kg krmné směsi s výslednou produkcí 1050 kg tržní ryby<sup>7</sup>.

3. Napište nebo načrtněte následující množiny:

- (a)  $\{2k\}_{k=1}^{\infty}$ ,  $k \in \mathbb{Z}$ ,  
(b)  $\{x \in \mathbb{R} : |x| \leq 2\}$ .

4. Jak by jste zapsali množinu:

- (a) lichých čísel v absolutní hodnotě menších než 10,  
(b) prvních sto přirozených čísel?

5. Nechtě jsou vektory<sup>8</sup>  $\mathbf{u} = (u_1, u_2, u_3, u_4, u_5)^T = (5, 7, 13, 0, 6)^T$ ,  $\mathbf{v} = (0, 0, 4, 9, 3)^T$  a  $\mathbf{w} = (2, 2, 3, 2, 7)^T$ . Jednotlivé složky vektoru udávají počet rybek různých druhů v soukromých sbírkách Uršuly, Viktora a Waltera. Přičemž složky po řadě určují počet Ramirézek (Cichlidka Ramirezova), Akar hnědých, Bojovnic pestrých, Čichavců zakrslých a Neonek červených. Vektor  $\mathbf{c} = (30, 25, 40, 25, 22)^T$  reprezentuje cenu jednotlivých druhů ryb v Kč.<sup>9</sup> Vypočtete:

- (a)  $u_1 + v_1 + w_1$  Kolik ramirézek mají dohromady?  
(b)  $(1, 1, 1, 1, 1) \cdot \mathbf{v} = \sum_{i=1}^5 v_i$  Kolik rybek má Viktor?  
(c)  $\mathbf{w}^T \cdot \mathbf{c} = \sum_{i=1}^5 w_i \cdot c_i$  Jakou cenu mají Walterovi rybičky?

6. Vytvořte matici  $\mathbf{R} = (\mathbf{u}, \mathbf{v}, \mathbf{w})$  z vektorů v úkolu 5.

<sup>7</sup>Zvažte, zda je nutné údaje převádět na gramy. Hmotnosti v gramech se uvažují z důvodu výpočtu krmného koeficientu pro jednotlivou rybu, pro celé obsádky je rozumnější používat „větší“ hmotnostní jednotky

<sup>8</sup>Vektor budeme uvažovat jako jednosloupcovou matici  $\mathbf{A} = (a_{ij})_{m1}$ . Kdyby jsme to takto neuvažovali, museli bychom rozlišovat skalární součin pro vektory a pak součin matic, tak jak jej znáte z maticové algebry. My se tedy omezíme na maticové násobení, u kterého musíme dbát na správné rozměry matice, při nichž je tento součin definován. Skalární součin dvou vektorů v maticovém podání pak musí být např.  $\mathbf{u}^T \cdot \mathbf{v} = (u_{ij})_{1n} \cdot (v_{ij})_{n1}$ , jehož výsledkem je číslo (skalár) – jednoprvková matice  $(a_{ij})_{11} = u_{11} \cdot v_{11} + u_{21} \cdot v_{12} + \dots + u_{n1} \cdot v_{1n}$ . Vzhledem k tomu, že matice vektoru má pouze jeden sloupec, zapisujeme jen číslo řádku a fakt, že se jedná o sloupcový vektor zvýrazňujeme pomocí symbolu  $^T$ , např.  $\mathbf{u} = (u_1, u_2, \dots, u_n)^T$ . Vektory se jako sloupcové v textu obvykle jako sloupce nevyepisují, neboť by zabíraly příliš mnoho místa. Běžně se vektor značí tučně a skolněně –  $\mathbf{v}$ , kdežto matice pouze tučně a svisle –  $\mathbf{v}$ .

<sup>9</sup>ceny k 8/2007





(a) Vypište tuto datovou matici  $\mathbf{R}$

$$(b) \sum_{j=1}^3 r_{1j}$$

Kolik ramirézek mají dohromady?

$$(c) \sum_{i=1}^5 r_{i2}$$

Kolik rybek má Viktor?

$$(d) \sum_{i=1}^5 \sum_{j=1}^3 r_{ij} = \sum_{j=1}^3 \sum_{i=1}^5 r_{ij} \quad |^{10}$$

Kolik rybek mají dohromady?

$$(e) \sum_{i=1}^5 \sum_{j=1}^3 r_{ij} \cdot c_i \quad |^{11}$$

Jakou cenu mají rybičky všech akvaristů?

<sup>10</sup>Pro konečné „rozumné“ součty lze pořadí sčítání přehazovat.

<sup>11</sup> $c_i$  pro  $i = 1, \dots, 5$  jsou opět příslušné ceny rybiček z vektoru  $\mathbf{c}$  v úkolu 5.

## 2.2 Kombinatorika

1. (a) Za lokomotivou jsou zapojeny 4 různé vagóny. První cisternový, druhý na uhlí, třetí na sypký materiál a poslední plošinový. Výpravčí má ve stanici k následující zboží připravené k přepravě: LTO<sup>12</sup>, brikety, naftu, palety tašek jednoho typu, koks, glycerol, mazut, hnědé uhlí, antracit, písek, dodávku aut jednoho typu, šterk a betonové kanalizační potrubí. Zboží se na vagóny nakládá vždy od jednoho druhu. Kolika způsoby může výpravčí různě naložit zboží na vagóny? Za různé naložení se pokládá změna bytí jediného zboží.  
(b) Na jídelním lístku jsou 3 aperitivy, 9 předkrmů, 5 polévek, 16 hlavních jídel, 2 moučníky, 4 alkoholické a 8 nealkoholických nápojů. Určete kolika způsoby lze sestavit menu sestávající se ze všech sedmi součástí.
2. (a) Restaurace nabízí několik pokrmů. V dalším textu jsou uvedeny druhy pokrmů, počty jídel jednotlivých druhů pokrmů jsou uvedeny v závorce. Sýry a bezmasá jídla (8), ryby (15), drůbež (22), vepřové maso (15), jehněčí maso (5), hovězí maso (8) a zeleninové hrnce (6). Kolik dní by jste mohli chodit do této restaurace, aby jste každý den jedli něco jiného, přičemž pokrm s různou přílohu nepovažujeme za různý?  
(b) Pět restaurací nabízí shodně ve svých jídelních lístečích po 20-ti hlavních jídlech. Kolik dní by jste mohli chodit do těchto pěti restaurací, aby jste každý den jedli něco jiného, přičemž pokrm s různou přílohu nepovažujeme za různý?
3. (a) Ve třídě je 27 studentů. Z celkového počtu studentů 12 mluví anglicky a 17 německy. Oběma jazyky mluví 7 studentů. Zodpovězte následující otázky:
  - i. Kolik studentů mluví nějakým cizím jazykem?
  - ii. Kolik studentů mluví pouze anglicky nebo pouze německy?
  - iii. Kolik studentů nemluví žádným z výše uvedených jazyků?  
(b) Na přednášku ze statistiky by studenti – co se barev svršků týče – dostavili následujícím způsobem pestří. V první lavici bylo možno nalézt černou, bílou, tmavě modrou, světle modrou a tyrkysovou barvu; v druhé lavici hnědou, oranžovou, tmavě a světle modrou, bílou, černou a zelenou; nakonec řada třetí zářila: fialovou, hnědou, černou, tyrkysovou, zelenou, bílou, červenou a žlutou. Ověřte, že pravidlo inkluze a exkluze dává opravdu správné výsledky pro počet barev ve sjednocení jak dvou, tak tří řad.
4. (a) V hotelu je 70 neekvivalentních lůžek (každé lůžko je jiné). Kolika způsoby lze lůžka přidělit 62 hostům z pohledu hostů?  
(b) Kolik různých zastupitelstev může mít obec o 15-ti volitelných občanech, obsazovali se post starosty, místostarosty, uklízečky, obecního blbečka a veřejné drbny?

---

<sup>12</sup>Lehký Topný Olej



- Uvažujte, že funkce nelze kumulovat, tj. každá osoba může mít nejvýše jednu funkci.
- (c) Na šachovnici osm kamarádů postupně rozmístí osm svých označených věží. Kolik variant uspořádání může nastat:
- stojí-li tak, aby se neohrožovaly (tj. žádné dvě věže neleží ve stejné řadě či sloupci),
  - nezáleží-li na tom, zda se ohrožují, či nikoliv.
5. (a) Kolika způsoby lze seřadit do fronty  $n$  zákazníků?
- (b) Kolika způsoby lze zapsat libovolnou posloupnost  $n$  navzájem různých znaků (tj. každý je k dispozici pouze jednou)?
- (c) Kolika způsoby si mohou stoupnout do fronty před Sněhurku její trpaslíci<sup>13</sup> tak, že:
- bez omezení (tj. každý může stát v zástupu kdekoliv),
  - Šmudla je jako obvykle poslední,
  - Šmudla kupodivu poslední není?
- (d) Na šachovnici se rozmístí osm věží. Kolik variant uspořádání může nastat:
- stojí-li tak, aby se neohrožovaly (tj. žádné dvě věže neleží ve stejné řadě či sloupci),
  - nezáleží-li na tom, zda se ohrožují, či nikoliv.
6. (a) Kolik různých zastupitelstev může mít obec o 15-ti volitelných občanech, obsazují-li se post starosty, místostarosty, uklízečky, obecního blbečka a veřejné drbny? Uvažujte, že funkce lze kumulovat, tj. jedna osoba může mít např. všechny funkce.
- (b) Anglická abeceda má 26 písmen. Kolik z ní lze teoreticky vytvořit šestipísmenných slov?
- (c) Kolik různých značek teoreticky existuje v Morseově abecedě, sestavují-li se tečky a čárky do skupin od jedné do pěti?
7. (a) Kolik různých slov vznikne přesmyčkou písmen ve slově:
- DAN,
  - RAMA a
  - RAMADAN?
- (b) Kolik různých pětímístných přirozených čísel lze vytvořit z číslic 2, 3, 3, 7 a 7?
- (c) Kolik různých slov vznikne přesmyčkou písmen ve slově POPOKATEPETL<sup>14</sup>?

<sup>13</sup>Pro neznalé pohádek, prameny uvádějí, že je jich sedm: Prófa, Kejchal, Rejpal, Štístko, Dřimal, Stydlín a v textu zmíněný Šmudla.

<sup>14</sup>(z aztéckého jazyka Nahuatl *popōka*, dýmání, a *tepētl*, hora), činná sopka ve střední části Mexika na okraji Mexické náhorní plošiny; 5 452 m n.m. Kráter je hluboký 200 m, sněžná čára ve výši 4 300 m n.m. Národní park. Pro příklad použit fonetický přepis oficiálního jména „Popocatepetl“.

## 2.2 Kombinatorika

8. (a) V hotelu je 70 neekvivalentních lůžek (každé lůžko je jiné). Kolika způsoby lze lůžka přidělit 62 hostům z pohledu pokojské?  
(b) V Matesu se tipuje 5 čísel z 35. Kolik je všech možných tipů?  
(c) Kolika způsoby lze na šachovnici vybrat tři pole tak, aby všechna neměla stejnou barvu?  
(d) Kolika způsoby lze rozdělit 8 chlapců a 4 dívky na dvě volejbalová družstva tak, aby v každém byla alespoň jedna dívka?  
(e) Sešlo se 11 kamarádů a na pozdrav si každý s každým potřásl rukou. Kolikrát se akt potřesení ruky zopakoval?
9. (a) Zjistěte, kolik takových různých kvádrů existuje, pro něž platí, že délka každé jejich hrany je přirozené číslo z intervalu  $\langle 2, 15 \rangle$ .  
(b) V pytlíku je 23 žlutých, 12 červených, 15 modrých a 20 zelených kuliček. Z pytlíku je náhodně vybráno 10 kuliček. Kolik různých seskupení lze získat?
10. Pokud si nevíte rady možný návod naleznete v příloze [A](#), nebo v doplňkovém textu *Kombinatorika letem světem*.



## 3 Základy teorie pravděpodobnosti

### 3.1 Náhodný jev

1. Definujte jevy opačné k následujícím jevům:
  - (a) Cestovní kanceláře mají zavřeno.
  - (b) Alespoň jedna cestovní kancelář má zavřeno.
  - (c) Nejvýše jedna cestovní kancelář má zavřeno.
2. Pravděpodobnost, že se jedinec z jisté populace dožije šedesáti let, je 0,8; pravděpodobnost, že se dožije sedmdesáti let, je 0,5. Jaká je pravděpodobnost, že jedinec zemře mezi šedesátým a sedmdesátým rokem svého života?
3. Házíme 2 hracími kostkami. Jaká je pravděpodobnost, že součet čísel na kostkách bude
  - (a) právě 5,
  - (b) větší než 3?
4. Jaká je pravděpodobnost, že při hodu pěti hracími kostkami
  - (a) padnou vzájemně různá čísla,
  - (b) padnou pouze lichá čísla?
5. O náhodných jevech  $A$  a  $B$  jsou známy následující skutečnosti:
  - (a) Pravděpodobnost, že nastane alespoň jeden z jevů  $A$  a  $B$ , je  $3/4$ .
  - (b) Pravděpodobnost, že oba jevy  $A$  a  $B$  nastanou současně, je  $1/4$ .
  - (c) Pravděpodobnost, že nenastane jev  $A$ , je  $2/3$ .

Určete pravděpodobnosti obou jevů  $A$  a  $B$ . Jaká je pravděpodobnost, že nastane jev  $A$  a přitom nenastane jev  $B$ ?

6. Pro naléhavou potřebu byli povoláni dobrovolní dárci krve. Dostavilo se 15 lidí, z nich 8 mělo krevní skupinu A, 5 krevní skupinu B, 2 krevní skupinu AB. Stanice má 5 odběrových boxů. Jaká je pravděpodobnost, že mezi prvními dárci krve je:
  - (a) všech 5 s krevní skupinou A,
  - (b) jsou tři s krevní skupinou A a dva s krevní skupinou B,
  - (c) jsou dva s krevní skupinou A, jeden s krevní skupinou B a dva s krevní skupinou AB?
7. Na šachovnici se náhodně rozmístí osm věží. Jaká je pravděpodobnost, že žádná z věží neohrožuje žádnou z ostatních (tj. žádné dvě věže neleží ve stejné řadě či sloupci)?

### 3.1 Náhodný jev

8. Z balíčku mariášových karet vytáhneme postupně osm karet; nevracíme je přitom zpět. Jaká je pravděpodobnost, že právě tři vytažené karty jsou srdcové?
9. Jaká je pravděpodobnost, že ve skupině  $n$  lidí se najdou alespoň dva, kteří mají narozeniny ve stejný den? (Předpokládejme, že rok má 365 dní a že porodnost se v průběhu roku nemění.) Speciálně určete tuto pravděpodobnost pro  $n = 23$  a pro  $n = 56$ .
10. Mezi 8 bezvadných výrobků byly zamíchány 3 zmetky. Náhodně byly vybrány dva výrobky. Jaká je pravděpodobnost, že
  - (a) jsou oba bezvadné,
  - (b) je jeden vadný,
  - (c) je alespoň jeden vadný?
11. Z karetní hry o 52 kartách náhodně vybereme tři karty. Určete pravděpodobnost toho, že to bude
  - (a) trojka, sedmička a eso;
  - (b) trojka, devítka a desítka;
  - (c) král a dvě esa.
12. V krabici je 5 koulí. 3 jsou bílé a 2 černé. Táhneme dvě koule za sebou (koule po vytažení do krabice nevracíme). Jaká je pravděpodobnost, že první vytáhneme bílou a druhou černou?
13. Ve skupině 25 studentů (15 děvčat a 10 chlapců) se losují tři studenti.
  - (a) Jaká je pravděpodobnost, že bude jako první vylosována dívka?
  - (b) První je vylosována dívka jaká je pravděpodobnost, že bude opět vylosována dívka?
  - (c) Jako první byly vylosovány 2 dívky, jaká je pravděpodobnost, že bude jako třetí vylosována opět dívka?
  - (d) Jaká je pravděpodobnost, že byly vylosovány tři dívky?
14. Pan Houška se snaží potmě otevřít dveře svého bytu. V kapse má deset klíčů, z toho právě jeden je od bytu. Náhodně vybírá a zkusíme jeden klíč po druhém. Narazí-li na klíč, který není od bytu, přendá jej do druhé kapsy, čímž zajišťuje, že žádný klíč nebude nezkouší opakovaně.
  - (a) Jaká je pravděpodobnost, že mu správný klíč padne do ruky při 6-tém pokusu?
  - (b) Jaká je pravděpodobnost, že mu správný klíč padne do ruky nejpozději při 3-tím pokusu?



15. I druhý den stále chybí žárovka na chodbě a pan Houška řeší opět ten samý problém – otevřít potmě dveře svého bytu. V kapse má deset klíčů, z toho právě dva jsou od bytu. (Po zkušenostech z předchozího dne, kdy mu otevírání bytu trvalo příliš dlouho, si totiž do kapsy vložil ještě jeden klíč od bytu navíc a vyndal z ní jeden nepříliš používaný klíč.) Znovu, tak jako předchozího dne, náhodně vybírá a poté zkouší jeden klíč po druhém a již vyzkoušené klíče přendává do druhé kapsy.
- (a) Jaká je pravděpodobnost, že správný klíč mu padne do ruky poprvé při 5-tém pokusu?
  - (b) Jaká je pravděpodobnost, že správný klíč mu padne do ruky nejpozději při 9-tém pokusu?
16. Jaká je pravděpodobnost, že meteor padne na tu část zeměkoule, kde je pevnina? Šance zásahu libovolných míst zeměkoule je shodná. Pevnina zabírá  $149\,000\,000\text{ km}^2$ , moře pak  $361\,000\,000\text{ km}^2$ .
17. Na dvoumetrové tyči jsou všechna místa zlomu stejně pravděpodobná. Předpokládejme, že se zlomí na dva kusy. Jaká je pravděpodobnost, že
- (a) se zlomí přesně v polovině,
  - (b) bude jeden díl po zlomení alespoň dvakrát tak delší než ten druhý,
  - (c) bude delší díl po zlomení o 10 cm delší než kratší,
  - (d) bude delší díl po zlomení maximálně o 10 cm delší než kratší.
18. Pravděpodobnost úspěchu určitého pokusu je 0,9. Jaká je pravděpodobnost, že při dvojnásobném nezávislém opakování pokusu bude dosaženo alespoň jednoho úspěchu?
19. Zahradník zasel 8 semen. Ví, že semínko vyklíčí s 40% pravděpodobností.
- (a) Určete pravděpodobnost, že vyklíčí alespoň jedno semínko.
  - (b) Kolik musí zahradník zasadit minimálně semen, aby vyklíčilo alespoň jedno s 80% pravděpodobností?
20. Operátor telemarketingu ze svých zkušeností ví, že oslovený zákazník nabízený produkt akceptuje z 5%.
- (a) S jakou pravděpodobností prodá nabízený produkt všem 20 osloveným zákazníkům?
  - (b) S jakou pravděpodobností prodá nabízený produkt alespoň jednomu z 20 oslovených zákazníků?
  - (c) Kolik zákazníků musí oslovit, aby alespoň jeden z nich produkt akceptoval s 80% pravděpodobností?

### 3.1 Náhodný jev

21. V pokusném rybníčku má váčkový kapří plůdek při přechodu z  $K_0$  na  $K_1$  vypočítávané ztráty 95 %.<sup>15</sup>
- (a) S jakou pravděpodobností přežije celá kapří obsádka, která má 200 kusů váčkového kapřího plůdku?
  - (b) S jakou pravděpodobností přežije alespoň jeden kus váčkového kapřího plůdku z 200 vysazených?
  - (c) Kolik kusů váčkového kapřího plůdku je třeba vysadit, pokud se alespoň jeden z nich má dožít jednoho roku s 99% pravděpodobností?
22. Pravděpodobnost narození chlapce je 0,515. Jaká je pravděpodobnost, že mezi čtyřmi po sobě narozenými dětmi budou:
- (a) první dva chlapci a další dvě dívky,
  - (b) první dva chlapci,
  - (c) právě dva chlapci,
  - (d) alespoň dva chlapci?
23. Rodina plánuje v červenci 3 dny dovolené. Pravděpodobnost, že v červencovém dni vůbec nezaprší je 0,62. Pravděpodobnost, že alespoň chvíli prší, je 0,38. Počasí jednotlivých dnů je nezávislé na počasí ostatních dní. Jaká je pravděpodobnost, že:
- (a) po celou dobu dovolené nebude pršet,
  - (b) každý den zaprší,
  - (c) budou převažovat dny bez deště,
  - (d) bude alespoň jeden den bez deště,
  - (e) po celou dobu dovolené se ráz počasí nezmění?
24. Aby měl student vyznamenání, musí mít průměr do 1,5. Víme, že uvedenému kritériu vyhovuje 38 % dívek a 29 % chlapců. Jaká je pravděpodobnost, že když se zeptáme smíšené dvojice vytvořené nezávisle na studijních úspěších:
- (a) budou mít oba vyznamenání,
  - (b) bude mít alespoň jeden z dvojice vyznamenání,
  - (c) bude mít vyznamenání pouze dívka,
  - (d) bude mít vyznamenání pouze hoch,
  - (e) bude mít vyznamenání právě jeden z páru,
  - (f) nebude mít vyznamenání nikdo?

---

<sup>15</sup>Normativní ztráty během vegetačního období jsou nižší – 85 %.





25. Necht' pravděpodobnostní prostor  $\Omega$  jsou přirozená čísla do padesáti ( $\Omega = \{1, 2, \dots, 50\}$ ). Definujme na tomto pravděpodobnostním prostoru jev  $A$  jako přirozená čísla menší nebo rovna 25 ( $A = \{1, 2, \dots, 25\}$ ), jev  $B$  jako sudá přirozená čísla  $B = \{x \in \Omega : x = 2k, k = 1, 2, \dots, 25\}$  a jev  $C$  jako čísla dělitelná čtyřmi ( $C = \{x \in \Omega : 4|x\}$ ). Jsou tyto jevy párově nezávislé?
26. Uvažte následující pravděpodobnostní prostor  $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$  a definujme náhodné jevy  $A = \{1, 2, 3, 4\}$ ,  $B = \{3, 4, 5, 6\}$  a  $C = \{1, 2, 5, 6\}$ . Jsou tyto jevy
- párově nezávislé, respektive
  - vzájemně (totálně) nezávislé?
27. Výrobní linka se sestává z 5-ti automatů. Aby byl výrobek dokončen musí projít všemi obráběcími procedurami. Žádná procedura neovlivňuje další. První automat produkuje zmetky s pravděpodobností 0,002, pro ostatní automaty jsou pravděpodobnosti produkce zmetků 0,01, 0,05, 0,008 a konečně 0,003. Jaká je pravděpodobnost, že
- linka vyrobila bezvadný výrobek,
  - výrobek, který prošel procedurou má vadu?
28. Nemocnice je z bezpečnostních důvodů zásobována elektrickou energií z 5-ti rovnocenných zdrojů. Každý zdroj je dostačující pro běh nemocnice. Porucha na prvním zdroji nastává s 0,2% pravděpodobností, pro další zdroje jsou pravděpodobnosti výpadku 0,01, 0,05, 0,008 a 0,003. Jaká je pravděpodobnost, že
- elektrická energie bude,
  - budou v nemocnici startovat záložní agregát?
29. Student jde na zkoušku, ale neví, který ze tří možných předmětů (RPZ, PV, PG) se zkouší. Ví, že neumí 40 % otázek z RPZ, 15 % z PV a 20 % z PG.<sup>16</sup>
- Jaká je pravděpodobnost, že bude vyhozen?
  - Jaká je pravděpodobnost, že bude vyhozen z RPZ, byl-li z RPZ zkoušen?
  - Bude-li vyhozen, jaká je pravděpodobnost toho, že to bude z RPZ?
30. Celostátní pozorování zvyklostí manželských párů ukázalo, že potraviny pro domácnost pravidelně nakupují v 60 % domácností manželky (z toho 40 % manželek bere s sebou na nákup dítě) a v 25 % domácností nakupují manželé (muži). Zároveň se však ukázalo, že chystá-li na nákup manželka, pak ve 30 % domácností jsou manželé (muži) ochotni se „obětovat“ a jdou na nákup spolu s manželkou. Jaká je pravděpodobnost, že v náhodně vybrané domácnosti:

<sup>16</sup>Pravděpodobnosti, jež jsou známy předem ( $P(A|B_i)$ ,  $P(B_i)$ , pro  $i = 1, \dots, n$ ) se označují jako apriorní pravděpodobnosti, pravděpodobnosti, které jsou získány jako výsledek Bayesova vzorce ( $P(B_i|A)$  pro  $i = 1, \dots, n$ ) se nazývají aposteriorní pravděpodobnosti.

### 3.1 Náhodný jev

- (a) chodí na nákup zásadně oba manželé,
  - (b) chodí na nákup alespoň jeden z manželů (neposílají děti, babičku apod.),
  - (c) nenakupuje žádný z manželů (muž + žena),
  - (d) jde-li na nákup manžel, doprovodí jej manželka,
  - (e) nakupuje manželka a bere s sebou dítě?
31. Ve skupině studentů je 18 „matematiků“, 12 „techniků“ a 7 „právníků“. Pravděpodobnost splnění IQ normy je 0,9 u matematika, 0,75 u technika a 0,65 u právníka.
- (a) S jakou pravděpodobností náhodně vybraný student splní tuto IQ normu?
  - (b) Náhodně vybraný student splnil IQ normu. S jakou pravděpodobností se jednalo o právníka?
32. Nemoc se vyskytuje u 1% lidí. Nemocný člověk má pozitivní nález s 95% pravděpodobností a zdravý s 30% pravděpodobností.<sup>17</sup>
- (a) Jaká je pravděpodobnost, že člověk je diagnostikovaný jako nemocný (pozitivní nález)?
  - (b) Jaká je pravděpodobnost, že je člověk správně diagnostikovaný, tj. nemocný jako nemocný a zdravý jako zdravý?
  - (c) Jaká je pravděpodobnost, že člověk je nemocný, má-li pozitivní nález?
  - (d) Jaká je pravděpodobnost, že je člověk zdravý i přes pozitivní nález?
  - (e) Jsou jevy „být zdrav“ a „být pozitivně diagnostikován“ nezávislé?

---

<sup>17</sup>Nechť existují doplňkové jevy – jev, že osoba nemoc má, a jev, že osoba sledovanou nemoc nemá. Pravděpodobnost výskytu nemoci v populaci se obvykle nazývá *prevalence* (apriorní pravděpodobnost, pretestová pravděpodobnost). Uvažme diagnostický test, pro nějž jsou známy pravděpodobnosti, se kterými odhaluje chorobu u zdravých a nemocných (též apriorní pravděpodobnosti). *Senzitivitou* testu se rozumí pravděpodobnost odhalení nemoci u nemocných, *specificitou* pravděpodobnost, že test je negativní u zdravých osob. Obdobně *nesprávná (falešná) pozitivita* je pravděpodobnost, že zdravá osoba je označena za nemocnou a naopak *nesprávná (falešná) negativita* je pravděpodobnost, že nemocná osoba je označena za zdravou.

Zajímavé jsou i aposteriorní pravděpodobnosti – pravděpodobnosti podmíněné výsledkem testu (potesťová pravděpodobnost). Hovoří se o tzv. *prediktivní hodnotě pozitivního testu*, což je pravděpodobnost, že osoba je opravdu nemocná, když byla pozitivně diagnostikována a *prediktivní hodnotě negativního testu*, což je pravděpodobnost, že osoba je opravdu zdravá, byla-li za zdravou označena. Podobně je tomu i v právnickém prostředí – nemocí bývá lež a diagnostický prostředkem například detektor lži.



## 3.2 Náhodná veličina

### 3.2.1 Diskrétní náhodná veličina

1. Dokažte pro diskrétní náhodnou veličinu, že rozptyl<sup>18</sup> lze vyjádřit jako rozdíl aritmetického průměru čtverce náhodné veličiny a čtverce aritmetického průměru hodnot, tj.  $DX = EX^2 - (EX)^2$  (k důkazu použijte vzorce pro výpočet rozptylu a střední hodnoty). ◊T◊  
Lehké
2. Alternativní rozdělení popisuje jednokolový pokus v němž úspěch nastává s pravděpodobností  $\pi$  a neúspěch s pravděpodobností  $1 - \pi$ ,  $\pi \in (0, 1)$ . Tento pokus je do řeči náhodné veličiny, označme ji například  $X$ , převeden následovně: ◊T◊  
Lehké

výsledek pokusu	zakódování	pravděpodobnost
úspěch	$X = 1$	$P(X = 1) = \pi$
neúspěch	$X = 0$	$P(X = 0) = 1 - \pi$

- (a) Odvoďte vzorec pro výpočet střední hodnoty a rozptylu tohoto rozdělení.
  - (b) Jak vypadá distribuční funkce alternativního rozdělení? Uveďte předpis a nakreslete ji.
3. V testu je 8 otázek. Na každou je nabídnuto 5 odpovědí, právě jedna je vždy správná.
    - (a) Jaká je pravděpodobnost, že při náhodné volbě odpovědi dostaneme  $k$ ,  $k = 0, \dots, 8$ , správných odpovědí?
    - (b) Jak vypadá distribuční funkce rozdělení, jež náhodná veličina popisující tento pokus sleduje?
    - (c) Jaký je střední hodnota počtu správných odpovědí?
    - (d) Jaká je pravděpodobnost, že ztratíme maximálně jeden bod za předpokladu, že jsme získali nadpoloviční počet bodů
  4. Vyjet z domečku – nasadit figurku do hry – při hře „Člověče nezlob se!“ lze při padnutí šestky na hrací kostce. Předpokládejte, že v případě padnutí šestky nasazujete figurku do hry a že Vás nikdo nevyhodil.
    - (a) Jaká je pravděpodobnost, že budete moc vyjet z domečku s poslední (čtvrtou) figurkou při desátém kole?
    - (b) Jaký je střední počet pokusů pro nasazení všech čtyř figurek?
  5. Pravděpodobnost, že stroj vyrobí zmetek je 0,001. V okamžiku, kdy je vadný výrobek vyroben, je stroj zastaven a seřízen. S jakou pravděpodobností bude stroj zastaven:
    - (a) bezprostředně po vyrobení stého kusu.

<sup>18</sup>samořejmě pokud  $DX$  existuje.

### 3.2 Náhodná veličina

- (b) nejpozději po vyrobení stého kusu.
- (c) Jaký je střední počet vyrobených výrobků do vyrobení prvního zmetku?
6. Nechť do ústředny přichází v průměru 120 hovorů za jednu hodinu. Počet hovorů za časový interval délky  $t$  se modeluje Poissonovým rozdělením  $Po(\lambda)$  s parametrem  $\lambda = a \cdot t$ , kde  $t$  je délka intervalu a  $a$  je parametr (v tomto případě je to, řekněme parametr četnosti příchozích hovorů).
- (a) Spočítejte jaká je pravděpodobnost, že za půl minuty nepřijde hovor.
- (b) Spočítejte s jakou pravděpodobností přijdou do ústředny za půl minuty méně než 3 hovory.
7. Ve spořitelně pracuje 15 mužů a 21 žen. 6 zaměstnanců z nich si sjednává stavební spoření.
- (a) Vypočítejte pravděpodobnost, že jsou mezi nimi právě dva muži, tj. dva muži a čtyři ženy.
- (b) Vypočítejte pravděpodobnost toho, že jsou to maximálně z jedné poloviny muži.
8. Mějme náhodnou veličinu  $X$ , jež sleduje Binomické rozdělení s parametry  $n = 9$  a  $\pi = 0,7$  (Symbolicky lze tento vztah vyjádřit  $X \sim \text{Bi}(9; 0,7)$ ). Zjistěte následující pravděpodobnosti:
- (a)  $P(X = 5)$ ,
- (b)  $P(X \leq 5)$ ,
- (c)  $P(X < 5)$ ,
- (d)  $P(X > 5)$ ,
- (e)  $P(X \geq 5)$ ,
- (f)  $P(3 < X \leq 5)$ ,
- (g)  $P(3 \leq X < 5)$ .
9. Mějme náhodnou veličinu  $Y$ , jež sleduje Poissonovo rozdělení s parametrem  $\lambda = 17$  (Symbolicky lze tento vztah vyjádřit  $Y \sim \text{Po}(17)$ ). Místo parametru často používáno udání střední hodnoty náhodné veličiny neboť  $EY = \lambda$ ). Zjistěte následující pravděpodobnosti:
- (a)  $P(Y = 12)$ ,
- (b)  $P(Y \leq 12)$ ,
- (c)  $P(Y > 12)$ ,
- (d)  $P(12 < Y \leq 20)$ .



10. Mějme náhodnou veličinu  $Z$ , jež sleduje Hypergeometrické rozdělení s parametry  $M = 15$ ,  $N = 36$  a  $n = 6$ . (Symbolicky lze tento vztah vyjádřit  $Z \sim H(15; 36; 6)$ . Značení není tak zažité jako u binomického a Poissonova rozdělení). Zjistěte následující pravděpodobnosti:
- (a)  $P(Z = 2)$ ,
  - (b)  $P(Z \leq 3)$ ,
  - (c)  $P(Z > 5)$ ,
  - (d)  $P(2 < Z \leq 6)$ .
11. Jaká je pravděpodobnost, že při deseti poctivých hodech poctivou hrací kostkou, tj. každá strana padá s pravděpodobností  $1/6$ ,
- (a) padnou samé šestky,
  - (b) nepadne ani jedna šestka,
  - (c) padne alespoň jedna šestka,
  - (d) padnou právě tři šestky?
  - (e) Jaký počet šestek je v sérii deseti hodů nejvíce pravděpodobný?
12. Co je pravděpodobnější, že
- (a) šesti kostkami hodíme alespoň jedenkrát šestku, nebo
  - (b) dvanácti kostkami hodíme alespoň dvakrát šestku?
13. V běžně velké populaci kryš se albín v průměru vyskytuje v 0,01 exemplářích, ostatní krysy jsou normálně pigmentované. Jaká je pravděpodobnost, že v běžně velké populaci
- (a) není žádný albín,
  - (b) je právě jeden albín,
  - (c) je alespoň jeden albín,
  - (d) jsou dva albíni za předpokladu, že je tam alespoň jeden?
  - (e) Jaká je pravděpodobnost, že pětkrát tak velká populace kryš, než byla uvažována doposud, skrývá jednoho albína?
14. V trolejbusu MHD je 30 cestujících. Z celkového počtu cestujících jede 8 „na černo“. Ostatní mají platný cestovní doklad. Do trolejbusu přistoupí revize jízdenek, která mezi zastávkami stihne zkontrolovat 12 náhodně vybraných cestujících. Jaká je pravděpodobnost, že mezi zkontrolovanými
- (a) je pět černých cestujících,
  - (b) jsou více jak tři černí cestující,

## 3.2 Náhodná veličina

(c) jsou pouze dva cestující s platným cestovním dokladem.

Teorie 15. Projděte příklady v oddílu 3.1 o náhodném jevu a vyhledejte ty, které je možno řešit pomocí známého rozdělení náhodné veličiny – diskrétní (alternativní, binomické, hypergeometrické, Poissonovo, negativněbinomické a geometrické). Hodnoty pravděpodobností přepočtete pomocí „excelovských“ funkcí BINOMDIST, HYPGEOMDIST, POISSON a NEGBINOMDIST.

Praxe

♣♦♥♠ 16. V souboru 08\_10\_22.pdf naleznete zadání průběžné písemky. Nebo viz Příloha D. Otestujte své znalosti ;-).



## 3.2.2 Spojitá náhodná veličina

1. Dokažte, že rozptyl<sup>19</sup> lze vyjádřit jako rozdíl aritmetického průměru čtverce náhodné veličiny a čtverce aritmetického průměru hodnot, tj.  $DX = EX^2 - (EX)^2$ . Toto tvrzení dokažte pro
- T○  
Střední
- (a) spojitou náhodnou veličinu (k důkazu použijte vzorce pro výpočet rozptylu a střední hodnoty) i
- (b) obecně (k důkazu použijte vlastnosti rozptylu a střední hodnoty).
2. Uvažujte rovnoměrné rozdělení na intervalu  $(0; 10)$ , symbolicky lze vyjádřit skutečnost, že náhodná veličina  $X$  sleduje toto rozdělení, takto –  $X \sim R(0; 10)$ .
- T○  
Střední
- (a) Napište předpis hustoty pravděpodobnosti tohoto rozdělení. Funkci zakreslete do grafu.
- (b) Napište předpis distribuční funkce tohoto rozdělení. Funkci zakreslete do grafu.
- (c) Vypočtěte střední hodnotu a rozptyl.
3. Mějme náhodnou veličinu  $X$ , jež sleduje Normální rozdělení s parametry  $\mu = 3$  a  $\sigma^2 = 3,61$  (Symbolicky lze tento vztah vyjádřit  $X \sim N(3; 3,61)$ ). Zjistěte následující pravděpodobnosti:
- (a)  $P(X = 4)$ ,
- (b)  $P(X \leq 4)$ ,
- (c)  $P(X > 4)$ ,
- (d)  $P(X \geq 4)$ ,
- (e)  $P(X < 4)$
- (f)  $P(3 < X \leq 5)$ ,
- (g)  $P(3 \leq X \leq 5)$ .

Dále zjistěte pro jakou hodnotu  $a$  je pravděpodobnost:

- (a)  $P(X \leq a) = 0,95$ ,<sup>20</sup>
- (b)  $P(X > a) = 0,25$ .
4. Mějme náhodnou veličinu  $Y$ , jež sleduje Studentovo rozdělení (nebo též  $t$ -rozdělení) s parametrem  $n = 43$  (říkáme s  $n$  stupni volnosti) (Symbolicky lze tento vztah vyjádřit  $Y \sim t(n)$ ). Zjistěte následující pravděpodobnosti:
- (a)  $P(Y = 0,5)$ ,

<sup>19</sup>samořejmě pokud  $DX$  existuje.

<sup>20</sup>pro náhodnou veličinu s rozdělením  $N(0; 1)$  se značí  $u_{0,95}$

### 3.2 Náhodná veličina

- (b)  $P(Y \leq 0,5)$ ,
- (c)  $P(Y \geq 0,5)$ ,
- (d)  $P(Y < -0,5)$ ,
- (e)  $P(-1 < Y \leq 0,5)$ .

Dále zjistěte pro jakou hodnotu  $a$  je pravděpodobnost:

- (a)  $P(Y \leq a) = 0,99$  (obvykle se značí  $t_{0,99}(43)$ ),
- (b)  $P(Y > a) = 0,12$ .

5. Mějme náhodnou veličinu  $Z$ , jež sleduje  $\chi^2$  rozdělení s parametrem  $n = 17$  (říkáme s  $n$  stupni volnosti) (Symbolicky lze tento vztah vyjádřit  $Z \sim \chi^2(n)$ ). Zjistěte následující pravděpodobnosti:

- (a)  $P(Z = 15)$ ,
- (b)  $P(Z \leq 15)$ ,
- (c)  $P(Z \geq 30)$ ,
- (d)  $P(Z < 5)$ ,
- (e)  $P(15 < Z \leq 30)$ .

Dále zjistěte pro jakou hodnotu  $a$  je pravděpodobnost:

- (a)  $P(Z \leq a) = 0,99$  (obvykle se značí  $\chi_{0,99}^2(17)$ ),
- (b)  $P(Z > a) = 0,12$ .

6. Mějme náhodnou veličinu  $W$ , jež sleduje Fisherovo-Snedecorovo rozdělení s parametry  $m = 11$ ,  $n = 19$  (říkáme o  $m$  a  $n$  stupních volnosti) (Symbolicky lze tento vztah vyjádřit  $W \sim F(m; n)$ ). Zjistěte následující pravděpodobnosti:

- (a)  $P(W = 15)$ ,
- (b)  $P(W \leq 12)$ ,
- (c)  $P(W \geq 13)$ ,
- (d)  $P(W < 5)$ ,
- (e)  $P(12 < W \leq 13)$ .

Dále zjistěte pro jakou hodnotu  $a$  je pravděpodobnost:

- (a)  $P(W \leq a) = 0,99$  (obvykle se značí  $F_{0,99}(11; 19)$ ),
- (b)  $P(W > a) = 0,05$ .

7. Najděte následující hodnoty:





- (a)  $u_{0,95}, u_{0,05}, u_{0,975}$ ,
- (b)  $t_{0,95}(10), t_{0,05}(10), t_{0,025}(16)$ ,
- (c)  $\chi_{0,9}^2(7), \chi_{0,1}^2(7), \chi_{0,99}^2(16)$ ,
- (d)  $F_{0,99}(5; 12), F_{0,01}(5; 12), F_{0,99}(12; 5), F_{0,01}(12; 5)$ .
8. O rozdělení IQ<sup>21</sup> u obyvatel je známo, že má normální rozdělení se střední hodnotou 100 a směrodatnou odchylkou 15.<sup>22</sup> Jaká je pravděpodobnost, že Váš kamarád(ka) má IQ:
- (a) větší než 85,
- (b) větší než 125,
- (c) mezi 90 a 110,
- (d) rovno 100?
- (e) Jaká je pravděpodobnost, že má IQ menší než 125 za předpokladu, že je inteligentně nadprůměrný?
- (f) Jakou by bylo nutno stanovit hranici pro přijetí do elitního „intoušského“ klubu<sup>23</sup>, aby do něj patřily jen 2 % lidí s nejvyšším IQ?
9. Pro oděvní továrnu je neziskové vyrábět pro velmi malé a velmi velké osoby. Ignoruje proto 7,5 % největších a 7,5 % nejmenších osob. Výška mužů se řídí normálním rozdělením  $N(69; 2,8^2)$  (míry jsou v palcích). Nalezněte nejmenší a největší výšku, pro kterou vyrábět.
10. Životnost žárovky v hodinách se řídí normálním rozdělením  $N(61; 6,3^2)$ . Výrobce garantuje, že pouze 3 % žárovek se spálí před garantovanou dobou. Určete tuto dobu.
11. Výtah má nosnost 700 kg. Průměrná váha osob v kg má rozdělení  $N(70; 400)$ .
- (a) Do výtahu náhodně nastoupilo šest osob. Jaká je pravděpodobnost, že výtah bude přetížen.<sup>24</sup>
- (b) Kolik osob musí nejvýše nastoupit, aby pravděpodobnost přetížení byla menší než 0,001? Těžké

<sup>21</sup>IQ – inteligentní kvocient – je číslo, popisující inteligenci člověka v poměru k populaci. Základní definici IQ vytvořil v roce 1912 německý psycholog WILLIAM STERN, když definoval míru inteligence jako poměr odhadnutého „mentálního“ věku a věku kalendářního. Hodnota IQ tedy vyjadřuje úroveň rozumových schopností jedince vzhledem k jeho věku a tvoří asi 20 % intelektové vybavenosti člověka.

<sup>22</sup>Většina testů inteligence ve Spojených státech inklinuje používat směrodatnou odchylku rovnou 15, nebo 16. Nicméně, evropská IQ test inklinují používat směrodatnou odchylku rovnou 24 nebo 25. Proto by IQ 130 (+2 směrodatné odchylky) v USA mělo odpovídat IQ 148–150 v Evropě.

<sup>23</sup>Mensa je mezinárodní společenská organizace založená roku 1946 v Oxfordu. Je to nevýdělečné apolitické sdružení nadprůměrně inteligentních lidí bez rozdílu rasy a vyznání. Více na Mensa ČR: <http://www.mensa.cz/>.

<sup>24</sup>Využijte vlastnosti:  $X_i \sim N(\mu; \sigma^2)$ ,  $i = 1, \dots, n$  a  $X_i$  jsou nezávislé veličiny, pak  $Y = \sum_{i=1}^n X_i \sim N(n \cdot \mu; n \cdot \sigma^2)$ .

## 3.2 Náhodná veličina

- Teorie 12. Projděte příklady v oddílu 3.1 o náhodném jevu a vyhledejte ty, které je možno řešit pomocí známého rozdělení náhodné veličiny – spojitě.
- ↔  
Praxe
- T○ 13. Nechť  $X$  je náhodná veličina sledující normální rozdělení s parametry  $\mu$  a  $\sigma^2$ . Dokažte, že náhodná veličina  $Y = \frac{X-\mu}{\sigma}$  sleduje normální normované rozdělení.
- Těžké
- T○ 14. Nechť  $X$  je náhodná veličina sledující normální rozdělení s parametry  $\mu$  a  $\sigma^2$ . Dokažte, že náhodná veličina  $Y = \frac{X-\mu}{s}$  sleduje studentovo rozdělení s  $(n - 1)$  stupni volnosti, kde  $s$  je výběrová směrodatná odchylka získaná z výběru o rozsahu  $n$ .
- Těžké
15. Uvažujme krychli s hranou délky  $X$ .  $X$  je náhodná veličina s rovnoměrným rozdělením na  $\langle 0; 10 \rangle$ . Určete střední hodnotu a rozptyl objemu krychle?
- ♣◇♥♠ 16. V souborech [08.12.04a.pdf](#) a [08.12.04b.pdf](#) naleznete zadání průběžné písemky. Nebo viz Příloha [E.1](#) a [E.2](#). Otestujte své znalosti ;-).



### 3.3 Zadání samostatné práce – Náhodná veličina

1. Zjistěte počet žen a mužů ve Vaší studijní skupině. Uvažujte, že ze svého středu budete vybírat prostřednictvím losovacích lístků a spočtete následující pravděpodobnosti.
  - (a) Spočtete jaká je pravděpodobnost, že z náhodně vybraných 10 studentek/studentů vylosujete pět žen, za předpokladu, že losovací lístky vracíte zpět?
  - (b) Za stejného předpokladu jako u 1a spočtete jaká je pravděpodobnost, že vylosujete méně než pět žen?
  - (c) Jaká je pravděpodobnost, že z náhodně vybraných 10 studentek/studentů budou právě dvě ženy, jestliže losovací lístky nevracíte?
  - (d) Za stejného předpokladu jako u 1c spočtete jaká je pravděpodobnost, že budou vybrány nejvýše dvě ženy a navíc víte, že tam alespoň jedna žena byla?<sup>25</sup>
  
2. Uvažujte normální rozdělení s parametry  $\mu$  a  $\sigma^2$  ( $N(\mu; \sigma^2)$ ) a náhodnou veličinu  $X$ , jež sleduje toto rozdělení, symbolicky  $X \sim N(\mu; \sigma^2)$ . Pro následující výpočty vymeze numericky  $\mu$  jako Váš věk v letech,  $\sigma^2$  jako Vaše číslo boty v číslování EUR a dvě hodnoty  $x_1$  a  $x_2$  tak, že  $x_1$  bude Vaše tělesná teplota ve stupních Celsia a  $x_2 = \mu + \sigma$ . Tedy např.  $\mu = 29$ ,  $\sigma^2 = 41$ ,  $x_1 = 36,7$  a pak nutně  $x_2 = 35,403$ .
  - (a) Vykreslete graf hustoty pravděpodobnosti tohoto rozdělení (STATISTICA, Maple, Excel) a vyznačte v něm střední hodnotu  $\mu$  a hodnoty  $\mu - \sigma$  a  $\mu + \sigma$ . Programem stačí vykreslit graf hustoty pravděpodobnosti. Zbytek můžete vyznačit vlastní rukou.
  - (b) Vypočtete pravděpodobnost, že náhodná veličina  $X$  nabyde hodnoty větší než je hodnota  $x_1$ . Symbolicky  $P(X > x_1)$ . Vyznačte tuto pravděpodobnost v grafu příslušné hustoty pravděpodobnosti – viz předchozí úkol (2a).
  - (c) Vypočtete pravděpodobnost, že náhodná veličina  $X$  je v absolutní hodnotě větší, než hodnota  $x_2$ , symbolicky  $P(|X| > x_2)$ . Vyznačte tuto pravděpodobnost v grafu příslušné hustoty pravděpodobnosti – křivka by měla být stále tatáž jako u 2a.
  - (d) Jakou největší hodnotu náhodné veličiny by jste mohli vzít tak, aby „pod ní leželo“ 75 % hodnot z veličiny  $X$ ,<sup>26</sup> symbolicky:  $P(X \leq ?) = F_X(?) = 0,75$ , či  $F_X^{-1}(0,75)$ ? Vyznačte tuto hodnotu v grafu příslušné hustoty pravděpodobnosti.

<sup>25</sup>Nápověda: využijte podmíněnou pravděpodobnost (viz náhodný jev), tj.  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  a dílčí pravděpodobnosti vypočtete skrze pravděpodobnost náhodné veličiny.

<sup>26</sup>aneb hledání kvantilu

### 3.3 Zadání samostatné práce I



## 4 Zpracování datového souboru

### 4.1 Třídění statistického souboru

1. V letním semestru školního roku 2004/2005 byl zaznamenáván zisk bodů ze zkouškové písemky předmětu Statistika v řádném termínu. Po prostém třídění dle statistického znaku „ $x_i$ “ = počet získaných bodů, byla získána následující data (viz tabulka 1). Doplňte tabulku četností o sloupce kumulativních (absolutních) četností, relativních četností a kumulativních relativních četností a pak ověřte, že:  $\sum_{i=1}^k n_i = N$ ,  $\sum_{i=1}^k p_i = 1$  (100 %),  $k_{n_k} = N$ ,  $k_{p_k} = 1$  (100 %).

Tabulka 1: Počet bodů získaných z písemné části zkoušky ze statistiky LS 2005 – řádný termín

$x_i$	$n_i$	$p_i$ (%)	$k_{n_i}$	$k_{p_i}$ (%)
0,0	27			
0,5	11			
1,0	20			
1,5	15			
2,0	14			
2,5	11			
3,0	22			
3,5	8			
4,0	7			
4,5	3			
5,0	1			
5,5	0			
6,0	0			
<b><math>\Sigma</math></b>	<b>139</b>			

2. Zvědavá studentka si ze statistické ročenky vypsala přehled o uzavřených sňatcích v roce 1998, roztržiděných podle věku nevěsty. Výpis se ztratil. Pamatovala si pouze údaje uvedené v tabulce (viz tabulka 2). Doplňte tabulku.
3. Ve dnech 10.–12. ledna 2003 se v Hradci Králové uskutečnil memoriál R. Myslíka, při jehož příležitosti proběhlo zjišťování antropometrických znaků a speciálních pohybových schopností u hráčů (ročník narození 1986 a mladší) volejbalu z příslušných krajů. Dodejme, že se jednalo o výběr nejlepších hráčů z krajů, kde se hraje a stabilně volejbal. Konkrétně se jednalo o následující statistické znaky: *hráčská specializace*<sup>27</sup>, *mateřský oddíl*, *datum narození*, *výška hráčů* (cm), *váha hráčů* (kg), *dosah jednou rukou i oběma rukama* (cm), *smečářský dosah* (cm), *blokařský dosah* (cm), *běh 4x10 m* (s), *hod 1kg plným míčem v sedu* (cm) a *skok snožný z místa* (cm). Data obsahuje soubor [memorial.xls](#).

<sup>27</sup>S = smečář, B = blokař, N = nahrávač, U = univerzál a L = libero

#### 4.1 Třídění statistického souboru

Tabulka 2: Počet uzavřených sňatků dle věku nevěsty (1998)

$x_i$	$n_i$	$p_i$ (%)	$k_{n_i}$
18	432		
19		4,176	
20			13 405
21	10 501		
22	9 116		
23		14,986	
24			
$\Sigma$	44 782		

- (a) Dopočítejte absolutní smečářský výskok, absolutní blokařský výskok a BMI index<sup>28</sup>. Přičemž absolutní blokařský výskok je rozdíl blokařského dosahu a dosahu oběma rukama, absolutní smečářský výskok je rozdíl smečářského dosahu a dosahu jednou rukou a BMI index je podíl hmotnosti v kilogramech a druhé mocniny výšky v metrech ( $BMI = \frac{m}{h^2}$  (kg/m<sup>2</sup>)).
- (b) Zařaďte volejbalisty dle BMI indexu do příslušných kategorií. Viz tabulka 3.

Tabulka 3: Zařazení do skupin dle hodnoty BMI indexu. Zdroj: iDnes

BMI	Popis
< 20	<b>Podváha</b> je typická pro modelky, je to však stav se zdravotními riziky.
<20, 25)	<b>Ideální</b> a vyvážený stav těla pro zdravý život.
<25, 30)	<b>Mírná nadváha</b> , která lehce zvyšuje zdravotní rizika.
<30, 40)	<b>Obezita</b> značí vysoká zdravotní rizika.
≥ 40	<b>Těžká obezita</b> je spojena s velmi vysokými zdravotními riziky.

- (c) Bez využití statistického softwaru postupně seřídíte statistický soubor dle hráčské specializace, mateřského oddílu, výšky hráčů a váhové skupiny určené BMI indexem. Mají všechny tabulky četností rozumnou vypovídací schopnost?
- (d) Pokud má vytvořené třídění smysl, vhodně jej vizualizujte např. pomocí polygonu četností, histogramu četnosti nebo výsečového grafu.
4. Pomocí programu STATISTICA komplet 6.1 Cz importujte „excelovská“ data některého ze souborů: [auta.xls](#), [student.xls](#) (FORMA: 81 = denní studium, 86 = dálkové studium; POHLAVÍ: 1 = muž, 2 = žena; SKTYP: 1 = učiliště s maturitou, 2 = gymnázium, 3 = střední odborná škola; PRUMSTR: průměrný prospěch na střední škole; MATUR: průměrný prospěch z maturity; PR1: získané body za matematiku; PR2: získané body za jazyk; PR3: získané body za ekonomický přehled. PRUMSTIP: stipendijní průměr; POCET4: počet čtyřek (neúspěchů) u zkoušek) a [memorial.xls](#).

<sup>28</sup>BMI = Body Mass Index – index tělesné hmotnosti



- (a) Dopačtete v programu STATISTICA komplet 6.1 Cz absolutní smečařský výskok, absolutní blokařský výskok a *BMI* index – viz úkol 3a.
  - (b) Zařadte prostřednictvím programu STATISTICA komplet 6.1 Cz volejbalisty dle BMI indexu do příslušných kategorií – opět podle tabulky 3.
  - (c) Pomocí programu STATISTICA komplet 6.1 Cz setřídte data. V případě intervalového třídění odhadněte počet intervalů pomocí Sturgesova pravidla.
  - (d) Výsledek vizualizujte (v programu STATISTICA komplet 6.1 Cz).
5. Pomocí programu STATISTICA komplet 6.1 Cz importujte data některého z textových souborů: [auta.txt](#), [student.txt](#) a [memorial.txt](#).

## 4.2 Charakteristiky úrovně, variability a koncentrace kvantitativního znaku

- T○  
Lehké
1. Dokažte, že platí vzorec pro výpočet rozptylu  $s_p^2 = \overline{x^2} - \bar{x}^2$  k důkazu použijte vzorce pro výpočet populačního rozptylu a aritmetického průměru).
- T○  
Lehké
2. Ověřte/dokažte, že platí vzorec pro Besselovu opravu (viz přednášky:  $s_p^2 = \frac{n-1}{n} s_v^2$ ).
  3. V následující tabulce (tabulka 4) najdete rozlohy krajů České republiky.

Tabulka 4: Rozloha krajů

Kraj	Rozloha (km <sup>2</sup> )
Pražský	496
Středočeský	11 015
Karlovarský	3 315
Plzeňský	7 560
Budějovický	10 056
Ústecký	5 335
Liberecký	3 163
Královéhradecký	4 757
Pardubický	4 519
Jihlavský	6 925
Brněnský	7 067
Zlínský	3 965
Olomoucký	5 139
Ostravský	5 554
<b>ČR celkem</b>	<b>78 866</b>

Vypočítejte následující číselné charakteristiky pro Českou Republiku: průměrnou velikost kraje, modus, medián, dolní, horní kvartil, maximum, minimum, variační rozpětí, kvartilové rozpětí, rozptyl, směrodatnou odchylku, výběrový rozptyl, výběrovou směrodatnou odchylku, variační koeficient, míru šikmosti, míru špičatosti.

4. Ve známých souborech: [auta.xls](#), [student.xls](#) a [memorial.xls](#) najdete statistické znaky reálných amerických aut, studentů této fakulty a účastníků Myslíkova memoriálu. Za pomoci programu MS Excel spočítejte nejen výběrové charakteristiky každého statistického znaku (viz úkol 3). Vypočtené charakteristiky komentujte.
5. Využijte naimportovaná data z oddílu 4.1 – úkol 4 nebo 5a pomocí programu STATISTICA komplet 6.1 Cz spočítejte popisné charakteristiky na těchto datech.
6. V tabulce 5 najdete informace o studentkách a studentech, kteří se nebáli dát všanc své osobní údaje, abychom mohli dělat trochu statistiky.





Tabulka 5: Osobní informace statečných studentů (přednáška ZS 2006/2007)

Výška (cm)	Číslo boty	Věk (roky)
173,0	39	20
178,0	37	20
164,0	37	20
179,0	43	21
164,0	38	20
163,5	37	20
182,0	44	20
187,0	45	21
169,0	38	20
180,0	41	21
170,0	38	24

Spočítejte číselné charakteristiky statistických znaků (viz úkol 3). Zamyslete se nad relevancí vypočtených statistik a komentujte je.

7. Manuálně, nebo za pomoci MS Excelu vypočtete na základě dat z oddílu 4.1 – úkol 2:
  - (a) průměrný věk nevěst,
  - (b) věk, ve kterém se dívky nejčastěji vdávají,
  - (c) variační koeficient stáří nevěst.
  - (d) Mladší kolika let je 50 % nevěst?
  - (e) Navíc pomocí programu STATISTICA komplet 6.1 Cz vypočtete  $\bar{x}$ ,  $s_v^2$ ,  $s_v$ . Použijte funkci *váhy*. Vycházejte z doplněné tabulky četností .
8. Vizualizujte popisné statistiky vypočtené pro studenty, auta, volejbalisty a nevěsty pomocí krabicového grafu (nastavení grafu zakládejte jak na „kvartilech“, tak na aritmetickém průměru a směrodatné odchylce.
9. Průměrná hrubá měsíční mzda<sup>29</sup> v České republice činila v roce 2005 přibližně 18 900 Kč. Předpokládejte, že u 5 % obyvatel ČR, z jejichž platu je tato průměrná mzda vypočítávána, je průměrná mzda 200 000 Kč. Kolik si tedy v průměru vydělává zbytek populace, aby byl republikový průměr právě 18 900 Kč?

<sup>29</sup>Průměrná hrubá měsíční mzda představuje podíl mezd bez ostatních osobních nákladů připadajících na jednoho zaměstnance evidenčního počtu za měsíc. Do mezd se zahrnují základní mzdy a platy, příplatky a doplatky ke mzdě nebo platu, prémie a odměny, náhrady mezd a platů, odměny za pracovní pohotovost a jiné složky mzdy nebo platu, které byly v daném období zaměstnancům zúčtovány k výplatě. Jedná se o hrubé mzdy, tj. před snížením o pojistné na všeobecné zdravotní pojištění a sociální zabezpečení, zálohové splátky daně z příjmů fyzických osob a další zákonné nebo se zaměstnancem dohodnuté srážky. Zdroj: ČSÚ

## 4.2 Charakteristiky úrovně, ...

10. Vypočtete průměrnou nezaměstnanost za roky 1993–2007 za Českou republiku. Nezaměstnanost za jednotlivé roky byla: 4,3; 4,3; 4,0; 3,9; 4,8; 6,5; 8,7; 8,8; 8,1; 7,3; 7,8; 8,3; 7,9; 7,1 a 5,3.<sup>30</sup>
11. Vypočtete průměrnou míru inflace za roky 1993–2007 za Českou republiku. Míry inflace za jednotlivé roky byly: 20,8; 10,0; 9,1; 8,8; 8,5; 10,7; 2,1; 3,9; 4,7; 1,8; 0,1; 2,8; 1,9; 2,5 a 2,8.<sup>31</sup>
- ♣♦♥♠ 12. V souborech [05\\_11\\_09.pdf](#), [06\\_04\\_03.pdf](#) a [06\\_04\\_05.pdf](#) (nebo viz Příloha [F.1–F.3](#)) naleznete zadání „přepadové“ písemky. Otestujte své znalosti ;-).

---

<sup>30</sup>Data byla získána z ČSU (<http://www.czso.cz>)

<sup>31</sup>Data byla získána z ČSU (<http://www.czso.cz>)



### 4.3 Zadání samostatné práce – Zpracování vlastního datového souboru

1. Vytvořte či jinak získajte datový soubor
  - (a) Pro potřeby následující analýzy získajte data, která jsou Vám blízká. Buďte provedte vlastní šetření, nebo využijte data, jež máte k dispozici z praxe, nebo z jiných zdrojů. Rozhodně si data nevymýšlejte! Datový soubor by měl mít alespoň 50 statistických jednotek. Vzhledem k tomu, že data budete zpracovávat s využitím softwaru, nebude pro Vás pracnější, budete-li mít soubor rozsáhlejší. Navíc na rozsáhleším souboru pro Vás možná budou závěry jednodušší, snadněji interpretovatelné a v neposlední řadě statisticky hodnotnější. Váš datový soubor by měl obsahovat několik statistických znaků pro každou statistickou jednotku:
    - i. alespoň jeden spojitý numerický znak,
    - ii. alespoň jeden diskrétní numerický znak,
    - iii. alespoň jeden ordinální znak (minimálně 3 kategorie),
    - iv. alespoň jeden další nominální znak (minimálně 2 kategorie).

Např. viz tabulka 6.

Tabulka 6: Příklad datového souboru daných parametrů

Věk	Pohlaví	Stav	Vzdělání	Zaměstnání	Příjem	Úvěr	Splátek	Pojištění
59	žena	ženatý	základní	důchodce	7 200 Kč	5 390 Kč	20	A
27	žena	ženatý	střední	dělník	7 000 Kč	7 542 Kč	20	A
50	muž	rozvedený	střední	kuchař	61 000 Kč	6 216 Kč	10	Ne
29	muž	svobodný	vyučený	dělník	10 000 Kč	7 002 Kč	20	A
31	muž	ženatý	vyučený	řidič	15 000 Kč	8 982 Kč	10	Ne
19	žena	druh	základní	mateř dovolená	5 500 Kč	6 696 Kč	10	A
22	muž	svobodný	vyučený	malíř, natěrač	10 000 Kč	4 621 Kč	20	A
34	muž	ženatý	střední	stát. zam.	15 159 Kč	7 624 Kč	30	A
45	žena	ženatý	vyučený	podnikatel	10 000 Kč	7 515 Kč	20	A
24	muž	rozvedený	vyučený	technik	12 000 Kč	6 680 Kč	20	A
30	muž	rozvedený	vyučený	pekař	12 500 Kč	3 228 Kč	20	A
25	muž	svobodný	střední	pol. inspektor	14 000 Kč	14 229 Kč	30	A
27	muž	druh	vyučený	technik	12 500 Kč	10 791 Kč	10	Ne
55	muž	ženatý	základní	topič	8 000 Kč	8 540 Kč	50	A
37	muž	ženatý	střední	pol. inspektor	19 000 Kč	4 203 Kč	10	Ne
36	žena	rozvedený	střední	účetní	11 000 Kč	5 841 Kč	10	Ne
:								

- (b) Zevrubně popište Vámi používaný datový soubor (původ, statistické znaky, ...).
  - (c) Zamyslete se nad tím, co je pro Vás statistická jednotka, jakou populaci Vaše šetření popisuje a z jaké populace by naopak mohlo být výběrem.
2. Deduktivní zhodnocení datového souboru. Pro korektní použití metod v tomto úkolu, předpokládejte, že datový soubor získaný v 1 je *úplným* šetřením, tj. že se jedná o základní soubor (populaci). Zpracujte datový soubor. Každý statistický znak musí být nějakým způsobem statisticky zpracován. Zejména proveďte následující:

### 4.3 Zadání samostatné práce II

- (a) Pomocí popisných statistik, jež znáte a jež jsou pro daný znak vhodné, zhodnoťte statistické znaky  $(\bar{x}, \hat{x}, V_X, \dots)$ . Výsledky komentujte.
- (b) Proveďte intervalové třídění spojitého numerického znaku. Počet intervalů odhadněte pomocí Sturgesova vzorce.
- (c) Vypočtěte váženou variantu aritmetického průměru, rozptylu a směrodatné odchylky na základě intervalového třídění vytvořeného v předchozím úkolu (2b). Proč se tyto hodnoty liší od hodnot vypočtených před tříděním (viz 2a)?
- (d) Zvolte vhodnou vizualizaci dat pomocí grafů – jak popisných charakteristik vypočtených v 2a, tak tabulek četností sestavených v 2b.



## 5 Základy teorie odhadu

### 5.1 Bodový a intervalový odhad $\mu$ , $\sigma^2$ a $\pi$

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. Vypočítejte<sup>32</sup> střední hodnotu a rozptyl výběrového průměru  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ . Předpokládejte, že populace, ze které byl (náhodný) výběr získán, sleduje libovolné rozdělení se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ . ○T○  
Těžké
2. Odvoďte<sup>33</sup> intervalový odhad pro střední hodnotu normálního rozdělení při známém rozptyle  $\sigma^2$ . Spolehlivost volte  $1 - \alpha$ . ○T○  
Střední
3. Odvoďte<sup>34</sup> intervalový odhad pro střední hodnotu normálního rozdělení při neznámém rozptyle, který je odhadnutý výběrovým rozptylem  $s^2$ . Spolehlivost volte  $1 - \alpha$ . ○T○  
Střední
4. Odvoďte intervalový odhad pro rozptyl normálního rozdělení. Spolehlivost volte  $1 - \alpha$ . ○T○  
Střední
5. Ve výrobě granulovaných krmiv byla u granulátu TL 700 zjišťována životnost matrice v provozních hodinách (Šlo o matrice s průměrem 5 mm používané na výrobu krmné směsi A1 – pro prasata). U 15 matric byla zjištěna životnost v pracovních hodinách (viz tabulka 7).

Tabulka 7: Životnost matric v provozních hodinách

440	490	605	510	540	470	580	600
550	620	575	660	560	540	480	

- (a) Bodově odhadněte střední hodnotu a směrodatnou odchylku životnosti matrice.
- (b) Stanovte 95% interval spolehlivosti pro průměrnou životnost matrice.
- (c) Stanovte 95% levostranný interval spolehlivosti pro průměrnou životnost matrice. Tj. kolik hodin minimálně vydrží 95 % průměrných matric (lepších).
- (d) Stanovte rozsah výběrového souboru, tak aby při 95% intervalovém odhadu průměrné životnosti matrice byla chyba nejvýše  $\pm 20$  provozních hodin (maximální přípustná chyba  $\Delta = 20$ ).

<sup>32</sup>Pro výpočet využijte vlastností střední hodnoty, rozptylu a vlastností náhodného výběru.

<sup>33</sup>Pro odvození využijte výsledků z úkolu 1 a vlastností náhodných veličin.

<sup>34</sup>Pro odvození využijte výsledků z úkolu 1 a vlastností náhodných veličin.

## 5.1 Bodový a intervalový odhad

- (e) Jaká je spolehlivost odhadu použijeme-li původní data (k dispozici je pouze zmíněných 15 matric), avšak maximální přípustná chyba nesmí převýšit hodnotu 10 provozních hodin?
6. Na základě příkladu 5 určete 95% interval spolehlivosti pro směrodatnou odchylku životnosti matrice.
7. Při výlovu rybníka bylo náhodně vybráno a zváženo 15 kaprů. Naměřené hmotnosti jednotlivých kaprů lze najít v tabulce 8.

Tabulka 8: Váha kaprů v gramech

3000	3100	2500	2500	4200	2100	3250	2500
4800	2300	4100	3600	3000	3600	4000	

- (a) S 99% spolehlivostí odhadněte průměrnou hmotnost kapra.
- (b) Jakou minimální garantovanou nosnost musí mít taška, aby unesla **průměrného** kapra s 95% spolehlivostí?
- (c) Jakou minimální garantovanou nosnost musí mít taška, aby unesla **libovolného** kapra s 95% spolehlivostí?<sup>35</sup>
- (d) Pro publikaci výsledků výlovu je třeba uvést výsledky s předem určenou přesností. Kolik je třeba zvážít kaprů, aby bylo možno odhadnout střední hodnotu váhy kapra s 99% spolehlivostí a maximální přípustnou chybou 0,25 kg?
- (e) S jakou spolehlivostí lze odhadnout střední hodnotu váhy kapra v rybníce při maximální přípustné chybě nutné pro publikaci (0,25 kg) pouze na základě stávajícího výběru čítajícího 15 kaprů?
8. Na přednáškách Zemědělské fakulty Jihočeské univerzity v Českých Budějovicích bylo provedeno mezi studenty náhodné šetření zjišťující poměr dívek a žen na přednáškách. Rozsah šetření byl  $n = 79$  studentů, z toho dívek a žen bylo 54.
- (a) Intervalově, s 90% spolehlivostí, odhadněte populační pravděpodobnost dívek na přednáškách.

<sup>35</sup>Chování náhodných veličin a jejich průměrů jsou poněkud odlišné. Přestože mají „shodnou“ střední hodnotu, rozptyly se nerovnájí. Navíc dvě pozorování významně vybočující z intervalu spolehlivosti pro střední hodnotu, tam mohou v průměru bez problémů patřit. Jako ilustraci jeden „statistický kameňák“: Tři kamarádi jdou na hon. Jeden je zatvrzelý republikán, druhý nepoučitelný komunista a třetí – statistik. Každý je svým způsobem postizen osobním přesvědčením, které se projeví na jejich střelbě. První střílí republikán a o jeden metr vpravo mine. Druhý, s ironickým úsměvem, vystřelí komunista. Kulka mine cíl o metr vlevo a republikán vrátí kolegovi úšklebek. Oběma střilejícím však vyrazí dech jejich kolega statistik, který složí zbraň a má se k odchodu. Na jednohlasnou otázku vystřelivších, „proč ... NESTŘÍLÍŠ?!?!“ stoicky klidně odpoví otázkou: „Proč bych střílel, přátelé, v průměru už jsme ho přece dostali?“



- (b) O kolik studentů by se musel výběr rozšířit, aby bylo možno s 90% odhadnout populační pravděpodobnost s maximální přípustnou chybou  $\Delta = 0,5\%$ ?
9. Firmy Bravo a Happy Day garantují zákazníkům minimálně 50% obsah pomerančové šťávy v pomerančových nektarech. Laboratorními zkouškami byly v osmi náhodně vybraných vzorcích naměřeny hodnoty, jež jsou uvedeny v tabulce 9.

Tabulka 9: Obsah pomerančové šťávy (%)

<b>Bravo</b>	58,8	50,2	53,5	49,4	51,7	47,9	50,3	53,3
<b>Happy Day</b>	51,1	49,7	53,8	52,3	50,5	49,1	52,8	57,1

Stanovte 95% interval spolehlivosti pro průměrnou hodnotu a směrodatnou odchylku obsahu pomerančové šťávy u jednotlivých značek.

10. Uvažujte náhodné výběry ze dvou různých populací.
- (a) Pro jakou populaci lze vytvořit přesnější (užší) 95% interval spolehlivosti pro populační pravděpodobnost, jsou-li odhady populačních pravděpodobností rovny  $p_1 = 0,01$  a  $p_2 = 0,48$  a rozsahy výběrů jsou  $n_1 = 78$  a  $n_2 = 100$ ?
- (b) Který interval spolehlivosti pro populační pravděpodobnost by byl užší, kdyby se odhady populační pravděpodobnosti rovnaly ( $p_1 = p_2$ ) a rozsahy výběrů byly  $n_1 = 78$  a  $n_2 = 100$ ?
- (c) [K zamyšlení] O kolik procent vyšší respektive nižší je třeba uskutečnit výběr z první populace, aby byla zachována stejná spolehlivost a přesnost intervalového odhadu populační pravděpodobnosti jako u výběru z druhé populace? Vyřešte obecně, uvažujte následující bodové odhady populačních pravděpodobností v jednotlivých populacích:
- $p_1 = 0,45$  a  $p_2 = 0,999$ ,
  - $p_1 = 0,001$  a  $p_2 = 0,999$ .
11. Opakovanými měřeními byla zjištěna tloušťka vlákna. Uvedené hodnoty v tabulce 10 jsou v  $10^{-6}$  m.

Tabulka 10: Tloušťky vláken ( $10^{-6}$  m)

210	217	209	216	216	215	220	214	213
-----	-----	-----	-----	-----	-----	-----	-----	-----

- (a) Je známo, že měření mají právě rozdělení  $N(\mu; 25)$ . Nalezněte 95% interval spolehlivosti pro  $\mu$ .
- (b) Nalezněte 95% interval spolehlivosti pro  $\mu$  za předpokladu, že rozptyl není znám a je ho proto nutno odhadnout výběrovým rozptylem  $s^2$ .

## 5.1 Bodový a intervalový odhad

- K zamyšlení
12. Ankety o legalizaci měkkých drog se zúčastnilo 500 osob. Jaká je maximální přípustná chyba tohoto šetření chceme-li 95% odhad procenta osob souhlasících s legalizací.
  13. Projděte „vzorově“ řešené příklady ([cv\\_5.pdf](#) a [cv\\_7.pdf](#)) viz podadresář tohoto cvičení – [rpzs](#)<sup>36</sup>. Zkuste vyřešit samostatně.
  - ♣♦♥♠ 14. V souboru [04\\_04\\_14.pdf](#) naleznete zadání průběžné písemky. Nebo viz Příloha G. Otestujte své znalosti ;-).

---

<sup>36</sup>resene príklady ze skript (lehce modifikované)





## 5.2 Zadání samostatné práce – Teorie odhadu

1. Induktivní zhodnocení datového souboru, který máte (je výsledkem úkolu 1). Pro korektní použití metod v tomto úkolu, předpokládejte, že hodnoty spojitého numerického znaku jsou *výběrovým* šetřením, tj. že se jedná o výběr ze základního souboru (populace), jež má normální rozdělení.
  - (a) Bodově odhadněte střední hodnotu ( $\mu$ ) a rozptyl ( $\sigma^2$ ) pro soubor (populaci), z něž pochází spojitý numerický znak.
  - (b) Intervalově odhadněte střední hodnotu souboru, z něž pochází spojitý numerický znak s 95% a 99% spolehlivostí. Napište odpověď, ze které bude patrné jak se intervalový odhad interpretuje.
  - (c) Intervalově odhadněte rozptyl souboru, z něž pochází spojitý numerický znak s 95% spolehlivostí.
  - (d) Určete potřebný rozsah výběru (kolik měření by bylo nutno provést) tak, aby hodnota maximální přípustné chyby pro 95% interval spolehlivosti střední hodnoty souboru, z něž pochází spojitý numerický znak byla rovna

$$\Delta = \frac{x_{max} - x_{min}}{\tilde{x}_{0,75} - \tilde{x}_{0,25}}.$$

Kolik jednotek výběru by musel obsahovat „dovýběr“ (viz dvoufázový výběr)?

- (e) Na základě hodnot spojitého numerického znaku spočítejte spolehlivost intervalového odhadu pro střední hodnotu s maximální přípustnou chybou rovnou čtvrtině délky 95% intervalu spolehlivosti (viz výpočet v úkolu 1b).
- (f) Pokud jste získali data s nominálním znakem o dvou kategoriích (např.: muž/žena, plátce DPH/neplátce DPH), označte počet pozorování jedné kategorie  $n_1$  a druhé  $n_2$ . Pokud máte data s více jak dvěma nominálními kategoriemi, pak pozorování logicky seskupte do dvou skupin (např.: {svobodný, druh} a {ženatý, rozvedený, vdovec}) a určete počet pozorování  $n_1$  a  $n_2$  v těchto kategoriích. V obou případech máte/získali jste alternativní statistický znak. Intervalově odhadněte populační pravděpodobnost (relativní četnost) alternativního statistického znaku s 90% spolehlivostí. Jaký je bodový odhad populačního pravděpodobnosti (poměru)?

## 5.2 Zadání samostatné práce III



## 6 Testování hypotéz

### 6.1 Testování na základě jednoho výběru – Testy $\mu$ , $\sigma^2$ a $\pi$ proti konstantě

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. V příloze C, nebo v doplňkovém textu [p-value.pdf](#) naleznete interpretaci hodnoty  $p$ -value. Pečlivě prostudujte.
2. Do obchodu jsou dodávány balení cukru o deklarované hmotnosti 500 g. Dávkování ve výrobním závodě probíhá automaticky. Bylo provedeno náhodné nezávislé šetření zaměřené na hmotnosti jednotlivých balení, jehož výsledky máte k dispozici (viz tabulka 11).

Tabulka 11: Výsledky měření dávkovacího automatu (g)

525	509	524	508	573	484	552	524	454	502	521	526
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- (a) Rozhodněte na hladině významnosti  $\alpha = 0,05$ , zda je automat dobře seřízen.
  - (b) Má pravdu obsluhující pracovník, který tvrdí, že automat dávkuje více, než je deklarovaná hmotnost? Pracovníkovu hypotézu otestujte na hladině významnosti  $\alpha = 0,05$ .
  - (c) Pan Volek se cítí podváděn, neboť mu cukr nevydrží tak dlouho jako dříve. Je jeho domněnka o „šizení“ zákazníků správná? Volte  $\alpha = 0,05$ .
3. Firmy Bravo a Happy Day garantují zákazníkům minimálně 50% obsah pomerančové šťávy v pomerančových nektarech. Laboratorními zkouškami byly v osmi náhodně vybraných vzorcích naměřeny hodnoty, jež jsou uvedeny v tabulce 9.
    - (a) Otestujete na hladině významnosti  $\alpha = 0,05$  zda se obsah šťávy v nektarech významně liší od 50 %.
    - (b) Otestujte, zda uvedené firmy dodržují své závazky vůči zákazníkům.
  4. Na základě výběrového šetření byla získána data o cenách bytů v Českých Budějovicích (příslušná data z dubna 2006 jsou v tabulce 12). Má pravdu ten, který se domnívá, že chce-li si pořídit byt, musí mít více než 890 000 Kč? Hladinu významnosti volte 5 %.

## 6.1 Testování hypotéz na základě jednoho výběru

Tabulka 12: Ceny bytů v Českých Budějovicích – duben 2006

Typ	Vlastnictví	cena (Kč)
byt 2+1/B	Družstvo vlastníků	980 000
byt 3+1/B	Osobní	1 050 000
byt 3+1	Družstvo vlastníků	1 100 000
byt 2+1	Osobní	1 000 000
byt 3+1/L	Družstevní	1 430 000
byt 3+1/B	Osobní	1 170 000
byt 3+1	Družstvo vlastníků	1 160 000
byt 1+0	Družstvo vlastníků	540 000
byt 3+1	Družstevní	1 200 000
byt 3,5+1	Družstevní	1 000 000
byt 3+1	Družstvo vlastníků	1 200 000
byt 3+1	Družstvo vlastníků	900 000
byt 2+1	Osobní	1 050 000
byt 1+1	Družstvo vlastníků	660 000
byt 2+1	Družstevní	970 000
byt 3+1	Družstevní	760 000
byt 2+1	Družstevní	1 000 000
byt 3+1	Družstvo vlastníků	980 000
byt 3+1	Družstevní	690 000
byt 3+1	Družstvo vlastníků	760 000

Tabulka 13: Pevnost vláken bavlněné příze

2,22	3,54	2,37	1,66	4,74	4,82	3,21	5,44
3,23	4,79	4,85	4,05	3,48	3,89	4,9	5,37

5. Pro bavlněnou přízi je předepsána horní mez variability pevnosti vlákna: rozptyl pevnosti nemá překročit  $\sigma_0^2 = 0,36$ . Při zkoušce 16 vzorků byly zjištěny výsledky shrnuté do tabulky 13.

Je důvod k podezření na vyšší nestejnoměrnost, než je stanoveno (pro  $\alpha = 0,01$ )?

6. Zástupci ekologického hnutí aktivně vystupují proti výstavbě nové továrny v oblasti, jejíž životní prostředí je již tak dost poznamenané průmyslovou činností. Předpokládají, že jedním z důsledků nadměrné industrializace je i nízká porodní váha novorozenců v dané oblasti. Má smysl, aby použili nižší porodní váhu jako argument proti výstavbě nové továrny, když vědí, že porodní váha zdravé populace má normální rozdělení se střední hodnotou 3300 g? Své tvrzení chtějí prokázat na souboru 25 náhodně vybraných novorozenců z této oblasti narozených v minulém roce, u nichž naměřili průměrnou váhu 3010 g a výběrovou směrodatnou odchylku 476 g.



## 6.1 Testování hypotéz na základě jednoho výběru

7. Na většině vysokých škol probíhá pravidelně hodnocení výuky. Výsledky pak slouží učitelům a grantům předmětů jako zpětná vazba o kvalitě výuky. Je velkou škodou, že studenti svého práva nevyužívají a nepodílí se tak na zkvalitnění výuky. Tabulka reftab:hodnoceni obsahuje počty studentů hodnotící výuku na Západočeské univerzitě v Plzni za zimní semestr 2002/2003. Výsledky jsou rozděleny dle jednotlivých fakult. Předpokládejme, že tyto výsledky reprezentují ochotu studentů jednotlivých fakult hodnotit výuku a považujeme je za náhodný vzorek za jednotlivé fakulty bez ohledu na rok, či semestr.

Tabulka 14: Počty studentů Západočeské univerzity v Plzni (dle jednotlivých fakult), kteří mohli/hodnotili výuku ve školním roce 2002/2003 – zimní semestr

Fakulta	Zapsaných studentů	Hodnotících studentů
Fakulta aplikovaných věd	1300	357
Fakulta ekonomická	1407	175
Fakulta elektrotechnická	1653	369
Fakulta humanitních studií	1718	163
Fakulta pedagogická	2886	70
Fakulta právnická	1904	61
Fakulta strojní	1212	170

- (a) Otestujete hypotézu (pro každou fakultu zvlášť), že k hodnocení přichází méně než 10 % studentek a studentů. Hladinu významnosti volte  $\alpha = 0,1$ .
- (b) Otestujete hypotézu, že k hodnocení na Západočeské univerzitě v Plzni (celé) přichází méně než 10 % studentek a studentů. Hladinu významnosti volte  $\alpha = 0,1$ .
8. Má pravdu porybný, který tvrdí, že jeho kapři mají v průměru 4 kg? Náhodný výběr kaprů – jejich váhy – je uvedený v tabulce 8.
9. Ve zimním semestru školního roku 2008/2009 byla získána data (tabulka 15), která ukazují, kolik peněz studentky a studenti utratí v restauračním zařízení během jedné návštěvy. Data jsou roztríděna v závislosti na studijním oboru.

Tabulka 15: Útraty studentek a studentů v restauračních zařízení – listopad 2008

<b>OP</b>	150	700	0	300	25	70	240	550	150	160	40	200				
<b>ÚFRP</b>	150	250	0	205	45	20	500	130	115	320	311	600	0	0	0	237

Rozhodněte na 0,1% hladině významnosti, zda studentky a studenti jednotlivých oborů utrácejí během návštěvy restauračního zařízení více než já ;- ) – tj. 50 Kč.

## 6.2 Testování na základě dvou výběrů – Testy na shodu středních hodnot, relativních četností a rozptylů

V následujících příkladech předpokládejte, že data jsou porížena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

- Byla porovnáována účinnost dvou protikorozních látek. První látka byla aplikována v  $n_1 = 10$  případech druhá v  $n_2 = 11$  případech. Po stanovené době byl zjištěn stupeň poškození s těmito výsledky (v mikrometrech):  $\bar{x}_1 = 82,4 \mu\text{m}$ ;  $s_1^2 = 12,1 \mu\text{m}^2$ ;  $\bar{x}_2 = 80,0 \mu\text{m}$ ;  $s_2^2 = 10,5 \mu\text{m}^2$ . Porovnejte průměrný účinek obou látek.
- Ve výrobě betonu bylo vyrobeno a vyzkoušeno 11 krychlí betonu ze dvou různých směsí. Otestujte na hladině významnosti  $\alpha = 0,05$ , zda jsou oba druhy betonu v průměru stejně kvalitní. Zjištěná pevnost je v MPa a údaje obsahuje následující tabulka (viz tabulka 16).

Tabulka 16: Pevnost betonových kvádrů (MPa)

<b>1. směs</b>	18	19	19	21	22	20	19	21	22	18	19
<b>2. směs</b>	19	20	21	20	18	22	21	19	21	20	21

- U deseti dojníc byla ve dvou obdobích měřena tučnost mléka. V období před zahájením zkrmování skrojků a v období při zkrmování skrojků. Na základě výsledků pokusů (viz tabulka 17) posuďte, zda při zkrmování řepných skrojků dochází:

Tabulka 17: Průměrná tučnost mléka (%)

Číslo dojnice	1	2	3	4	5	6	7	8	9	10
<b>Před zkrmováním</b>	4,0	3,7	3,8	3,5	4,2	3,8	3,9	4,0	4,2	3,9
<b>Při zkrmování</b>	4,1	4,1	4,0	3,8	4,2	4,0	4,3	4,3	4,1	4,1

- k průkazné změně v tučnosti mléka,
  - k průkaznému zvýšení tučnosti mléka.
- Je třeba „porovnat“ dva způsoby měření kompresního tlaku (v  $10^5$  Pa) ve spalovacím motoru. Pro 6 motorů, u nichž bylo provedeno měření kompresního tlaku oběma způsoby, byly stanoveny difference získaných výsledků. Výsledky uvádí tabulka 18.
    - Porovnejte oba způsoby měření, tj. jejich shodu či rozdílnost (v průměru) na hladině významnosti  $\alpha = 0,01$ .



Tabulka 18: Diference kompresních tlaků ( $10^5$  Pa)

Číslo motoru	1	2	3	4	5	6
diference $d_i$	-0,1	-0,2	0,2	0,1	-0,2	0,1

- (b) Za předpokladu, že diference vznikaly jako rozdíl naměřených hodnot druhého měření od prvního  $d_i = x_{1,i} - x_{2,i}$ , pro  $i = 1, \dots, 6$ , rozhodněte o platnosti tvrzení, že měření druhým způsobem dává signifikantně vyšší průměrné výsledky na hladině významnosti  $\alpha = 0,1$ .
- Na základě dat o pomerančových nektarech (viz tabulka 9) otestujte na hladině významnosti  $\alpha = 0,05$ , zda pomerančové nektary firmy Bravo obsahuje v průměru více pomerančové šťávy než výrobky Happy Day.
  - V Litvínově byla naměřena rychlost (km/hod) u náhodně vybraných aut kolem poledne a kolem půlnoci. Pomocí testu na hladině významnosti  $\alpha = 0,1$  zjistěte, zda je v noci signifikantně vyšší průměrná rychlost než ve dne. Naměřené hodnoty naleznete v tabulce 19.

Tabulka 19: Rychlosti naměřené v Litvínově (km/hod)

Poledne	55	90	50	65	63	72	69	59	50	52	70	85	68
	71	70	51										
Půlnoc	75	82	52	65	58	72	90	110	56	55	85	100	77

- Jak již víte z příkladu 7 v kapitole 6.1 probíhá na většině vysokých škol pravidelně hodnocení výuky. Na 5% hladině významnosti rozhodněte zda fakulty, u kterých k hodnocení přichází méně než 10% studentek a studentů – viz motivace, data (tabulka 14) a výsledky z příkladu 7 v kapitole 6.1 – mají stejný poměr hodnotících studentů.
- U řady odrůd byla vyšetřena potencionální možnost zásobování živinami měřená průřezem mízních pletiv. Porovnejte výsledky ve dvou po sobě následujících sezónách (Zůstaly hodnoty průřezy v meziročním srovnání shodné?). Výsledky měření naleznete v tabulce 20.
- V rámci výuky byl uspořádán následující pokus. Náhodně byla vybrána šestice studentů, kteří měli co nejrychleji (libovolně) nasadit deset kancelářský sponek podél strany listu papíru velikosti A4. Pokus byl prováděn s dvojnásobným opakováním. Jednou s otevřenými a jednou se zavřenými očima. Polovina studentů začínala „poslepu“, druhá polovina „poslepu“ prováděla opakovaný pokus. Dosažené časy jsou uvedeny v tabulce 21. Dá se říci, že vyřazení zraku, v průměru zhoršuje výsledky testu?
- Výsledky dovednostní soutěže, při níž jde o zasouvání tužky zavěšené na provázku do hrdla láhve obsahuje tabulka 22. Pokus byl proveden na přednášce z předmětu

## 6.2 Testování hypotéz na základě dvou výběrů

Tabulka 20: Zásobování živinami

Odrůda	1985	1986	Odrůda	1985	1986
Mexico 50	1,45	1,19	NA – 4	1,66	1,64
Super Zlatana	1,67	1,44	Weihenstephan	1,58	1,71
Kosutka	1,45	1,27	Maris Huntsman	1,77	1,75
Sava	1,46	1,60	Kormoran	1,55	1,56
Vala	1,47	1,56	Margin	1,46	1,86
Karlik	1,41	1,32	Caribo	1,53	1,77
Roazon	1,52	1,58	Kavkaz	1,80	1,71
BU 17	1,67	1,73	Iljicovka	1,50	1,61
Maris Fundin	1,58	1,36	Marinovska 10	1,70	1,57
Maris Marksman	2,10	1,69	Marinovska	1,45	1,48
UH 1072	1,48	1,32	UH 1072	1,98	1,37
Slavia	1,30	1,41	Slavia	1,41	1,46
Amika	1,54	1,55	Amika	1,19	1,11

Tabulka 21: Doba nutná k nasazení sponek (s)

Student(ka) č.	1	2	3	4	5	6
Zrak – ANO	31,46	26,50	42,38	30,03	35,02	20,75
Zrak – NE	36,06	44,30	57,12	40,46	34,02	38,01

Statistika (ZS 2006/2007) na ZF JU v Českých Budějovicích. Náhodný výběr aktérů byl proveden zvláště (odděleně) u studentek a u studentů. Rozsah výběru byl v obou skupinách shodně 9 osob. Hodnoty v řádcích tabulky přináší dobu po „úspěšné zasunutí“ ve vteřinách: **ZZ** – časy studentek s tužkou zavěšenou na provázku na bedrech v místě pasu, **MZ** – časy studentů s tužkou zavěšenou na provázku na bedrech v místě pasu, **ML** – časy týž studentů s tužkou zavěšenou na provázku v místě loketní jamky. Studenti tedy postoupili opakované měření.

Tabulka 22: Doba nutná k zasunutí tužky do láhve (s)

Student(ka) č.	1	2	3	4	5	6	7	8	9
<b>ZZ</b>	15,9	17,9	25,6	29	35	45	50,7	87	200
<b>MZ</b>	35,4	18,4	41,4	59	27,1	53,5	31,3	28,5	16,6
<b>ML</b>	6	4,9	7,2	11,5	10,4	4	8,9	8	13

- Rozhodněte, na jakou populaci by teoreticky bylo možno výsledky testu zobecnit.
- Jsou studentky v populaci určené výše v průměru „šikovnější“ než studenti?
- Je závěs v místě loketní jamky výhodnější (s ohledem na rychlost zasunutí tužky), než závěs na bedrech v místě pasu? Změní se hypotetický základní soubor?





## 6.2 Testování hypotéz na základě dvou výběrů

11. Na základě dat o útratách studentů v restauračních zařízeních (viz tabulka 15) otestujte na hladině významnosti  $\alpha = 0,1$ , zda studenti obou studijních oborů utrácení v průměru jinak.

## 6.3 Testování hypotéz na základě více jak dvou výběrů

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

⊙T⊙  
Těžké a  
pracné

1. Myšlenka analýzy rozptylu je postavena na rozkladu celkového rozptylu  $s_c^2$  na rozptyl způsobený efektem – rozptyl mezi skupinami  $s_x^2$  a rozptyl způsobený náhodným působením – rozptyl uvnitř tříd, tzv. reziduální rozptyl  $s_r^2$ . Odchylka pozorování od celkové střední hodnoty  $(Y_{ij} - \bar{Y}_\bullet)$  se dá rozepsat jako odchylka pozorování od průměru ve skupině  $(Y_{ij} - \bar{Y}_i)$  a odchylka průměru skupiny od celkového průměru  $(\bar{Y}_i - \bar{Y}_\bullet)$ , tj.  $(Y_{ij} - \bar{Y}_\bullet) = (Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \bar{Y}_\bullet)$ . Pro součty čtverců těchto odchylek pak platí:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_\bullet)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^k n_i \cdot (\bar{Y}_i - \bar{Y}_\bullet)^2,$$

tj.  $(n-1) \cdot s_c^2 = (n-k) \cdot s_r^2 + (k-1) \cdot s_x^2$ . Dokažte, že tento vztah platí (třeba pomocí důkazu sporem).<sup>37</sup>

2. Společnost s ručením omezeným prodávající sportovní obuv vlastní v ČR 30 obchodů. K dispozici jsou dostupné údaje o průměrném počtu kusů prodaných za měsíc (viz tabulka 23). Uvedené údaje byly získány v šesti náhodně vybraných prodejnách. Pro každou značku obuvi se jedná o jiný výběr, tj. například prodejna označená číslem jedna je jiná v případě obuvi značky Adidas a jiná v případě obuvi značky Nike apod.

Tabulka 23: Prodávánost obuvi

Značka	Prodejna					
	1	2	3	4	5	6
<b>Adidas</b>	25	29	28	27	30	24
<b>Nike</b>	22	19	23	24	30	17
<b>Reebok</b>	24	27	28	27	30	28

- (a) Lze předpokládat, že je variabilita prodeje obuvi sledovaných značek stejná?
- (b) Zjistěte, zda zákazníci prokazatelně preferují některou značku.

<sup>37</sup>Důkaz sporem (reductio ad absurdum) je typ logického důkazu (ano matematika i statistika jsou logické obory), ve kterém se snažíme prokázat, že předpoklad vede k nesmyslnému výsledku (ke sporu), což znamená, že předpoklad je nepravdivý, a tedy platí jeho negace. Jelikož důkaz sporem je založen na zákonu o vyloučení třetího, lze jej použít pouze v těch logických systémech, ve kterých tento zákon platí – třeba v aritmetice. Chceme-li pak dokázat, že z bezsporných tvrzení  $A_1, A_2, \dots, A_n$  plyne tvrzení  $B$ , pak se obvykle „sporem“ předpokládá, že platí tvrzení  $A_1, A_2, \dots, A_n$  a  $\neg B$ . Pokud zjistíme, že dojdeme k nesmyslu (např.  $x^2 = -8$  pro  $x \in \mathbb{R}$ ), pak musí dohromady platit tvrzení  $A_1, A_2, \dots, A_n$  a  $B$ , čímž byl důkaz proveden.



### 6.3 Testování hypotéz na základě více jak dvou výběrů

- (c) Jak by se situace změnila, kdyby průměrné hodnoty prodaných kusů obuvi za měsíc pocházely pouze ze šesti prodejen tak, že zjištěné údaje za jednotlivé prodejny by byly vždy rozděleny podle značky? Otestujete stejnou hypotézu jako v 2b) i pro tuto situaci.
3. Bylo vybráno 15 studentů a ti byli rozděleni do tří skupin o stejném počtu studentů. Každý student nezávisle na ostatních měl vyřešit týž zadaný problém. U každého se zjišťoval čas v minutách potřebný k vyřešení zadaného úkolu. Všechny skupiny měly stejné podmínky, rozdíl byl jen v navození situace. V jedné skupině byli studenti pochváleni za dosavadní vynikající práci ve škole a bylo jim zdůrazněno, že předložený problém se dá snadno vyřešit. U další skupiny byly konstatovány různé nedostatky mladé generace a bylo prohlášeno, že takoví lidé mohou jen stěží vyřešit předložený problém. U poslední skupiny nebyla situace navozena žádným komentářem. Výsledky pokusu jsou uvedeny v následující tabulce (viz tabulka 24).

Tabulka 24: Časy nutné k vyřešení problému

Skupina	Čas nutný k řešení problému (min)				
A (pochvala)	2	4	5	2	3
B (konstatování nedostatků)	1	4	4	3	2
C (kontrolní skupina)	5	2	4	5	3

Měla počáteční motivace vliv na výsledný čas řešení v jednotlivých skupinách?

4. Bylo sledováno procento niklu v tavební analýze legované oceli. Analýza se prováděla u 4 pecí a u každé pece bylo odebráno 5 vzorků. Data jsou uvedena v následující tabulce (viz tabulka 25).
- (a) Má se zjistit, zda procento niklu je u všech pecí stejné, nebo zda se některé pece od sebe liší.

Tabulka 25: % Ni v tavební analýze legované oceli

1. pec	2. pec	3. pec	4. pec
4,15	4,38	4,23	4,41
4,26	4,40	4,16	4,31
4,10	4,29	4,20	4,42
4,30	4,39	4,24	4,37
4,25	4,45	4,27	4,43

- (b) Pro otestování stejné hypotézy použijte také Kruskalův-Wallisův test<sup>38</sup>.

<sup>38</sup>O normalitě údajů mnoho nevíme, podle Levenova testu však nelze vyloučit homoskedasticitu, jak jste se měli možnost přesvědčit, když jste ji ověřovali pro ANOVAu.

### 6.3 Testování hypotéz na základě více jak dvou výběrů

5. Soutěž o nejlepší jakost výrobků obeslali čtyři výrobci A, B, C, D celkem 26 výrobky. Porota sestavila toto pořadí (uveden pouze původ výrobku od nejlepšího k nejhoršímu) (viz tabulka 26).

Tabulka 26: Výsledky soutěže

<b>Pořadí</b>	1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Výrobce</b>	B	C	C	A	B	D	D	C	A	B	B	D	C
<b>Pořadí</b>	14	15	16	17	18	19	20	21	22	23	24	25	26
<b>Výrobce</b>	D	C	B	C	A	C	D	D	C	C	A	C	A

Na základě těchto údajů posudte, zda původ výrobků má vliv na jeho jakost<sup>39</sup>.

6. Na základě dat o bytech (viz tabulka 12) otestujte na hladině významnosti  $\alpha = 0,05$ , zda cena bytu v Českých Budějovicích závisí na typu vlastnictví.
7. Pět ras psů se účastnilo armádního výcviku v Grabštejně. Po ukončení výcviku byly provedeny zkoušky u vybraných psů z jednotlivých ras. Jednalo se o zkoušky: chůze u nohy, hledání stopy, cviky sedni–lehni–vstaň, aportování, zadržení pachatele a překonání překážky. Jednotliví psi byli v každé disciplíně bodováni. Ukazatelem úrovně výcviku byl součet získaných bodů. Rozhodněte, zda lze předpokládat, že kvalita výcviku závisí nejenom na tom, kdo výcvik vede, ale i na tom, jaká rasa se cvičí, máte-li k dispozici výsledky zkoušek (viz tabulka 27).

Tabulka 27: Výsledky hodnocení psích dovedností

Rasa	Pes					
	1	2	3	4	5	6
<b>Německý ovčák</b>	98	95	89	90	97	95
<b>Dobrman</b>	80	85	76	56	80	70
<b>Rotvailer</b>	95	90	93	89	87	75
<b>Velký knírač</b>	69	81	78	85	89	90
<b>Briard</b>	78	69	65	80	78	78

8. U 15 osob byly zkoušeny 3 vyučovací metody: samostudium, studium prostřednictvím počítače a studium s pomocí učitele. Pokus probíhal tak, že každý člověk nastudoval prostřednictvím každé metody určitou oblast a poté se podrobil testu sestávajícího ze 20 otázek. Počet správných odpovědí je uveden v tabulce (viz tabulka 28).

Lze předpokládat, že je efektivnost všech tří vyučovaných metod stejná (volte  $\alpha = 0,05$ )?

<sup>39</sup>Pořadí, které je známo ze zadání příkladu, jistě nesplňuje předpoklad normality. Máte však k dispozici test, který normalitu dat nevyžaduje, dokonce je přímo na pořadí založen. Použijte ten.



Tabulka 28: Ověřování efektivnosti vyučovacích metod

	osoby														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<b>1. metoda</b>	13	15	12	14	12	13	11	13	11	12	9	11	10	7	10
<b>2. metoda</b>	15	13	14	15	14	13	14	12	12	14	10	14	9	9	6
<b>3. metoda</b>	16	17	15	16	15	17	9	13	15	13	13	14	9	14	8

9. Na 10% hladině významnosti rozhodněte zda ovlivňuje počet vytvořených čárek („čím více tím lépe“ způsob jejich zápisu za předem určený časový limit 10 s. Nezávislý pokus (tj. každá osoba byla zařazena právě do jedné skupiny) byl proveden na přednášce z předmětu Statistika (ZS 2007/2008) na EF JU v Českých Budějovicích. V tabulce 29 indikuje **R** výsledky při použití „šikovnější“ ruky, **L** při použití „méně šikovné“ ruky k psaní čárek a **B** indikuje dosažené počty čárek vytvořených křídou na tabuli.

Tabulka 29: Počty vytvořených čárek za časový limit 10 s v závislosti na provedení

Počet	25	32	42	45	54	54	55	62	64	66	42	43	46	48	53
Provedení	L	L	L	L	L	P	P	P	P	P	T	T	T	T	T

10. Na ekonomické fakultě cvičí statistiku 4 pedagogové: Č, B, K a R. Statistická témata jsou rozdělena do 4 částí: *Pravděpodobnost, deskriptivní metody, induktivní metody a regresní analýzy*. Jednotliví pedagogové cvičí jednotlivá témata. Mají přitom individuální pedagogické metody. Na konci každého tématu se píše test, který má 20 otázek. Správně zodpovězená otázka je za 1 bod. Aby se eliminoval nezodpovědný přístup studentů k přípravě na test, je na testy vázán zápočet: Zápočet ze statistiky dostane student, u kterého počet bodů ani u jednoho z testů neklesne pod 10 bodů. Opravné testy se nepřipouštějí.

Otestujte, zda účinnost různých pedagogických metod je stejná či nikoli (účinnost metod je měřena počtem bodů z testu – viz tabulka 30). Ze všech studentů bylo náhodně vybráno 10 studentů a u nich zjišťován počet bodů.

Lze předpokládat, že efektivnost metod všech vyučujících je stejná (volte  $\alpha = 0,01$ )?

11. Úlohu 4 a 5 naleznete vzorově vyřešenou v dokumentu [nonparametrics.pdf](#), jež připravila RNDr. Marie Kletečková. Zkonfrontujte její a Vaše řešení.

## 6.4 Neparametrické testy

Není hotovo, Friedmanův a Kruskalův-Wallisův test součástí kapitoly 6.3.

## 6.4 *Neparametrické testy*

Tabulka 30: Ověřování efektivnosti vyučovacích metod jednotlivých pedagogů

<b>Student</b> č.	<b>Vyučující</b>			
	<b>C</b>	<b>B</b>	<b>K</b>	<b>R</b>
<b>1</b>	13	15	16	12
<b>2</b>	15	15	15	12
<b>3</b>	12	15	12	12
<b>4</b>	17	15	16	14
<b>5</b>	12	12	12	12
<b>6</b>	19	19	19	12
<b>7</b>	17	13	17	13
<b>8</b>	11	14	10	10
<b>9</b>	18	18	18	18
<b>10</b>	17	17	17	17



## 6.5 Zadání samostatné práce – Testování hypotéz All in One

1. Pro korektní použití metod v tomto úkolu, opět předpokládejte, že hodnoty spojitého numerického znaku jsou *výběrovým* šetřením ze souboru s normálním rozdělením. Na hladině významnosti  $\alpha = 0,05$  rozhodněte o následujících hypotézách:
  - (a) Střední hodnota souboru, z něž pochází spojitý numerický znak, je různá od hodnoty  $\mu_0 = \frac{\bar{x}_{0,50} + \bar{x}_{0,75}}{2}$ .
  - (b) Střední hodnota souboru, z něž pochází spojitý numerický znak, je větší než hodnota  $\mu_0 = \frac{2\bar{x}_{0,50} + \bar{x}_{0,75}}{3}$ .
  - (c) Střední hodnota souboru, z něž pochází spojitý numerický znak, je větší nebo rovna hodnotě  $\mu_0 = \frac{2\bar{x}_{0,50} + \bar{x}_{0,75}}{3}$ .
  - (d) Jaký je rozdíl mezi výsledky předchozích dvou úkolů ((1b) a (1c))?

Všechny hypotézy запиšte jak symbolicky, tak slovně v souvislosti s tím jaký spojitý statistický znak zpracováváte (příjem, náklady, ...). Stejně takovou pozornost věnujete slovní odpovědi, která by neměla být redukována jen např. na to, že se nulová hypotéza na zadané hladině významnosti zamítá nebo nezamítá.

2. Pro korektní použití metod v tomto úkolu, předpokládejte, že hodnoty spojitého numerického znaku jsou *výběrovým* šetřením ze souboru s normálním rozdělením. Statistické jednotky jste v úkolu 1f rozdělili do dvou skupin. Předpokládejte, že tyto dvě skupiny nyní reprezentují výběry z dvou základních souborů (např.: mužů a žen) s normálním rozdělením. Na hladině významnosti  $\alpha = 0,05$  rozhodněte o následujících hypotézách:
  - (a) Střední hodnota souboru, z něž pochází spojitý numerický znak první kategorie je různá od střední hodnoty souboru, z něž pochází spojitý numerický znak kategorie druhé. Jinými slovy řečeno, zda existuje statisticky významný rozdíl mezi výběrovými průměry obou skupin.
  - (b) Střední hodnota souboru, z něž pochází spojitý numerický znak kategorie s vyšším výběrovým průměrem, je větší než střední hodnota souboru, z něž pochází spojitý numerický znak kategorie s nižším výběrovým průměrem.

Všechny hypotézy opět запиšte jak symbolicky, tak slovně v souvislosti s tím jaký spojitý statistický znak zpracováváte a do jakých skupin jste data rozdělili (tj. to, co ovlivňuje rozdíly ve středních hodnotách a rozptylech těchto dvou skupin). Stejně takovou pozornost věnujete slovní odpovědi, která by neměla být redukována jen např. na to, že se alternativní hypotéza na zadané hladině významnosti prokázala nebo neprokázala.

3. Pro korektní použití metod v tomto úkolu, stejně jako dříve předpokládejte, že hodnoty spojitého numerického znaku jsou *výběrovým* šetřením ze souboru s normálním rozdělením. Podle zadání obsahují Vaše data alespoň jeden ordinální statistický znak s nejméně třemi kategoriemi (např.: základní/vyučen/středosškolské/...). Data lze tím pádem rozdělit přímo do minimálně tří skupin – obecně  $k$  skupin.

## 6.5 Zadání samostatné práce IV

Předpokládejte, že těchto  $k$  skupin nyní reprezentuje výběry z  $k$  základních souborů (např.: základní/vyučen/středosškolské/...) s normálním rozdělením. Na hladině významnosti  $\alpha = 0,05$  rozhodněte, zda se střední hodnoty souborů respektive rozdělení souborů, z nichž pochází spojitý numerický znak  $i$ -té kategorie, pro  $i = 1, \dots, k$ , liší v závislosti na rozdělení do těchto minimálně tří skupin. Pokud bude záhodno, pomocí testů mnohonásobného srovnávání najděte skupiny s odlišnou střední hodnotou respektive rozdělením.

Zapište hypotézu naznačenou výše opět jak symbolicky, tak slovně v souvislosti s tím jaký spojitý statistický znak zpracováváte a do jakých skupin jste data rozdělili (tj. to, co ovlivňuje rozdíly ve středních hodnotách a rozptylech těchto skupin). Stejně takovou pozornost věnujete slovní odpovědi, která by neměla být redukována jen např. na to, že se střední hodnoty na zadané hladině významnosti závisí respektive nezávisí na rozdělení do skupin.





## 7 Regresní a korelační analýza

### 7.1 Lineární regresní a korelační analýza

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. Pomocí metody nejmenších čtverců odvoďte odhady koeficientů lineárního regresního modelu mezi proměnnými  $x$  (nezávislá proměnná) a  $y$  (závislá proměnná)<sup>40</sup>.
2. Zjistěte závislost počtu pracovních hodin za měsíc spojených s provozováním anesteziologické služby na velikosti spádové populace nemocnice (viz tabulka 31). Údaje byly získány ve 12 nemocnicích ve Spojených státech.

OT  
Těžké a  
pracné

Tabulka 31: Nemocnice

Nemocnice	Celkový počet pracovních hodin	Spádová populace (v tisících osob)
1	304,37	25,5
2	2616,32	294,3
3	1139,12	83,7
4	285,43	30,7
5	1413,77	129,8
6	1555,68	180,8
7	383,78	43,4
8	2174,27	165,2
9	845,3	74,3
10	1125,28	60,8
11	3462,6	319,2
12	3682,33	376,2

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- (c) Vypočítejte parametry regresního modelu( $\hat{u}$ ).
- (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.

<sup>40</sup>Pro snazší kontrolu výsledků označte absolutní člen  $\beta_0$  a regresní koeficient i lineárního členu  $\beta_1$

## 7.1 Lineární regresní a korelační analýza

- (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
  - (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
  - (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
  - (h) Odhadněte v závislosti na spádové populaci celkový počet pracovních hodin pro 100 000 lidí velkou spádovou oblast.
  - (i) Určete 95% interval spolehlivosti celkového počtu pracovních hodin pro 100 000 lidí velkou spádovou oblast.
  - (j) Odhadněte v závislosti na spádové populaci **průměrný** celkový počet pracovních hodin pro 100 000 lidí velkou spádovou oblast.
  - (k) Určete 95% interval spolehlivosti pro odhad **průměrného** celkového počtu pracovních hodin pro 100 000 lidí velkou spádovou oblast.
  - (l) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
3. U 44 náhodně vybraných studentek a studentů byla zaznamenáno pohlaví, velikost nohy (UK) a tělesná výška (cm), viz tabulka 35.
- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
  - (b) Zobraďte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
  - (c) Vypočítejte parametry regresního modelu(ů).
  - (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
  - (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
  - (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
  - (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
  - (h) Odhadněte v závislosti na velikosti nohy tělesnou výšku postavy s nohou číslo 38 a 46.



Tabulka 32: Velikosti nohou a výšky studentek a studentů – OP 2007/2008 ZS

Pohlaví	Číslo nohy (UK)	Výška (cm)	Pohlaví	Číslo nohy (UK)	Výška (cm)
žena	38	163	žena	38	165
žena	38	168	žena	37	167
žena	38	164	žena	40	165
žena	37	162	muž	45	191
žena	38	168	žena	41	171
žena	41	165	žena	38	168
žena	39	173	žena	40	165
žena	38	172	žena	39	170
žena	38	168	muž	48	197
žena	38	165	žena	39	179,5
žena	39	170	žena	39	175
žena	39	165	muž	43	186
žena	41	170	muž	42	180
žena	40	168	muž	46	192
žena	41	168	žena	40	173
žena	38	170	žena	39	170
žena	40	173	žena	40	174
žena	40	172	žena	40	175
žena	39	162	žena	38	168
žena	38	159	žena	38	174
žena	39	173	žena	38	170
žena	40	172	žena	38	167

- (i) Určete 95% interval spolehlivosti pro tento odhad.
- (j) Odhadněte v závislosti na velikosti nohy **průměrnou** tělesnou výšku postavy s nohou číslo 39.
- (k) Určete 95% interval spolehlivosti pro odhad **průměrné** tělesnou výšku postavy s nohou číslo 38 a 46.
- (l) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
- (m) S ohledem na vybočující hodnoty, které ovlivňují výsledky celé analýzy (hodnoty pro muže), vyzkoušejte rozdělit soubor na muže a ženy a odpovězte znovu na otázky 2b–2l. O výsledcích přemýšlejte!
4. Byly sledovány výdaje (v tisících Kč za 3 měsíce) za potraviny v jednotlivých domácnostech v závislosti jednak na počtu členů domácnosti a na celkovém čistém příjmu domácnosti (v tisících Kč za 3 měsíce). Údaje jsou v následující tabulce (viz tabulka 33).

## 7.1 Lineární regresní a korelační analýza

Tabulka 33: Závislost výdajů v jednotlivých domácnostech

<b>Početnost domácnosti</b>	4	3	5	3	6	2	5
<b>Čistý příjem (Kč/3 měsíce)</b>	60	42	60	75	63	36	66
<b>Výdaje za potraviny (Kč/3 měsíce)</b>	21	5	24	30	27	12	27

- Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
- Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- Vypočítejte parametry regresního modelu(ů).
- Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
- Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
- Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
- Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
- Odhadněte bodově i intervalově (95% interval spolehlivosti) v závislosti na čistém příjmu domácnosti výdaje za potraviny pro příjmy Kč 45 000,- a 65 000,-.
- Odhadněte bodově i intervalově (99% interval spolehlivosti) v závislosti na čistém příjmu domácnosti **průměrné** výdaje za potraviny pro příjmy Kč 45 000,- a 65 000,-.
- Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.



## 7.2 Nelineární regresní a korelační analýza

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. U patnácti náhodně vybraných jatečných prasat byla při kontrole jatečné užitkovosti zjišťována délka jatečného trupu a výška špeku. Výsledky jsou uvedeny v tabulce 34).

Tabulka 34: Výsledky kontroly jateční užitkovosti prasat

Zvíře	Délka jatečného trupu (cm)	Výška špeku (cm)
1	92,7	3,80
2	93,2	3,90
3	94,0	3,70
4	95,0	3,65
5	93,5	3,80
6	96,5	3,65
7	95,5	3,70
8	92,7	3,70
9	93,5	3,60
10	95,5	3,65
11	95,2	3,60
12	94,7	3,70
13	96,5	3,55
14	97,2	3,50
15	98,0	3,48

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- (c) Vypočítejte parametry regresního modelu( $\hat{u}$ ).
- (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
- (e) Prostřednictvím testu zjistěte zda lze regresní model( $y$ ) zjednodušit. Pokud ano, zjednodušte.
- (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.

## 7.2 Nelineární regresní a korelační analýza

- (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
- (h) Určete jakou výšku špeku lze v průměru očekávat u prasete, jehož délka trupu je 96 cm.
- (i) Určete 95% interval spolehlivosti pro tyto odhady.
- (j) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
2. U 44 náhodně vybraných studentek a studentů byla zaznamenáno pohlaví, velikost nohy (UK) a tělesná výška (cm), viz tabulka 35.

Tabulka 35: Velikosti nohou a výšky studentek a studentů – OP 2007/2008 ZS

Pohlaví	Číslo nohy (UK)	Výška (cm)	Pohlaví	Číslo nohy (UK)	Výška (cm)
žena	38	163	žena	38	165
žena	38	168	žena	37	167
žena	38	164	žena	40	165
žena	37	162	muž	45	191
žena	38	168	žena	41	171
žena	41	165	žena	38	168
žena	39	173	žena	40	165
žena	38	172	žena	39	170
žena	38	168	muž	48	197
žena	38	165	žena	39	179,5
žena	39	170	žena	39	175
žena	39	165	muž	43	186
žena	41	170	muž	42	180
žena	40	168	muž	46	192
žena	41	168	žena	40	173
žena	38	170	žena	39	170
žena	40	173	žena	40	174
žena	40	172	žena	40	175
žena	39	162	žena	38	168
žena	38	159	žena	38	174
žena	39	173	žena	38	170
žena	40	172	žena	38	167

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.



- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
  - (c) Vypočítejte parametry regresního modelu(ů).
  - (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
  - (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
  - (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
  - (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
  - (h) Odhadněte v závislosti na velikosti nohy tělesnou výšku postavy s nohou číslo 38 a 46.
  - (i) Určete 95% interval spolehlivosti pro tento odhady.
  - (j) Odhadněte v závislosti na velikosti nohy **průměrnou** tělesnou výšku postavy s nohou číslo 39.
  - (k) Určete 95% interval spolehlivosti pro odhad **průměrné** tělesnou výšku postavy s nohou číslo 38 a 46.
  - (l) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
  - (m) S ohledem na vybočující hodnoty, které ovlivňují výsledky celé analýzy (hodnoty pro muže), vyzkoušejte rozdělit soubor na muže a ženy a odpovězte znovu na otázky 2b–2l. O výsledcích přemýšlejte!
3. Při sledování závislosti těla tloušťě na jeho stáří bylo zváženo 14 tloušťů (viz tabulka 36).
- (a) Vyjádřete funkční závislost mezi sledovanými znaky pomocí paraboly, hyperboly a exponenciální funkce.
  - (b) Pro každý regresní model spočtete ukazatel těsnosti korelační závislosti a na základě věcného rozboru problému spolu s přihlédnutím k hodnotě příslušného ukazatele vyberte nejvhodnější model. Jaký z nich to je?
  - (c) [Opakování:] Otestujte zda hmotnost tloušťě závisí na jeho stáří (test na shodu středních hodnot). Hladinu významnosti volte  $\alpha = 0,01$ .
4. Při volejbalovém memoriálu R. Myslíka byly získány antropometrické údaje (viz soubor [memorial.sta](#)), které již znáte např z úkolu 3 v oddílu 4.1.

Tabulka 36: Tloušťky tlouště

Tloušť číslo	Věk tlouště (roky)	Hmotnost tlouště (g)
1	1	10
2	1	12
3	1	15
4	2	28
5	2	29
6	2	29
7	3	60
8	3	55
9	4	100
10	4	115
11	4	98
12	5	160
13	5	168
14	5	170

Pro další úvahy vycházejte z následujících možných závislostí (nezávislá  $\rightarrow$  závislá proměnná):

- |                                     |   |
|-------------------------------------|---|
| 1. Výška $\rightarrow$ Váha         | 3. Výška $\rightarrow$ Dosah 1 stoj     |
| 5. Výška $\rightarrow$ Dosah 2 stoj | 2. Výška $\rightarrow$ Dosah rozb.      |
| 4. Výška $\rightarrow$ Abs. rozb.   | 6. Dosah rozb. $\rightarrow$ Abs. rozb. |

- (a) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- (b) Vypočítejte parametry regresního modelu(ů).
- (c) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
- (d) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
- (e) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
- (f) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
- (g) Určete jakou(ý) ... (viz vysvětlovaná proměnná) lze očekávat u volejbalisty vysokého 180 cm, resp. jaký absolutní smečářský výskok bude mít, má-li smečářský dosah 321 cm.





- (h) Určete 99% interval spolehlivosti pro tyto odhady.
- (i) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.

### 7.3 Vícenásobná regresní a korelační analýza

1. Na základě příkladu o rodinných výdajích z oddílu 7.1, úkol 4 – tabulka 33 vyřešte následující úkoly.
  - (a) Vysvětlete výdaje za potraviny v závislosti jak na počtu osob v domácnosti, tak příjmu na domácnost. Volte lineární aditivní regresní model.
  - (b) Prostřednictvím testu zjistěte zda je tento model celkově statisticky významný.
  - (c) Prostřednictvím testu zjistěte zda lze regresní model zjednodušit, tj. jsou-li oba regresory a absolutní člen v modelu statisticky významné. Pokud nejsou, zjednodušte.
  - (d) Interpretujte hodnotu(y) regresních koeficientu(ů) tohoto modelu.
  - (e) Odhadněte bodově i intervalově (95% interval spolehlivosti) v závislosti na čistém příjmu domácnosti a počtu osob v domácnosti výdaje za potraviny pro příjem Kč 50 000,- a 4 osoby.
2. Rozhodněte, zda spotřeba automobilu závisí na jeho výkonu motoru, když navíc přihlédnete k hmotnosti a maximální rychlosti (viz soubor [auta.sta](#)), které již znáte např. z úkolu 5 v oddílu 4.2.



## 7.4 Zadání samostatné práce – Regresní a korelační analýza

1. Z internetu (či jinak) získejte denní kurzy dvou měn vůči České koruně. Mělo by se jednat o minimálně 20 hodnot. Měny vybírejte tak, aby jejich mezinárodní značky obsahovaly ve svém kódu počáteční písmena Vašeho jména a příjmení (např.:

**R**oman → **EUR** (euro) 1. měna  
**B**iskup → **GBP** (britská libra) 2. měna)

- Přestože rozdělení měn na závislou a nezávislou měnu (proměnou) nemá logické opodstatnění, dodržujte jej. Zjistěte míru korelace mezi jednotlivými kurzy.
- Na základě korelačního pole a věcné analýzy problému zvolte vhodný typ závislosti (lineární, polynomiální, ...) kurzu druhé měny na první. Vždy vyzkoušejte minimálně závislost modelovat lineárně a pomocí funkce  $y = a \cdot e^{b \cdot x}$ .
- Metodou nejmenších čtverců odhadněte regresní koeficienty (zapište předpisy regresních funkcí) a uveďte hodnoty korelačních koeficientů/indexů, pro Vámi vytvářené modely.
- Jsou všechny vytvořené modely vhodné/významné? Dále pracujte jen s těmi vhodnými.
- Pokud je nutné, modely zjednodušte.
- Zvolte libovolně hodnotu ( $x_0$ ) kurzu první měny v rozsahu od nejnižšího kurzu k nejvyššímu (ne nutně hodnotu, jež byla mezi daty).
  - Bodově a intervalově s 99% spolehlivostí odhadněte na základě Vámi vytvořených (vhodných) modelů směnný kurz pro měnu druhou ( $\hat{y}(x_0)$ ).
  - S tou samou spolehlivostí bodově i intervalově odhadněte střední hodnotu druhé měny na úrovni  $x_0$ .

#### 7.4 Zadání samostatné práce V



## 8 Poděkování, reference a realizace

### 8.1 Poděkování za inspiraci

Na tomto místě by čtenář nejspíše očekával soupis literatury, ze které bylo čerpáno pro vytvoření zadání příkladů v tomto textu. Těžko by mně kdokoliv z Vás věřil, že jsem všechny příklady vymyslel sám. To samozřejmě není pravdou stejně jako to, že jsem vše jen opsal. V současné době bylo vymyšleno prakticky všechno, co se dá v úvodu do předmětů předkládat. Originální jsou většinou pouze zpracování. Vzhledem k tomu, že obvykle nepřístupuji k textu nekriticky a zadání i otázky modifikuji, mohl bych být teoreticky z obliga a předkládat Vám text za svůj.

Ve skutečnosti jsem člověk vděčný a proto bych poněkud netradičním způsobem poděkoval autorům zdrojů, ze kterých jsem čerpal. Následující lidé figurují jako autoři publikací, výukových materiálů kurzů (kterých jsem měl možnost se účastnit), webových stránek. Do svého výčtu jsem zahrnul i ty, které osobně znám a kteří nějakým způsobem byli motivací pro tvorbu „originálních“ příkladů.

Mnoho osob z mého seznamu (abecedně seřazeno) je notoricky známých, přesto (aby nedošlo k nedorozumění) jsem prostřednictvím poznámek pod čarou všechny trochu specifikoval. Každého jsem uvedl s tituly, které si vydobyl, a institucí, se kterou je jeho jméno alespoň nějaký čas svázáno. Oč by bylo jednodušší klasické uvedení literatury, že! Nicméně vězte, že valná většina materiálů, které jsem využil, neměla ani ISBN či jiný oficiální identifikátor a přesto (alespoň pro mne) byla více, než hodnotným zdrojem. Mnoho příkladů jsem tak znal dříve, než jsem otevřel publikace, ve kterých se k nim někdo hlásil, tak jako jako já nyní :-). Tedy inspirací mi byli . . .

### 8.2 Reference trochu jinak

Jiří ANDĚL<sup>41</sup>, Vladimír BRABENEC<sup>42</sup>, Anna ČERMÁKOVÁ<sup>43</sup>, Ludvík FRIEBEL<sup>44</sup>, Jana FRIEBELOVÁ<sup>45</sup>, Michal FRIESEL<sup>46</sup>, Daniel HLUBINKA<sup>47</sup>, Michal HOUDA<sup>48</sup>, Marie HUŠKOVÁ<sup>49</sup>, Bohumil KÁBA<sup>50</sup>, Marie KLETEČKOVÁ<sup>51</sup>, Jana KLICNAROVÁ<sup>52</sup>, Renata KLUFOVÁ<sup>53</sup>, Pa-

---

<sup>41</sup>prof. RNDr. Jiří ANDĚL, DrSc. – MFF CUNI v Praze

<sup>42</sup>prof. Ing. Vladimír BRABENEC, CSc. – PEF ČZU v Praze

<sup>43</sup>prof. RNDr. Anna ČERMÁKOVÁ, CSc. – EF JU v Českých Budějovicích

<sup>44</sup>Ing. Ludvík FRIEBEL, Ph.D. – EF JU v Českých Budějovicích

<sup>45</sup>Ing. Jana FRIEBELOVÁ, Ph.D. – EF JU v Českých Budějovicích

<sup>46</sup>Mgr. Michal FRIESEL, Ph.D. – FAV ZČU v Plzni

<sup>47</sup>doc. RNDr. Daniel HLUBINKA, Ph.D. – MFF CUNI v Praze

<sup>48</sup>Mgr. Michal HOUDA, Ph.D. – EF JU v Českých Budějovicích

<sup>49</sup>prof. RNDr. Marie HUŠKOVÁ, DrSc. – MFF CUNI v Praze

<sup>50</sup>doc. RNDr. Bohumil KÁBA, CSc. – PEF ČZU v Praze

<sup>51</sup>RNDr. Marie KLETEČKOVÁ – EF JU v Českých Budějovicích

<sup>52</sup>RNDr. Jana KLICNAROVÁ, Ph.D. – EF JU v Českých Budějovicích

<sup>53</sup>RNDr. Renata KLUFOVÁ, Ph.D. – EF JU v Českých Budějovicích

### 8.3 Realizace

vel LEISCHNER<sup>54</sup>, Otakar MACHÁČEK<sup>55</sup>, Martin MARŠÍK<sup>56</sup>, Kamil NAVRÁTIL<sup>57</sup>, Václav NÝDL<sup>58</sup>, Petr OTIPKA<sup>59</sup>, Vladimíra PETRÁŠKOVÁ<sup>60</sup>, Jan POPELKA<sup>61</sup>, Michael ROST<sup>62</sup>, Jitka RUTKAYOVÁ<sup>63</sup>, Libuše SVATOŠOVÁ<sup>64</sup>, Vladislav ŠMAJSTRLA<sup>65</sup>, Josef ŠTĚPÁN<sup>66</sup>, Marek ŠULISTA<sup>67</sup>, Pavel TLUSTÝ<sup>68</sup>, Zdeněk TOMŠÍČEK<sup>69</sup>, Karel ZVÁRA<sup>70</sup> a Jana ZVÁROVÁ<sup>71</sup>.

Upřímně děkuji všem z předchozího výčtu a doufám, že jsem svým didaktickým záměrem (zde se projevil vliv PF JU v Českých Budějovicích), se kterým jsem „přeoral“ (a zde ZF JU v Českých Budějovicích ;-)) jejich zadání, nezničil ten jejich, který vytvořením výukových příkladů sledovali. Zároveň stále ještě nejsem příliš ovlivněn EF JU v Českých Budějovicích, či komerční sférou (zmiňuji svého laskavého zaměstnavatele – Essox a tento text se Vám dostává do rukou bezplatně.

### 8.3 Realizace (moje, jenom moje :o))

Není-li zcela má obsahová část tohoto textu, pak forma podání je mým – snad přívětivým – přínosem. Pro realizaci jsem využil sazečího programu L<sup>A</sup>T<sub>E</sub>X v distribuci MiK<sub>T</sub>TeXu a několika málo stažených i vlastnoručně vytvořených features [čti: 'fi:ʃərs], které s T<sub>E</sub>Xovským vzhledem pionýra pročistil největší T<sub>E</sub>Xista v Jižních Čechách, kterého znám – Houdík.

Děkuji.  
Autor

---

<sup>54</sup>RNDr. Pavel LEISCHNER, Ph.D. – PF JU v Českých Budějovicích

<sup>55</sup>doc. Ing. Otakar MACHÁČEK, CSc. – PEF ČZU v Praze

<sup>56</sup>Ing. Martin MARŠÍK, Ph.D. – ZF JU v Českých Budějovicích

<sup>57</sup>MUDr. Kamil NAVRÁTIL, Ph.D. – LF UP v Olomouci

<sup>58</sup>doc. RNDr. Václav NÝDL, CSc. – EF JU v Českých Budějovicích

<sup>59</sup>Mgr. Petr OTIPKA – VŠB – TU v Ostravě

<sup>60</sup>RNDr. Vladimíra PETRÁŠKOVÁ, Ph.D. – PF JU v Českých Budějovicích

<sup>61</sup>Ing. Jan POPELKA, Ph.D. – UJEP v Ústí nad Labem

<sup>62</sup>Ing. Michael ROST, Ph.D. – EF JU v Českých Budějovicích

<sup>63</sup>Ing. Jitka RUTKAYOVÁ – ZF JU v Českých Budějovicích

<sup>64</sup>prof. Ing. Libuše SVATOŠOVÁ, CSc. – PEF ČZU v Praze

<sup>65</sup>doc. PaedDr. Vladislav ŠMAJSTRLA – VŠB – TU v Ostravě

<sup>66</sup>prof. RNDr. Josef ŠTĚPÁN, DrSc. – MFF CUNI v Praze

<sup>67</sup>PhDr. Marek ŠULISTA, Ph.D. – EF JU v Českých Budějovicích

<sup>68</sup>prof. RNDr. Pavel TLUSTÝ, CSc. – PF JU v Českých Budějovicích

<sup>69</sup>Mgr. Zdeněk TOMŠÍČEK – PF JU v Českých Budějovicích

<sup>70</sup>doc. RNDr. Karel ZVÁRA, CSc. – MFF CUNI v Praze

<sup>71</sup>prof. RNDr. Jana ZVÁROVÁ, DrSc. – Evropské centrum pro lékařskou informatiku, statistiku a epidemiologii Univerzity Karlovy a Akademie věd České republiky



## Seznam tabulek

1	Bodový zisk z písemky ze Statistiky . . . . .	27
2	Počet uzavřených sňatků . . . . .	28
3	BMI index . . . . .	28
4	Rozloha krajů . . . . .	30
5	Osobní informace statečných studentů . . . . .	31
6	Příklad datového souboru . . . . .	33
7	Životnost matric . . . . .	35
8	Váha kaprů . . . . .	36
9	Pomerančové šťávy . . . . .	37
10	Tloušťky vláken . . . . .	37
11	Dávkovací automat . . . . .	41
12	Ceny bytů v Českých Budějovicích . . . . .	42
13	Pevnost vláken bavlněné příze . . . . .	42
14	Počty studentů hodnotící výuku na ZCU v ZS 2002/2003 . . . . .	43
15	Útraty studentek a studentů v restauračních zařízení . . . . .	43
16	Pevnost betonových kvádrů . . . . .	44
17	Průměrná tučnost mléka . . . . .	44
18	Diference kompresních tlaků . . . . .	45
19	Rychlosti naměřené v Litvínově . . . . .	45
20	Zásobování živinami . . . . .	46
21	Výsledky testu zručnosti . . . . .	46
22	Výsledky dovednostní soutěže . . . . .	46
23	Prodávánost obuvi . . . . .	48
24	Časy nutné k vyřešení problému . . . . .	49
25	% Ni v tavební analýze legované oceli . . . . .	49
26	Výsledky soutěže . . . . .	50
27	Testování psích dovedností . . . . .	50
28	Ověřování efektivnosti vyučovacích metod . . . . .	51
29	Počty vytvořených čárek . . . . .	51
30	Ověřování efektivnosti vyučovacích metod jednotlivých pedagogů . . . . .	52
31	Nemocnice . . . . .	55
32	Velikosti nohou a výšky studentek a studentů . . . . .	57
33	Závislost výdajů v jednotlivých domácnostech . . . . .	58
34	Výsledky kontroly jateční užitkovosti prasat . . . . .	59
35	Velikosti nohou a výšky studentek a studentů . . . . .	60
36	Tloušťky tlouště . . . . .	62
38	Modifikace oboustranného $p$ -value . . . . .	84

## Seznam obrázků

1	Rozhodovací strom pro výběr oblečení (různé kombinace) . . . . .	72
2	Logický strom výsledků tenisového zápasu . . . . .	74
3	Vzorový zkouškový test ze Statistiky – rok 2005 . . . . .	93
4	Vzorový zkouškový test ze Statistiky – rok 2009 . . . . .	95





## A Kombinatorika letem světem – *Odvození vzorců*

Cílem této přílohy je ozřejmění základních kombinatorických vztahů, jež mnozí znají ze střední školy, ale většinou je nedokáží používat již proto, že neví, kde se dané vztahy vzaly a „Proč se tam něco může opakovat a něco ne a někde je pořadí a někde je to bez pořadí, ...“. V textu většinu vztahů odvodíme a to na jednoduchých modelech.

### A.1 Základní kombinatorické principy

#### A.1.1 Princip bijekce

Princip bijekce je založen na vzájemně jednoznačném přiřazení prvků dvou množin. Jedna množina pro nás může být nepřehledná a vztahy v ní dokážeme těžko postihnout, zatímco druhá množina je pro nás přehledná a jsme na ní schopni daný problém vyřešit (ne nutně vždy kombinatorický problém). Známe-li tedy řešení na množině pro nás přehledné, známe i řešení na množině druhé. Kombinatoricky: *Má-li situace na jedné množině právě  $m$  řešení, pak stejná situace na vzájemně jednoznačně přiřazené množině (byť nepřehledné) má také  $m$  řešení.*

Vzájemně jednoznačné přiřazení znamená, že každému prvku z jedné množiny (označme ji  $A$ ) odpovídá (je přiřazen) právě jeden prvek z množiny druhé (tu označme  $B$ ). Pokud má existovat vzájemně jednoznačné přiřazení mezi množinami  $A$  a  $B$ , pak obě množiny musí být stejně početné, tj.  $\#A = \#B$ . Symbolem  $\#A_i$  rozumíme kardinalitu (mohutnost) množiny  $A_i$  pro  $i = 1, \dots, k$ . Konkrétně pro nás se můžeme omezit na to, že budeme symbol  $\#A$  chápat jako počet prvků množiny  $A$ .

S využitím tohoto principu můžeme například kombinatorické problémy převést na situace, kdy rozlišitelné objekty přiřazujeme přihrádkám (ať již rozlišitelným nebo nerozlišitelným), ale o tom dále. S principem bijekce se setkáme také například při odvození vzorce (17) pro výpočet kombinací s opakováním, viz podkapitola A.2.6.

#### A.1.2 Kombinatorické pravidlo o násobení

Předpokládejme, že máme vybrat  $k$ -tici prvků, přičemž první prvek této  $k$ -tice vybíráme z konečné neprázdné množiny, druhý z konečné neprázdné množiny, atd., až poslední,  $k$ -tý, vybíráme z konečné neprázdné množiny. V případě, že výběr každého z prvků je nezávislý na výběru ostatních prvků, je celkem

$$\#A_1 \cdot \#A_2 \cdot \dots \cdot \#A_k \quad (1)$$

různých možností ( $k$ -tic), jak vybrat tyto prvky.

Výše uvedený vztah je, řekl bych, zřejmý, ale přece jen provedeme jeho přiblížení příkladem. Představme si, že máme jen dva druhy ponožek, troje kalhoty a dvě košile. Chceme-li zjistit kolika různými způsoby se můžeme obléci bez ohledu na módní trendy, pak uvažujeme asi následovně. Ke každému druhu ponožek si můžeme vzít jedny z trojice kalhot.

## A.1 Základní kombinatorické principy

K jedné kombinaci kalhot s ponožkami si můžeme vzít jednu ze dvojice košil. Kolik je to možností? Právě 12. Konstrukci kombinací ukazuje následující rozhodovací strom na obrázku (viz obrázek 1). Výše uvedený vztah je jen zobecněním tohoto příkladu.

1. druh ponožek	1. kalhoty	1. košile	1.
		2. košile	2.
	2. kalhoty	1. košile	3.
		2. košile	4.
	3. kalhoty	1. košile	5.
		2. košile	6.
2. druh ponožek	1. kalhoty	1. košile	7.
		2. košile	8.
	2. kalhoty	1. košile	9.
		2. košile	10.
	3. kalhoty	1. košile	11.
		2. košile	12.
2×	3×	2×	= 12

Obrázek 1: Rozhodovací strom pro výběr oblečení (různé kombinace)

### A.1.3 Kombinatorické pravidlo o součtu

Předpokládejme, že máme  $k$ -tici disjunktních množin<sup>72</sup>, potom sjednocení těchto množin,  $\bigcup_{i=1}^k A_i$  má právě  $\#A_1 + \#A_2 + \dots + \#A_k$  prvků, tj.:

$$\#\bigcup_{i=1}^k A_i = \#A_1 + \#A_2 + \dots + \#A_k. \quad (2)$$

Uvědomte si, že podmínka disjunktnosti je podstatná. Uvažujme pro jednoduchost dvě nedisjunktní množiny  $A_{\heartsuit} = \{m, o, s, t\}$  a  $A_{\spadesuit} = \{k, o, s, t\}$ , jenž mají shodně 4 prvky. Sjednocení množin  $A_{\heartsuit} \cup A_{\spadesuit}$  má pouze 5 prvků ( $\{m, o, s, t, k\}$ ). Kdybychom zapomněli na podmínku disjunktnosti a jen tupě dosadili do vzorce výše (2), dostali bychom nesprávný výsledek 8 prvků ( $4 + 4$ ). Jak se vypořádat s případem, kdy o disjunktnosti nelze rozhodnout, si ukážeme v následující oddíle (oddíl A.1.4).

### A.1.4 Princip exkluze a inkluze

Pro začátek uvažujme opět množiny  $A_{\spadesuit}$  a  $A_{\heartsuit}$  definované výše. Uvědomme si, že pokud sečteme počet prvků obou množin, chováme se tak, jako kdybychom do sjednocení zařadili všechny prvky, což je ve sporu s definicí množiny, neboť jsme některé prvky zařadili do sjednocení množin  $A_{\heartsuit} \cup A_{\spadesuit}$  dvakrát. Je zřejmé, že právě společné prvky jsou „problémové“.

<sup>72</sup>disjunktní množiny jsou takové množiny, jenž mají po dvou prázdný průnik, tj. množiny  $A_i$ , pro  $i = 1, \dots, k$  jsou disjunktní  $\Leftrightarrow A_i \cap A_j = \emptyset$  pro  $i \neq j$  a  $i, j = 1, \dots, k$



Společné prvky, jsou, jak jistě víte, právě všechny prvky průniku  $A_{\text{✕}} \cap A_{\text{✗}}$ , tj.  $A_{\text{✕}} \cap A_{\text{✗}} = \{o, s, t\}$ . Protože prvky průniku jsme prostým součtem počtů prvků množin  $A_{\text{✗}}$  a  $A_{\text{✕}}$  započítali dvakrát, jednou s množinou  $A_{\text{✕}}$  a podruhé s množinou  $A_{\text{✗}}$ , stačí od prostého součtu prvků množin odečíst počet prvků jejich průniku:

$$\#(A_{\text{✕}} \cup A_{\text{✗}}) = \#A_{\text{✕}} + \#A_{\text{✗}} - \#(A_{\text{✕}} \cap A_{\text{✗}}). \quad (3)$$

Nyní se podívejme jak to vypadá v případě 3 množin  $A_1, A_2$  a  $A_3$ . Zkusme aplikovat postup popsany výše a od celkového součtu prvků množin  $A_1, A_2$  a  $A_3$  odečíst počty prvků jednotlivých průníků  $A_1 \cap A_2, A_1 \cap A_3$  a  $A_2 \cap A_3$ . Tím jsme se sice zbavili navíc napočtených prvků průníků. Co se však stalo s průnikem všech třech množin  $A_1 \cap A_2 \cap A_3$ ? Prostým součtem byl na začátku počet prvků průniku  $A_1 \cap A_2 \cap A_3$  zahrnut třikrát. Následně byl zase třikrát odečten. Proto musíme počet prvků průníků  $A_1 \cap A_2 \cap A_3$  přičíst:

$$\begin{aligned} \#(A_1 \cup A_2 \cup A_3) &= \#A_1 + \#A_2 + \#A_3 - \\ &\quad - \#(A_1 \cap A_2) - \#(A_2 \cap A_3) - \#(A_1 \cap A_3) + \\ &\quad + \#(A_1 \cap A_2 \cap A_3). \end{aligned} \quad (4)$$

Pro snažší pochopení zápisu počtu prvků sjednocení pro obecně  $k$  nedisjunktních množin přepíšeme výraz (4) s využitím sumačního symbolu a symbolu sjednocení přes indexovou množinu:

$$\#\bigcup_{i=1}^3 A_i = \sum_{i=1}^3 \#A_i - \sum_{i \neq j} \#(A_i \cap A_j) + \#\bigcap_{i=1}^3 A_i. \quad (5)$$

Pro  $k$  množin  $A_1, \dots, A_k$  ne nutně disjunktních můžeme počet prvků tohoto sjednocení  $\bigcup_{i=1}^k A_i$  vypočíst podle následujícího vzorce:

$$\begin{aligned} \#\bigcup_{i=1}^k A_i &= \sum_{i=1}^k \#A_i - \sum_{i \neq j} \#(A_i \cap A_j) + \\ &\quad + \sum_{i \neq j \neq l} \#(A_i \cap A_j \cap A_l) + \dots \\ &\quad + (-1)^{k-1} \cdot \#\bigcap_{i=1}^k A_i. \end{aligned} \quad (6)$$

Předpis (6) říká: „Sečti počty prvků všech množin a odečti od nich počty prvků všech různých průníků dvou množin. K výsledku přičti počty prvků všech různých průníků tří množin a tak dále, až nakonec k výsledku přičti respektive odečti počet prvků průniku všech  $k$  množin“. Ve vzorci je pomocí výrazu  $(-1)^{k-1}$  ošetřeno to, aby se znaménka pravidelně střídala a poslední člen vzorce (6) do této řady zapadl.

Následující vzorec (7) shrnuje v poměrně krátkém zápise rozepsaný vzorec (6). Podle mého soudu je však méně přehledný. Pokuste se však nahlédnout, že jeho zápis je korektní:

$$\#\bigcup_{i=1}^k A_i = \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{(\#I-1)} \cdot \#\bigcap_{i=1}^k A_i. \quad (7)$$

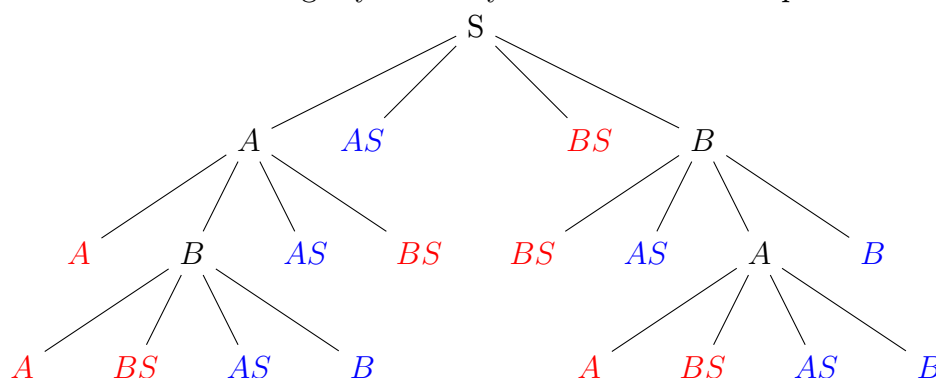
Dále si uvědomte, že vzorec (2) a vzorce (3) až (7) jsou ve shodě. Je-li splněna podmínky disjunktnosti množin, jsou všechny průniky prázdné a tudíž je jejich kardinalita nulová.

### A.1.5 Pravidlo logického stromu

Při odvozování kombinatorického pravidla (podkapitola A.1.2 o násobení jsme se k výsledku dostali prostřednictvím zjištěním počtu všech možných výsledku. Na příkladě bylo ukázáno jak obecně dojít k výsledku shrnutého ve vzorci (1). Pro přehlednost byl výčet uspořádán do logického stromu, jehož jednotlivé úrovně reprezentovali rozhodnutí dalšího oblečení (ponožek, kalhot a košil). Tento strom byl svým způsobem „ideální“. Na každé úrovni byl „rovnoměrně“ rozvětvený.

Uvažme však jiný případ, který již nebude tak „symetrický“. Jak může proběhnout tenisové utkání, které se hraje na dva vyhrané sety. Každý set tedy může vyhrát hráč  $A$  nebo hráč  $B$ . A aby to nebylo tak jednoduché uvažme, že hráč  $A$  či hráč  $B$  může odstoupit, tj. utkání skrečovat. První set tedy může vyhrát hráč  $A$  nebo hráč  $B$  a utkání pokračuje, nebo vzdá hráč  $A$  nebo  $B$  a utkání končí. Po druhém setu utkání skončí jen v případě, že vyhrál stejný hráč jako v setu prvním, nebo opět někdo skrečuje. To se to komplikuje, že? Pojďme možné výsledky uspořádat do logického stromu, kde  $A$  bude znamenat vítězství hráče  $A$ ,  $AS$  skreč hráče  $A$  u hráče  $B$  obdobně.

Obrázek 2: Logický strom výsledků tenisového zápasu



Spojnice mezi jednotlivými stavy nazýváme větve a stavy, které jsou konečné, pak listy. Listy na obrázku 2, které jsou červené, zobrazují výsledek, který skončil vítězstvím hráče  $A$ , modré pak symbolizují vítězství hráče  $B$ . Počet listů (16) je počtem všech možných výsledků tenisového zápasu. V polovině vyhrává hráč  $A$  v polovině hráč  $B$ .

Kombinatoricky: *Počet různých řešení dané situace zaznamenávané prostřednictvím logického stromu je roven počtu listů tohoto stromu.* Jistě Vás napadnou i jiné způsoby využití logického stromu (výsledky her, vícekolové rozhodování, ...).



## A.2 Permutace, variace a ty další

Pro konstrukci variací, permutací a kombinací, tak jak je znáte, využijeme následující modelovou úlohu:

Mějme  $n$ , u některých příkladů  $l$ , rozlišitelných předmětů, které chceme po jednom umístit do  $k$  přihrádek.

Otázkou je, kolika různými způsoby to lze učinit. S ohledem na to, zda nás zajímá pořadí přihrádek, tj. přihrádky jsou rozlišitelné, a zda se rozlišitelné předměty vyskytují současně ve více přihrádkách, tj. máme-li více předmětů stejného druhu, vytvoříme několik modelů, kterými se budeme snažit ozřejmit pro mnohé tajemné „kombinatorické vzorečky“.

### A.2.1 Rozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují

Mějme  $n$  rozlišitelných předmětů, jež chceme umístit do  $k$  očíslovaných přihrádek a to tak, že do každé přihrádky jen jeden předmět. Dále předpokládejme, že každý předmět máme jen jednou. Pak je zřejmé, že do první přihrádky můžeme umístit jeden z  $n$  předmětů, které máme. Do druhé přihrádky již můžeme umístit jen  $n - 1$  předmětů, protože jsme již jeden vyčerpali. Do poslední  $k$ -té přihrádky pak můžeme vybírat jen z  $n - k + 1$  předmětů, jež nám zbyly. Z tohoto důvodu je rozumné požadovat, abychom měli více, nebo alespoň stejně předmětů jako přihrádek, tedy  $n \geq k$ <sup>73</sup>.

1	2	3	...	$k - 1$	$k$
$n$	$n - 1$	$n - 2$	...	$n - k + 2$	$n - k + 1$

S odvoláním na pravidlo o násobení A.1.2 můžeme vyjádřit počet způsobů, jimiž lze danou úlohu vyřešit a stanovit výsledek:

$$V_k(n) = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k). \quad (8)$$

To odpovídá následujícímu zápisu, ve kterém využíváme faktoriály čísel:

$$V_k(n) = \frac{n!}{(n - k)!}. \quad (9)$$

Uvědomte si, že jsme právě odvodili vzorec pro počet variací  $k$ -té třídy z  $n$  prvků bez opakování.

### A.2.2 Rozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují \*

Speciálním typem předcházející úlohy je případ, kdy je počet rozlišitelných předmětů stejný jako počet očíslovaných přihrádek. Tento společný počet označme, jak bývá zvykem,  $n$ . Pak je zřejmé, že do první přihrádky můžeme umístit jeden z  $n$  předmětů, které máme. Do druhé přihrádky můžeme umístit jen  $n - 1$  předmětů, protože jsme již jeden vyčerpali. Do poslední  $n$ -té přihrádky již nemůžeme vybírat a dáme tam ten, jež nám zbyl.

<sup>73</sup>V případě, kdyby bylo přihrádek více než objektů, tj.  $n \leq k$ , se úloha převádí na přiřazení přihrádek předmětům.

## A.2 Permutace, variace a ty další

1	2	3	...	$n - 1$	$n$
$n$	$n - 1$	$n - 2$	...	2	1

S odvoláním na pravidlo o násobení A.1.2 můžeme vyjádřit počet způsobů, jimiž lze danou úlohu vyřešit a stanovit výsledek:

$$P(n) = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 \quad (= V_n(n)). \quad (10)$$

To odpovídá následujícímu zápisu, ve kterém využíváme faktoriál:

$$P(n) = n! \quad (= V_n(n)). \quad (11)$$

Uvědomte si, že jsme právě odvodili vzorec pro počet permutací  $n$  prvků (bez opakování).

### A.2.3 Rozlišitelné přihrádky a rozlišitelné předměty, jež se mohou opakovat

Mějme  $n$  rozlišitelných předmětů, jež chceme umístit do  $k$  očíslovaných přihrádek. Dále předpokládejme, že každý předmět máme tolikrát, že bychom s ním mohli zaplnit všechny přihrádky. Pak je zřejmé, že do první přihrádky můžeme umístit jeden z  $n$  druhů předmětů, které máme. Do druhé přihrádky můžeme umístit opět  $n$  předmětů různého druhu, protože jich máme dostatek, aby se mohli opakovat. Do poslední  $k$ -té přihrádky můžeme vybírat stále z  $n$  druhů předmětů z toho samého důvodu.

1	2	3	...	$k - 1$	$k$
$n$	$n$	$n$	...	$n$	$n$

S odvoláním na pravidlo o násobení A.1.2 můžeme vyjádřit počet způsobů, jimiž lze danou úlohu vyřešit a stanovit výsledek:

$$V_k^*(n) = \underbrace{n \cdot n \cdot n \cdot \dots \cdot n}_{k \times} \quad (12)$$

To odpovídá následujícímu zápisu:

$$V_k^*(n) = n^k. \quad (13)$$

Uvědomte si, že jsme právě odvodili vzorec pro počet variací  $k$ -té třídy z  $n$  prvků s opakováním.

### A.2.4 Rozlišitelné přihrádky a rozlišitelné předměty, jež se opakují \*

Speciálním typem předcházející úlohy je případ, kdy každý z předmětů máme tolikrát, že s ním sice nemůžeme zaplnit všechny přihrádky. Celkem však máme právě tolik předmětů kolik je přihrádek. Počet přihrádek označme, jak bývá zvykem,  $n$ . Počet různých druhů předmětů označme  $l$ . V souvislosti předchozím jistě platí:  $n_1 + n_2 + \dots + n_l = n$ ,



kde  $n_i$  pro  $i = 1, \dots, l$  jsou počty jednotlivých druhů předmětů. Konstrukce, jíž bychom došli k správnému výsledku vychází z myšlenky použité v jednom z následujících oddílů (viz oddíl A.2.5).

Zkuste si odvození rozmyslet sami a uvědomte si, že princip použitý ke kompenzaci výpočtu u úlohy A.2.5 lze jen opakovaně použít pro všech  $l$  druhů předmětů. Vycházejte z výpočtu pro permutace bez opakování.

$$P^*(n) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_l!}. \quad (14)$$

Uvědomte si, že jsme právě odvodili vzorec pro počet permutací  $n$  prvků s opakováním.

### A.2.5 Nerozlišitelné přihrádky a rozlišitelné předměty, jež se neopakují

Mějme  $n$  rozlišitelných předmětů, jež chceme umístit do  $k$  nerozlišitelných přihrádek. Dále předpokládejme, že každý předmět máme jen jednou a ve skutečnosti nás zajímá, jen jaké předměty vybereme, neboť přihrádky jsou nerozlišitelné. Z tohoto důvodu je rozumné požadovat, abychom měli více, nebo alespoň stejně předmětů jako je přihrádek, tedy  $n \geq k$ . Vyjdeme z modelu A.2.1 a vzorec upravíme, tak aby vyhovoval této specifikaci. Uvědomte si, že v případě A.2.1 záleželo na pořadí a proto jsme do výsledků zahrnovali všechny možné permutace výběru se stejnými předměty. Kdybychom měli například jen tři přihrádky a do nich umísťovali symboly ♣, ♠ a ♥. Pak bychom dostali jako různá řešení všechny jejich permutace, tj.:

1.	♣	♠	♥
2.	♣	♥	♠
3.	♠	♣	♥
4.	♠	♥	♣
5.	♥	♠	♣
6.	♥	♣	♠

Tyto případy, jež byly z hlediska úlohy A.2.1 různé, se v případě neočíslovaných přihrádek neliší, neboť nezáleží na pořadí. Těchto případů je pro  $k$  přihrádek právě  $k!$ , neboť se jedná o permutace  $n$  prvků bez opakování, viz oddíl A.2.2. To znamená, že hodnota vypočtená podle vzorce (9) je  $k!$  krát větší než hodnota, kterou chceme spočítat. Proto  $V_k(n)$  vydělíme  $k!$  a získáme počet způsobů, jimiž lze úlohu A.2.5 vyřešit:

$$C_k(n) = \frac{n!}{(n-k)! \cdot k!}. \quad (15)$$

To odpovídá následujícímu zápisu, ve kterém využíváme kombinačních čísel:

$$C_k(n) = \binom{n}{k}. \quad (16)$$

Uvědomte si, že jsme právě odvodili vzorec pro počet kombinací  $k$ -té třídy z  $n$  prvků bez opakování.

### A.2.6 Nerozlišitelné přihrádky a rozlišitelné předměty, jež se mohou opakovat

Mějme  $n$  druhů rozlišitelných předmětů, jež chceme umístit do  $k$  přihrádek. Dále předpokládejme, že každý předmět máme tolikrát, že bychom s ním mohli zaplnit všechny přihrádky (tj. počet předmětů jednoho druhu je větší nebo roven než počet přihrádek). Ve skutečnosti nás zajímá, jen jaké předměty vybereme a kolikrát. Pro rozřešení této úlohy je využito pěkného nápadu se zakódováním každého možného výběru (vzpomeňte na princip bijekce – podkapitola A.1.1).

Představte si, že máte k dispozici „basu“<sup>74</sup>. Ve Vašem oblíbeném obchodu je k dispozici 5 druhů piva. Řekněme Budvar, Staropramen, Plzeň, Nektar a Krušovice. Basu můžete z pohledu konzumenta naplnit mnoha způsoby. Kolika? Na to budete schopni za chvíli schopni odpovědět. Některými z možných nákupů jsou například (viz tabulka níže):

Případ	Budvar	Staropramen	Plzeň	Nektar	Krušovice
1.	20	0	0	0	0
2.	4	4	4	4	4
3.	2	0	3	10	5
4.	0	12	0	8	0
5.	3	5	3	4	5
6.	1	7	0	6	6
⋮					

Představte si, že jste si vybrali a rádi by jste sdělili někomu blízkému z čeho, že si večer bude moci vybrat. Jelikož jste oba (obě) ve skrze hravé osoby nebojící se výzev, rozhodli jste se skladbu večerní rehydratace popsat jen pomocí, dejme tomu, fazolí a zápalek. Nebudu Vás dlouho napínat a prozradím Vám, že pro takovéto popsání budete potřebovat právě 20 fazolí a 4 zápalky. Domluva je následující:

1. Předměty budete skládat do řady za sebe.
2. Každá z fazolí představuje jednu láhev s pivem<sup>75</sup>.
3. Každá zápalka představuje oddělovač mezi druhy piv. Vzhledem k tomu, že předměty skládáme v řadě stačí nám o jednu zápalku méně než máme druhů piv. Tj. fazole před 1. zápalkou reprezentují jeden druh piva, fazole za 1. a před 2. zápalkou reprezentují druhý druh piva. A tak stále dál, až fazole za 4. zápalkou reprezentují pátý druh piva.
4. Druhy piv mají v našem zakódování pevně dané pořadí Budvar, Staropramen, Plzeň, Nektar a nakonec Krušovice.

Jak tedy budou vypadat jednotlivá zakódování počtů piv uvedených možných řešeních? Přesně v souladu s vytvořenou úmluvou jsou vidět na následujícím obrázku zakódování prvních čtyř z výše uvedených možných nákupů.

<sup>74</sup>V tomto kontextu: Basa = plastová přepravka na láhve s pivem respektive od piva, do které se obvykle vejde 20 již zmíněných lahví.

<sup>75</sup>Předpokládáme, že obsah bude skutečně konzumován až večer.





Případy

1.	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q					
2.	Q	Q	Q	Q		Q	Q	Q	Q		Q	Q	Q	Q		Q	Q	Q	Q		Q	Q	Q	Q
3.	Q	Q			Q	Q	Q		Q	Q	Q	Q	Q	Q	Q	Q	Q	Q		Q	Q	Q	Q	Q
4.		Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q			Q	Q	Q	Q	Q	Q	Q	Q	Q	
:																								

Naopak z tohoto zakódování jasně vidíme, že v 1. případě bylo koupeno 20 Budvarů a žádné jiné pivo, v 2. případě byly koupeny 4 Budvary, 4 Staroprameny, 4 Plzně, 4 Nektary a 4 Krušovice. Ve 3. případě se jedná o 2 Budvary, žádné Staroprameny, 3 Plzně, 10 Nektarů a 5 Krušovic. Ostatní nepopsané příklady jsou nechány laskavému čtenáři k procvičení kódovacího algoritmu.

Opusťme však pro tuto chvíli pivní tematiku a vraťme se zpět k obecnému příkladu. Využijeme-li poznatky z předchozího kódování, vidíme, že každý z kódů obsahoval právě  $n+k-1$  pozic.  $k$  pozic pro počet vybíraných předmětů a  $n-1$  pozic pro oddělovníky druhů. Vzhledem k tomu, že proces kódování je vzájemně jednoznačný, pak možných řešení je právě tolik, kolik je různých kódů. Konečně, různých kódů je tolik, kolika způsoby lze různě umístit  $n-1$  nerozlišitelných oddělovníků druhů na  $n+k-1$  rozlišitelných pozic, respektive je třeba vybrat  $n-1$  pozic bez ohledu na pořadí, neboť na ně budeme umísťovat nerozlišitelné oddělovníky, z  $n+k-1$  všech možných. Pečlivý čtenář již jistě poznal, že vhodným zakódováním jsme tuto úlohu převedli na úlohu A.2.5, tj. kombinace bez opakování. Po dosažení do vzorce (16) získáváme vzorec pro počet způsobů, jimiž lze tuto úlohu vyřešit a stanovit výsledek:

$$C_k^*(n) = \binom{n+k-1}{n-1} \tag{17}$$

Uvědomte si, že jsme právě uvedli vzorec pro počet kombinací  $k$ -té třídy  $n$  prvků s opakováním.

Vrátíme-li se k ilustračnímu příkladu, jistě již nebude problém spočítat, kolika způsoby lze různě naplnit basu z pohledu uživatele, kterého zřejmě obsah lahví zajímá více než jejich umístění<sup>76</sup>.

<sup>76</sup> pro jistotu výsledek je  $\binom{24}{19} = 42504$

## A.3 Shrnutí

Obecné pojmenování	výpočetní vzorec	odkaz
Faktoriál čti [en faktoriál]	$n! = n \cdot (n - 1) \cdot \dots \cdot 1$	
Kombinační číslo čti [en nad ká]	$\binom{n}{k} = \frac{n!}{(n - k)! \cdot k!}$	
Komb. princip o násobení	$\#A_1 \cdot \#A_2 \cdot \dots \cdot \#A_k$	<a href="#">A.1.2</a>
Komb. princip o součtu	$\# \bigcup_{i=1}^k A_i = \#A_1 + \dots + \#A_k$	<a href="#">A.1.3</a>
Princip exkluze a inkluze	$\# \bigcup_{i=1}^k A_i = \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{(\#I-1)} \cdot \# \bigcap_{i=1}^k A_i$	<a href="#">A.1.4</a>
Variace $k$ -té třídy z $n$ prvků bez opakování	$V_k(n) = \frac{n!}{(n - k)!}$	<a href="#">A.2.1</a>
Variace $k$ -té třídy z $n$ prvků s opakováním	$V_k^*(n) = n^k$	<a href="#">A.2.3</a>
Permutace $n$ prvků	$P(n) = n!$	<a href="#">A.2.2</a>
Permutace $n$ prvků s opako- váním	$P^*(n) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_l!}$	<a href="#">A.2.4</a>
Počet kombinací $k$ -té třídy z $n$ prvků bez opakování	$C_k(n) = \binom{n}{k}$	<a href="#">A.2.5</a>
Počet kombinací $k$ -té třídy z $n$ prvků s opakováním	$C_k^*(n) = \binom{n + k - 1}{n - 1}$	<a href="#">A.2.6</a>



## B Jak počítá statistik ryby v rybníku

V rybníku je neznámý počet ryb (označme ho  $N$ ), o němž bychom se rádi něco dozvěděli. Metoda přímého, vyčerpávajícího měření (populační šetření), vyžadující, aby byl rybník vypuštěn, není přitom často použitelná.

Teorie  
Praxe

### Statistik navrhuje nepřímý postup, opírající se náhodný o výběr:

Z rybníka se vyloví  $M$  ryb, které se označí a pustí do rybníka zpět. Po nějaké době (zaručující promíchání  $M$  označených ryb mezi ostatní) se provede další částečný výlov; řekněme, že při něm bylo vyloveno  $n$  ryb, mezi nimiž se rozpoznalo  $k$  ( $0 \leq k \leq \min(M; n)$ ) ryb dříve označených. Přijme-li se princip, že podíl označených ve vzorku  $n$  vylovených ryb se rovná podílu označených v celém rybníku, měl by platit vztah

$$\frac{k}{n} = \frac{M}{N}$$

v němž  $k$ ,  $M$  a  $n$  jsou veličiny známé, což umožňuje odhadovat počet ryb v rybníku ze vzorce po jednoduché úpravě:

$$N = \frac{M \cdot n}{k}.$$

Profesionální statistický přístup tímto výpočtem ovšem nekončí – spíše teprve začíná. Získaný výsledek  $N$  je třeba považovat za odhad, který je vystaven riziku omylu. Bylo by přece naivně optimistické předpokládat, že v rybníku žije právě tolik ryb, kolik ukazuje náš odhad. Podstatné je, že velikost chyby (odchylky odhadu od neznámé skutečnosti) se dá rovněž statistickými metodami posoudit a ukázat její závislost na rozsahu výběru  $n$  a podílu označených ryb  $M/N$ .

Z toho vyplývá, že počet  $k$  označených ryb ve druhém výlovu je náhodná veličina  $K$  s hypergeometrickým rozdělením  $K \sim H(M; N; n)$  pro  $k \in (\max(0; n - (N - M)); \min(M; n))$ , kde pravděpodobnost, že náhodná veličina  $K$  nabyde hodnoty  $k$  spočteme jako

$$P(K = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

a střední hodnotou jako  $EK = \frac{M \cdot n}{N}$ , což naznačuje, že výše uvedený postup odhadování byl rozumný.

Volně podle <http://medinfo.upol.cz/PUVODNI/statistika/>

## C Vyhodnocování výsledků testování hypotéz na základě „ $p$ -value“

OTO

### C.1 Klasická metodologie vyhodnocování testování hypotéz

Uvědomme si na, jakém principu je založeno testování hypotéz: Za platnosti *nulové hypotézy* a splnění předpokladů pro použití testu tvrdíme, že bude *testové kritérium* – jakžto náhodná veličina – sledovat určité rozdělení. Testovým kritériem rozumíme hodnotu, vypočtenou na základě náhodného výběru z dat, jež testujeme. Proti nulové hypotéze ( $H_0$ ) obvykle stojí *hypotéza alternativní* ( $H_A$ ), která svým způsobem popírá hypotézu nulovou. Ty hodnoty testového kritéria, které jsou více méně málo pravděpodobné za platnosti nulové hypotézy, pak „obětujeme“ a nulovou hypotézu zamítneme, i když by mohla platit, a tvrdíme, že se nám podařilo prokázat hypotézu alternativní. *Chyba prvního druhu*, která je vlastně touto obětí (nulová hypotéza platí a my ji zamítáme) je kvantifikována<sup>77</sup> prostřednictvím *hladiny významnosti*  $\alpha$ , obvykle volené 0,1, 0,05, 0,01 či 0,001. Obětované hodnoty testového kritéria vymezuujeme v *kritických oborech* pomocí *kritických hodnot*. Kritické hodnoty jsou vlastně kvantily rozdělení, jež testové kritérium sleduje, a odpovídají hladině významnosti. Kritické obory jsou navíc konstruovány tak, aby hodnoty v nich obsažené byly přijatelné pro alternativní hypotézu a pro nulovou málo pravděpodobné.

Pokud tedy hodnota testového kritéria spadne do kritického oboru, zamítáme nulovou hypotézu ve prospěch alternativní na hladině významnosti  $\alpha$ . Čím nižší je hladina významnosti, tím obtížnější je zamítnutí nulové hypotézy (kritické obory jsou menší), ale test zároveň ztrácí svou *sílu*<sup>78</sup>. Srovnání tedy probíhá v hodnotách náhodné veličiny a hladina významnosti je pro tento účel převáděna na kritické hodnoty. Konstrukce kritických oborů však není cílem tohoto textu.

Nešlo by tedy naopak modifikovat hodnotu testového kritéria a tu pak porovnávat přímo s hladinou významnosti? Srovnání by tedy probíhalo na úrovni pravděpodobnosti. Samozřejmě to jde. Modifikovaná hodnota testového kritéria bývá obvykle označována *p-value* a její možná interpretace a úskalí budou vyložena dále.

### C.2 Interpretace významu hodnoty „ $p$ -value“

Má-li být hodnota *p-value* jasně definována, potřebujeme vedle testu, který provádíme, znát hodnotu testového kritéria a alternativní hypotézu.

Podstata hodnoty *p-value* a hladiny významnosti  $\alpha$  je v podstatě stejná (jedná se o pravděpodobnosti, vypovídající cosi o nulové hypotéze). Hladina významnosti je předpokládaná pravděpodobnost zamítnutí nulové hypotézy za předpokladu, že byla správná (pravděpodobnost chyby prvního druhu) a určujeme ji vždy před výpočtem testového kritéria (tj. před testem). *p-value* je taková nejnižší možná hladina významnosti – určená

<sup>77</sup>Je vyjádřena její pravděpodobnost:  $P(H_A|H_0) = \alpha$ .

<sup>78</sup>Síla testu je pravděpodobnost, že je testem zamítnutá nulová hypotéza, když skutečně neplatí:  $P(H_A|H_A) = 1 - \beta$ .



na základě hodnoty testového kritéria (tj. na základě výsledku kvantifikace výběru) –, při které lze ještě zamítnout nulovou hypotézu.

$p$ -value nám poskytuje obecněji více informací o výsledku statistického testování než pouhé zamítnutí nebo nezamítnutí nulové hypotézy. Předpokládejme, že  $p$ -value vyjde rovna 0,05. Z toho lze usoudit, že nulovou hypotézu lze zamítnout například na hladině významnosti  $\alpha = 0,1$ , ale již ne na hladině významnosti  $\alpha = 0,01$  a  $\alpha = 0,001$ . Nejnižší možnou hladinu významnosti, na které ještě můžeme nulovou hypotézu zamítnout je právě  $\alpha = 0,05$ . Ještě jednou bude-li  $p$ -value rovna 0,03, potom můžeme nulovou hypotézu zamítnout na hladině významnosti  $\alpha = 0,1$  nebo  $\alpha = 0,05$ . Na hladině významnosti nižší než  $\alpha = 0,03$  nulovou hypotézu nezamítáme (tedy např. na hladině významnosti  $\alpha = 0,01$ ). Čím nižší vyjde  $p$ -value, tím více jsme přesvědčeni, že nulová hypotéza není správná a je třeba jí zamítnout.

Nejjednodušší způsob, jak rozhodovat o výsledku testu spočívá v porovnání  $p$ -value (vypočte počítač – např. program STATISTICA) a hladiny významnosti  $\alpha$  (určíme před testem sami). Platí následující pravidlo:

- a) Zamítni nulovou hypotézu  $H_0$ , když  $p\text{-value} \leq \alpha$ .
- b) Nezamítej nulovou hypotézu  $H_0$ , když  $p\text{-value} > \alpha$ .

Tento způsob testování je nejrychlejší a nejpohodlnější, pokud máme k dispozici počítač. Většina statistických programů počítá  $p$ -value, ať už je v softwarovém prostředí označováno symboly  $p$ , *Significance level* apod.

### C.3 Není všechno zlato, co ...

Vše co bylo zmíněno výše platí, musíme však připustit, že i zde existuje **ALE**. Tím ALE je způsob výpočtu  $p$ -value pomocí statistického softwaru. Jak už jste jistě zjistili, existují z určitého hlediska dva typy alternativních hypotéz – *jednostranná* a *oboustranná* alternativní hypotéza. Některé testy hypotéz jsou používány jen pro jednostranné resp. oboustranné alternativní hypotézy. Je-li tomu tak, pak statistický software dává jako výsledek  $p$ -value, kterou můžeme interpretovat tak, jak bylo zmíněno výše.

Pro testy hypotézy, kde se běžně používají jak pro oboustranné tak jednostranné varianty alternativních hypotéz, vykalkuluje program hodnotu  $p$ -value pro oboustrannou hypotézu. To se týká především takzvaných  $t$ -testů. Software vypočte pravděpodobnost s jakou by náhodná veličina sledující teoretické rozdělení nabyla hodnotu větší, než je absolutní hodnota z hodnoty testového kritéria, a tuto hodnotu vynásobí dvěma. Vynásobená hodnota je následně uživateli předložena jako  $p$ -value. Proto je výše popsán způsob interpretace  $p$ -value (ve většině programech) bez modifikace možný jen pro takzvaný oboustranný test, tedy když dvojice hypotéz  $H_0$  a  $H_A$  má např. následující matematické předpisy<sup>79</sup>:

$$H_0 : \mu = \mu_0, \quad H_A : \mu \neq \mu_0. \quad (18)$$

<sup>79</sup>Jedna se o jednovýběrový  $t$ -test.

## C.4 Modifikace oboustranného $p$ -value

Lze tedy tuto hodnotu  $p$ -value využít k rozhodnutí o výsledku testování i pro jednostranné testy? I v tomto případě lze naznačený problém vyřešit. Dokonce několikerým způsobem.

### C.4 Modifikace oboustranného $p$ -value

Vzhledem k tomu, že se tento text elegantně vyhýbá výpočetnímu pozadí testů, dokonce jsme v něm ani přesně nespecifikovali, jak se tvoří kritické obory, existuje jediná možnost, jak vyžrát na oboustranné  $p$ -value, aniž by jsme tento trend porušili. Dříve než tak učiníme, je nutnou dodat, že hodnotu  $p$ -value – příslušnou jakékoliv alternativní hypotéze – lze pomocí vhodného software vypočítat manuálně<sup>80</sup>, nebo na základě znalosti výpočtu testového kritéria a předpisu pro kritický obor odvodit pravidlo jak z oboustranného  $p$ -value získat správné „jednostranné“  $p$ -value. Úmyslně na tomto místě není napsáno jednoduché pravidlo, protože – byť je triviální – musíme k výpočtu „jednostranného“  $p$ -value přistupovat s rozmyslem. Zde toto pravidlo odvozovat nebudeme a omezíme se na popsání tohoto pravidla modifikace  $p$ -value pomocí známých pojmů.

Označme alternativní hypotézu  $H_A : \theta_1 < \theta_2$  jako levostrannou a  $H_A : \theta_1 > \theta_2$  jako pravostrannou. Přičemž  $\theta_1$  a  $\theta_2$  označují testované parametry (například střední hodnoty hodnoty u  $t$ -testů). Záludností pravidla, které popisujeme, je skutečnost, že modifikace  $p$ -value vypočteného softwarem pro oboustrannou alternativní hypotézu nezáleží jen na námi testované alternativní hypotéze, ale také na datech – tedy hodnotě testového kritéria. Pravidlo v sobě tedy kombinuje jak typ alternativní hypotézy tak hodnotu testového kritéria.

Pokud budeme uvažovat **levostrannou** alternativní hypotézu, pak oboustranné  $p$ -value **dělíme dvěma** v případě **záporné** hodnoty testového kritéria, respektive **dělíme dvěma a odečítáme od jedné** v případě **kladné** hodnoty testového kritéria.

Naopak v případě **pravostranné** alternativní hypotézy, oboustranné  $p$ -value **dělíme dvěma** v případě **kladné** hodnoty testového kritéria a respektive **dělíme dvěma a odečítáme od jedné** v případě **záporné** hodnoty testového kritéria.

Přehledně toto pravidlo shrňme v tabulce 38, v níž je naznačena i triviální situace pro oboustrannou alternativní hypotézu.

Tabulka 38: Modifikace oboustranného  $p$ -value. Modifikace oboustranného  $p$ -value,  $p$ -value je v tabulce pro přehlednost značenou pouze  $p$

		test. kritérium < 0	test. kritérium > 0
$H_A : \theta_1 < \theta_2$	$t_{\text{hyp}} < 0$	$p/2$	$1 - p/2$
$H_A : \theta_1 > \theta_2$	$t_{\text{hyp}} > 0$	$1 - p/2$	$p/2$
$H_A : \theta_1 \neq \theta_2$	$t_{\text{hyp}} \neq 0$	$p$	$p$

<sup>80</sup>V programu STATISTICA lze hodnotu  $p$ -value dopočítat například pomocí pravděpodobnostního kalkulátoru rozdělení.



## D Průběžný test Statistika I



1. Kolik musí být  $p_2 = P(X = 3)$  (viz tabulka níže), aby se jednalo o rozdělení pravděpodobnosti?

$x_i$	1	3	5	6
$p_i$	0,2	??	0,3	0,15

2. Určete:  $F(0,3)$ ,  $F(3)$  a  $F(7)$
3. Napište jaká diskrétní rozdělení pravděpodobnosti znáte.





## E Průběžný test Statistika II



### E.1 Průběžný test Statistika II/a



1. Náhodná veličina  $X$  sleduje Poissonovo rozdělení s parametrem  $\lambda = 6$ , vypočtěte následující pravděpodobnosti:

- (a)  $P(X \leq 4)$ ,
- (b)  $P(X > 4)$ ,
- (c)  $P(X = 4)$ ,
- (d)  $P(X < 4)$ .

2. Zemědělec, který vždy na podzim zvažuje, zda má dále ponechat na poli řepu (nepřijdou-li mrazy, vyzraje, přijdou-li mrazy, hrozí, že o úrodu přijde) bedlivě sleduje již dvacet let počasí. V tabulce má zapsáno, kolikrát za poslední roky byla předpověď, že má mrznout nebo ne a kolikrát skutečně mrzlo či bylo pěkně.

	předpověď mráz	předpověď bez mrazu
mrzlo	10	2
nemrzlo	4	4

Na základě těchto údajů, určete, jaká je v daném období pravděpodobnost, že

- (a) bude mrznout,
  - (b) budou předpovídat mráz,
  - (c) bude mrznout, pokud předpovídají mráz.
  - (d) Jsou jevy „bude mrznout“ a „předpovídá se mráz“ nezávislé? A proč?
  - (e) Co je jevem opačným k jevu „alespoň jedno ráno klesne teplota na bod mrazu“?
  - (f) Co je průnikem jevů „bude mráz“ a „předpovídají mráz“?
  - (g) Co je sjednocením jevů „bude mráz“ a „předpovídají mráz“?
3. V testu z matematiky je 15 otázek. Na každou je nabízeno pět možných odpovědí, přičemž pouze jedna je správná. K úspěšnému napsání testu je třeba mít alespoň 8 správných odpovědí.
- (a) Jaká je pravděpodobnost, že test úspěšně napíše student, který nemá o matematice ani páru?
  - (b) Jaká je pravděpodobnost, že tento student získá zápočet (bez využití opravných termínů) z matematiky, potřebuje-li k tomu napsat dva testy a opět nemá o matematice ani páru?
  - (c) Jaká je pravděpodobnost, že tento student projde přes první test, tzn. napíše ho buď na poprvé nebo v opravném termínu (na který se opět nepřipravil)?



## E.2 Průběžný test Statistika II/b

- Náhodná veličina  $X$  sleduje normální rozdělení s parametry  $\mu = 5$  a  $\sigma^2 = 4$ , vypočtete následující pravděpodobnosti:
  - $P(X \leq 4)$ ,
  - $P(X > 4)$ ,
  - $P(X = 4)$ ,
  - $P(X < 4)$ .
- Zemědělec, který vždy na podzim zvažuje, zda má dále ponechat na poli řepu (nepřijdou-li mrazy, vyzraje, přijdou-li mrazy, hrozí, že o úrodu přijde) bedlivě sleduje již dvacet let počasí. V tabulce má zapsáno, kolikrát za poslední roky byla předpověď, že má mrznout nebo ne a kolikrát skutečně mrzlo či bylo pěkně.

	předpověď mráz	předpověď bez mrazu
mrzlo	10	2
nemrzlo	4	4

Na základě těchto údajů, určete, jaká je v daném období pravděpodobnost, že

- nebude mrznout,
  - budou předpovídat, že mráz nebude,
  - bude mrznout, pokud předpovídají nemá mrznout.
  - Jsou jevy „nebude mrznout“ a „nepředpovídá se mráz“ nezávislé? A proč?
  - Co je jevem opačným k jevu „alespoň jedno ráno klesne teplota na bod mrazu“?
  - Co je průnikem jevů „nebude mráz“ a „nepředpovídají mráz“?
  - Co je sjednocením jevů „nebude mráz“ a „nepředpovídají mráz“?
- K panu Čáblíkovi chodí do holičství průměrně pět lidí za hodinu. Mistr Čáblík už se na tolik vypracoval, že jednoho zákazníka dělá právě dvanáct minut (rád si povídá, takže i když by mohl být hotov dříve, raději přidělený čas „prokecá“ se zákazníkem).
    - Za mistrem Čáblíkem přijde návštěva, která má hodinu čas. Jaká je pravděpodobnost, že si na ni dokáže pan holič udělat čas?
    - Jaká je pravděpodobnost, že by se podařilo mistru Čáblíkovi si nějaký čas vyšetřit dvakrát po sobě?
    - A jaká je pravděpodobnost, že ve dvou po sobě jdoucích hodinách by měl pan holič čas, alespoň v jedné z těchto dvou hodin?



## F Průběžný test Statistika III



### F.1 Průběžný test Statistika III/a



1. Vypočtete:  $\binom{5}{3}$
2. Uvažujte dva neslučitelné (vzájemně se vylučující) jevy  $A$  a  $B$ . Doplňte následující vyjádření co nejpřesněji:
  - (a)  $A \cap B =$ ,
  - (b)  $P(A \cup B) =$ .
3. Náhodná veličina  $X$  sleduje binomické rozdělení s parametry  $n = 8$ ,  $p = 0,4$ . Určete:
  - (a)  $P(X \leq 5)$ ,
  - (b)  $P(X = 10)$ .
4. Náhodná veličina  $Y$  sleduje  $\chi^2$  rozdělení se 17 stupni volnosti. Určete:
  - (a)  $P(Y = 12)$ ,
  - (b)  $P(X \leq 17)$ .
5. Spočtete průměrnou hodnotu statistického znaku  $X$  zadaného posloupností  $\{2i + 1\}_{i=1}^5$ .

### F.2 Průběžný test Statistika III/b



1. Vypočtete:  $\binom{5}{3} - \binom{5}{2}$ .
2. Uvažujte takové jevy, že jev  $A$  je podjevem jevu  $B$  ( $A \subset B$ ). Vyjádřete následující pravděpodobnosti jen s využitím  $P(A)$  a  $P(B)$  – lze-li to:
  - (a)  $P(A \cap B) =$  ,
  - (b)  $P(B - A) =$  .
3. Náhodná veličina  $X$  sleduje Poissonovo rozdělení ( $X \sim \text{Po}(\lambda)$ )s rozptylem  $DX = 7$ . Určete:
  - (a)  $P(X \geq 3)$
  - (b)  $P(X < 0)$
4. Určete hodnotu:  $\Phi(u_{0,975} + t_{0,026}(139))$
5. Spočtete populační rozptyl statistického znaku  $X$  zadaného posloupností  $\{(-1)^n\}_{i=1}^4$ .

### F.3 Průběžný test Statistika III/c



1. Symbolicky vyznačte, že dva jevy jsou neslučitelné.
2. Napište značení (symbol) pro:

F.3 Průběžný test Statistika III/c

- (a) střední hodnotu náhodné veličiny  $X$ ,      (c) distribuční funkci,  
(b) rozptyl náhodné veličiny  $X$ ,                      (d) (aritmetický) průměr.

3. Napište jinak  $P(X \leq x)$ .
4. Pro jaké hodnoty je definována náhodná veličina s binomickým rozdělením  $X \sim \text{Bi}(14; 0,7)$ , tj. pro jaké hodnoty  $x$  je pravděpodobnost  $P(X = x)$  nenulová?
5. Spočtete (populační) rozptyl ze 4 hodnot: 1, 1, 1, 1.



## G Průběžný test Statistika IV



1. Vypište již probrané oblasti, jimiž jste se v přednáškách respektive ve cvičeních zabývali.
2. Na večírek přijde v průměru 60 hostů za hodinu. Jaká je pravděpodobnost, že během půl minuty nikdo nepřijde, řídí-li se náhodná veličina – počet příchozích – Poissonovým rozdělením?
3. Zjistěte kvantily v následujících příkladech:
  - (a)  $t_{0,95}(45)$ ,  $t_{0,05}(45)$ ;
  - (b)  $\chi^2_{0,95}(17)$ ,  $\chi^2_{0,05}(17)$ ;
  - (c)  $F_{0,95}(17; 26)$ ,  $F_{0,05}(26; 17)$ ;
  - (d)  $u_{0,95}$ ,  $u_{0,01}$ .
4. Napište předpisy pro (populační) směrodatnou odchylku a pro výběrovou směrodatnou odchylku při prostém třídění. Dále napište:
  - (a) kdy jakou z výše uvedených měr variability používáme,
  - (b) jakou míru variability Vám počítá implicitně program STATISTICA komplet 6.1 Cz,
  - (c) jaké jiné míry variability znáte.
5. Zjistěte podle jakého vzorce MS Excel počítá šikmost nebo špičatost.
6. Popište rozdíl mezi bodovým a intervalovým odhadem.
  - (a) Jaké vlastnosti oceňujeme u bodového odhadu? (nápopověda: *bodový odhad je . . .*) Alespoň jednu vlastnost striktně definujte.
  - (b) Jaký je nutný předpoklad na základní soubor, ze kterého činíme výběr, pro výpočet intervalu spolehlivosti pro střední hodnotu, který znáte z přednášek?
7. Na základě dat určete 95% interval spolehlivosti pro směrodatnou odchylku životnosti matrice. U náhodně vybraných matric byla životnost následující – viz tabulka 7.



# H Zkouškový test Statistika

## H.1 Zkouškový test Statistika – 2005



Obrázek 3: Vzorový zkouškový test ze Statistiky – rok 2005

**Jméno a příjmení:** \_\_\_\_\_ **Obor:** \_\_\_\_\_ **Hodnocení:** \_\_\_\_\_

Počístič: Písemná část trvá 60 min. Povolené pomůcky: software STATISTICA komplet 6.1 Cx, MS Excel, Maple, R a kapesní kalkulačtor. Vždy uvádějte podstatné hodnoty, na základě nichž jste se rozhodovali. Z výsledku **musí** být zřejmý postup. V průběhu práce není povoleno odcházet z místnosti.

**Test STAT**  
16. čerence 2005

1. Popište princip párového  $t$ -testu. Pro jaké soubory se tento test používá?

2. Četnost kazů skla při výrobě okenních tabulí se řídí Poissonovým rozdělením. Dlouhodobým měřením bylo zjištěno, že na jednu tabuli skla odpovídá v průměru  $1/4$  kazu. Okenní tabule bez kazu jsou označovány jako výběrové, okenní tabule s jedním nebo dvěma kazů jsou označovány jako partiové. Okenní tabule s více kazů jsou vyřazovány jako nevyhovující.

(a) Spočítejte pravděpodobnost, že vyrobená tabule je partiová.  
(b) Spočítejte pravděpodobnost, že v okenní tabuli bylo nejvýše pět kazů za předpokladu, že tato tabule byla vyřazena.

3. V továrně jsou šrouby vyráběny na dvou strojích. Na prvním stroji dosahují šrouby požadované kvality s pravděpodobností 0,92. Na druhém stroji jsou šrouby s požadovanou kvalitou vyráběny s pravděpodobností 0,8. Přitom první stroj má větší výkonnost než druhý, tj. vyrobí třikrát více výrobků, než stroj druhý. Ze skladu byl vybrán zaměstnancem jeden šroub. Určete pravděpodobnost, že šroub má předepsanou kvalitu.

© Biskup, Rost 2005–2007

První stránka

4. Co rozumíte příjmutnou chybou odhadu střední hodnoty při neznámém rozptylu? Napište vzorec a vypište význam jednotlivých symbolů ve vzorci.

5. Zjistěte závislost počtu pracovních hodin ( $PPH$ ) za měsíc spojených s provozováním anesteziologické služby na velikosti spádové populace (v tisících lidech) nemocnice ( $SP$ ). Údaje byly získány ve 12 nemocnicích ve Spojených státech.

Nemocnice	1.	2.	3.	4.	5.	6.
$SP$	25,5	294,3	83,7	30,7	129,8	180,8
$PPH$	304,37	2616,32	1139,12	285,43	1413,77	1555,68

Nemocnice	7.	8.	9.	10.	11.	12.
$SP$	43,4	165,2	74,3	60,8	319,2	376,2
$PPH$	383,78	2174,27	845,3	1125,28	3402,6	3682,33

Zvolte vhodný model, pokud bude třeba tak zjednodušte. Výsledný model zapište. Jaký počet pracovních hodin lze očekávat při spádové populaci velké 100 000 lidí? Udejte bodový i intervalový odhad (95%) pro predikci.

6. V nejmenovaném městě se na třech základních školách v páté třídě měřilo IQ žáků (viz tabulka):

Škola	Dosažené výsledky							
Horní	128	110	98	101	103	112	110	105
Dolní	95	102	92	118	105	107	89	
Kostelní	132	112	109	111	98	118	113	111

Dokazují data odlišnost v IQ mezi školami? Tj. mají žáci alespoň jedné z nich IQ odlišné od žáků ostatních škol? Výsledky interpretujte na hladině významnosti  $\alpha = 0,05$ .

© Biskup, Rost 2005–2007

Druhá stránka

1. Popište princip párového  $t$ -testu. Pro jaké soubory se tento test používá?
2. Četnost kazů skla při výrobě okenních tabulí se řídí Poissonovým rozdělením. Dlouhodobým měřením bylo zjištěno, že na jednu tabuli skla odpovídá v průměru  $1/4$  kazu. Okenní tabule bez kazu jsou označovány jako výběrové, okenní tabule s jedním nebo dvěma kazů jsou označovány jako partiové. Okenní tabule s více kazů jsou vyřazovány jako nevyhovující.
  - (a) Spočítejte pravděpodobnost, že vyrobená tabule je partiová.
  - (b) Spočítejte pravděpodobnost, že v okenní tabuli bylo nejvýše pět kazů za předpokladu, že tato tabule byla vyřazena.

3. V továrně jsou šrouby vyráběny na dvou strojích. Na prvním stroji dosahují šrouby požadované kvality s pravděpodobností 0,92. Na druhém stroji jsou šrouby s požadovanou kvalitou vyráběny s pravděpodobností 0,8. Přitom první stroj má větší výkonnost než druhý, tj. vyrobí třikrát více výrobků, než stroj druhý. Ze skladu byl vybrán zaměstnancem jeden šroub. Určete pravděpodobnost, že šroub má předepsanou kvalitu.
4. Co rozumíte přípustnou chybou odhadu střední hodnoty při neznámém rozptylu? Napište vzorec a vypište význam jednotlivých symbolů ve vzorci.
5. Zjistěte závislost počtu pracovních hodin (*PPH*) za měsíc spojených s provozováním anesteziologické služby na velikosti spádové populace (v tisících lidech) nemocnice (*SP*). Údaje byly získány ve 12 nemocnicích ve Spojených státech.

Nemocnice	1.	2.	3.	4.	5.	6.
<b>SP</b>	25,5	294,3	83,7	30,7	129,8	180,8
<b>PPH</b>	304,37	2 616,32	1 139,12	285,43	1 413,77	1 555,68

Nemocnice	7.	8.	9.	10.	11.	12.
<b>SP</b>	43,4	165,2	74,3	60,8	319,2	376,2
<b>PPH</b>	383,78	2 174,27	845,3	1 125,28	3 462,6	3 682,33

Zvolte vhodný model, pokud bude třeba tak zjednodušte. Výsledný model zapište. Jaký počet pracovních hodin lze očekávat při spádové populaci velké 100 000 lidí? Udejte bodový i intervalový odhad (95%) pro predikci.

6. V nejmenovaném městě se na třech základních školách v páté třídě měřilo IQ žáků (viz tabulka):

Škola	Dosažené výsledky							
<b>Horní</b>	128	110	98	101	103	112	110	105
<b>Dolní</b>	95	102	92	118	105	107	89	
<b>Kostelní</b>	132	112	109	111	98	118	113	111

Dokazují data odlišnost v IQ mezi školami? Tj. mají žáci alespoň jedné z nich IQ odlišné od žáku ostatních škol? Výsledky interpretujte na hladině významnosti  $\alpha = 0,05$ .





## H.2 Zkouškový test Statistika – 2009



Obrázek 4: Vzorový zkouškový test ze Statistiky – rok 2009

<b>Jméno a příjmení:</b> _____	<b>Obor:</b> _____	<b>Hodnocení:</b> _____
--------------------------------	--------------------	-------------------------

POUČENÍ: Písemná část trvá 60 min. Povolené pomůcky: software STATISTICA komplet 6.1 Cx, MS Excel, Maple, R a kapesní kalkulačtor. Vždy uvádějte podstatné hodnoty, na základě nichž jste se rozhodovali. Z výsledku musí být zřejmý postup. V přílohu práce není povoleno odcházet z místnosti.

**Test STAT**  
*27. ledna 2009*

1. Dvanácti pacientům s vysokým tlakem byl na začátku léčby změřen systolický tlak. Následná léčba se skládala z vhodné diety a pravidelných dávek antihypertensiv. Po šesti měsících byl učiněn nový záznam o systolickém tlaku jednotlivých pacientů. Výsledky byly následující:

<b>Před</b>	150	141	142	145	150	143	143	137	146	147	133	154
<b>Po</b>	114	134	148	144	128	141	139	135	129	140	122	126

Lze tvrdit, že navržená léčba příznivě působí na zdravotní stav pacientů (snižuje krevní tlak)? Napište potřebné hypotézy, uveďte vzorec pro testovací kritérium a jeho hodnotu. Formulujte závěry, ke kterým jste dospěli. Výsledky interpretujte s 95% spolehlivostí.

2. Uveďte klasickou a statistickou definici pravděpodobnosti. Uveďte předpoklady, za kterých lze podle těchto definic vypočítat pravděpodobnost nastání daného jevu. Případně uveďte na příkladě.

3. Jsou dána hodnoty statistického znaku  $X$ :  $x_1 = 1$ ,  $x_2 = 2$ ,  $x_3 = 3$ ,  $x_4 = 5$ ,  $x_5 = a^2$ . Určete číslo  $a$  tak, aby platilo  $\bar{x} = 5,4$ .

© Biskup, Rost 2005-2009

První stránka

4. Zjišťování sklízových ztrát u obilí se provádí výběrovým způsobem. Prvky výběrového souboru jsou plošky o výměře  $1 \text{ m}^2$ , na nichž se sbírají nesklizené klasy i jednotlivá zrna a zjišťuje se jejich přesná hmotnost. Z následujících údajů v gramech (20 hodnot) odhadněte průměr v základním souboru a mez, o které můžeme s 95% pravděpodobností prohlásit, že je průměrné ztráty na  $1 \text{ m}^2$  v základním souboru nepřekračují:

3,2	6,1	8,0	6,5	5,5	5,5	3,2	6,2	8,7	5,6
7,8	5,6	9,5	6,2	8,5	6,3	6,2	6,8	6,6	6,0

5. Zkoušku ze statistiky zatím neabsolvovalo 120 studentů. Z tohoto počtu je 18 studentů oboru BT. Jaká je pravděpodobnost, že z dvaceti náhodně vybraných studentů:

(a) nebude žádný studentem oboru BT,  
(b) bude více jak 18 studentů studenty oboru BT?

6. Definujte *homoskedasticitu* a *heteroskedasticitu*. Jmenujte testy, kterými lze o homoskedasticitě a heteroskedasticitě rozhodnout.

© Biskup, Rost 2005-2009

Druhá stránka

1. Dvanácti pacientům s vysokým tlakem byl na začátku léčby změřen systolický tlak. Následná léčba se skládala z vhodné diety a pravidelných dávek antihypertensiv. Po šesti měsících byl učiněn nový záznam o systolickém tlaku jednotlivých pacientů. Výsledky byly následující:

<b>Před</b>	150	141	142	145	150	143	143	137	146	147	133	154
<b>Po</b>	114	134	148	144	128	141	139	135	129	140	122	126

Lze tvrdit, že navržená léčba příznivě působí na zdravotní stav pacientů (snižuje krevní tlak)? Napište potřebné hypotézy, uveďte vzorec pro testovací kritérium a jeho hodnotu. Formulujte závěry, ke kterým jste dospěli. Výsledky interpretujte s 95% spolehlivostí.

2. Uveďte klasickou a statistickou definici pravděpodobnosti. Uveďte předpoklady, za kterých lze podle těchto definic vypočítat pravděpodobnost nastání daného jevu. Případně uveďte na příkladě.

## H.2 Zkouškový test Statistika – 2009

3. Jsou dána hodnoty statistického znaku  $X$ :  $x_1 = 1$ ,  $x_2 = 2$ ,  $x_3 = 3$ ,  $x_4 = 5$ ,  $x_5 = a^2$ . Určete číslo  $a$  tak, aby platilo  $\bar{x} = 5,4$ .
4. Zjišťování sklizňových ztrát u obilí se provádí výběrovým způsobem. Prvky výběrového souboru jsou plošky o výměře  $1 \text{ m}^2$ , na nichž se sbírají nesklizené klasy i jednotlivá zrna a zjišťuje se jejich přesná hmotnost. Z následujících údajů v gramech (20 hodnot) odhadněte průměr v základním souboru a mez, o které můžeme s 95% pravděpodobností prohlásit, že ji průměrné ztráty na  $1 \text{ m}^2$  v základním souboru nepřekračují:

3,2	6,1	8,0	6,5	5,5	5,5	3,2	6,2	8,7	5,6
7,8	5,6	9,5	6,2	8,5	6,3	6,2	6,8	6,6	6,0

5. Zkoušku ze statistiky zatím neabsolvovalo 120 studentů. Z tohoto počtu je 18 studentů oboru BT. Jaká je pravděpodobnost, že z dvaceti náhodně vybraných studentů:
- (a) nebude žádný studentem oboru BT,
  - (b) bude více jak 18 studentů studenty oboru BT?
6. Definujte *homoskedasticitu* a *heteroskedsticitu*. Jmenujte testy, kterými lze o homoskedasticitě a heteroskedsticitě rozhodnout.

