



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **DIPLOMOVÁ PRÁCE**

Jméno Příjmení

**Název práce**

Název katedry nebo ústavu

Vedoucí diplomové práce: Vedoucí práce

Studijní program: studijní program

Studijní obor: studijní obor

Praha ROK

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Poděkování.

Název práce: Název práce

Autor: Jméno Příjmení

Katedra: Název katedry nebo ústavu

Vedoucí diplomové práce: Vedoucí práce, katedra

Abstrakt: Abstrakt.

Klíčová slova: klíčová slova

Title: Name of thesis

Author: Jméno Příjmení

Department: Name of the department

Supervisor: Vedoucí práce, department

Abstract: Abstract.

Keywords: key words

# Obsah

<b>Úvod</b>	<b>2</b>
<b>1 Teoretický úvod</b>	<b>3</b>
1.1 Výpočetní model . . . . .	3
1.2 Rankově vyvážené stromy . . . . .	4
1.3 Meze optimality . . . . .	4
1.4 Tango stromy . . . . .	4
1.5 Multisplay stromy . . . . .	4
<b>2 Odkazy na literaturu</b>	<b>5</b>
2.1 Několik ukázek . . . . .	5
<b>3 Tabulky, obrázky, programy</b>	<b>6</b>
3.1 Tabulky . . . . .	6
3.2 Obrázky . . . . .	7
3.3 Programy . . . . .	7
<b>4 Formát PDF/A</b>	<b>12</b>
<b>Závěr</b>	<b>13</b>
<b>Seznam použité literatury</b>	<b>14</b>
<b>Seznam obrázků</b>	<b>15</b>
<b>Seznam tabulek</b>	<b>16</b>
<b>Seznam použitých zkratk</b>	<b>17</b>
<b>A Přílohy</b>	<b>18</b>
A.1 První příloha . . . . .	18

# Úvod

Následuje několik ukázkových kapitol, které doporučují, jak by se měla diplomová práce sázet. Primárně popisují použití T<sub>E</sub>Xové šablony, ale obecné rady poslouží dobře i uživatelům jiných systémů.

# 1. Teoretický úvod

V této kapitole zavedeme formální definici binárního vyhledávacího stromu. Poté se podíváme na třídu rankově vyvážených stromů a jejich speciální případ Weak AVL strom. Potom představíme několik mezí optimality, vlastností, které musí mít optimální binární vyhledávací strom. Nakonec se podíváme na Tango stromy a Multisplay stromy a dokážeme, že tyto meze splňují (nebo, v případě tango stromu, že je téměř splňují).

## 1.1 Výpočetní model

Chceme-li mluvit o optimálním stromu, musíme nejprve specifikovat, co přesně budeme za binární vyhledávací strom považovat. Definici, kterou zde představíme, používali implicitně už Sleator a Tarjan (1985), formalizoval ji však až Demaine a kol. (2007).

**Definice 1.** *Mějme statické univerzum klíčů  $\mathcal{U} = \{1, 2, \dots, n\}$ . Dále mějme binární strom nad těmito klíči takový, že pro každé dva vrcholy  $v_1, v_2$  platí, že je-li vrchol  $v_1$  v podstromu vrcholu  $v_2$ , pak je v jeho levém podstromu, právě když je jeho klíč menší než klíč vrcholu  $v_2$ . Nakonec mějme přístupovou posloupnost  $x_1, x_2, \dots, x_m$ , kde  $\forall i \in \mathbb{N}, i \leq m$  platí  $x_i \in \mathcal{U}$ . Pak přístupový algoritmus binárního vyhledávacího stromu je algoritmus, který postupně provede přístupy ke vrcholům s klíči  $x_1, x_2, \dots, x_m$ .*

*Přístup probíhá tak, že algoritmus smí mít vždy právě jeden ukazatel na vrchol stromu, který na počátku každého přístupu ukazuje na kořen stromu. Dále v každém kroku smí provést právě jednu z následujících operací:*

- Přesunout ukazatel na levého syna aktuálního vrcholu,
- přesunout ukazatel na pravého syna aktuálního vrcholu,
- přesunout ukazatel na rodiče aktuálního vrcholu,
- provést rotaci hrany mezi aktuálním vrcholem a jeho rodičem.

*Řekneme, že čas běhu algoritmu je počet těchto operací, které za sekvenci přístupů provede, plus jedna. O vrcholu stromu řekneme, že jsme se ho při daném přístupu dotkli, pokud na něj někdy během tohoto přístupu ukazoval ukazatel algoritmu.*

Takovému přístupovému algoritmu se někdy také říká *offline přístupový algoritmus*. V praxi ale potřebujeme přístupy provádět online.

**Definice 2.** *Online přístupový algoritmus je tekový přístupový algoritmus, jehož rozhodnutí během  $i$ -tého přístupu nijak neovlivňují hodnoty  $x_j$  z přístupové posloupnosti pro  $j > i$ . Na druhou stranu si tento algoritmus smí v každém vrcholu uložit až  $\mathcal{O}(1)$  slov paměti informací (nikoli však ukazatele na vrcholy).*

Všimneme si, že běžné algoritmy binárních vyhledávacích stromů tuto definici splňují – Například červenočerné a AVL stromy potřebují v každém vrcholu jediný bit informace, splay strom se obejde zcela bez dalších informací.

Pro danou přístupovou sekvenci  $X$  existuje přístupový algoritmus, který ji vykoná optimálně, tedy v nejkratším čase ze všech možných algoritmů. Tento počet operací označíme  $\text{OPT}(X)$ . Zde předpokládáme, že je strom na začátku v nejlepší možné konfiguraci. Tím však nesnížíme potřebný čas na přístupy o více než aditivní  $\mathcal{O}(n)$ , protože z libovolného BVS je možné pomocí  $\mathcal{O}(n)$  rotací vytvořit libovolný jiný (nad tou samou množinou klíčů). Proto budeme dále zkoumat pouze přístupové posloupnosti  $X$  takové, že  $|X| \in \Omega(n)$ . Vzhledem k tomu, že nahlédneme, že  $\text{OPT}(x) \geq |X|$ , je tento faktor asymptoticky zanedbatelný.

**Definice 3.** *O přístupovém algoritmu řekneme, že je  $f(n)$ -kompetitivní, pokud každou posloupnost přístupů  $X$  nad univerzem velikosti  $n$  vykoná v čase  $\{f(n) \cdot \text{OPT}(X)\}$ . O online přístupovém algoritmu řekneme, že je dynamicky optimální, pokud je  $\mathcal{O}(1)$ -kompetitivní.*

Předtím, než si popíšeme některé konkrétní stromy, už pouze podotkneme, že existence dynamicky optimálního algoritmu je otevřeným problémem. Na druhou stranu například o splay stromech vyslovili Sleator a Tarjan (1985) hypotézu, že jsou dynamicky optimální.

## 1.2 Rankově vyvážené stromy

## 1.3 Meze optimality

## 1.4 Tango stromy

## 1.5 Multisplay stromy



## 2. Odkazy na literaturu

Odkazy na literaturu vytváříme nejlépe pomocí příkazů `\citet`, `\citep` atp. (viz L<sup>A</sup>T<sub>E</sub>Xový balíček `natbib`) a následného použití BibT<sub>E</sub>Xu. V matematickém textu obvykle odkazujeme stylem „Jméno autora/autorů (rok vydání)“, resp. „Jméno autora/autorů [číslo odkazu]“. V českém/slovenském textu je potřeba se navíc vypořádat s nutností skloňovat jméno autora, respektive přechylovat jméno autorky. Je potřeba mít na paměti, že standardní příkazy `\citet`, `\citep` produkují referenci se jménem autora/autorů v prvním pádě a jména autorek jsou nepřechýlena.

Pokud nepoužíváme bibT<sub>E</sub>X, řídíme se normou ISO 690 a zvyklostmi oboru. Jména časopisů lze uvádět zkráceně, ale pouze v kodifikované podobě.

### 2.1 Několik ukázek

Mezi nejvíce citované statistické články patří práce Kaplana a Meiera a Coxe (Kaplan a Meier, 1958; Cox, 1972). Student (1908) napsal článek o t-testu.

Prof. Anděl je autorem učebnice matematické statistiky (viz Anděl, 1998). Teorii odhadu se věnuje práce Lehmann a Casella (1998). V případě odkazů na specifickou informaci (definice, důkaz, ...) uvedenou v knize bývá užitečné uvést specificky číslo kapitoly, číslo věty atp. obsahující požadovanou informaci, např. viz Anděl (2007, Věta 4.22) nebo (viz Anděl, 2007, Věta 4.22).

Mnoho článků je výsledkem spolupráce celé řady osob. Při odkazování v textu na článek se třemi autory obvykle při prvním výskytu uvedeme plný seznam: Dempster, Laird a Rubin (1977) představili koncept EM algoritmu. Respektive: Koncept EM algoritmu byl představen v práci Dempstera, Lairdové a Rubina (Dempster, Laird a Rubin, 1977). Při každém dalším výskytu již používáme zkrácenou verzi: Dempster a kol. (1977) nabízejí též několik příkladů použití EM algoritmu. Respektive: Několik příkladů použití EM algoritmu lze nalézt též v práci Dempstera a kol. (Dempster a kol., 1977).

U článku s více než třemi autory odkazujeme vždy zkrácenou formou: První výsledky projektu ACCEPT jsou uvedeny v práci Genbergové a kol. (Genberg a kol., 2008). V textu *nenapíšeme*: První výsledky projektu ACCEPT jsou uvedeny v práci Genberg, Kulich, Kawichai, Modiba, Chingono, Kilonzo, Richter, Pettifor, Sweat a Celentano (2008).

## 3. Tabulky, obrázky, programy

Používání tabulek a grafů v odborném textu má některá společná pravidla a některá specifická. Tabulky a grafy neuvádíme přímo do textu, ale umístíme je buď na samostatné stránky nebo na vyhrazené místo v horní nebo dolní části běžných stránek. L<sup>A</sup>T<sub>E</sub>X se o umístění plovoucích grafů a tabulek postará automaticky.

Každý graf a tabulku očíslováme a umístíme pod ně legendu. Legenda má popisovat obsah grafu či tabulky tak podrobně, aby jim čtenář rozuměl bez důkladného studování textu práce.

Na každou tabulku a graf musí být v textu odkaz pomocí jejich čísla. Na příslušném místě textu pak shrneme ty nejdůležitější závěry, které lze z tabulky či grafu učinit. Text by měl být čitelný a srozumitelný i bez prohlížení tabulek a grafů a tabulky a grafy by měly být srozumitelné i bez podrobné četby textu.

Na tabulky a grafy odkazujeme pokud možno nepřímou v průběhu běžného toku textu; místo „*Tabulka 3.1 ukazuje, že muži jsou v průměru o 9,9 kg těžší než ženy*“ raději napíšeme „*Muži jsou o 9,9 kg těžší než ženy (viz Tabulka 3.1)*“.

### 3.1 Tabulky

U **tabulek** se doporučuje dodržovat následující pravidla:

- Vyhýbat se svislým linkám. Silnějšími vodorovnými linkami oddělit tabulku od okolního textu včetně legendy, slabšími vodorovnými linkami oddělovat záhlaví sloupců od těla tabulky a jednotlivé části tabulky mezi sebou. V L<sup>A</sup>T<sub>E</sub>Xu tuto podobu tabulek implementuje balík `booktabs`. Chceme-li výrazněji oddělit některé sloupce od jiných, vložíme mezi ně větší mezeru.
- Neměnit typ, formát a význam obsahu políček v tomtéž sloupci (není dobré do téhož sloupce zapisovat tu průměr, onde procenta).
- Neopakovat tentýž obsah políček mnohokrát za sebou. Máme-li sloupec *Rozptyl*, který v prvních deseti řádcích obsahuje hodnotu 0,5 a v druhých deseti řádcích hodnotu 1,5, pak tento sloupec raději zrušíme a vyřešíme to jinak. Například můžeme tabulku rozdělit na dvě nebo do ní vložit popisné řádky, které informují o nějaké proměnné hodnotě opakující se v následujícím oddíle tabulky (např. „*Rozptyl = 0,5*“ a níže „*Rozptyl = 1,5*“).
- Čísla v tabulce zarovnávat na desetinnou čárku.

Efekt	Odhad	Směrod. chyba <sup>a</sup>	P-hodnota
Abs. člen	−10,01	1,01	—
Pohlaví (muž)	9,89	5,98	0,098
Výška (cm)	0,78	0,12	< 0,001

Pozn: <sup>a</sup> Směrodatná chyba odhadu metodou Monte Carlo.

Tabulka 3.1: Maximálně věrohodné odhady v modelu M.

- V tabulce je někdy potřebné používat zkratky, které se jinde nevyskytují. Tyto zkratky můžeme vysvětlit v legendě nebo v poznámkách pod tabulkou. Poznámky pod tabulkou můžeme využít i k podrobnějšímu vysvětlení významu některých sloupců nebo hodnot.

## 3.2 Obrázky

Několik rad týkajících se obrázků a grafů.

- Graf by měl být vytvořen ve velikosti, v níž bude použit v práci. Zmenšení příliš velkého grafu vede ke špatné čitelnosti popisků.
- Osy grafu musí být řádně popsány ve stejném jazyce, v jakém je psána práce (absenci diakritiky lze tolerovat). Kreslíme-li graf hmotnosti proti výšce, nenecháme na nich popisky **ht** a **wt**, ale osy popíšeme *Výška [cm]* a *Hmotnost [kg]*. Kreslíme-li graf funkce  $h(x)$ , popíšeme osy  $x$  a  $h(x)$ . Každá osa musí mít jasně určenou škálu.
- Chceme-li na dvourozměrném grafu vyznačit velké množství bodů, dáme pozor, aby se neslily do jednolitě černé tmy. Je-li bodů mnoho, zmenšíme velikost symbolu, kterým je vykresluje, anebo vybereme jen malou část bodů, kterou do grafu zaneseme. Grafy, které obsahují tisíce bodů, dělají problémy hlavně v elektronických dokumentech, protože výrazně zvětšují velikost souborů.
- Budeme-li práci tisknout černobíle, vyhneme se používání barev. Čáry rozlišujeme typem (plná, tečkovaná, čerchovaná, ...), plochy dostatečně rozdílnými intenzitami šedé nebo šrafováním. Význam jednotlivých typů čar a ploch vysvětlíme buď v textové legendě ke grafu anebo v grafické legendě, která je přímo součástí obrázku.
- Vyhýbejte se bitmapovým obrázkům o nízkém rozlišení a zejména JPEGům (zuby a kompresní artefakty nevypadají na papíře pěkně). Lepší je vytvářet obrázky vektorově a vložit do textu jako PDF.

## 3.3 Programy

Algoritmy, výpisy programů a popis interakce s programy je vhodné odlišit od ostatního textu. Jednou z možností je použití L<sup>A</sup>T<sub>E</sub>Xového balíčku **fancyvrb** (fancy verbatim), pomocí něhož je v souboru **makra.tex** nadefinováno prostředí **code**. Pomocí něho lze vytvořit např. následující ukázky.

```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```

Menší písmo:

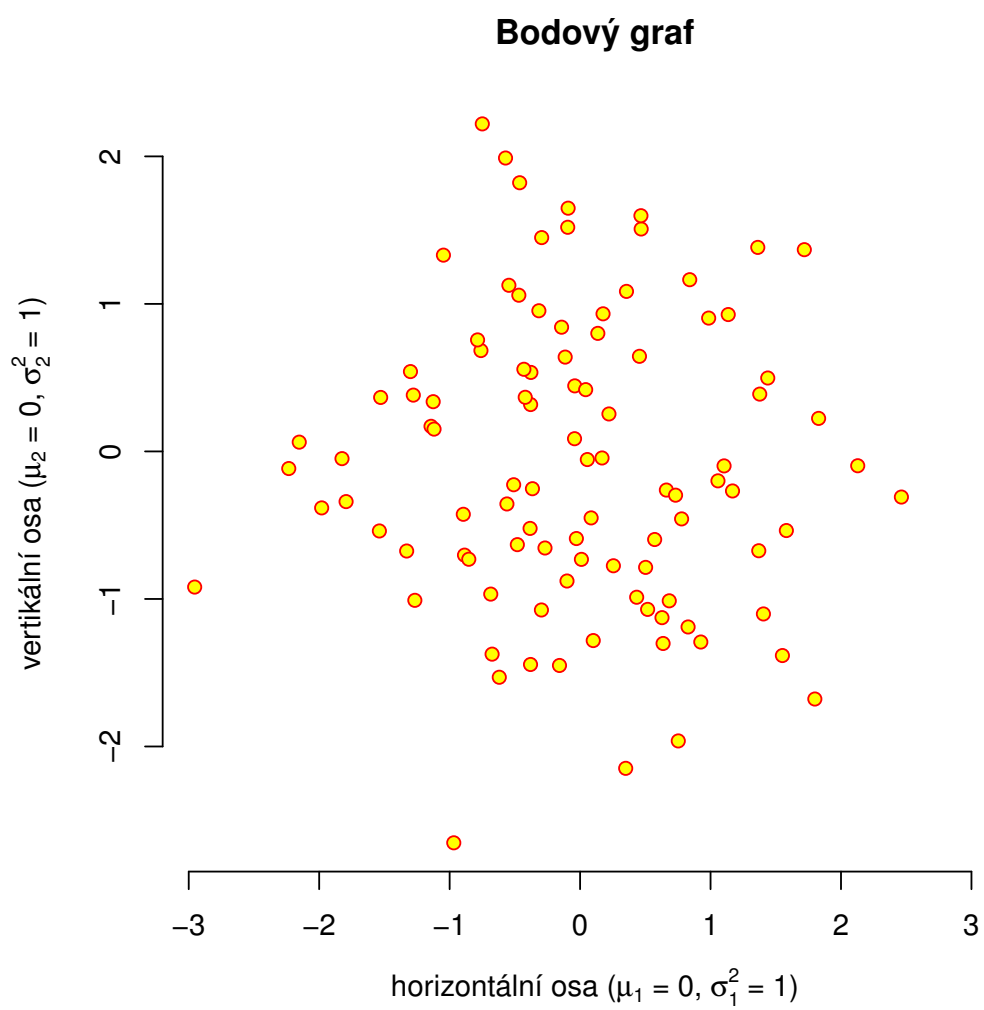
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```

Bez rámečku:

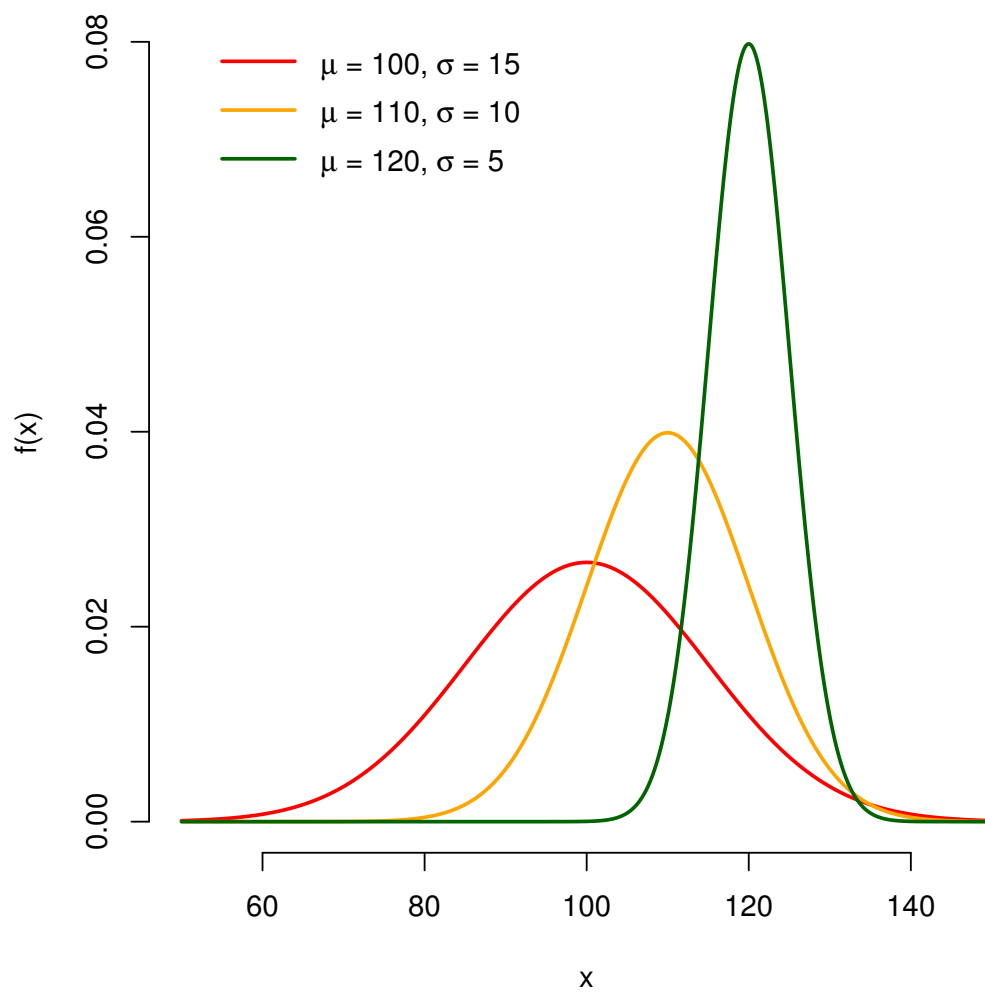
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```

Užší rámeček:

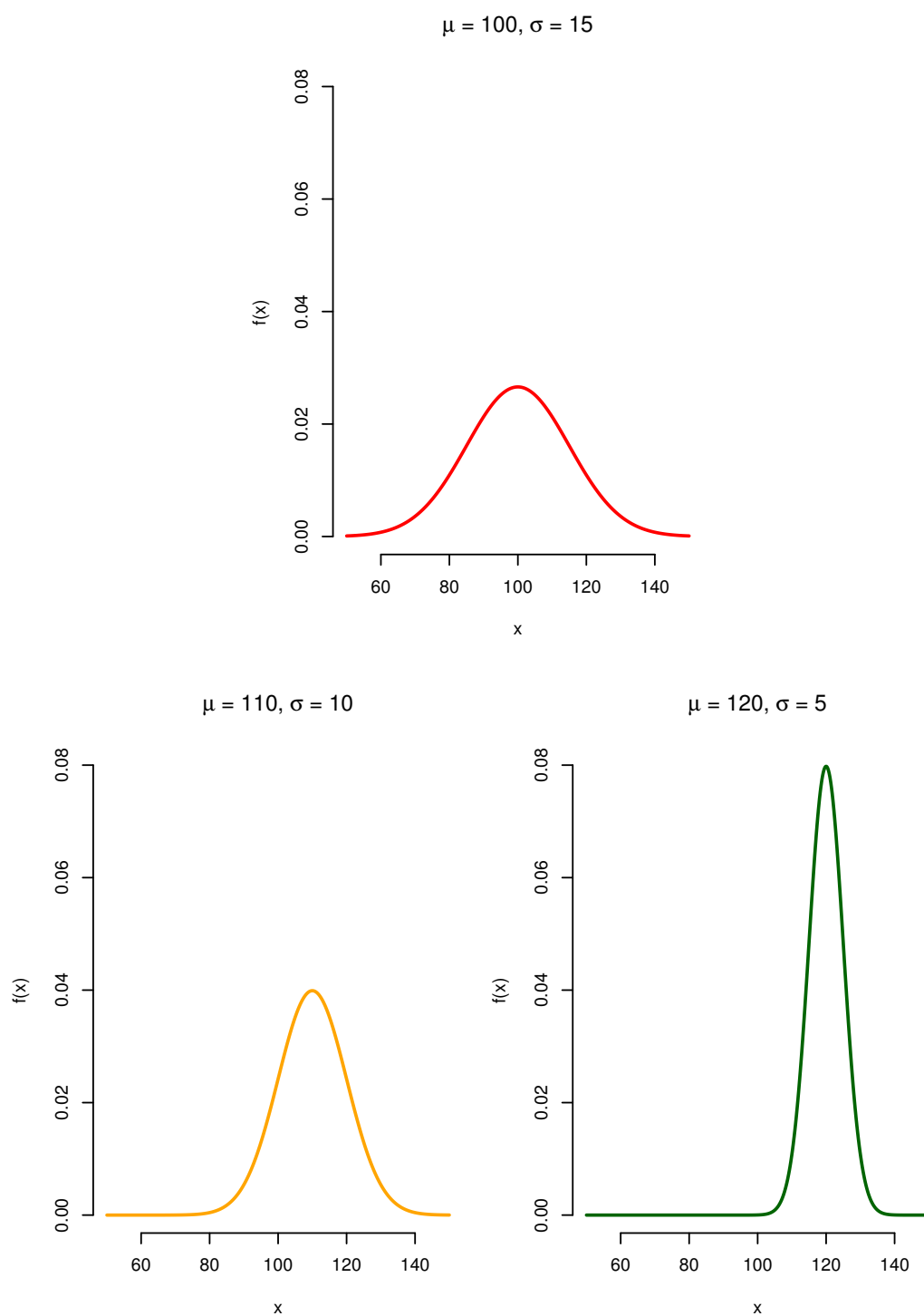
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```



Obrázek 3.1: Náhodný výběr z rozdělení  $\mathcal{N}_2(\mathbf{0}, I)$ .



Obrázek 3.2: Hustoty několika normálních rozdělení.



Obrázek 3.3: Hustoty několika normálních rozdělení.

## 4. Formát PDF/A

Opatření rektora č. 13/2017 určuje, že elektronická podoba závěrečných prací musí být odevzdávána ve formátu PDF/A úrovně 1a nebo 2u. To jsou profily formátu PDF určující, jaké vlastnosti PDF je povoleno používat, aby byly dokumenty vhodné k dlouhodobé archivaci a dalšímu automatickému zpracování. Dále se budeme zabývat úrovní 2u, kterou sázíme  $\text{\LaTeX}$ .

Mezi nejdůležitější požadavky PDF/A-2u patří:

- Všechny fonty musí být zabudovány uvnitř dokumentu. Nejsou přípustné odkazy na externí fonty (ani na „systémové“, jako je Helvetica nebo Times).
- Fonty musí obsahovat tabulku ToUnicode, která definuje převod z kódování znaků použitého uvnitř fontu to Unicode. Díky tomu je možné z dokumentu spolehlivě extrahovat text.
- Dokument musí obsahovat metadata ve formátu XMP a je-li barevný, pak také formální specifikaci barevného prostoru.

Tato šablona používá balíček `pdfx`, který umí  $\text{\LaTeX}$  nastavit tak, aby požadavky PDF/A splňoval. Metadata v XMP se generují automaticky podle informací v souboru `prace.xmpdata` (na vygenerovaný soubor se můžete podívat v `pdfa.xmpi`).

Validitu PDF/A můžete zkontrolovat pomocí nástroje VeraPDF, který je k dispozici na <http://verapdf.org/>.

Pokud soubor nebude validní, mezi obvyklé příčiny patří používání méně obvyklých fontů (které se vkládají pouze v bitmapové podobě a/nebo bez unicodových tabulek) a vkládání obrázků v PDF, které samy o sobě standard PDF/A nesplňují.

Další postřehy o práci s PDF/A najdete na <http://mj.ucw.cz/vyuka/bc/pdfaq.html>.



# Závěr

# Seznam použité literatury

- ANDĚL, J. (1998). *Statistické metody*. Druhé přepracované vydání. Matfyzpress, Praha. ISBN 80-85863-27-8.
- ANDĚL, J. (2007). *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.
- COX, D. R. (1972). Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, **34**(2), 187–220.
- DEMAINE, E. D., HARMON, D., IACONO, J. a P<sup>✓</sup> ATRAȘCU, M. (2007). Dynamic optimality—almost. *SIAM Journal on Computing*, **37**(1), 240–251.
- DEMPSTER, A. P., LAIRD, N. M. a RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**(1), 1–38.
- GENBERG, B. L., KULICH, M., KAWICHAI, S., MODIBA, P., CHINGONO, A., KILONZO, G. P., RICHTER, L., PETTIFOR, A., SWEAT, M. a CELENTANO, D. D. (2008). HIV risk behaviors in sub-Saharan Africa and Northern Thailand: Baseline behavioral data from project Accept. *Journal of Acquired Immune Deficiency Syndrome*, **49**, 309–319.
- KAPLAN, E. L. a MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.
- LEHMANN, E. L. a CASELLA, G. (1998). *Theory of Point Estimation*. Second Edition. Springer-Verlag, New York. ISBN 0-387-98502-6.
- SLEATOR, D. D. a TARJAN, R. E. (1985). Self-adjusting binary search trees. *Journal of the ACM (JACM)*, **32**(3), 652–686.
- STUDENT (1908). On the probable error of the mean. *Biometrika*, **6**, 1–25.

# Seznam obrázků

3.1	Náhodný výběr z rozdělení $\mathcal{N}_2(\mathbf{0}, I)$ . . . . .	9
3.2	Hustoty několika normálních rozdělení. . . . .	10
3.3	Hustoty několika normálních rozdělení. . . . .	11

# Seznam tabulek

3.1	Maximálně věrohodné odhady v modelu M. . . . .	6
-----	--	---

# Seznam použitých zkratek

# A. Přílohy

## A.1 První příloha