

1 \TeX

\TeX is a typesetting program which was designed and written by computer scientist and *Stanford University* professor **Donald Knuth** and first released in 1978. The term now refers to the system of extensions – which includes software programs called \TeX engines, sets of \TeX macros, and packages which provide extra typesetting functionality – built around the original \TeX language. \TeX is a popular means of typesetting complex mathematical formulae; it has been noted as one of the most sophisticated digital typographical systems.

<https://en.wikipedia.org/wiki/TeX>

2 \LaTeX

\LaTeX is a software system for typesetting documents. \LaTeX markup describes the content and layout of the document, as opposed to the formatted text found in WYSIWYG word processors like *Google Docs*, *LibreOffice Writer*, and *Microsoft Word*. The writer uses markup tagging conventions to define the general structure of a document, to stylize text throughout a document (such as bold and italics), and to add citations and cross-references. A \LaTeX distribution such as $\TeX Live$ or *MiKTeX* is used to produce an output file (such as PDF or DVI) suitable for printing or digital distribution.

\LaTeX was originally written in the early 1980s by **Leslie Lamport** at *SRI International*. The current version is $\LaTeX 2\epsilon$, first released in 1994 but incrementally updated starting in 2015. This update policy replaced earlier plans for a separate release of $\LaTeX 3$, which had been in development since 1989.

\LaTeX was created in the early 1980s by Leslie Lamport when he was working at *Stanford Research Institute* (SRI). He needed to write \TeX macros for his own use and thought that with a little extra effort, he could make a general package usable by others. Peter Gordon, an editor at *Addison-Wesley*, convinced him to write a \LaTeX user's manual for publication (Lamport was initially skeptical that anyone would pay money for it); it came out in 1986 and sold hundreds of thousands of copies. Meanwhile, Lamport released versions of his \LaTeX macros in 1984 and 1985. On 21 August 1989, at a \TeX Users Group (*TUG*) meeting at Stanford, Lamport agreed to turn over maintenance and development of \LaTeX to **Frank Mittelbach**. Frank Mittelbach, along with Chris Rowley and Rainer Schöpf, formed the $\LaTeX 3$ team; in 1994, they released $\LaTeX 2\epsilon$, the current standard version. $\LaTeX 3$ has since been cancelled with features intended for that version being back-ported to $\LaTeX 2\epsilon$ since 2018.

<https://en.wikipedia.org/wiki/LaTeX>

3 Náповěda k sazбě bakalářské práce

Aktuální verzi šablony pro sazbu bakalářské práce najdete v gitovém repozitáři Martina Mareše (<https://gitlab.mff.cuni.cz/teaching/thesis-templates/thesis-cs>). Také se může hodit prohlédnout si další návody udržované stejným autorem (<https://mj.ucw.cz/vyuka/bc/>).

3.1 Úprava práce

Vlastní text práce je uspořádaný hierarchicky do kapitol a podkapitol, každá kapitola začíná na nové straně. Text je zarovnán do bloku. Nový odstavec se obvykle odděluje malou vertikální mezerou a odsazením prvního řádku. Grafická úprava má být v celém textu jednotná.

Práce se tiskne na bílý papír formátu A4. Okraje musí ponechat dost místa na vazbu: doporučen je horní, dolní a pravý okraj 25 mm, levý okraj 40 mm. Číslojí se všechny strany kromě obálky a informačních stran na začátku práce; první číslovaná strana bývá obvykle ta s obsahem.

Písmo se doporučuje dvanáctibodové (12 pt) se standardní vzdáleností mezi řádky (pokud píšete ve Wordu nebo podobném programu, odpovídá tomu řádkování 1,5; v \TeX u není potřeba nic přepínat).

Primárně je doporučován jednostranný tisk (příliš tenkou práci lze obtížně svázat). Delší práce je lepší tisknout oboustranně a přizpůsobit tomu velikosti okrajů: 40 mm má vždy *vnitřní* okraj. Rub titulního listu zůstává nepotíštěný.

Zkratky použité v textu musí být vysvětleny vždy u prvního výskytu zkratky (v závorce nebo v poznámce pod čarou, jde-li o složitější vysvětlení pojmu či zkratky). Pokud je zkratek více, připojuje se seznam použitých zkratek, včetně jejich vysvětlení a/nebo odkazů na definici.

Delší převzatý text jiného autora je nutné vymežit uvozovkami nebo jinak vyznačit a řádně citovat.

3.2 Jednoduché příklady

K různým účelům se hodí různé typy písma. Pro běžný text používáme vzpřímené patkové písmo. Chceme-li nějaký pojem zvýraznit (třeba v okamžiku definice), používáme obvykle *kurzívu* nebo **tučné písmo**. Text matematických vět se obvykle tiskne pro zdůraznění *skloněným (slanted)* písmem; není-li k dispozici, může být zastoupeno *kurzívou*. Text, který je chápán doslova (například ukázky programů) píšeme psacím strojem. Důležité je být ve volbě písma konzistentní napříč celou prací.

Čísla v českém textu obvykle sázíme v matematickém režimu s desetinnou čárkou: $\pi \doteq 3,141\,592\,653\,589$. V matematických textech je často lepší používat desetinnou tečku (pro lepší odlišení od čárky v roli oddělovače). Nestřídejte však obojí. Numerické výsledky se uvádějí s přiměřeným počtem desetinných míst.

Mezi číslo a jednotku patří úzká mezera: šířka stránky A4 činí 210 mm, což si pamatuje pouze 5 % autorů. Pokud ale údaj slouží jako přívlástek, mezera vynecháváme: 25mm okraj, 95% interval spolehlivosti.

Rozlišujeme různé druhy pomlček: červeno-černý (krátká pomlčka), strana 16–22 (střední), 45 – 44 (matematické minus), a toto je — jak se asi dalo čekat — vložená věta ohraničená dlouhými pomlčkami.

V českém textu se používají „české“ uvozovky, nikoliv „anglické“.

Na některých místech je potřeba zabránit lámání řádku (v \TeX u značíme vlnovkou): u předložek (neslabičných, nebo obecně jednopísmenných), vrcholů, před k kroky, a proto, ... obecně kdekoliv, kde by při rozlomení čtenář „šokbrtnul“.

3.3 Matematické vzorce a výrazy

Proměnné sázíme kurzívou (to \TeX v matematickém módu dělá sám, ale nezapomínejte na to v okolním textu a také si matematický mód zapněte). Názvy funkcí sázíme vzpřímeně. Tedy například: $\text{var}(X) = \mathbb{E} X^2 - (\mathbb{E} X)^2$.

Zlomky uvnitř odstavce (třeba $\frac{5}{7}$ nebo $\frac{x+y}{2}$) mohou být příliš stísněné, takže je lepší sázet jednoduché zlomky s lomítkem: $5/7$, $(x + y)/2$.

Není předepsáno, jakým písmem označovat jednotlivé druhy matematických objektů (matice, vektory, atd.), ale značení pro tentýž druh objektu musí být v celé práci používáno stejně. Podobně používáte-li více různých typů závorek, je třeba dělat to v celé práci konzistentně.

Nechť

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}.$$

Povšimněme si tečky za maticí. Byť je matematický text vysázen ve specifickém prostředí, stále je gramaticky součástí věty a tudíž je zapotřebí neopomenout patřičná interpunkční znaménka. Obecně nechceme sázet vzorce jeden za druhým a raději je propojíme textem. **Pozor na prázdný řádek za vzorcem. Způsobí vytvoření nového odstavce, což často není žádoucí.**

Výrazy, na které chceme později odkazovat, je vhodné očíslovat:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}. \tag{1}$$

Výraz (1) definuje matici \mathbf{X} . Pro lepší čitelnost a přehlednost textu je vhodné číslovat pouze ty výrazy, na které se autor někde v další části textu odkazuje. To jest, nečíslojte automaticky všechny výrazy vysázené některým z matematických prostředí.

Zarovnání vzorců do několika sloupečků:

$$\begin{aligned} S(t) &= P(T > t), & t > 0 & \quad (\text{zprava spojitá}), \\ F(t) &= P(T \leq t), & t > 0 & \quad (\text{zprava spojitá}). \end{aligned}$$

Dva vzorce se spojovníkem:

$$\left. \begin{aligned} S(t) &= P(T > t) \\ F(t) &= P(T \leq t) \end{aligned} \right\} t > 0 \quad (\text{zprava spojité}). \tag{2}$$

Dva centrované nečíslované vzorce:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_1^\top \\ \vdots & \vdots \\ 1 & \mathbf{x}_n^\top \end{pmatrix}.$$

Dva centrované číslované vzorce:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3}$$

$$\mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_1^\top \\ \vdots & \vdots \\ 1 & \mathbf{x}_n^\top \end{pmatrix}. \tag{4}$$

Definice rozdělená na dva případy:

$$P_{r-j} = \begin{cases} 0, & \text{je-li } r - j \text{ liché,} \\ r!(-1)^{(r-j)/2}, & \text{je-li } r - j \text{ sudé.} \end{cases}$$

Všimněte si použití interpunkce v této konstrukci. Čárky a tečky se dávají na místa, kam podle jazykových pravidel patří.

$$\begin{aligned} x &= y_1 - y_2 + y_3 - y_5 + y_8 - \dots = && \text{z (3)} \\ &= y' \circ y^* = && \text{podle (4)} \\ &= y(0)y' && \text{z Axiomu 1.} \end{aligned} \tag{5}$$

Dva zarovnané vzorce nečíslované (povšimněte si větších závorek, aby se do nich vešel vyšší vzorec):

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f_i(y_i; \boldsymbol{\theta}),$$

$$\ell(\boldsymbol{\theta}) = \log\{L(\boldsymbol{\theta})\} = \sum_{i=1}^n \log\{f_i(y_i; \boldsymbol{\theta})\}.$$

Dva zarovnané vzorce, první číslovaný:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f_i(y_i; \boldsymbol{\theta}), \tag{6}$$

$$\ell(\boldsymbol{\theta}) = \log\{L(\boldsymbol{\theta})\} = \sum_{i=1}^n \log\{f_i(y_i; \boldsymbol{\theta})\}.$$

Vzorec na dva řádky, první řádek zarovnaný vlevo, druhý vpravo, nečíslovaný:

$$\begin{aligned} \ell(\mu, \sigma^2) = \log\{L(\mu, \sigma^2)\} &= \sum_{i=1}^n \log\{f_i(y_i; \mu, \sigma^2)\} = \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2. \end{aligned}$$

Vzorec na dva řádky, zarovnaný na =, číslovaný uprostřed:

$$\begin{aligned} \ell(\mu, \sigma^2) = \log\{L(\mu, \sigma^2)\} &= \sum_{i=1}^n \log\{f(y_i; \mu, \sigma^2)\} = \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2. \end{aligned} \tag{7}$$

3.4 Definice, věty, důkazy, ...

Konstrukce typu definice, věta, důkaz, příklad, ... je vhodné odlišit od okolního textu a případně též číslovat s možností použití křížových odkazů. Pro každý typ těchto konstrukcí je vhodné mít v souboru s makry (`makra.tex`), **resp. v hlavičce dokumentu**, nadefinované jedno prostředí, které zajistí jak vizuální odlišení od okolního textu, tak automatické číslování s možností křížově odkazovat.

Definice 1. *Nechť náhodné veličiny X_1, \dots, X_n jsou definovány na témž pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Pak vektor $\mathbf{X} = (X_1, \dots, X_n)^\top$ nazveme náhodným vektorem.*

Definice 2 (náhodný vektor). *Nechť náhodné veličiny X_1, \dots, X_n jsou definovány na témž pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Pak vektor $\mathbf{X} = (X_1, \dots, X_n)^\top$ nazveme náhodným vektorem.*

Definice 1 ukazuje použití prostředí pro sazbu definice bez titulku, definice 2 ukazuje použití prostředí pro sazbu definice s titulkem.

Věta 1. *Náhodný vektor \mathbf{X} je měřitelné zobrazení prostoru (Ω, \mathcal{A}, P) do $(\mathbb{R}_n, \mathcal{B}_n)$.*

Lemma 2 (Anděl (2007), str. 29). *Náhodný vektor \mathbf{X} je měřitelné zobrazení prostoru (Ω, \mathcal{A}, P) do $(\mathbb{R}_n, \mathcal{B}_n)$.*

Důkaz. Jednotlivé kroky důkazu jsou podrobně popsány v práci Anděl (2007, str. 29). □

Věta 1 ukazuje použití prostředí pro sazbu matematické věty bez titulku, lemma 2 ukazuje použití prostředí pro sazbu matematické věty s titulkem. Lemmata byla zavedena v hlavním souboru tak, že sdílejí číslování s větami.

4 Odkazy na literaturu

Při zpracování bibliografie (přehledu použitých zdrojů) se řídíme normou ISO 690 a zvyklostmi oboru. V \LaTeX u nám pomohou balíčky `biblatex`, resp. `biblatex-iso690`. Zdroje definujeme v souboru `Literatura.bib`

a pak se na ně v textu práce odkazujeme pomocí makra `\cite`. Tím vznikne odkaz v textu a odpovídající položka v seznamu literatury.

V matematickém textu obvykle odkazy sázíme ve tvaru „Jméno autora/autorů (rok vydání)“, **případně** „Jméno autora/autorů [číslo odkazu]“. V českém/slovenském textu je potřeba se navíc vypořádat s nutností skloňovat jméno autora, respektive přechylovat jméno autorky. K doplňování jmen se hodí příkazy `\citet`, `\citep` z balíčku `natbib`, ale je třeba mít na paměti, že produkuje referenci se jménem autora/autorů v prvním pádě a jména autorek jsou nepřechýlena.

Jména časopisů lze uvádět zkráceně, ale pouze v kodifikované podobě. **U matematických prací spíše nepodporováno.**

Při citování je třeba se vyhnout neověřitelným, nedohledatelným a nestálým zdrojům. Doporučuje se pokud možno necitovat osobní sdělení, náhodně nalezené webové stránky, poznámky k přednáškám apod. Citování spolehlivých elektronických zdrojů (maji ISSN nebo DOI) a webových stránek oficiálních institucí je zcela v pořádku. Citujeme-li elektronické zdroje, je třeba uvést URL, na němž se zdroj nachází, a datum přístupu ke zdroji.

4.1 Několik ukávek

Porovnejte rozdíl při použití `\citep`/`\citet` a pouhého `\cite`.

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.

Profesor Anděl je autorem učebnice matematické statistiky ([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 \(2007, Věta 4.22\)](#).

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 a kol., 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 \(2008\)](#).

5 Tabulky, obrázky, programy

Používání tabulek a grafů (**obrázků**) 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

Tabulka 1: Maximálně věrohodné odhady v modelu M.

Efekt	Odhad	Směrod. chyba ^a	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: ^a Směrodatná chyba odhadu metodou Monte Carlo.

místo v horní nebo dolní části běžných stránek. \LaTeX se o umístění plovoucích grafů a tabulek postará automaticky.

Každý graf a tabulku očíslováme a umístíme pod ně (**nad 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 nepřímo v průběhu běžného toku textu; místo „*Tabulka 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 1)*“.

5.1 Tabulky

AK má raději tabulky na horním okraji stránky a popis tabulky nad ní.

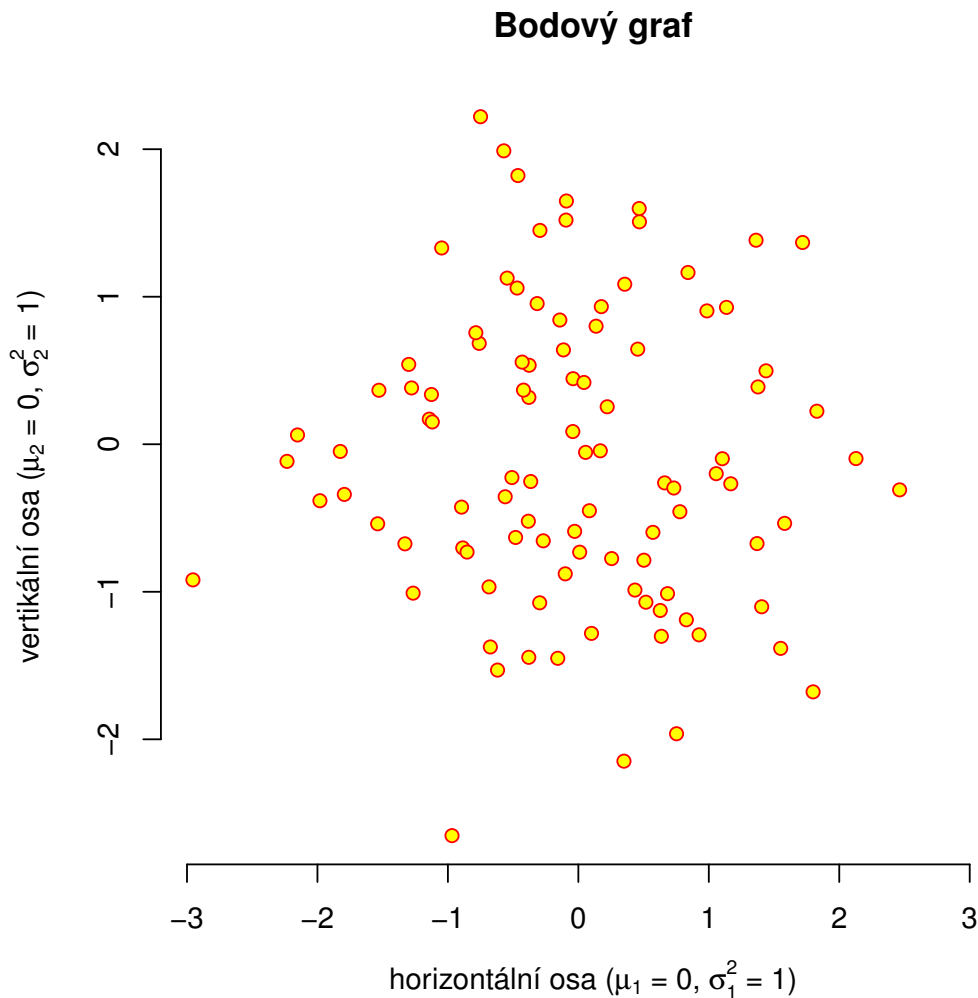
U **tabulek** se doporučuje dodržovat následující pravidla:

- 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 \LaTeX u 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 (tečku).
- 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.

5.2 Obrázky

Dodejme ještě 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 jednoduté 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

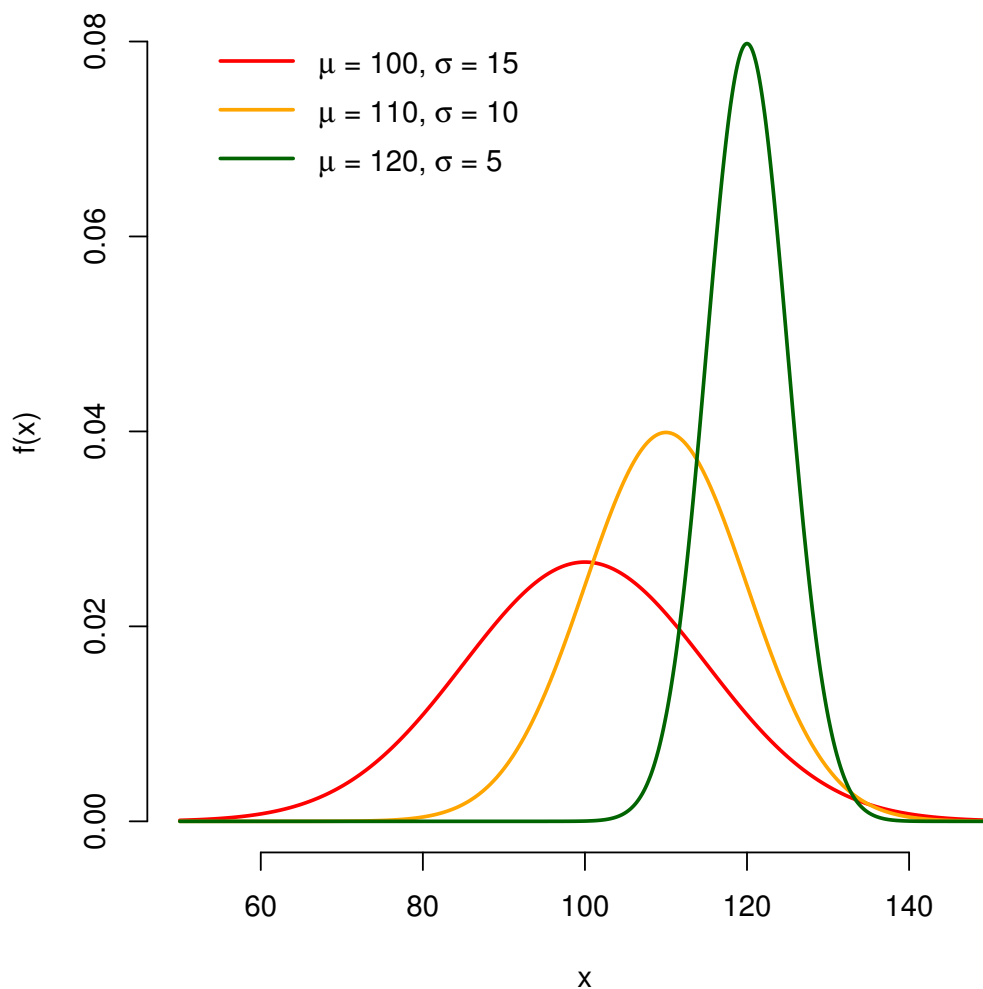


Obrázek 1: Náhodný výběr z rozdělení $\mathcal{N}_2(\mathbf{0}, I)$.

problémy hlavně v elektronických dokumentech, protože výrazně zvětšují velikost souborů.

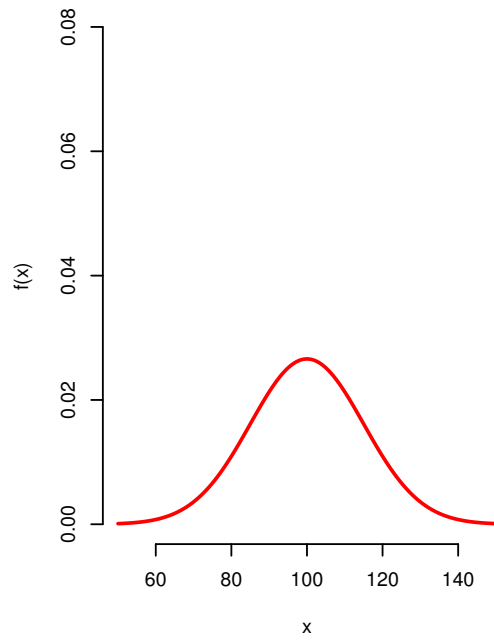
- Budeme-li práci tisknout černobíle, vyhneme se používání barev. Čáry rozlišujeme typem (plná, tečkovaná, čerchovaná, ...), plochy dostatečně rozdílnými intenzitami šedé nebo šrafováním. Význam jednotlivých typů čar a ploch vysvětlíme buď v textové legendě ke grafu anebo v grafické legendě, která je 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.

Co to jsou plovoucí prostředí a jejich umístění v rámci zdrojového souboru. Argumenty `width` a `height` u příkazu `\includegraphics`.

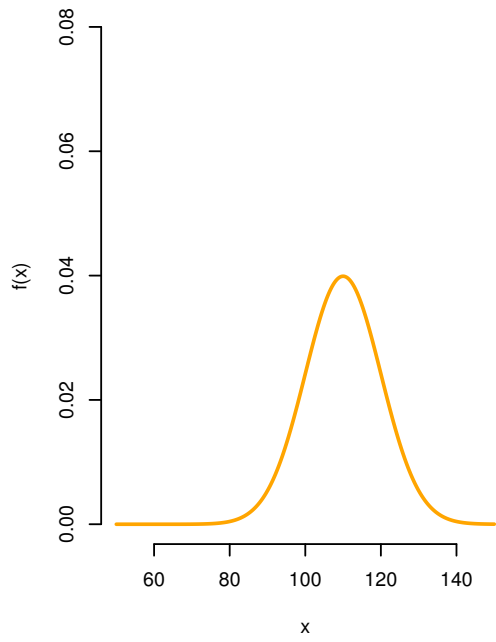


Obrázek 2: Hustoty několika normálních rozdělení.

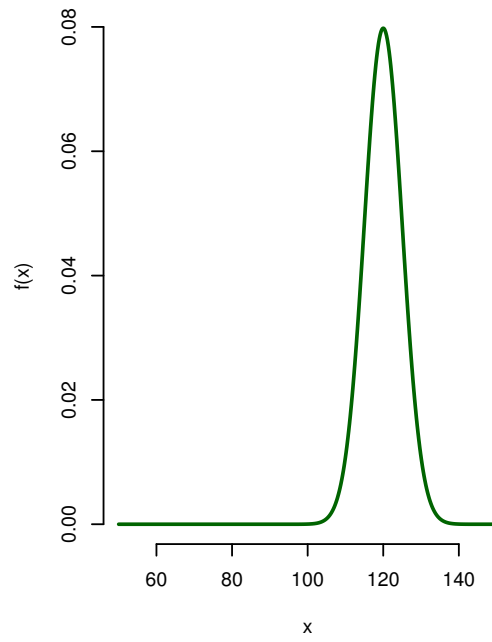
$\mu = 100, \sigma = 15$



$\mu = 110, \sigma = 10$



$\mu = 120, \sigma = 5$



Obrázek 3: Hustoty několika normálních rozdělení.

5.3 Programy

Algoritmy, výpisy programů a popis interakce s programy je vhodné odlišit od ostatního textu. Pro programy se hodí prostředí `lstlisting` z \TeX ového balíčku `listings`, které umí i syntakticky zvýrazňovat běžné programovací jazyky. Většinou ho chceme obalit prostředím `listing`, čímž z něj uděláme plovoucí objekt s popiskem (více viz šablona).

Další možností je použití \TeX ového balíčku `fancyvrb` (fancy verbatim), pomocí něhož je v souboru `makra.tex` nadefinováno prostředí `code`. Pomocí něho lze vytvořit např. následující ukázky.

V základním nastavení dostaneme:

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

Můžeme si říci o menší písmo:

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

Nebo vypnout rámeček:

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

Případně si říci o užší rámeček:

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

6 Formát PDF/A

Opatření rektora č. 13/2017 určuje, že elektronická podoba závěrečných prací musí být odevzdávána ve formátu PDF/A úrovně la 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 \TeX em.

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 do 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.

Šablona pro sazbu bakalářské práce používá balíček pdfx, který umí \LaTeX nastavit tak, aby požadavky PDF/A splňoval. Metadata v XMP se generují automaticky podle informací v souboru práce.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>.

Reference

- 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. doi: 10.1111/j.2517-6161.1972.tb00899.x.
- 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. doi: 10.1111/j.2517-6161.1977.tb01600.x.
- 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. doi: 10.1097/QAI.0b013e3181893ed0.
- KAPLAN, E. L. a MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481. doi: 10.1080/01621459.1958.10501452.
- LEHMANN, E. L. a CASELLA, G. (1998). *Theory of Point Estimation*. Second Edition. Springer-Verlag, New York. ISBN 0-387-98502-6.
- STUDENT (1908). On the probable error of the mean. *Biometrika*, **6**, 1–25. doi: 10.1093/biomet/6.1.1.