



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

ROČNÍKOVÁ PRÁCE

Jméno Příjmení

Závěrečná zpráva z ročníkového projektu

Název katedry nebo ústavu

Vedoucí bakalářské práce: Vedoucí práce

Studijní program: studijní program

Studijní obor: studijní obor

Praha 2024

Obsah

Úvod	3
1 Sada nástrojů	4
1.1 MM-cat	4
1.2 MM-infer	4
1.3 MM-evocat	4
1.4 MM-quecat	4
2 Cílové skupiny uživatelů	5
2.1 Definice cílových skupin	5
2.2 Popis cílových skupin	5
2.2.1 Skupina mladých lidí a dětí	6
2.2.2 Skupina začínajících ve věku 40-50 let	6
2.2.3 Skupina expertních uživatelů	6
3 Tabulky, obrázky, programy	8
3.1 Tabulky	8
3.2 Obrázky	9
3.3 Programy	9
4 Formát PDF/A	14
Závěr	15
Seznam použité literatury	16
Seznam obrázků	17
Seznam tabulek	18
Seznam použitých zkratk	19
A Přílohy	20
A.1 První příloha	20

Úvod

V rámci GAČR projektu Unified Management of Multi-Model Data (č. 20-22276S) byla navržena sada nástrojů pro modelování a správu multi-modelových dat, např. MM-cat a MM-evocat. Nástroje vznikaly postupně a za podpory různých lidí. Díky tomu nemají jednotné rozhraní pro práci. Navíc jsou současná rozhraní často nepřehledná a uživatelsky nepřívětivá. Proto se v rámci ročníkového projektu zabýváme návrhem sjednoceného rozhraní. Důraz je kladen na návrh vhodného uživatelského rozhraní a jeho udržitelnost.

Práce je soustředěna hlavně na nástroj MM-evocat. A to tak, aby bylo možné rozšířit aplikaci o další nástroje. Výsledkem práce je návrh low-fidelity prototypu MM-evocat a jeho otestování na vybraném vzorku reálných uživatelů.

TODO: nahradit mezery za no-break space všude v dokumentu... pomocí ~

1. Sada nástrojů

TODO: Rychlý úvod do všech nástrojů. Rozebrat hlavně mm-evocat, protože ten je základem návrhu.

1.1 MM-cat

KOUPIL, Pavel; SVOBODA, Martin; HOLUBOVÁ, Irena. MM-cat: A tool for modeling and transformation of multi-model data using category theory. In: 2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C). IEEE, 2021. p. 635-639.

1.2 MM-infer

KOUPIL, Pavel; HRICKO, Sebastián; HOLUBOVÁ, Irena. MM-infer: A Tool for Inference of Multi-Model Schemas. In: EDBT. 2022. p. 566-2.

1.3 MM-evocat

KOUPIL, Pavel; BÁRTÍK, Jáchym; HOLUBOVÁ, Irena. MM-evocat: A tool for modelling and evolution management of multi-model data. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022. p. 4892-4896.

1.4 MM-quecat

KOUPIL, Pavel; CRHA, Daniel; HOLUBOVÁ, Irena. MM-quecat: A Tool for Unified Querying of Multi-Model Data. 2023.

2. Cílové skupiny uživatelů

Pro dobré nastavení projektu si definujeme cílové skupiny, které budou naši aplikaci používat. Pokud už k předchozím projektům, které se snažíme sjednotit, cílové skupiny existují, je dobré se s nimi alespoň seznámit.

Jakmile nastavíme potenciální cílové skupiny uživatelů, zvolíme si konkrétní persony. Pod personou si můžeme představit detailní popis fiktivní osoby, která reprezentuje cílovou skupinu. Persony nám pomohou lépe rozebrat ekonomický status a vlastnosti lidí ve skupinách. Nezapomínáme ani na jejich motivace a zájmy. Nesnažíme se vystihnout všechny ze skupiny, ale vybíráme si hlavně stereotypní vlastnosti a chování. Stejně jako kladné rysy a zájmy chceme vystihnout i co dotyčný nemá rád, případně vůči čemu je úplně odmítavý.

Pro práci si definujeme jenom persony, které reprezentují uživatele na které cílíme. Lze ale i definovat vylučující persony. Tedy takové, na které cílit nechceme.

Vytváření cílových uživatelů je důležitým nástrojem, který nás bude provádět v dalších krocích projektu. Díky konkrétním představám nás budou lépe napadat konkrétní řešení a realizace. Snažíme se uzpůsobit návrh personám a vyvolat v nich kladné emoce. Díky tomu získáme konkrétně zaměřený projekt, který nebude tolik odtržený od reálných uživatelů.

2.1 Definice cílových skupin

Protože k předchozím projektům, které se snažíme sjednotit, persony a skupiny neexistují, definovali jsme skupiny nové.

Nejdřív jsme potenciální uživatele rozdělili na začínající a expertní. Jedni očekávají jednoduché koncepty, protože neznají ty složitější. Seznamují se s doménou poprvé a proto bývají mnohem rychleji frustrovaní, či odrazení. Nemají zkušenost ani nadhled, kterým by se učili aplikaci používat rychleji. Tak by se dala v krátkosti vystihnout skupina začínajících. Druzí už mají nějakou zkušenost v oboru, jsou zvyklí na dotazovací jazyky. Nerozumí přímo teorii co stojí za aplikací, ale spíše se opírají o již zmíněnou zkušenost v oboru.

Bylo těžké najít jednotné zástupce skupiny začínajících. I proto jsme se rozhodli skupinu rozdělit na podskupinu mladých lidí a dětí, kteří mají více sklony k hravosti a experimentování při objevování nových konceptů a na podskupinu lidí ve věku 40 až 50 let, kteří se chtějí rekvalifikovat pro práci s daty, ale zatím nemají vhodné zkušenosti.

Nakonec jsme si ještě pohrávali s myšlenkou, že by aplikaci využíval stroj. Jednotlivé úkony by byly prováděny pomocí skriptů. Ačkoli se taková skupina nedá zařadit do skupin potenciálních uživatelů, je potřeba i s takovým využitím počítat. Nicméně ji dál nebudeme rozebírat.

2.2 Popis cílových skupin

S charakterizací skupin nám pomůže představa konkrétních person. Do skupiny mladých začínajících můžeme zařadit bakalářského studenta informatického oboru. Pod starším začínajícím uživatelem si představíme padesátiletou pracov-

nici České pošty, která se chce rekvalifikovat pro práci s daty. Člověk co vystudoval MFF UK a dále pracuje v oboru s daty, třeba učitel Datového inženýrství, bude reprezentovat zástupce expertní skupiny uživatelů.

Za pomoci person rozebereme jednotlivé skupiny potenciálních uživatelů. Každou skupinu krátce charakterizujeme, následně se snažíme vystihnout její chování. Nakonec se zamyslíme nad úkoly, které budou uživatelé v aplikaci dělat.

2.2.1 Skupina mladých lidí a dětí

Je skupinou začínajících laiků, kteří se s aplikací seznamují úplně poprvé. Nemají příliš zkušeností v oboru, neznají dotazovací jazyk. Budou pracovat hlavně s grafickým prostředím, které by jim mělo práci usnadnit a motivovat je.

Chovají se spontánně, rozhodují se impulzivně a experimentují. Z toho co dělají vyzařuje hravost. Tímto přístupem velmi rychle prozkoumají různá zákoutí aplikace a dobře otestují její funkčnost. Učí se velmi rychle, ale stejně tak je dokáže rychle odradit i maličkost. Lépe si představují věci vizuálně a ocení odezvu nebo odměnu za úkol, který provedou. Odměna pro ně může být rychlá zpětná vazba, třeba odezva na úkony v grafickém prostředí. Pro pohyb v prostředí využívají hlavně počítačovou myš. Ocení barevnost, různorodost a zajímavé podání.

V aplikaci budou využívat možnosti dotazování nad daty, jednoduché modelování schematických kategorií, bez přidávání mapování a dalších složitějších operací. Nebude se jich týkat převod konceptuálního znázornění dat do multimodelové databáze a funkce nástroje MM-infer.

2.2.2 Skupina začínajících ve věku 40-50 let

Je druhá skupina začínajících uživatelů. Je to skupina, co se typicky chce rekvalifikovat do práce s daty, ale mají zatím jen nevhodné zkušenosti. Jsou seznámeni se základní prací na počítači, umí používat webový prohlížeč, tabulkový a textový procesor. Aplikace by pro ně měla být ideální na přeučení. Nemusí se učit komplexní technologie a dál budou dělat úkony přes grafické rozhraní.

Narozdíl od mladých začínajících mají metodický přístup chování na webu, přistupují k věci konzervativně. Upřednostňují umírněné podání grafiky. Jsou pro ně důležité popsané konkrétní kroky a argumenty. Nepracují spontánně, dělají zadanou práci. Je potřeba servírovat jim množství práce postupně v dávkách, aby nepřišlo informační zahlcení a odrazení od aplikace. Stejně jako mladí začínající jsou klikací typ, k práci využívají počítačovou myš.

Úkony prováděné v aplikaci jsou podobné skupině mladých začínajících lidí.

2.2.3 Skupina expertních uživatelů

Poslední skupina, která už zná koncepty ve světě databází. Zvyklá na textové dotazovací jazyky, databázové modely. Jsou schopní psát si vlastní skripty a jinak automatizovat a zefektivnit práci.

Nechovají se spontánně a k aplikaci přistupují hlavně metodicky. Dokážou se lépe orientovat v aplikaci, díky zkušenosti z oboru. Jsou spíše tolerantní, nemají přehnané nároky na aplikaci. Vědí, že je složité přecházet na jiný systém, narozdíl od začínajících uživatelů, které odradí i maličkost. Více si potrpí na funkce, které

jim urychlí práci. Myšleno možnost používat skripty a klávesové zkratky. Ocení strážlivé podání a strohá data. Pracuje jak s grafickým rozhraním, tak i textovým.

Expertní uživatel využije více funkcí aplikace. Není omezen jen na tvorbu schematické kategorie. Může přidávat mapování a joby, převádět konceptuální znázornění dat do multimodelové databáze. Stejně tak bude využívat funkce nástroje MM-infer.

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^AT_EX se o umístění plovoucích grafů a tabulek postará automaticky.

Každý graf a tabulku očíslovujeme 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:

- Vyhybat 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^AT_EXu 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 ^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.

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^AT_EXového 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:

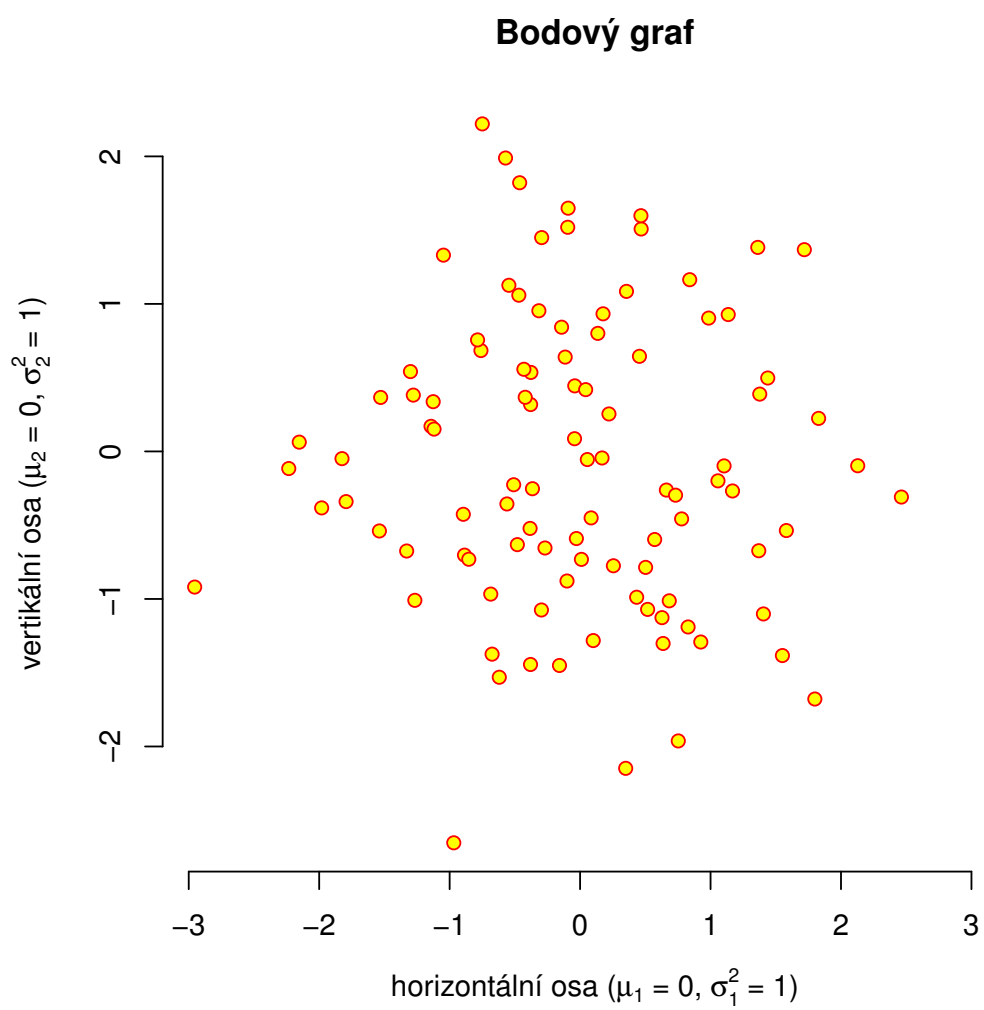
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```

Bez rámečku:

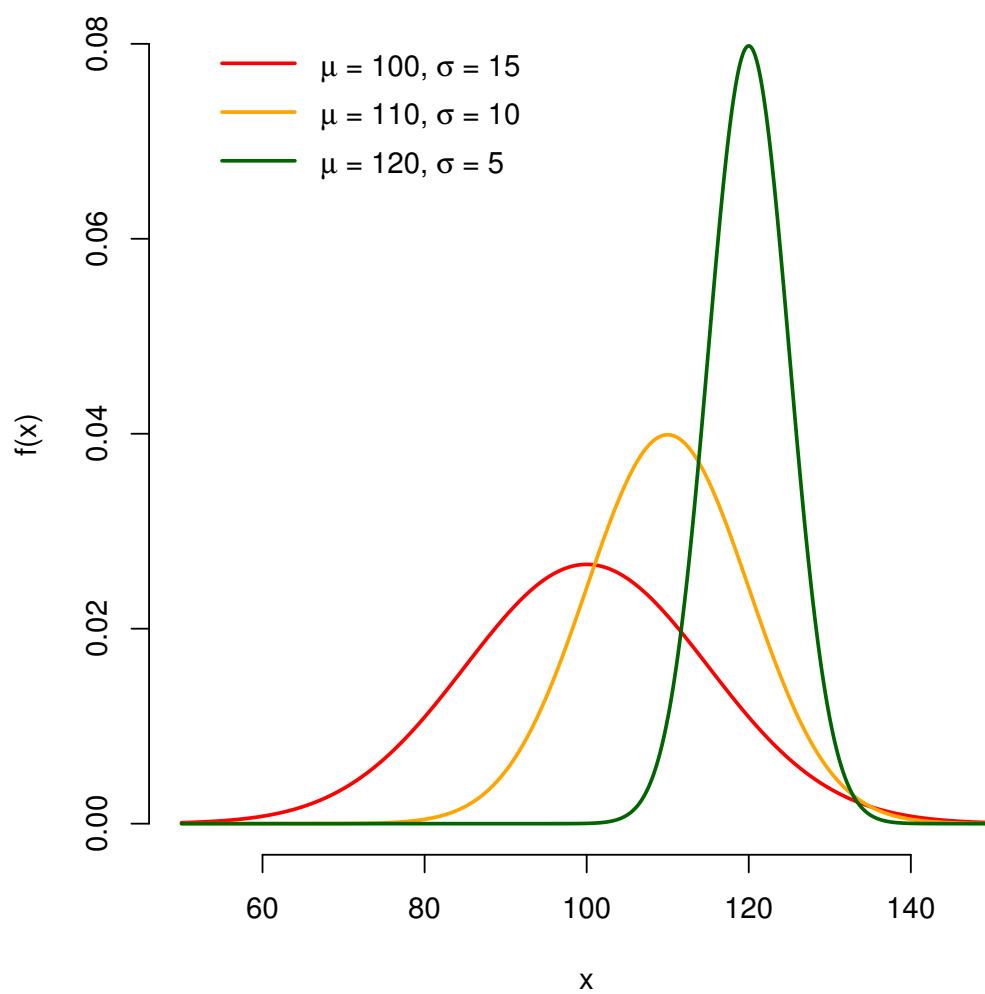
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```

Užší rámeček:

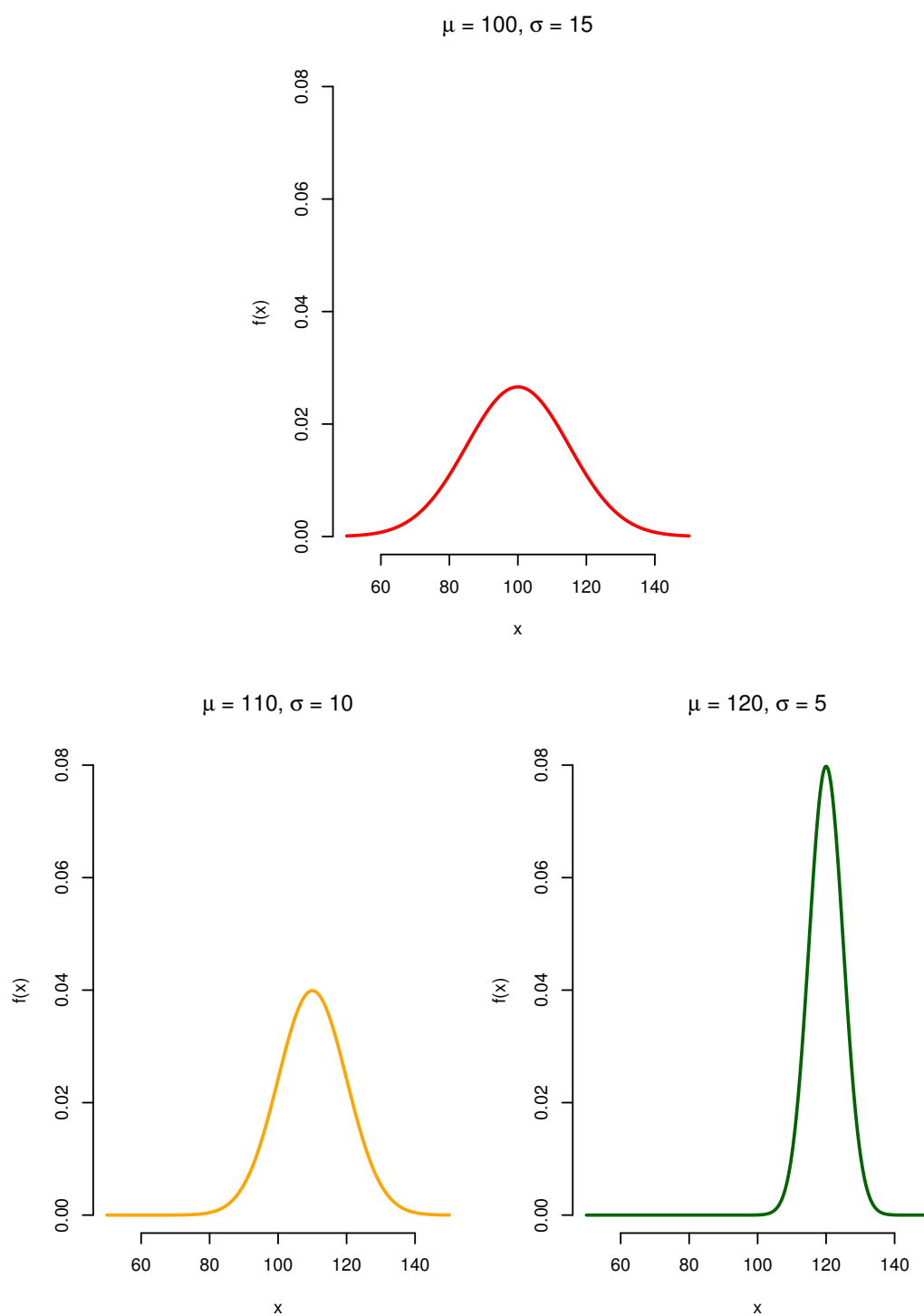
```
> mean(x)
[1] 158.90
> objekt$prumer
[1] 158.90
```



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 \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 \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

Seznam obrázků

3.1	Náhodný výběr z rozdělení $\mathcal{N}_2(\mathbf{0}, I)$	11
3.2	Hustoty několika normálních rozdělení.	12
3.3	Hustoty několika normálních rozdělení.	13

Seznam tabulek

3.1	Maximálně věrohodné odhady v modelu M.	8
-----	--	---

Seznam použitých zkratek

A. Přílohy

A.1 První příloha