

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Sokolová** Jméno: **Nikola** Osobní číslo: **453422**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávací katedra/ústav: **Katedra počítačů**  
Studijní program: **Softwarové inženýrství a technologie**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Datová analýza studentského hodnocení výuky**

Název bakalářské práce anglicky:

**Data analysis of student evaluation of courses**

Pokyny pro vypracování:

- 1) Analyzujte datové zdroje hodnocení vědecko-výzkumné činnosti, studijních výsledků a studentského hodnocení výuky.
- 2) Navrhněte způsob uložení datových zdrojů pro další statistickou analýzu.
- 3) Definujte hypotézy korelace vědeckovýzkumných výsledků vyučujících a jejich hodnocení studenty a korelace hodnocení předmětů a studijních výsledků studentů.
- 4) Ověřte hypotézy vhodnými statistickými metodami, které algoritmizujete.
- 5) Výsledky vyhodnoťte.

Seznam doporučené literatury:

Pokorný J., Valenta M.: Databázové systémy, Nakladatelství ČVUT, Praha, 2013, ISBN 978-80-01-05212-9  
Anděl, J.: Statistické metody. 2. vyd., Matfyzpress, Praha, 1998.  
Chatfield, C.: Statistics for Technology. 3rd ed., Chapman & Hall, London, 1992.  
Zvára, K., Štěpán, J.: Pravděpodobnost a matematická statistika (2. vydání). Matfyzpress, MFF UK, Praha, 2002.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

**Ing. Lukáš Zoubek, Centrum znalostního managementu FEL**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **12.02.2020**

Termín odevzdání bakalářské práce: **22.05.2020**

Platnost zadání bakalářské práce: **30.09.2021**

Ing. Lukáš Zoubek  
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.  
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studentky



České vysoké učení technické v Praze  
Fakulta elektrotechnická  
Katedra počítačů



Bakalářská práce

**Datová analýza studentského hodnocení výuky**

*Nikola Sokolová*

Vedoucí práce: Ing. Lukáš Zoubek

Studijní program: Softwarové inženýrství a technologie, Bakalářský

20. května 2020



## Poděkování

Moja vďaka pri vypracovaní tejto práce patrí najmä dvom ľuďom. Vedúcemu mojej práce Ing. Lukášovi Zoubkovi a Doc. RNDr. Kateřine Helisové, Ph.D. za ich poskytnutú pomoc, ich čas a rady.

V neposlednom rade by som sa chcela poďakovať svojim starým rodičom a mame, vďaka ktorým mám možnosť študovať vysokú školu.



## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 15. 5. 2020

.....





# Abstract

The goal of this thesis was to analyze data from the evaluation of scientific research, study results and student evaluation of teaching and based on the findings define hypotheses of the correlation between these variables. This research focuses only on simple correlation which describes the relationship between two random variables. We defined hypotheses of the dependency between publication activity of a teacher and their cumulative rating, the impact of their teaching workload on their publication activity and finally we studied the dependency between the students' ratings of the teachers and their grade average. To verify our first two hypotheses, we used methods suitable for verifying dependency of random continuous variables, such as visualization of the variables in the form of a scatter plot, covariance values and correlation coefficient values. Based on results of testing for zero correlation we either rejected or failed to reject the hypothesis. The last hypothesis was verified using the test of independence in a contingency table. We have found that the publication results of a teacher positively correlate with their student poll rating as well as with their teaching workload. The existence of dependency of the student's rating of teaching and their grade average was not rejected.

**Keywords:** scientific research, study results, student evaluation of teaching, data analysis, hypothesis, correlation, hypothesis testing

# Abstrakt

Cieľom našej práce bolo zanalyzovať dátové zdroje hodnotenia vedecko-výskumnej činnosti, študijných výsledkov a študentského hodnotenia výuky, a na ich základe definovať hypotézy korelácie vedecko-výskumných výsledkov vyučujúcich, hodnotenia vyučujúcich v ankete a študijných výsledkov študentov. V práci sme sa obmedzili na skúmanie jednoduchej korelácie, ktorá popisuje vzťah medzi dvomi náhodnými veličinami. Stanovili sme hypotézy závislosti publikačnej činnosti vyučujúceho a jeho celkového hodnotenia v ankete, vplyvu pedagogickej vyťaženia vyučujúceho na jeho publikačnú činnosť a nakoniec sme skúmali závislosť študentmi udelených hodnotení v ankete od ich prospechu. Pre overenie prvých dvoch uvedených hypotéz, sme využili metódy, ktoré sú vhodné pre overenie závislosti spojených náhodných veličín a to vizualizáciu skúmaných veličín formou bodového grafu, hodnoty kovariancie, ako aj hodnoty korelačného koeficientu. Hypotézu sme zamietli, alebo nezamietli na základe testu o nulovej korelácií dvoch náhodných veličín. Poslednú hypotézu sme overili testom nezávislosti v kontingenčnej tabuľke. Zistili sme, že publikačné výsledky vyučujúceho kladne korelujú s jeho hodnotením v študentskej ankete, ako aj s jeho pedagogickou vyťaženosťou, a nezamietame existenciu závislosti medzi študentom udeleným hodnotením v študentskej ankete a intervalom jeho študijného priemeru.

**Kľúčové slová:** vedecko-výskumná činnosť, študijné výsledky, študentské hodnotenie výuky, dátová analýza, hypotéza, korelácia, testovanie hypotéz



# Obsah

<b>1</b>	<b>Úvod</b>	<b>1</b>
1.1	Motivácia . . . . .	1
1.2	Ciele práce . . . . .	1
1.3	Štruktúra . . . . .	2
<b>2</b>	<b>Hodnotenie výuky na vysokých školách</b>	<b>3</b>
2.1	Spôsob hodnotenia kvality výuky na vysokých školách . . . . .	3
2.2	Anketa hodnotenia výuky . . . . .	4
2.2.1	Štruktúra Ankety . . . . .	4
2.2.2	Anketné otázky . . . . .	5
2.2.2.1	Typy otázok . . . . .	5
2.2.2.2	Zoznam anketných otázok . . . . .	6
2.2.3	Popis schémy Ankety . . . . .	7
<b>3</b>	<b>Hodnotenie vedecko výskumnej-činnosti na vysokých školách</b>	<b>9</b>
3.1	Spôsob hodnotenia vedecko-výskumnej činnosti na vysokých školách . . . . .	9
3.1.1	Informačný systém výskumu a vývoja . . . . .	9
3.1.1.1	RIV . . . . .	10
3.1.2	Hodnotenie výsledkov výskumu a vývoja . . . . .	10
3.1.2.1	Kritéria hodnotenia . . . . .	11
3.2	Kritéria hodnotenia VVČ na FEL . . . . .	11
3.2.1	Komponenta V3S . . . . .	12
<b>4</b>	<b>Študijné výsledky a vyťaženosť vyučujúcich</b>	<b>13</b>
4.1	Študijný informačný systém KOS . . . . .	13
4.1.1	Študijné výsledky študentov . . . . .	13
4.2	Metodika pre rozdeľovanie finančných prostriedkov . . . . .	13
4.2.1	Vyťaženosť vyučujúcich . . . . .	14
<b>5</b>	<b>Štatistické metódy testovania závislosti javov</b>	<b>15</b>
5.1	Definícia náhodnej veličiny . . . . .	15
5.1.1	Náhodná veličina . . . . .	15
5.1.2	Delenie náhodných veličín . . . . .	16
5.1.3	Charakteristiky náhodných veličín . . . . .	16
5.2	Definícia náhodného výberu . . . . .	17
5.2.1	Rozsah výberu . . . . .	17
5.2.2	Náhodný vektor . . . . .	17
5.2.3	Náhodný výber . . . . .	17
5.2.4	Funkcie náhodného výberu . . . . .	17
5.3	Definícia korelácie . . . . .	17

5.3.1	Združená hustota náhodných veličín . . . . .	18
5.3.2	Nezávislosť náhodných veličín . . . . .	18
5.3.3	Kovariancia . . . . .	18
5.3.3.1	Vlastnosti kovariancie . . . . .	18
5.3.3.2	Interpretácia kovariancie . . . . .	19
5.3.4	Korelácia . . . . .	19
5.3.4.1	Vlastnosti korelácie . . . . .	19
5.3.4.2	Interpretácia korelácie . . . . .	20
5.3.5	Typy korelačných koeficientov . . . . .	20
5.3.5.1	Pearsonov korelačný koeficient . . . . .	20
5.3.5.2	Spearmanov korelačný koeficient . . . . .	20
5.4	Stanovenie hypotézy . . . . .	21
5.4.1	Hypotéza . . . . .	21
5.4.2	Postup testovania hypotéz . . . . .	21
5.4.3	Chyby pri testovaní štatistických hypotéz . . . . .	21
5.4.4	Hladina významnosti testovania štatistických hypotéz . . . . .	21
5.5	Normalita dátového súboru . . . . .	22
5.5.1	Normálne rozdelenie . . . . .	22
5.5.2	Testovanie normality dát . . . . .	23
5.5.2.1	Analytické testy . . . . .	23
5.5.2.2	Grafické testy . . . . .	23
5.6	Metódy testovania hypotéz . . . . .	23
5.6.1	$\beta$ -kvantil pravdepodobnostného rozdelenia . . . . .	23
5.6.2	$\chi^2$ rozdelenie . . . . .	23
5.6.3	Studentovo t rozdelenie . . . . .	24
5.6.4	P-hodnota . . . . .	24
5.6.4.1	Interpretácia p-hodnoty . . . . .	24
5.6.5	Testy nezávislosti . . . . .	24
5.6.5.1	Test hypotézy o nulovej korelácii dvoch náhodných veličín . . . . .	24
5.6.5.2	Test nezávislosti v kontingenčnej tabuľke . . . . .	25
<b>6</b>	<b>Návrh riešenia</b> . . . . .	<b>27</b>
6.1	Návrh databázového modelu . . . . .	27
6.1.1	Popis návrhu schémy . . . . .	27
6.2	Stanovené hypotézy . . . . .	29
6.3	Postup overenia hypotéz . . . . .	29
6.3.1	Vplyv publikačnej činnosti vyučujúceho na jeho hodnotenie v ankete . . . . .	29
6.3.1.1	Postup vyhodnotenia závislosti hodnotenia v ankete a publikačnej činnosti pracovníka . . . . .	30
6.3.2	Závislosť hodnotenia v ankete od prospechu študenta . . . . .	31
6.3.2.1	Postup vyhodnotenia hypotézy o nezávislosti udeleného hodnotenia a priemeru študenta . . . . .	32
6.3.3	Vyťaženosť vyučujúcich a ich vplyv na publikačnú činnosť . . . . .	33
<b>7</b>	<b>Implementácia a testovanie</b> . . . . .	<b>35</b>
7.1	Implementácia . . . . .	35
7.1.1	Databáza . . . . .	35
7.1.1.1	Perzistentná vrstva . . . . .	35
7.1.1.2	Algoritmizácia postupu vyhodnocovania . . . . .	36
7.2	Testovanie správnosti výsledkov . . . . .	36

<b>8</b>	<b>Vyhodnotenie výsledkov</b>	<b>39</b>
8.1	Vyhodnotenie závislosti publikačnej činnosti a celkového hodnotenia vyučujúceho . . .	39
8.2	Vyhodnotenie závislosti hodnotenia v ankete od prospechu študenta . . . . .	42
8.3	Vyhodnotenie závislosti vyťaženia vyučujúcich a ich publikačnej činnosti . . . . .	43
<b>9</b>	<b>Záver</b>	<b>47</b>
<b>A</b>	<b>Obsah přiloženého CD</b>	<b>51</b>



# Seznam obrázků

2.1	SQL schéma využívané v aplikácií Ankety . . . . .	7
6.1	Schéma využívané v rámci práce pre uloženie dátových zdrojov z aplikácií Ankety, V3S a KOMETY. . . . .	28
8.1	Histogram dát publikačnej činnosti vyučujúceho 2016-2018. . . . .	40
8.2	Histogram dát celkového hodnotenia 2016-2018. . . . .	40
8.3	Bodový graf dát publikačnej činnosti a celkového hodnotenia 2016-2018. . . . .	41
8.4	Histogram dát publikačnej činnosti vyučujúceho 2017-2018. . . . .	43
8.5	Histogram dát vyťažnosti vyučujúceho 2017-2018. . . . .	44
8.6	Bodový dát publikačnej činnosti a vyťažnosti vyučujúceho 2017-2018. . . . .	45





# Seznam tabulek

6.1	Skúmané náhodné veličiny vplyvu publikačnej činnosti vyučujúceho na jeho hodnotenie v ankete . . . . .	30
6.2	Skúmané náhodné veličiny závislosť hodnotenia v ankete od prospechu študenta . . . . .	32
6.3	Kontigenčná tabuľka zostavená pre overenie závislostí veličín $X$ a $Y$ . . . . .	32
6.4	Skúmané náhodné veličiny vyťaženosť vyučujúcich a ich vplyvu na publikačnú činnosť . . . . .	33
8.1	Charakteristika dát publikačnej činnosti 2016-2018. . . . .	40
8.2	Charakteristika dát celkového hodnotenia 2016-2018 . . . . .	41
8.3	Charakteristika dát zjednotenia publikačnej činnosti a celkového hodnotenia 2016-2018. . . . .	42
8.4	Kontigenčná tabuľka zostavená pre overenie závislostí veličín Priemer študenta a Udelného hodnotenia. . . . .	42
8.5	Charakteristiky publikačnej činnosti 2017-2018. . . . .	44
8.6	Charakteristiky dát vyťaženosť 2017-2018. . . . .	45
8.7	Zjednotenie dát publikačnej činnosti a vyťaženosť za roky 2017-2018. . . . .	45



# Kapitola 1

## Úvod

V tejto práci sa budeme zaoberať dátovou analýzou študentského hodnotenia výuky, hodnotenia vedecko-výskumnej činnosti a študijných výsledkov študentov na Fakulte Elektrotechnickej ČVUT v Prahe.

Najskôr popíšeme dátové zdroje, ktoré budeme analyzovať v rámci práce a všeobecne popíšeme teoretický podklad pre hodnotenie výuky a vedecko-výskumnej činnosti na vysokých školách. Následne popíšeme základné definície nutné k pochopeniu problematiky testovania hypotéz a stanovíme hypotézy korelácie vedeckovýskumných výsledkov vyučujúcich a ich hodnotenia študentami a korelácie hodnotenia predmetov a študijných výsledkov študentov, ktorých pravdivosť budeme v rámci práce overovať.

### 1.1 Motivácia

Vo vedeckom, kultúrnom, sociálnom a ekonomickom rozvoji spoločnosti hrajú vysoké školy kľúčovú úlohu. Sú to práve tieto školy, ktoré sú nazývané vrcholnými centrami vzdelanosti a tvorivej činnosti. Aby tieto školy mohli poskytovať kvalitné vzdelanie a zároveň sa jej pracovníci mohli venovať výskumnej činnosti, potrebujú finančné prostriedky.

Výsledky hodnotenia vedy a výskumu sú rozhodujúcim kritériom pre pridelovanie finančných prostriedkov jednotlivým vysokým školám a očakáva sa, že tieto školy budú zároveň poskytovať čo najkvalitnejšie vzdelávanie študentov. Väčšina vedecko-výskumných pracovníkov vysokých škôl sa podieľa aj na výuke študentov. Potom je vhodné pokladať si otázky *"Ovplyvňuje ich čas venovaný výskumu samotné vzdelávanie študentov?"*, *"Sú lepší vedci lepšími učiteľmi?"*... a mnohé iné.

Preto je naším zámerom odhaliť, či existujú závislosti medzi vedeckou a pedagogickou činnosťou pracovníkov, ktoré by mohli priniesť nový pohľad na to, ako čo najviac zefektívniť výuku a vedeckú činnosť na vysokých školách.

### 1.2 Ciele práce

Pre vypracovanie práce boli stanovené nasledujúce ciele, ktoré budeme postupne rozpracúvať v jednotlivých kapitolách práce.

- Zanalyzovať dátové zdroje hodnotenia vedecko-výskumnej činnosti, študijných výsledkov a študentského hodnotenia výuky.
- Navrhnuť spôsob uloženia dátových zdrojov pre ďalšiu štatistickú analýzu.

- Definovať hypotézy korelácie vedecko-výskumných výsledkov vyučujúcich a ich hodnotenie predmetov a študijných výsledkov študentov.
- Overiť hypotézy vhodnými štatistickými metódami, ktoré algoritmizujem.
- Vyhodnotiť výsledky práce.

### 1.3 Štruktúra

Bakalárska práca je rozdelená do jednotlivých kapitol, pričom každá popisuje dielčie úlohy tak, aby boli dosiahnuté definované ciele práce. V kapitolách 2, 3, 4 analyzujeme dátové zdroje hodnotenia vedecko-výskumnej činnosti, študijných výsledkov, študentského hodnotenia výuky a obecné popisujeme hodnotenie výuky a vedecko-výskumnej činnosti na vysokých školách.

V rámci kapitoly 5 popisujeme teoretický podklad pre vykonanie štatistickej analýzy dátových zdrojov, na základe ktorých sme vyhodnocovali stanovené hypotézy.

Nasledujúce kapitoly 6 a 7, ktoré popisujú stanovené hypotézy, spôsob riešenia a použité technológie pre vypracovanie práce. Výsledky skúmaných hypotéz a ich vyhodnotenie vyhodnocujeme v kapitole 8.

## Kapitola 2

# Hodnotenie výuky na vysokých školách

Táto kapitola si kladie za cieľ oboznámiť čitateľa so základnými informáciami o priebehu hodnotenia výuky na vysokých školách a následne bude popisovať jeden zo zdrojov dát, ktorým je anketa hodnotenia výuky na ČVUT.

### 2.1 Spôsob hodnotenia kvality výuky na vysokých školách

Vysoké školy hrajú kľúčovú úlohu vo vedeckom, kultúrnom, sociálnom a ekonomickom rozvoji spoločnosti. Keďže sa považujú za centrum vzdelanosti, každá z nich je podľa zákona o vysokých školách<sup>1</sup> povinná vykonávať vnútorné hodnotenie kvality vzdelávacej, tvorivej a s nimi súvisiacich činností.

Tieto činnosti podliehajú pravidelnému hodnoteniu. Hodnotenie vzdelávacej činnosti sa delí nasledujúcim spôsobom.

- **Vnútorné hodnotenie**

Aby bola vysoká škola schopná doložiť informácie o kvalite zaisťovanej vzdelávacej činnosti vypracúva správy o vnútornom hodnotení. Správa popisuje dosiahnuté kvalitatívne výstupy vysokej školy a opatrenia prijaté k odstráneniu nedostatkov. Vypracúva sa pravidelne v termínoch stanovených vnútornými predpismi školy, avšak, najmenej raz za päť rokov s tým, že každý rok je správa aktualizovaná o dodatok popisujúci zmeny dosiahnuté v kvalite a v riadiacich opatreniach. Následne je správa sprístupnená orgánom a členom orgánov vysokej školy, Akreditačnému úradu<sup>2</sup> a ministerstvu.

- **Vonkajšie hodnotenie**

Vonkajšie hodnotenie vysokej školy vykonáva Akreditačný úrad, toto hodnotenie si však môže škola zaistiť tiež na vlastné náklady u všeobecne uznávanej hodnotiacej agentúry. Hodnotenie Akreditačným úradom sa vykonáva na základe podnetu ministra, alebo ak úrad uzná, že existujú závažné dôvody pre mimoriadne hodnotenie.

---

<sup>1</sup>Legislatíva a metodické pokyny pre vysoké školy

<sup>2</sup>Národný akreditačný úrad pre vysoké školstvo

## 2.2 Anketa hodnotenia výuky

Ako sme uviedli v predchádzajúcej sekcii kapitoly, dôležitou spätnou väzbou na vysokých školách je hodnotenie výuky, podľa zákona č. 111/1998 Sb., o vysokých školách. Pre tento účel vznikla na ČVUT aplikácia Anketa, ktorá sa používa na všetkých jej fakultách.

Zmyslom ankety je získanie spätnej väzby pre vyučujúcich a vedenie celej fakulty, ktorí ho používajú ako nástroj pre zvyšovanie kvality pedagogického procesu a zlepšenia prostredia na fakulte. Avšak, nie je hodnotná len pre nich. Študenti môžu vďaka nej využívať skúsenosti svojich kolegov, ktorí daný predmet absolvovali v predchádzajúcich obdobiach, a na základe týchto informácií sa zapísať do paralelky vyučovanej dobre hodnoteným vyučujúcim, alebo zvoliť vhodný spôsob prípravy na predmet.

Anketa sa spúšťa dva razy do roka, po zimnom a letnom semestri, a je dostupná študentom a pracovníkom školy. Každá anketa, vzťahujúca sa ku konkrétnej fakulte ČVUT, prebieha v troch etapách.

### 1. Príprava ankety

V kroku prípravy ankety, fakultný administrátor elektronicky pripraví a realizuje anketu hodnotenia výuky.

### 2. Vyplňanie anketných lístkov

Študenti ČVUT vyplňajú anketné lístky, ktoré majú k dispozícii. Po jeho vyplnení už anketný lístok nevidia.

### 3. Vyjadrenia vyučujúcich k študentskej ankete

V priebehu vyhodnocovania ankety sa spracujú anketné lístky, ktoré vyplnili študenti. Prístup ku študentským komentárom ku svojej osobe a k predmetu, má len vyučujúci, ktorý učil daný predmet. Následne sa každý z vyučujúcich k spomínaným komentárom vyjadrí, a výsledky sú zverejnené študentom a pracovníkom školy.

### 4. Vyjadrenia vedúcich katedier

Ku študentskej ankete sa vyjadria vedúci katedier a zároveň sa zverejnia odpovede na celofakultné otázky.

### 5. Vyjadrenia garantov študijných programov

K výsledkom sa vyjadria garanti študijných programov.

### 6. Vyhodnotenie dekanom

Prebehne fáza vyhodnotenia vyjadrení vedúcich katedier dekanom.

### 7. Zverejnenie výsledkov ankety

Zverejnia sa vyjadrenia vedúcich katedier, garantov programov a dekana.

Výsledky ankiet v jednotlivých semestroch sú dostupné na stránkach ankety<sup>3</sup>.

### 2.2.1 Štruktúra Ankety

V rámci ankety bolo vyhodnotenie spracované formou statického webu, ktorý sa snažil sprehľadniť zobrazenie hodnotenia výuky a vyučujúcich študentmi. Študenti a zamestnanci školy majú možnosť zobrazenia vizualizácie výsledkov hodnotenia učiteľov, predmetov a celofakultného hodnotenia. Predmety hodnotené v rámci ankety bolo možné zobrazit' podľa názvu, katedier a ich poradia podľa hodnotenia. Učiteľov podľa mena, katedier a poradia hodnotenia.

Štruktúra vyhodnotenie ankety<sup>4</sup> pozostáva z nasledujúcich častí.

<sup>3</sup><https://www.fel.cvut.cz/cz/aktuality/2020/anketa-zima-slosovani.html>

<sup>4</sup>Štruktúra študentskej ankety

- **Celkové hodnotenie predmetu**

Celkové hodnotenie predmetu obsahuje počty hodnotiacich a pomer hodnotiacich ku počtu študentov zapísaných na predmet, úspešnosť absolvovaných predmetov, otázky a jednotlivé komentáre k nim a nakoniec vyjadrenia vyučujúcich.

- **Celkové hodnotenie učiteľa**

Hodnotenie učiteľa obsahuje počty hodnotiacich a pomer hodnotiacich ku študentom učiteľa, jeho celkové ohodnotenie, celkový počet hodnotiacich, celkový počet hodnotení.

- **Hodnotenie učiteľa v predmete**

Hodnotenie učiteľa v role prednášajúceho alebo cvičiaceho v predmete obsahuje počty hodnotiacich a pomer hodnotiacich ku študentom učiteľa, jeho celkové ohodnotenie v predmete, počet hodnotení, otázky, komentáre a celkovú udelenú známku študentmi.

- **Poradie učiteľov podľa celkového hodnotenia**

Učители, ktorí mali aspoň 10 hodnotiacich študentov sú ďalej zobrazení v celkovom poradí hodnotenia. To obsahuje počet udelených známok, počet hodnotiacich študentov a ich pomer ku možným hodnotiacim a priemernú známku, ktorá mu bola udelená študentmi.

- **Celofakultné hodnotenie**

Celofakultné hodnotenie pozostáva z odpovedí študentov na celofakultné otázky.

## 2.2.2 Anketné otázky

Anketné otázky študentskej ankety môžu byť dvojakého druhu a tie sa ďalej delia na otázky otvorené a uzavreté. Študenti v rámci ankety majú možnosť vyplniť otázky k hodnoteniu predmetov a k hodnoteniu vyučujúcich. Vyučujúci je vždy hodnotený celkovo a v jednotlivých predmetoch, v ktorých daný semester pôsobil, a to v role prednášajúceho alebo cvičiaceho.

### 2.2.2.1 Typy otázok

Cieľom študentskej ankety je získanie spätnej väzby od študentov s cieľom zvýšenia kvality výuky na ČVUT. Pre tento účel, v rámci ankety, sa získavajú dva základné typy dát, a to dáta kvantitatívne a kvalitatívne. Otázky na kvantitatívne údaje sú formulované tak, aby bolo možné získať ako odpoveď numerické a číselné premenné, zatiaľ čo otázky na kvalitatívne údaje poskytujú kategoriálne, čiže verbálne premenné. Ako príklad kvantitatívnych dát môžeme uviesť počet odpracovaných hodín v práci, alebo výšku mesačného platu a ako príklad kvalitatívnych údajov, napríklad spokojnosť s výberom vysokej školy, alebo úroveň dosiahnutého vzdelania.

Študentská anketa je spracovaná formou dotazníku, ktorý je prístupný študentom a každý z nich má na výber, či sa zúčastní alebo nie. Otázky v dotazníku môžeme deliť podľa toho, či chceme získať kvalitatívne alebo kvantitatívne údaje [9].

- **Otvorené otázky**

V rámci otvorených otázok sa snažíme získať kvalitatívne údaje a respondent v nich vyjadrí vlastný názor. Používané sú najmä pre získavanie spontánnych odpovedí respondentov a v kvalitatívnych prieskumoch.

- **Uzavreté otázky**

Opakom otvorených otázok sú otázky uzavreté. Na výber ponúkajú niekoľko možností, ktoré sa vzájomne vylučujú, a ich výstupom sú kvantitatívne dáta. Pre jednoznačnosť interpretácie odpovedí na daný typ otázok sú využívané hlavne v kvalitatívnych prieskumoch.

Existuje niekoľko typov uzavretých otázok a tými sú otázky výberové, kde respondent si má možnosť na výber z niekoľkých alternatívnych možností odpovede, ďalej otázky ponúkajúce zoznam možných odpovedí, pričom respondent môže vybrať viac ako jednu odpoveď, otázky

poradové, kde respondent môže uviesť poradie odpovedí a nakoniec, typ uzavretých otázok využívaných v študentskej ankete, škálovacie otázky. Respondent na škálovacie otázky odpovedá výberom z množiny ponúknutých výrokov, ktoré sú bodom na predom vytvorenej číselnej stupnici.

- **Polouzavreté otázky**

Polouzavreté otázky ponúkajú predom dané odpovede, a navyše možnosť vlastnej. Jedná sa o kombináciu predchádzajúcich dvoch typov otázok, hoci, uvedený typ otázok sa v študentskej ankete hodnotenia výuky nevyskytuje.

### 2.2.2.2 Zoznam anketných otázok

Ako bolo uvedené v úvode, otázky v študentskej ankete sa delia na otvorené otázky, a na otázky uzavreté. Študenti na otvorené otázky odpovedajú textovým komentárom, relevantným k danej otázke, kde majú možnosť vyjadriť svoj názor a kvantitatívne otázky hodnotia na stanovenej škále.

- **Otázky k hodnoteniu predmetov**

- **Otvorené otázky**

- \* Váš komentár k predmetu (podrobnejší a konkrétnejší komentár čo sa vám na predmete páčilo, čo by šlo zlepšiť, ako ste boli spokojný(á) so študijnými materiálmi a pod.)
- \* Konkrétne pripomienky ku skúškam a skúšajúcim (objektivita, férovosť, a pod.)

- **Uzavreté otázky**

- \* Predmet bol pre mňa prínosom.
- \* Študijné materiály boli kvalitné.
- \* Ako vidíte obtiažnosť predmetu.
- \* Na predmet som sa pripravoval(a).

Prvé dve uzavreté otázky sú hodnotené na škále 1-5, kde 1 = rozhodne áno, 2 = skôr áno, 3 = skôr nie, 4 = rozhodne nie, 5 = neviem sa vyjadriť. Tretia otázka je taktiež hodnotená na škále 1-5, kde 1 = príliš vysoká, 2 = trochu vysoká, 3 = primeraná, 4 = mohlo to byť ťažšie, 5 = bolo to veľmi ľahké a posledná na škále 1-4, kde 1 = pravidelne v priebehu semestra, 2 = občas v priebehu semestra, 3 = v priebehu semestra (skoro) vôbec, prevažne iba v skúškovom období, 4 = (skoro) vôbec.

- **Otázky k hodnoteniu vyučujúcich**

- **Rola prednášajúcich**

- \* **Otvorené otázky**

- Napíšte, v čom by sa podľa vás mohol(a) prednášajúci zlepšiť, čo na prednášajúcom oceňujete a pod.

- \* **Uzavreté otázky**

- Prednášajúci prednášal(a) látku dobre a zrozumiteľne.
- Prednášajúci bol(a) ústretový(a).
- Organizácia predmetu, komunikácia a informovanosť študentov boli dobré.

- **Rola cvičiaci**

- \* **Otvorené otázky**

- Napíšte, v čom by sa podľa vás mohol(a) cvičiaci zlepšiť, čo na cvičiacom oceňujete a pod.



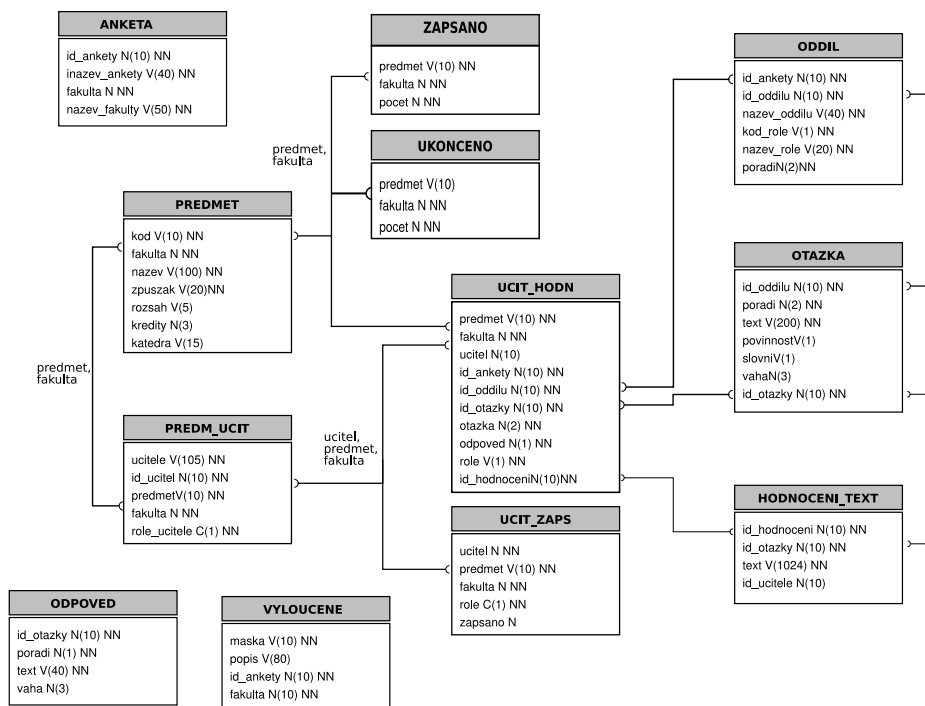
\* **Uzavrené otázky**

- Cvičiaci vedel(a) učiť.
- Cvičiaci bol(a) ústretový(a).
- Cvičiaci bol(a) férový(á)(objektivita, rovný prístup, viz komentár).

Prednášajúcich a cvičiacich hodnotia študenti na škále od 1-5, kde čím menšiu známku vyučujúcemu udelia, tým menej súhlasia s daným výrokom.

### 2.2.3 Popis schémy Ankety

Nasledujúci text obsahuje popis dostupných dát používaných v aplikácii anketa, ktoré sa preberajú z databázy študijného informačného systému KOS.



Obrázek 2.1: SQL schéma využívajúce v aplikácii Ankety

Schéma ankety je tvorená dvanástimi entitami, pričom hlavnou entitou, ktorá identifikuje jednotlivé entity, je *ANKETA*. V rámci každého behu ankety, študenti hodnotia predmety *PREDMET*, ktoré mali zapísané v semestri, v ktorom bola anketa spustená, a vyučujúcich v danom predmete *PREDM\_UCIT*. O tom, koľko študentov malo zapísaných daný predmet a koľko z nich predmet ukončilo sa dozvieme z entít *ZAPSANO* a *UKONCENO*, a to koľko študentov učiteľ v role cvičiaceho alebo prednášajúceho vyučoval, z entity *UCIT\_ZAPS*. Avšak, nie všetky predmety sú v ankete hodnotené. Tie sú reprezentované entitou *VYLOUCENE*, ktorej atribút maska nám hovorí, aké predmety budú na základe kódu predmetu vylúčené z ankety.

V rámci každej ankety študenti zodpovedajú sadu otázok, ktoré sú popísané v časti Anketné otázky 2.2.2. Jednotlivé oddiely otázok určuje entita *ODDIL* a entita *OTAZKA* reprezentuje jednu anketnú otázku. Študenti hodnotia vyučujúcich podľa ich role cvičiaceho alebo prednášajúceho v

danom predmete. Toto hodnotenie predstavuje entita *UCIT\_HODN*. Textové hodnotenie vyučujúceho predstavuje entita *HODNOCENI\_TEXT*. Entita *ODPOVED* reprezentuje odpoveď študenta na anketnú otázku.

## Kapitola 3

# Hodnotenie vedecko výskumnej-činnosti na vysokých školách

Táto kapitola popisuje hodnotenie vedecko-výskumnej činnosti na vysokých školách a predstavuje metodiku hodnotenia vedecko-výskumnej činnosti na ČVUT FEL, skrátene VVČ. Tiež veľmi stručne popisuje komponentu V3S<sup>1</sup>, ktorá obsahuje dáta o výsledkoch hodnotenia, ktoré používame pre ďalšiu analýzu.

### 3.1 Spôsob hodnotenia vedecko-výskumnej činnosti na vysokých školách

V posledných storočiach si ľudia uvedomili, že investícia do vedy zvyšuje ľudský potenciál, čo sa mnohokrát potvrdilo v praxi. Bez toho aby sme investovali financie do vedy a techniky, by ľudia nepristáli na mesiaci a nemali by napríklad počítače a mobilné telefóny [4].

Za vrcholné centrá vzdelanosti, nezávislého poznania a tvorivej činnosti sú považované práve vysoké školy. V predchádzajúcej kapitole sme popísali Metódy hodnotenia kvality výuky na vysokých školách 2. Avšak, preto aby sme hodnotili kvalitu vysokých škôl a vzdelávacích inštitúcií musíme podľa zákona č. 111/1998 Sb. merať a vyhodnocovať aj vedu a výskum. Výsledky hodnotenia vedy a výskumu sú totiž rozhodujúcim kritériom pre rozdeľovanie disponibilných finančných prostriedkov.

#### 3.1.1 Informačný systém výskumu a vývoja

Zo zákona plynie, že každá vysoká škola je povinná vykonávať ako vnútorné hodnotenie kvality vzdelávacej, tak aj tvorivej činnosti, a to znamená nezávislé a transparentné hodnotenie všetkých aktivít výskumu a vývoja. Podľa uznesenia vlády Českej republiky k hodnoteniu výskumu a vývoja a jeho výsledkov<sup>2</sup> je základným nástrojom k hodnoteniu výskumu a vývoja, podporovaného z verejných prostriedkov, Informačný systém výskumu a vývoja, skrátene IS VaVaI<sup>3</sup>. Ten spravuje Rada pre výskum a vývoj a údaje doň sú poskytované jednotlivými rezortmi.

---

<sup>1</sup>Aplikácia V3S

<sup>2</sup>Uznesenie vlády zo dňa 23. júna 2004 č. 644 k hodnoteniu výskumu a vývoja a jeho výsledkov

<sup>3</sup>Informačný systém IS VaVaI

Do uvedeného informačného systému sa vkladajú výsledky vedecko-výskumnej činnosti a to s cieľom rozlíšenia kvality jednotlivých výsledkov a efektívnosti poskytovania verejnej podpory. Podľa povahy hodnoteného výskumu sú to nasledujúce výsledky.

- články v impaktovaných a vo vybraných recenzovaných vedeckých časopisoch, odborné knihy, či články v odborných knihách alebo recenzovaných zborníkoch v cudzom jazyku, v prípade spoločenských vied aj v jazyku českom
- "patenty", čiže výsledky chránené priemyslovou právnou ochranou a ďalšie, podobne chránené výsledky a to napríklad prihlášky vynálezov, uznané odrody plodín, uznané plemená hospodárskych zvierat a iné
- aplikačné výsledky
- ostatné výstupy, to znamená špecifické aktivity výskumu a vývoja, a to napríklad výskumná správa a podobne

Jednotlivé rezorty tieto informácie odosielajú do RIV, teda Registra informácií o výsledkoch MŠMT ČR<sup>4</sup> a výsledky hodnotenia údajov zhromaždených v registri sa využívajú pre spracovanie analýzy stavu vedeckých a výskumných aktivít v ČR.

### 3.1.1.1 RIV

RIV je súčasťou informačného systému výskumu, experimentálneho vývoja a inovácií, popísaného v predchádzajúcej sekcii, ktorý slúži k zhromažďovaniu výsledkov projektov výskumu, vývoja a výskumných zámerov z verejných prostriedkov podľa zákona č. 130/2002 Sb.. o podpore výskumu a vývoja z verejných prostriedkov. V danom zákone je tiež stanovený obsah RIV, postup pri predávaní, zaradenie a spracovanie poskytovaných údajov, nariadením vlády č. 397/2009Sb., o informačnom systéme výskumu a vývoja, zvláštnymi právnymi predpismi a prevádzkovým rádom IS VaV.

Údaje do Registra informácií o výsledkoch predávajú územné samosprávne celky alebo poskytovatelia účelovej a inštitucionálnej podpory z verejných prostriedkov, ktorými sú napríklad Akadémia vied Českej republiky, Grantová agentúra českej republiky a ústredné orgány štátnej správy. RIV obsahuje údaje o vedecko-výskumnej činnosti uplatnených od roku 1993.<sup>5</sup>

### 3.1.2 Hodnotenie výsledkov výskumu a vývoja

Vedecko-výskumná činnosť v Českej republike sa hodnotí podľa kvantifikovateľných výsledkov. Pre kvantifikáciu vedecko-výskumných aktivít sa používa stupnica v podobe známok. Každá známka predstavuje závažnosť prínosu danej aktivity a odvíja sa od nej postup a náročnosť hodnotenia a tiež úroveň výsledkov hodnotenia. Tie sa ďalej premietajú do rozdeľovania výdavkov na výskum a vývoj a ako jedno z hlavných kritérií dokladania kvality pracovísk pri hodnotení návrhov projektov, výskumných zámerov a iných aktivít spojených s vedecko-výskumnou činnosťou.

Výsledky hodnotenia, odoslané do RIV, vyhodnocuje hodnotiaca komisia, ktorej hlavnými úlohami sú tieto výsledky používať pri hodnotení nových návrhov projektov, výskumných zámerov a ďalších aktivít výskumu a vývoja a následne tieto výsledky vnieť do IS VaV. Aby bola dosiahnutá efektívnosť prostriedkov vynaložených z verejných zdrojov, sa vstupy do RIV hodnotia súhrnne. Všetky výskumné inštitúcie, ktoré tieto prostriedky využívajú, podliehajú pravidelnému hodnoteniu, pričom hodnotenie vedecko-výskumných pracovníkov je vlastnou záležitosťou inštitúcií.

---

<sup>4</sup>Ministerstvo školstva, mládeže a telovýchovy

<sup>5</sup>Register informácií o výsledkoch a Informácie o predávaní údajov

### 3.1.2.1 Kritéria hodnotenia

Najdôležitejším kritériom hodnotenia vedecko-výskumnej činnosti je kvalita výskumnej práce. Okrem nej sa však hodnotí aj zapojenie sa do medzinárodných výskumov a tímov. Vedľa publikačnej aktivity, ktoré sa hodnotia podľa renomé časopisov a vydavateľstiev, sú tiež významným kritériom hodnotenia aplikačné výstupy, napríklad patenty. Je dôležité, aby kritéria hodnotenia výsledkov boli jasné, kvantifikovateľné a predom dané. V nasledujúcom texte uvedieme niektoré kritéria pre hodnotenie vedecko-výskumných aktivít v jednotlivých oblastiach.

- **Technika a inžinierstvo**

Pre techniku a inžinierstvo sú hodnotiacimi kritériami aplikované výsledky, patenty, licencie, nové technologické produkty, články v impaktovaných a vo vybraných vedeckých časopisoch, odborné knihy či články v odborných knihách, alebo recenzovaných zborníkoch v cudzom jazyku.

- **Živá príroda**

Vedy ako biológia, medicína, poľnohospodárstvo a životné prostredie sú hodnotené na základe nových liečebných postupov, uznaných plemien hospodárskych zvierat alebo odrôd rastlín, článkoch v impaktovaných a vo vybraných recenzovaných vedeckých časopisoch, odborných knihách, článkoch v odborných knihách alebo recenzovaných zborníkoch v cudzom jazyku, aplikovaných výsledkov, patentov, licencií a nových technologických produktov.

- **Spoločenské vedy**

Pod pojmom spoločenské vedy rozumieme napríklad humanitné, sociálne, ekonomické, právne a historické vedy, ktoré sú hodnotené podľa článkov v impaktovaných a vo vybraných recenzovaných vedeckých časopisoch, odborných knihách, článkoch v odborných knihách a aplikovaných výsledkov.

- **Matematické vedy**

Matematika, kybernetika, informatika, chemické vedy, fyzika a jej aplikácia sa hodnotia podľa článkov v impaktovaných a vo vybraných recenzovaných vedeckých časopisoch, odborných kníh, článkoch v odborných knihách, aplikovaných výsledkov, patentov a nových technológií.

## 3.2 Kritéria hodnotenia VVČ na FEL

Veda a výskum na ČVUT FEL sa hodnotí vždy za posledné tri roky, pričom sa časom menia aj kritéria hodnotenia. V rámci každého z nasledujúcich oddielov, je popísaný súbor kritérií, za ktoré sú vedecko-výskumným pracovníkom ČVUT FEL udeľované body za ich výsledky alebo aktivity v danom roku. Kompletná metodika hodnotenia, ktorá bola platná v dobe ich uplatnenia, sa nachádza na stránkach fakulty, a to za roky 2013<sup>6</sup>, 2015<sup>7</sup> a 2016<sup>8</sup>.

- **Oddiel A - Kvalifikácia pracovníkov**

V oddiele kvalifikácia pracovníkov sa hodnotí titul, ktorý vedecko-výskumný pracovník dosiahol.

- **Oddiel B - Publikácie**

V kategórii publikácia sa pracovníkom pridelujú body za ich publikačnú činnosť. Tou môže byť napríklad kapitola v knihe, odborná kniha alebo článok v periodiku.

---

<sup>6</sup>Kritéria hodnotenia VVČ na FEL za rok 2013

<sup>7</sup>Kritéria hodnotenia VVČ na FEL za rok 2015

<sup>8</sup>Kritéria hodnotenia VVČ na FEL za rok 2016

- **Oddiel BX** - Aplikačné výsledky  
V aplikačných výsledkoch sa hodnotí napríklad, či bol vedecko-výskumnému pracovníkovi udeľený národný patent, európsky alebo medzinárodný patent, alebo či vydal preklad knižnej publikácie.
- **Oddiel C** - Externé granty a doplnková činnosť  
V rámci tejto kategórie sa hodnotia granty a doplnková činnosť.
- **Oddiel D** - Doktorandi v DSP na FEL  
U vedecko-výskumných pracovníkov, ktorí sú zároveň doktorandmi, sa v kategórii D hodnotí, či obhájili dizertačnú prácu, či ju odovzdali v štandardnej dobe štúdia a tiež či zložili SDZ.
- **Oddiel E** - Uznanie vedeckou komunitou  
V rámci oddielu E, uznanie vedeckou komunitou, sa hodnotí napríklad, či vedecko-výskumný pracovník bol hosťujúcim profesorom na zahraničnej univerzite, predsedom alebo členom redakčnej rady časopisu, odborného orgánu od úrovne fakulty vyššie, výboru medzinárodnej vedeckej spoločnosti a výboru ostatnej vedeckej spoločnosti.
- **Oddiel Q** - Vybrané výsledky  
Vo vybraných výsledkoch sa hodnotí, či pracovník obdržal medzinárodný licencovaný patent alebo vydal časopisecký článok v impaktovanom periodiku, ktorý sa nachádza v hornej štvrtine časopisov oboru podľa WOS.<sup>9</sup>

### 3.2.1 Komponenta V3S

Evidencia výsledkov vedecko-výskumnej činnosti na ČVUT a ďalšie aktivity vedecko-výskumných pracovníkov školy vo vedeckej komunite, sú umiestnené v komponente V3S. Komponenta taktiež slúži na odosielaní výsledkov do RIV, exporty pre štatistické analýzy a pre ďalšie interné hodnotenie vedecko-výskumnej činnosti. Užívateľmi systému sú doktorandi, zamestnanci školy a emeritní profesori.

---

<sup>9</sup>Web of Science

## Kapitola 4

# Študijné výsledky a vyťaženosť vyučujúcich

V rámci tejto kapitoly budeme popisovať študijný informačný systém KOS, ktorý je používaný v rámci ČVUT. V krátkosti si tiež predstavíme metodiku pre stanovenie pedagogického výkonu katedier, materiálovej náročnosti štúdia a pre rozdeľovanie finančných prostriedkov za výuku na jednotlivé katedry.

Databáza uvedeného informačného systému je dôležitým zdrojom dát o študentoch a predmetoch na ČVUT a výsledky z výpočtov pre rozdeľovanie finančných prostriedkov podľa metodiky KOMETA2<sup>1</sup> budeme ďalej využívať pre určenie celkovej vyťaženosť vyučujúcich na fakulte ČVUT FEL.

### 4.1 Študijný informačný systém KOS

KOS, alebo aj komponenta štúdia, je centrálny informačný systém, ktorý slúži k podpore študijných agend na ČVUT. Prostredníctvom aplikácie KOS-u, umiestnenej na webovej adrese <https://www.kos.cvut.cz>, môžu študenti pristupovať do rozhrania systému, kde následne majú možnosť tvorby rozvrhu, zápisu na skúšky a tiež prezerania ich študijných výsledkov.

#### 4.1.1 Študijné výsledky študentov

Pre účely vypracovania bakalárskej práce nebudeme pracovať priamo s dátami nachádzajúcimi sa v KOS. Dáta o študijných výsledkoch študentov získame z aplikácie študentskej ankety, ktorá je popísaná v kapitole 2, vo forme intervalu ich študijného priemeru.

### 4.2 Metodika pre rozdeľovanie finančných prostriedkov

Ako bolo uvedené v úvode kapitoly, metodika KOMETA2 slúži na ČVUT FEL pre stanovenie kvantitatívneho pedagogického výkonu katedier, materiálovej náročnosti štúdia a pre rozdeľovanie fakultných finančných prostriedkov za výuku, a to za príslušný kalendárny rok.

Samotné výkony vyučujúcich, náročnosť štúdia a finančné prostriedky sa vypočítavajú podľa metodiky v aplikácii KOMETA, ktorá funguje interne v rámci fakulty.

---

<sup>1</sup>KOMETA2 - Metodika pre stanovenie pedagogického výkonu katedier, materiálovej náročnosti štúdia a pre rozdeľovanie finančných prostriedkov za výuku na katedry

### 4.2.1 Vyťaženosť vyučujúcich

Pre účely bakalárskej práce sme vyťaženosť vyučujúcich definovali ako súčet započítateľných hodín za ich pedagogický výkon v jednotlivých predmetoch, ktoré vyučujú. Pedagogický výkon vyučujúceho v predmete je určený na základe pomeru všetkých podielov vyučujúceho na predmete, za prednášky aj cvičenia, a celkového počtu všetkých prednášok a cvičení daného predmetu.

Z databázy aplikácie KOMETA budeme využívať dáta o započítateľných hodinách za pedagogický výkon pre jednotlivé predmety, z ktorých určíme celkovú vyťaženosť vyučujúceho.



## Kapitola 5

# Štatistické metódy testovania závislosti javov

Kľúčovým aspektom štatistickej analýzy je porozumenie vzťahu medzi dvomi a viacerými javmi. To, či spolu javy súvisia alebo nie, môžeme vyhodnotiť pomocou jedného zo základných nástrojov matematickej štatistiky, korelačnou analýzou.

V nasledujúcom texte popisujeme pravidlá pre rozhodovanie sa v štatistických rozhodovacích problémoch a základné definície nutné k pochopeniu metód testovania hypotéz a korelácie.

### 5.1 Definícia náhodnej veličiny

V rámci analýzy dát získaných pre testovanie stanovených hypotéz je prvým krokom určiť typ veličín, ktoré budeme skúmať. V nasledujúcom texte si preto definujeme náhodnú veličinu, jej delenie a popíšeme všeobecné charakteristiky náhodnej veličiny.

#### 5.1.1 Náhodná veličina

Majme pravdepodobnostný priestor  $(\Omega, \mathcal{A}, P)$ , náhodná veličina [2] je reálnou funkciou  $X$  definovanou na pravdepodobnostnom priestore, ktorá zobrazuje výsledky náhodného pokusu do priestoru reálnych čísel  $X : (\Omega, \mathcal{A}) \rightarrow (R, \mathcal{B})$ , tj.

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \quad (5.1)$$

kde  $B$  je ľubovoľná borelovská množina  $B \in \mathcal{B}$ .

Distribučná funkcia náhodnej veličiny [2] je reálna funkcia  $F_X$  premennej  $x \in R$ , ktorá je definovaná nasledovne.

$$F_X(x) = P(X \leq x) = P(\omega : X(\omega) \leq x) \quad (5.2)$$

Distribučná funkcia má nasledujúce vlastnosti.

- Je neklesajúca.
- Je sprava spojitá v bode  $x \in R$
- Platí, že  $\lim_{x \rightarrow \infty} F_X(x) = 1$  a  $\lim_{x \rightarrow -\infty} F_X(x) = 0$

### 5.1.2 Delenie náhodných veličín

Náhodné veličiny rozdeľujeme podľa oboru hodnôt na dva typy.

- **Diskrétné náhodné veličiny**

Obor hodnôt diskretných veličín je konečná alebo spočetná množina  $M$ , napríklad počet členov domácnosti  $M = \{1, 2, \dots\}$ . To znamená, že náhodnú veličinu môžeme nazvať diskretnou [2], ak existuje postupnosť reálnych čísel  $\{x_n\}$  a postupnosť nezáporných čísel  $\{p_n\} = P(X = x_n)$ , kde  $p_n > 0$ . Pre postupnosť čísel potom platí nasledujúci vzťah.

$$\sum_{n=1}^{\infty} p_n = 1 \quad (5.3)$$

Distribučná funkcia diskretnej náhodnej veličiny  $X$  má nasledujúci tvar.

$$F_X(x) = P(X \leq x) = \sum_{\{n: x_n \leq x\}} P(X = x_n) = \sum_{\{n: x_n \leq x\}} p_n \quad (5.4)$$

Pre ľubovoľné čísla  $a, b \in R$ , kde  $a \leq b$  platí nasledujúci vzťah.

$$P(a < X \leq b) = F_X(b) - F_X(a) = \sum_{\{n: a \leq x_n \leq b\}} P(X = x_n) = \sum_{\{n: a \leq x_n \leq b\}} p_n \quad (5.5)$$

- **Spojité náhodné veličiny**

Obor hodnôt spojitých náhodných veličín je uzavretý alebo otvorený interval napríklad hmotnosť čokolády, kde  $M = (0, \infty)$ .

Náhodná veličina je spojitá [2], ak existuje funkcia  $f_X$  taká, že platí nasledujúci vzťah.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt, x \in (-\infty, \infty). \quad (5.6)$$

Funkcia  $f_X$  sa nazýva hustota rozdelenia pravdepodobnosti a má nasledujúce vlastnosti.

- $f_X(x) = \frac{d}{dx} F_X(x)$ , to znamená, že geometricky je distribučná funkcia náhodnej veličiny plochou pod grafom hustoty pravdepodobnosti [8].
- $\int_{-\infty}^{\infty} f_X(x) dx = 1$
- $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx, x \in (-\infty, \infty)$ , kde  $a, b \in R$  a  $a \leq b$ .

### 5.1.3 Charakteristiky náhodných veličín

Pre popis náhodných veličín sa používajú najčastejšie dve charakteristiky, a to stredná hodnota  $EX$  a rozptyl  $var X$ .

Stredná hodnota náhodnej veličiny popisuje oblasť reálnej osi, ktorú má náhodná veličina tendenciu realizovať [8]. Pre obecnú náhodnú veličinu je stredná hodnota nasledujúci integrál, za predpokladu, že integrál existuje [2].

$$EX = \int_{-\infty}^{\infty} x dF_X(x) \quad (5.7)$$

Rozptyl náhodnej veličiny nám hovorí ako veľmi kolísajú hodnoty okolo  $EX$  [8] a vypočítame podľa nasledujúceho vzťahu [2].

$$var X = E(X - EX)^2 \quad (5.8)$$

## 5.2 Definícia náhodného výberu

Aby sme mohli definovať pojem náhodný výber, je vhodné predtým zdefinovať pojmy rozsah výberu a náhodný vektor, ktoré sú nutné pre jeho pochopenie a ďalej si popíšeme základné charakteristiky náhodného výberu.

### 5.2.1 Rozsah výberu

Predstavme si, že máme štatistický súbor dát rozsahu  $N$ , ktoré analyzujeme a z neho vyberieme menší súbor dát rozsahu  $n < N$ , kde  $n \in N$ . Číslo  $n$  označuje rozsah výberu [10], potom súbor, z ktorého sme vybrali náhodné veličiny sa nazýva základný súbor a vybraný súbor je ten, do ktorého vyberáme.

### 5.2.2 Náhodný vektor

Majme pravdepodobnostný priestor  $(\Omega, \mathcal{A}, P)$  a na tomto priestore definované náhodné veličiny  $X_1, \dots, X_n$ . Potom vektor  $\vec{X} = (X_1, \dots, X_n)^T$  nazývame náhodný vektor[2].

### 5.2.3 Náhodný výber

Náhodný vektor  $\vec{X} = (X_1, \dots, X_n)^T$  nezávislých a rovnako rozdelených náhodných veličín  $X_1, \dots, X_n$  nazývame náhodný výber [2], kde  $n$  je rozsah výberu.

### 5.2.4 Funkcie náhodného výberu

Výberový priemer náhodného výberu sa v štatistike používa ako odhad strednej hodnoty a vypočítame podľa nasledujúceho vzťahu [2].

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n (X_i) \quad (5.9)$$

Výberový rozptyl náhodného výberu odráža ako sa pohybujú hodnoty okolo výberového priemeru a vypočítame podľa nasledujúceho vzťahu [2].

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (5.10)$$

A nakoniec výberovú smerodtnú odchýlku podľa nasledujúceho vzťahu [2].

$$S_n = \sqrt{S_n^2} \quad (5.11)$$

## 5.3 Definícia korelácie

Korelácia, nazývaná aj korelačný koeficient, vyjadruje silu vzájomného vzťahu medzi dvomi a viacerými veličinami a teda popisuje vplyv zmeny jednej veličiny na druhú, či opačne. Podľa počtu náhodných veličín rozlišujeme tri typy korelácie.

- **Jednoduchá**

Jednoduchá korelácia popisuje vzťah dvoch veličín.

- **Mnohonásobná**

Mnohonásobná korelácia popisuje vzťahy viac ako dvoch veličín.

- **Parciálna**

Parciálna korelácia popisuje závislosť dvoch veličín vo viacrozmernom štatistickom súbore, a to pri vylúčení vplyvu ostatných znakov na skúmanú závislosť.

V rámci bakalárskej práce sa obmedzíme iba na skúmanie jednoduchej korelácie a predtým, než si zdefinujeme koreláciu a korelačný koeficient, definujeme nezávislosť náhodných veličín, kovarianciu, význam kovariancie a popíšeme spôsob interpretácie kovariancie a korelačného koeficientu.

### 5.3.1 Združená hustota náhodných veličín

Náhodný vektor  $\vec{X} = (X_1, \dots, X_n)^T$  má spojité rozdelenie, ak existuje nezáporná funkcia  $f_X$  n reálnych premenných taká, že platí

$$F_{\vec{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\vec{X}}(t_1, \dots, t_n) dt_1, \dots, dt_n, \quad (5.12)$$

kde funkcia  $f_{\vec{X}}$  je hustota rozdelenia pravdepodobnosti náhodného vektoru  $\vec{X}$ , alebo tiež združená hustota náhodných veličín  $X_1, \dots, X_n$ .

### 5.3.2 Nezávislosť náhodných veličín

Majme náhodné veličiny  $X$  a  $Y$ . Veličiny sú nezávislé, ak združená distribučná funkcia veličín  $X, Y$  je rovná súčinu marginálnych distribučných funkcií  $X$  a  $Y$  [7].

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad (5.13)$$

Definíciu nezávislosti dvoch náhodných veličín môžeme popísať zvlášť pre diskkrétne a pre spojité veličiny.

- Diskkrétne náhodné veličiny sú nezávislé, ak pre ich združené rozdelenie a marginálne rozdelenie platí nasledujúci vzťah.

$$p_{ij} = p_i \cdot p_j, \forall i, j, \quad (5.14)$$

kde  $p_{ij} = P(X = i, Y = j)$ ,  $p_i = P(X = i)$ ,  $p_j = P(Y = j)$ .

- Spojité náhodné veličiny sú nezávislé, ak ich združená hustota je rovná súčinu marginálnych hustôt.

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad (5.15)$$

### 5.3.3 Kovariancia

Závislosť náhodných veličín  $X$  a  $Y$  popisujeme kovarianciou [2], ktorá vyjadruje vzájomný vzťah týchto dvoch veličín. Kovariancia náhodných veličín  $X$  a  $Y$  je hodnota  $cov(X, Y) \in R$  vypočítaná podľa nasledujúceho vzťahu [5].

$$cov(X, Y) = E[(X - EX)(Y - EY)] \quad (5.16)$$

#### 5.3.3.1 Vlastnosti kovariancie

Pre kovarianciu dvoch náhodných veličín  $X$  a  $Y$  platia nasledujúce vlastnosti [5].

- $cov^2(X, Y) \leq varXvarY$
- $cov(X, Y) = EXY - EXEY$
- Ak sú veličiny  $X$  a  $Y$  nezávislé, potom  $cov(X, Y) = 0$ .

### 5.3.3.2 Interpretácia kovariancie

Kovarianciu medzi dvomi náhodnými veličinami  $X$  a  $Y$  interpretujeme podľa výslednej hodnoty nasledovne [5].

- $cov(X, Y) > 0$   
Hodnota kovariancie  $> 0$  znamená, že vyššie hodnoty náhodnej veličiny  $X$  sú zviazané s vyššími hodnotami  $Y$  a analogicky pre nižšie hodnoty veličín  $X$  a  $Y$ .
- $cov(X, Y) < 0$   
Hodnota kovariancie  $< 0$  znamená, že vyššie hodnoty náhodnej veličiny  $X$  sú zviazané s nižšími hodnotami  $Y$  a analogicky pre nižšie hodnoty veličiny  $X$  sú zviazané s vyššími hodnotami  $Y$ .
- $cov(X, Y) = 0$   
Nulová hodnota kovariancie nám nehovorí, či je alebo nie je vzťah medzi hodnotami  $X$  a  $Y$ .

Z výslednej hodnoty kovariancie poznáme, či sú náhodné veličiny  $X$  a  $Y$  závislé, avšak nedokážeme z nej interpretovať silu závislosti týchto veličín, ktorú zistíme z hodnoty korelácie.

### 5.3.4 Korelácia

Korelácia [6]  $cor(X, Y)$  náhodných veličín  $X, Y$ , ktoré majú nenulový rozptyl je stredná hodnota súčinu odpovedajúcich normovaných veličín  $\frac{X-EX}{var_X}$  a  $\frac{Y-EY}{var_Y}$ , kde

$$cor(X, Y) = \frac{E((X - EX)(Y - EY))}{var_X var_Y} \in \langle -1, 1 \rangle.$$

Pre odhad rozptylov veličín  $X, Y$   $var_X, var_Y$  [7] použijeme výberový rozptyl, ktorý vypočítame podľa vzťahu 5.10 a pre odhad strednej hodnoty  $EX, EY$  použijeme výberový priemer, ktorý určíme podľa vzťahu 5.9.

Podobne definujeme výberovú kovarianciu, ktorá je definovaná nasledujúcim vzťahom.

$$S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = cov(X, Y) \quad (5.17)$$

Za predpokladu, že platí  $S_X^2, S_Y^2 > 0$  definujeme výberový korelačný koeficient  $r$  nasledovne.

$$r = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}} \quad (5.18)$$

Pre nezávislé náhodné veličiny a tiež pre tie nekorelované, je korelácia nulová. Avšak, nulová korelácia neznamená, že  $X$  a  $Y$  sú nutne nezávislé. Výsledná hodnota korelácie udáva len mieru lineárnej závislosti.

#### 5.3.4.1 Vlastnosti korelácie

Pre koreláciu dvoch náhodných veličín  $X$  a  $Y$  platia nasledujúce vlastnosti [5].

- Výsledná hodnota  $cor(X, Y) \in \langle -1, 1 \rangle$  a  $cor(X, Y) = 0 \iff cov(X, Y) = 0$ .
- Ak sú veličiny  $X$  a  $Y$  nezávislé, potom  $cor(X, Y) = 0$ .
- $cor(X, Y) = 1$ , platí práve vtedy keď  $Y = a + bX$ , pre  $a \in R, b > 0$ .
- $cor(X, Y) = -1$ , platí práve vtedy keď  $Y = a + bX$ , pre  $a \in R, b < 0$

### 5.3.4.2 Interpretácia korelácie

Koreláciou meriame silu lineárnej závislosti medzi náhodnými veličinami  $X$  a  $Y$ . Znamienko výslednej hodnoty korelačného koeficientu nám udáva smer závislosti veličín [5].

- Kladné znamienko nám hovorí, že so zvyšovaním sa hodnôt veličiny  $X$  sa zvyšujú hodnoty veličiny  $Y$  a analogicky.
- Záporné znamienko hovorí, že so zvyšovaním sa hodnôt  $X$  sa znižujú hodnoty  $Y$  a analogicky.
- Hodnota  $cor(X, Y) = 0$  automaticky neznamena nezávislosť náhodných veličín  $X$  a  $Y$ . Hovorí nám, že veličina  $X$  nemá tendenciu voči  $Y$  rásť a ani klesať a že medzi nimi neexistuje lineárna závislosť.

Verbálne môžeme hodnoty korelácie popísať nasledovne pre absolútne hodnoty korelačného koeficientu [3].

- 0,00 – 0,19 - veľmi slabá
- 0,20 – 0,39 - slabá
- 0,40 – 0,59 - stredná
- 0,60 – 0,79 - silná
- 0,80 – 1,00 - veľmi silná

### 5.3.5 Typy korelačných koeficientov

Pre jednoduchú koreláciu môžeme vyjadriť mieru lineárnej závislosti medzi náhodnými veličinami  $X$  a  $Y$  napríklad Pearsonovým a Spearmanovým korelačným koeficientom, ktoré si popíšeme v nasledujúcom texte.

#### 5.3.5.1 Pearsonov korelačný koeficient

Majme náhodný výber  $(X_1, Y_1), \dots, (X_n, Y_n)$  z nejakého dvojrozmerného normálneho rozdelenia. Potom Pearsonov korelačný koeficient vypočítame podľa vzťahu 5.18 a slúži pre kvantifikáciu lineárneho vzťahu náhodných veličín. Predpoklad pre výpočet Pearsonovho korelačného koeficientu je nenulový rozptyl [8].

#### 5.3.5.2 Spearmanov korelačný koeficient

Spearmanov korelačný koeficient je neparametrický odhad korelačného koeficientu [7]. Spearmanov korelačný koeficient [8] použijeme za predpokladu, že nie sú splnené predpoklady pre výpočet Pearsonovho korelačného koeficientu. Spearmanov koeficient sa používa v prípade, že hodnoty medzi sebou nie sú lineárne závislé. Pre jeho zostavenie nám stačí znalosť poradí  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_n$  z dvojrozmerného náhodného výberu  $(X_1, Y_1), \dots, (X_n, Y_n)$  z nejakého dvojrozmerného rozdelenia [7].

Predpokladom pre jeho výpočet je teda realizácia dvojrozmerného náhodného vektoru rozsahu  $n$  a vypočítame ho podľa nasledujúceho vzťahu [8]. Potom nech  $R_1, \dots, R_n$  označuje poradie  $X_1, \dots, X_n$  a nech  $Q_1, \dots, Q_n$  označuje poradie  $Y_1, \dots, Y_n$ . Spearmanov korelačný koeficient potom vypočítame podľa vzťahu [7]

$$r = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - Q_i)^2. \quad (5.19)$$

Spearmanov korelačný koeficient pracuje s poradím hodnôt a je tak odolný voči odchýlkam od normality náhodných veličín. Ak sú si poradie veľmi podobné, ukazuje to závislosť medzi veličinami [8].

## 5.4 Stanovenie hypotézy

Nasledujúci text popisuje pojem hypotéza a jednotlivé metódy testovania hypotéz, ktoré môžu byť využité pre ďalšie štatistické testovanie v rámci práce.

### 5.4.1 Hypotéza

Štatistická hypotéza [10] je predpoklad o rozdelení jednej, alebo viacerých náhodných veličín a týka sa nejakej náhodnej veličiny  $X$  na základe realizácie náhodného výberu. Hypotéza je síce predpokladanou výpoveďou, avšak je formulovaná tak, aby ju bolo možné buď nezamietnuť, alebo zamietnuť. Pre účely štatistického testovania formulujeme nulovú hypotézu  $H_0$  a alternatívnu hypotézu  $H_A$ , ktorú prijímame v prípade, že zamietneme testovanú hypotézu  $H_0$ .

### 5.4.2 Postup testovania hypotéz

Formu štatistického usudzovania, ktorá hľadá odpoveď na výskumné otázky, v podobe áno, alebo nie, nazývame Testovanie hypotéz.

Samotné testovanie prebieha v nasledujúcich krokoch.

- Formulácia otázky formou nulovej a alternatívnej štatistickej hypotézy.
- Zvolíme si vhodnú úroveň hladiny chyby rozhodovania  $\alpha$  (kde  $\alpha$  je pravdepodobnosť chyby rozhodovania I. druhu viz. 5.4.4).
- Výpočet testovacej štatistiky pomocou zvolenej metódy testovania hypotéz.
- Vyhodnotenie výsledkov rozhodovanie, kde nastane niektorý z nasledujúcich prípadov.
  1.  $H_0$  nezamietame v prospech alternatívnej hypotézy  $H_A$ .
  2.  $H_0$  zamietame v prospech alternatívnej hypotézy  $H_A$ .

### 5.4.3 Chyby pri testovaní štatistických hypotéz

V rámci testovania štatistických hypotéz pri overovaní hypotézy, ktorá sa týka náhodnej veličiny  $X$  na základe realizácie náhodného výberu, môže dôjsť k nesprávnemu rozhodnutiu dvomi spôsobmi. V prípade, že zamietneme správnu hypotézu  $H_0$ , hovoríme o *chybe I. druhu*. Ak sa rozhodneme nezamietnuť hypotézu  $H_0$  v prospech alternatívnej hypotézy  $H_A$ , keď je správna práve alternatívna hypotéza, hovoríme o *chybe II. druhu*.

- **Chyba I. druhu**

Hypotéza  $H_0$  platí a na základe zistených hodnôt sme ju zamietli v prospech alternatívnej hypotézy  $H_A$ .

- **Chyba II. druhu**

Hypotéza  $H_0$  neplatí a na základe zistených hodnôt sme ju nezamietli v prospech alternatívnej hypotézy  $H_A$ .

### 5.4.4 Hladina významnosti testovania štatistických hypotéz

Pravdepodobnosť zamietnutia hypotézy  $H_0$  v prospech  $H_A$ , aj keď platí sa nazýva hladina významnosti testovania štatistický hypotéz [10]. Hladina  $\alpha$  sa zvyčajne volí nízka, napríklad  $\alpha = 0,05$ .

## 5.5 Normalita dátového súboru

Testy štatistických hypotéz často predpokladajú normálne rozdelenie dátového súboru. V nasledujúcom texte si definujeme normálne rozdelenie a popíšeme niektoré metódy testovania normality dát.

### 5.5.1 Normálne rozdelenie

Normálne rozdelenie  $N(\mu, \sigma^2)$  [8] je spojité pravdepodobnostné rozdelenie s parametrami  $\mu$  a  $\sigma^2$ , kde  $\mu = EX$  a  $\sigma^2 = \text{var}X$ . Veličiny s normálnym rozdelením sa symetricky zhľukujú okolo  $EX$  a ich hustota má charakteristický tvar Gaussovej krivky. Obecné normálne rozdelenie [2]  $N(\mu, \sigma^2)$  je definované hustotou

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, \quad (5.20)$$

kde  $\mu \in R$ ,  $\sigma^2 > 0$ .

Distribučná funkcia je definovaná ako

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, -\infty < x < \infty. \quad (5.21)$$

V teórii pravdepodobnosti a v matematickej štatistike má veľký význam. Normálne rozdelenie je vhodným modelom pre technické, biologické a fyzikálne veličiny akým je napríklad výška a váha jedinca, alebo chyby merania [2].

Potom normované normálne rozdelenie [2] je špeciálny prípad obecného normálneho rozdelenia pre  $\mu = 0$  a  $\sigma^2 = 1$ . Toto rozdelenie je definované hustotou

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty, \quad (5.22)$$

kde  $\mu \in R$ ,  $\sigma^2 > 0$ .

Distribučná funkcia, značená  $\Phi$ , je definovaná ako

$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, -\infty < x < \infty. \quad (5.23)$$

Stredná hodnota  $EX = 0$  a rozptyl  $\text{var}X = 1$ .

Vlastnosťou normálneho rozdelenia je zachovanie normality pri zmene meradla osi, na ktorej meriame náhodné veličiny [8]. Pre náhodné veličiny s normálnym rozdelením platia nasledujúce vlastnosti [2].

- Ak má  $X$  normované normálne rozdelenie a  $Y = \mu + \sigma X$ , potom  $Y$  má normálne rozdelenie s parametrami  $\mu$  a  $\sigma^2$ .
- Ak má  $X$  normálne rozdelenie s parametrami  $\mu$  a  $\sigma^2$ , a ak  $Y = a + bX$ , potom  $Y$  má opäť normálne rozdelenie s parametrami  $a + b\mu$  a  $b^2\sigma^2$ .
- Nech  $X, Y$  sú náhodné veličiny s normálnym rozdelením, kde  $X \sim N(\mu_1, \sigma_1^2)$  a  $Y \sim N(\mu_2, \sigma_2^2)$ , a  $\text{cov}(X, Y) = 0$ . Potom  $Z = X + Y$  má normálne rozdelenie  $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ .

Na základe poslednej vlastnosti môžeme transformovať náhodnú veličinu  $X$  s obecným normálnym rozdelením na veličinu s normovaným normálnym rozdelením [8].

Hustota normovaného normálneho rozdelenia je symetrická [8] a platí, že  $\Phi(x) = 1 - \Phi(-x)$  [2]. Pomocou kvantilu  $100(1 - \alpha)$  môžeme plochu pod hustotou rozdeliť na dve časti  $\alpha$  a  $1 - \alpha$ . Delenie odpovedá pravdepodobnosti, že náhodná veličina nadobúda hodnotu menšiu než  $100(1 - \alpha)$  s pravdepodobnosťou  $1 - \alpha$  a hodnotu vyššiu s pravdepodobnosťou  $\alpha$ . Tento fakt využívame najmä v testovaní hypotéz, kde na základe oblasti realizácie hodnoty testovacej štatistiky rozhodujeme o platnosti alebo neplatnosti sledovanej hypotézy [8].



### 5.5.2 Testovanie normality dát

V testoch normality dát [2] overujeme, že dáta pochádzajú zo súboru s normálnym rozdelením. Testujeme hypotézy  $H_0$ : *Dáta majú normálne rozdelenie* proti alternatívnej hypotéze  $H_A$ : *Dáta nemajú normálne rozdelenie*.

Pre testovanie normality dát sú využívané dva typy testov. Keďže analytické testy nebudeme v rámci práce využívať, len ich spomenieme a popíšeme interpretáciu grafických testov normality dát.

#### 5.5.2.1 Analytické testy

Vhodný analytický test zvolíme podľa rozsahu dátového súboru  $n$  [1].

- Ak  $n > 100$  použijeme  $\chi^2$ -test dobrej zhody.
- Ak platí, že  $30 \leq n \leq 100$  zvolíme Shapiro-Wilkov test.
- Pre Ak  $7 \leq n < 30$  normalitu otestujeme D'Agostinovým testom.

#### 5.5.2.2 Grafické testy

Využívané grafické testy normality dát sú najmä Histogram a Q-Q plot. V Q-Q plote sú porovnané teoretické kvantily testovaného rozdelenia 5.6.1 s empirickými kvantilmi získanými z dát. Dáta potom považujeme za výber z normálneho rozdelenia, ak je graf približne lineárny. Histogram by pri normálnom rozdelení mal pripomínať približne Gaussovu krivku.

## 5.6 Metódy testovania hypotéz

V nasledujúcom texte si podrobnejšie popíšeme metódy a definície nutné k pochopeniu vybraných metód testovania hypotéz, ktoré môžeme využiť na testovanie hypotéz stanovených v rámci práce pre spojité náhodné veličiny a test nezávislosti v kontingenčnej tabuľke, ktorý využijeme na testovanie nezávislosti diskretných veličín.

### 5.6.1 $\beta$ -kvantil pravdepodobnostného rozdelenia

Majme distribučnú funkciu  $F$ , ktorá je spojitá, rýdzo monotónna a nech  $0 < \beta < 1$ . Potom číslo  $z_\beta$  také, že  $F(z_\beta) = \beta$  sa nazýva  $\beta$ -kvantil [2] pravdepodobnostného rozdelenia s distribučnou funkciou  $F$ . Kvantil je číslo, ktoré rozdeľuje pozorované hodnoty na dve časti podľa pravdepodobnosti.  $\beta$ -percentný kvantil ich delí na  $\beta$  hodnôt a  $(100 - \beta)$  hodnôt, kde  $\beta$  hodnôt je menších než  $\beta$ -percentný kvantil a  $(100 - \beta)$  je väčších [8].

### 5.6.2 $\chi^2$ rozdelenie

$\chi^2(n)$  rozdelenie [8] je spojité pravdepodobnostné rozdelenie s parametrom  $n$ ,  $n \in N$ , kde  $n$  je počet stupňov voľnosti. Vzniká ako súčet mocnín  $n$  nezávislých náhodných veličín  $X_i$ , kde  $i \in N$ , s normovaným normálnym rozdelením  $N(0, 1)$ .

$$K = \sum_{i=1}^n X_i^2 \sim \chi^2(n) \quad (5.24)$$

Využíva sa ako modelové rozdelenie pravdepodobnosti testovej štatistiky pri testovaní hypotéz o nezávislosti náhodných veličín a testoch dobrej zhody.

### 5.6.3 Studentovo t rozdelenie

Studentovo t rozdelenie [8]  $t(n)$  je spojité pravdepodobnostné rozdelenie s parametrom  $n$ ,  $n \in N$ , kde  $n$  je počet stupňov voľnosti. Parameter  $n$  preberá studentovo t rozdelenie od  $\chi^2$  rozdelenia. Vzniká ako podiel dvoch nezávislých náhodných veličín, kde  $X$  má normované normálne rozdelenie a veličina  $Y$  má  $\chi^2$  rozdelenie.

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n) \quad (5.25)$$

Jeho limitným prípadom je štandardné normálne rozdelenia a s rastúcou veľkosťou súboru sa hustota Studentovho t rozdelenia približuje hustote normálneho rozdelenia [8].

### 5.6.4 P-hodnota

Pre rozhodnutie o platnosti stanovenej hypotézy môžeme použiť p-hodnotu [8]. P-hodnota nám hovorí, s akou pravdepodobnosťou môžeme prijať platnosť nulovej hypotézy a je definovaná ako najmenšia hladina významnosti testu, pri ktorej zamietneme nulovú hypotézu.

#### 5.6.4.1 Interpretácia p-hodnoty

Štandardne nulovú hypotézu zamietame pri veľmi nízkej hodnote [8] p-hodnoty. Nízka hodnota p-hodnoty nám hovorí, že nulová hypotéza má veľmi nízku pravdepodobnosť platnosti. Nulovú hypotézu zamietneme ak platí, že hladina významnosti  $\alpha >$  p-hodnota.

### 5.6.5 Testy nezávislosti

V rámci nasledujúceho textu popíšeme metódy testovania hypotéz, ktoré budeme ďalej využívať v rámci práce.

Menovite využijeme test nezávislosti v kontingenčnej tabuľke, ktorý vyhodnocuje, či je možné dva výbery z veličín s diskretným rozdelením považovať za nezávislé. V teste z dát vypočítame hodnotu testovej štatistiky  $\chi^2$  a túto hodnotu následne porovnáme s príslušným kvantilom  $\chi^2$  rozdelenia. Využijeme tiež test hypotézy o nulovej korelácii dvoch náhodných veličín.

#### 5.6.5.1 Test hypotézy o nulovej korelácii dvoch náhodných veličín

V rámci testu hypotézy o nulovej korelácii dvoch náhodných veličín [8], testujeme nulovú hypotézu  $H_0: r = 0$  proti alternatívnej hypotéze  $H_0: r \neq 0$  a to nasledujúcim postupom.

1. Určíme hodnotu veličiny  $T_0$ .

$$T_0 = r \sqrt{\frac{n-2}{1-r^2}} \quad (5.26)$$

2. Ak platí nasledujúci vzťah potom zamietame  $H_0$  v prospech  $H_A$ .

$$|T_0| \geq t_{1-\alpha/2}(n-2) \quad (5.27)$$

Štatistika  $T_0$  má studentovo t rozdelenie pravdepodobnosti s  $n-2$  stupňami voľnosti. Predpokladom pre test je realizácia dvojrozmerného náhodného vektora rozsahu  $n$  a test môžeme využiť ako pre Pearsonov korelačný koeficient, tak aj pre Spearmanov korelačný koeficient, za predpokladu, že rozsah výberu  $n > 30$  [7].

## 5.6.5.2 Test nezávislosti v kontingenčnej tabuľke

Majme náhodný výber  $(X_1, Y_1), \dots, (X_n, Y_n)$ , kde  $X_k$  nadobúda hodnoty  $1, \dots, r$  a  $Y_k$  hodnôt  $1, \dots, c$ . Ak označíme počet výskytov dvojice  $n_{ij}$  združené početnosti, potom maticu o rozmere  $r \times c$  s prvkami  $n_{ij}$  nazývame kontingenčná tabuľka. Marginálne početnosti sú dané nasledujúcimi vzťahmi [2].

$$n_{i.} = \sum_j n_{ij} \quad (5.28)$$

$$n_{.j} = \sum_i n_{ij} \quad (5.29)$$

Hypotézu  $H_0 = X$  a  $Y$  sú závislé veličiny potom testujeme proti alternatívnej hypotéze  $H_A = X$  a  $Y$  nie sú závislé veličiny tak, že určíme hodnotu  $\chi^2$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}} \quad (5.30)$$

Ak platí nasledujúci vzťah, potom hypotézu  $H_0$  zamietame v prospech alternatívnej hypotézy  $H_A$ .

$$\chi^2 \geq \chi_{1-\alpha}^2((r-1)(c-1)) \quad (5.31)$$



# Kapitola 6

## Návrh riešenia

Cieľom nasledujúcej kapitoly je predstavenie stanovených hypotéz, ktoré budeme testovať v rámci bakalárskej práce a popis návrhu uloženia dátových zdrojov pre ďalšiu štatistickú analýzu. Následne hypotézy hlbšie rozoberieme a popíšeme postup, akým sme overovali ich platnosť.

### 6.1 Návrh databázového modelu

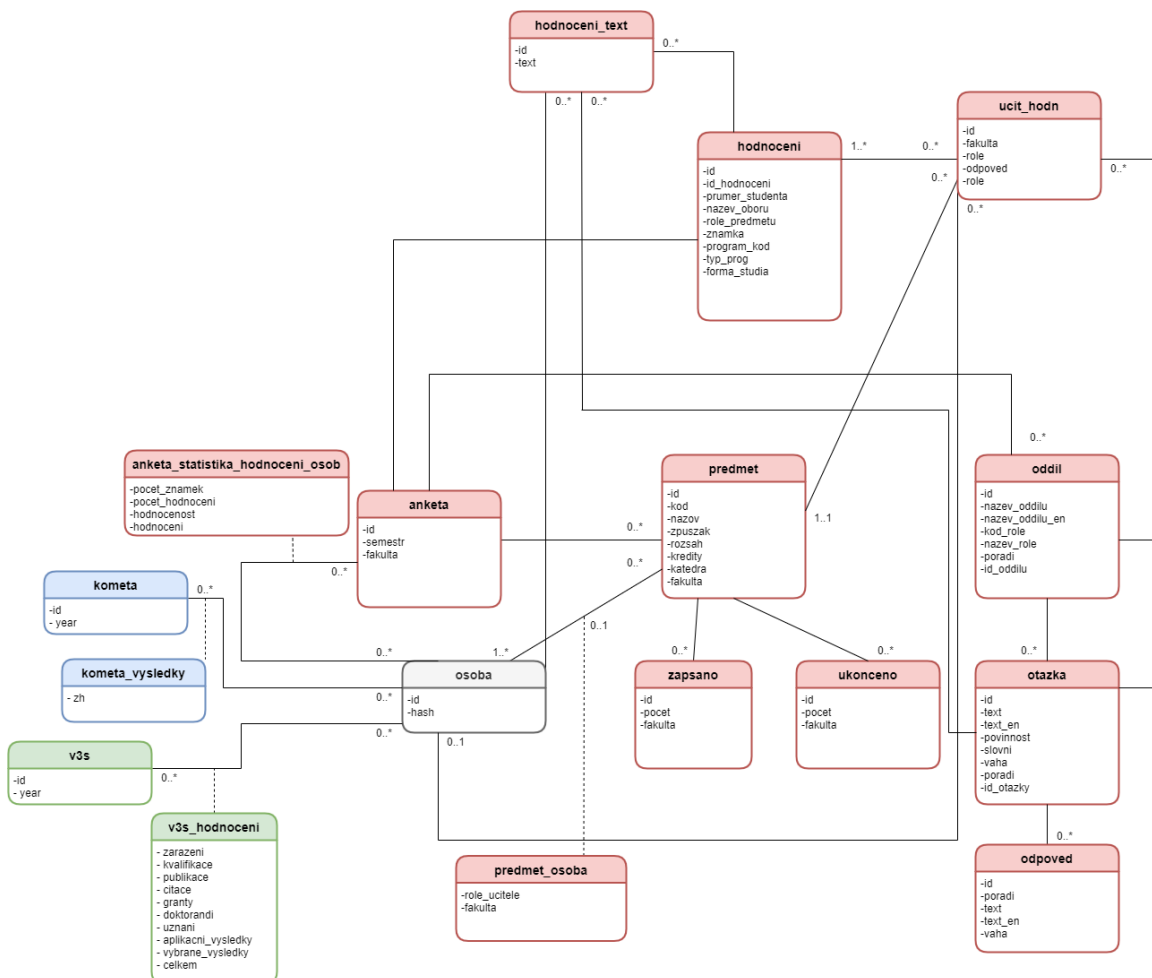
Jedným z cieľov bakalárskej práce je popis návrhu uloženia dátových zdrojov pre ďalšiu štatistickú analýzu. V štatistickom rozhodovaní o hypotézach závislosti veličín sú vhodné dáta pre analýzu základom úspechu. Na základe nevhodných dát, môžeme totiž urobiť mylný záver.

Z dostupných dát z aplikácie ankety študentské hodnotenia výuky na ČVUT 2, aplikácie V3S hodnotenia vedecko-výskumnej činnosti pracovníkov ČVUT FEL 3 a dát získaných z aplikácie KOMETA, ktorú sme popísali v kapitole Študijné výsledky a vyťaženosť vyučujúcich 4, sme navrhli schému konceptuálneho modelu, ktorá popisuje zjednotenie jednotlivých dátových zdrojov. Návrhom konceptuálneho modelu databázy popíšeme, aké dáta a v akej forme sú dostupné, pričom na základe nich stanovíme hypotézy, ktoré budeme overovať v rámci práce.

#### 6.1.1 Popis návrhu schémy

Podľa dostupných dátových zdrojov z aplikácie Ankety, V3S a KOMETY, relevantné pre náš databázový model, sme navrhli vhodnú konceptuálnu databázovú schému, ktorá sa nachádza na obrázku 6.1. Návrh sme vypracovali formou UML modelu, ktorým sme popísali identifikované entity, vzťahy a ich charakteristiky vo forme tried, asociácií a atribútov.

Dostupné dátové zdroje z aplikácie Ankety, V3S a KOMETY sú prepojené entitou *osoba*. Tabuľka *osoby* predstavuje pracovníka školy, ktorého atribútom je jeho anonymizované osobné číslo, ktoré v rámci ČVUT funguje ako jedinečný identifikátor jednotlivých osôb. Podľa nej si potom môžeme popis návrhu rozdeliť do nasledujúcich troch kategórií.



Obrázek 6.1: Schéma využívané v rámci práce pre uloženie dátových zdrojov z aplikácií Ankety, V3S a KOMETY.

• **Anketa**

Dostupné dáta z aplikácie ankety majú ako centrálnu tabuľku, *anketu*. Hodnotenie ankety, ako sme už popísali, sa vykonáva vždy za jednotlivé semestre akademického roku osobitne, pričom sa vyhodnocuje samostatne pre každú fakultu ČVUT. Preto *anketa* obsahuje atribúty *anketa.semestr* a *anketa.fakulta*. V nasledujúcom texte si popíšeme, aké dáta sú dostupné o študentskom hodnotení výuky na ČVUT FEL. Časť schémy ankety môžeme rozdeliť do dvoch častí a to vyhodnotenú výsledky a študentmi udelené hodnotenia.

Prvá časť, vyhodnotenú výsledky, je vzťah medzi tabuľkami *osoba* a *anketa* predstavuje celkové hodnotenie vyučujúceho v ankete za daný semester.

Druhá časť schémy ankety predstavu hodnotenia v ankete vyučujúcich v jednotlivých predmetoch a hodnotenia predmetov, reprezentované tabuľkou *hodnoceni*, ktoré reprezentujú hodnotenie jedného študenta. Môže mať dve podoby, a to hodnotenie otvorených otázok *hodnoceni\_text* a uzavretých otázok *ucit\_hodn*.

Okrem hodnotení v ankete a výsledkov vyučujúceho sa v každej ankete nachádzajú *otázky*,

ktoré patria do jednotlivých *oddielov*, pričom tabuľka *odpoved'* predstavuje možné odpovede na jednotlivé uzavreté otázky.

V každej ankete sú tiež dostupné informácie o *predmetoch*. Vzťah medzi *predmetom* a *osobou* predstavuje vyučujúceho v predmete za daný semester akademického roku. O každom predmete tiež poznáme štatistiky o počte zapísaných študentov a počte študentov, ktorí ukončili predmet.

- **Kometa**

Dáta, dostupné z aplikácie KOMETA, obsahovali počet započítateľných hodín vyučujúceho za jednotlivé roky. Výsledky z aplikácie KOMETA za daný rok sú reprezentované tabuľkou *kometa*. Vzťah medzi *kometou* a *osobou*, *kometa\_vysledky* obsahuje atribút *kometa\_vysledky.zh*, ktorý reprezentuje počet započítateľných hodín vyučujúceho za daný rok.

- **V3S**

Výsledky hodnotenia vedecko-výskumnej činnosti na ČVUT FEL z aplikácie V3S sú vyhodnocované za jeden rok. Výsledky pre daný rok sú popísané tabuľkou *v3s*. Vzťah medzi *v3s* a *osobou*, *v3s\_hodnoceni* obsahuje atribúty, ktoré predstavujú výsledky vedecko-výskumnej činnosti pracovníkov školy.

## 6.2 Stanovené hypotézy

Čiastkovou úlohou vytýčenou v rámci práce je definícia hypotéz korelácie vedecko-výskumných výsledkov vyučujúcich a ich hodnotenia v ankete, a študijných výsledkov študentov. Z dostupných dátových zdrojov sme pre ďalšiu analýzu stanovili nasledujúce hypotézy, ktoré si presnejšie rozoberieme v sekcii Postup pre overenie hypotéz 6.3.

1. *H01: "Publikačné výsledky vyučujúceho, ako vedecko-výskumného pracovníka, nekorelujú s jeho hodnotením v študentskej ankete."*
2. *H02: "Študenti s horším študijným priemerom sú v hodnotení vyučujúceho v ankete kritičkejší."*
3. *H03: "Vyučujúci, ktorí sú viac vyťažení výukou, majú horšiu publikačnú činnosť."*

## 6.3 Postup overenia hypotéz

V rámci nasledujúceho textu sa zameriame na ďalší cieľ práce, a tým je overenie hypotéz vhodnými štatistickými metódami. Hypotézy, jednu po druhej, postupne rozoberieme a predstavíme spôsob nášho riešenia, ktorým sme overovali ich platnosť.

### 6.3.1 Vplyv publikačnej činnosti vyučujúceho na jeho hodnotenie v ankete

Pripomeňme si znenie prvej stanovenej hypotézy:

“Publikačné výsledky vyučujúceho, ako vedecko-výskumného pracovníka, nekorelujú s jeho hodnotením v študentskej ankete.”

Z dostupných dátových zdrojov z aplikácie V3S a z aplikácie Ankety sme zostavili hypotézu o nezávislosti publikačných výsledkov vyučujúceho a jeho celkovým hodnotením v ankete študentského hodnotenia výuky. Nepredpokladáme teda, že publikačná činnosť vyučujúceho ako vedecko-výskumného pracovníka by nejak ovplyvňovala schopnosti pracovníka ako vyučujúceho a názor študentov naňho a analogicky.

V rámci kapitoly Anketa 2 sme si bližšie popísali otázky 2.2.2.2, ktorými študenti hodnotia vyučujúceho, a to buď v role prednášajúceho, alebo cvičiaceho v jednotlivých predmetoch. Študenti,

ktorí vyplnia anketné lístky, hodnotia uzavrené otázky o vyučujúcom známami na škále od 1-5, ktoré majú nasledujúce slovné vyjadrenie.

- 1 - Rozhodne áno.
- 2 - Skôr áno.
- 3 - Skôr nie.
- 4 - Rozhodne nie.
- 5 - Nevie sa vyjadriť.

Celkové hodnotenie vyučujúceho v ankete je potom určené na základe odpovedí na tieto otázky v jednotlivých predmetoch a počtu hodnotení vyučujúceho. Celková známka vyučujúceho je číslo  $z$ , kde  $z \in \langle 1, 5 \rangle$ .

Ako sme popísali v kapitole Hodnotenie vedecko-výskumnej činnosti na vysokých školách 3, pracovníkom školy sú pridelované body za ich publikačnú činnosť. Publikačné výsledky pracovníka sú potom číslo  $p$ , kde  $p \in \langle 0, \infty \rangle$ .

Budeme teda skúmať závislosť medzi veličinami  $X$ , kde  $X = \text{Publikačná činnosť vyučujúceho}$  a  $Y$ , kde  $Y = \text{Celková známka vyučujúceho}$ . Obe veličiny  $X, Y$  sú spojité náhodné veličiny, čo plynie z faktu, že ich oborom hodnôt nie je spočetná množina, ale interval.

Označenie	Slovné vyjadrenie	Obor hodnôt	Typ náhodnej veličiny
X	Publikačná činnosť	$\langle 0, \infty \rangle$	Spojité
Y	Celkové hodnotenie	$\langle 1, 5 \rangle$	Spojité

Tabulka 6.1: Skúmané náhodné veličiny vplyvu publikačnej činnosti vyučujúceho na jeho hodnotenie v ankete

Závislosť vplyvu publikačnej činnosti na celkové hodnotenie vyučujúceho v ankete budeme skúmať z dostupných dát za roky 2016 až 2018. Dáta o celkovom hodnotení vyučujúceho v študentskej ankete, teda vyhodnotenie za všetky sledované roky dohromady, spriemerujeme. Dôvodom je, že nepredpokladáme výrazné zmeny hodnotenia jednotlivých vyučujúcich. Z dostupných dát o hodnotení publikačnej činnosti vedecko-výskumných pracovníkov školy odstránime hodnoty  $p = 0$ , pretože v dátach sa nachádzajú pracovníci, ktorí sa nevenujú vedecko-výskumnej činnosti, doktorandi a iní zamestnanci školy, čo by mohlo skresliť výsledky v procese overenia hypotézy. Tieto dáta, vyhodnocované za roky 2016 - 2018 taktiež spriemerujeme, a to z rovnakého dôvodu, aký bol v prípade celkového hodnotenia pracovníka v ankete študentského hodnotenia výuky.

Stanovená nulová hypotéza  $H_0$  voči alternatívnej hypotéze  $H_A$  znie nasledovne.

- $H_0$ : *Publikačná činnosť pracovníka nekoreluje s jeho hodnotením v študentskej ankete.*
- $H_A$ : *Publikačná činnosť pracovníka koreluje s jeho hodnotením v študentskej ankete.*

### 6.3.1.1 Postup vyhodnotenia závislosti hodnotenia v ankete a publikačnej činnosti pracovníka

Náhodné veličiny  $X = \text{Publikačná činnosť}$  a  $Y = \text{Celkové hodnotenie vyučujúceho}$  sú obe spojité náhodné veličiny. Postup vyhodnotenia závislosti hodnotenia v ankete a publikačnej činnosti je rozvrhnutý do nasledujúcich krokov.



- 1. Pred samotným zhodnotením, či má alebo nemá zmysel koreláciu medzi dvomi náhodnými veličinami skúmať je nutné zistiť charakter dát, ktoré máme k dispozícii. Testy štatistických hypotéz tiež často predpokladajú normálne rozdelenie dátového súboru, a tak pre obe náhodné veličiny  $X$  a  $Y$  za roky 2016 - 2018 vykreslíme histogram, ktorý je vhodným nástrojom pre vizualizáciu intervalových dát a overíme ním, či dáta pochádzajú zo súboru s normálnym rozdelením. V rámci tohoto kroku budeme testovať nulovú hypotézu o normalite dát proti alternatívnej hypotéze, že dáta normálne rozdelenie nemajú. Histogram by pri normálnom rozdelení dát mal pripomínať približne Gaussovu krivku. Tiež si vypočítame základné charakteristiky náhodných výberov akými sú výberový priemer, výberový rozptyl a odchýlka a medián, na základe ktorých zhodnotíme súbor dát pre obe náhodné veličiny  $X$  a  $Y$ .
- 2. V druhom kroku si vykreslíme bodový graf, ktorým zobrazíme každú hodnotu ako bod plochy a je tak vhodným nástrojom pre vizualizáciu vzájomného vzťahu dvoch náhodných veličín, kde zobrazíme veličinu  $X$  na osi  $x$  a  $Y$  na osi  $y$ . Z vykresleného grafu následne preskúmame, či existuje medzi veličinami kladná, záporná, alebo nulová korelácia. Vieme, že pri kladnej korelácii má veličina  $X$  tendenciu rásť so zvyšujúcimi sa hodnotami veličiny  $Y$ , pri zápornej klesáť a analogicky. V prípade nulovej korelácie nemajú tieto veličiny vzájomne tendenciu rásť ani klesáť. Z bodového grafu tiež poznáme, či je splnený predpoklad pre výpočet Pearsonovho korelačného koeficientu, dvojrozmerné normálne rozdelenie. Ak vykreslené hodnoty pripomínajú približne elipsu, môžeme považovať predpoklad za splnený.
- 3. Vieme, že dve spojitú náhodné veličiny  $X$  a  $Y$  sú vzájomne nezávislé, ak združená hustota je rovná súčinu ich marginálnych hustôt. Ak tento vzťah platí, podľa vlastností korelácie 5.3.4.1 vieme, že  $cov(X, Y) = 0, cor(X, Y) = 0$ . Nezávislosť však nejdeme overovať týmto spôsobom a popíšeme vzájomný vzťah dvoch veličín kovarianciou. Tú určíme podľa vzťahu 5.17, kde za stredné hodnoty dosadíme ich výberové priemery. Výslednú hodnotu kovariancie náhodných veličín  $X, Y$  interpretujeme podľa vlastností 5.3.3.1 a interpretácie kovariancie 5.3.3.2 popísanej v predchádzajúcej kapitole.
- 4. Za predpokladu, že budú splnené podmienky pre výpočet Pearsonovho korelačného koeficientu ho použijeme pre výpočet korelačného koeficientu medzi náhodnými veličinami  $X$  a  $Y$  a vypočítame ho podľa vzťahu 5.18. Spearmanov korelačný koeficient použijeme za predpokladu, ak podmienky pre výpočet Pearsonovho korelačného koeficientu splnené nebudú a vypočítame ho podľa vzťahu 5.19. Výsledky pre oba koeficienty interpretujeme zhodne, a to podľa interpretácie korelácie popísanej v predchádzajúcej kapitole 5.3.4.2.
- 5. Na záver vypočítame testovú štatistiku pre test hypotézy o nulovej korelácii dvoch veličín podľa vzťahu 5.26 a výsledok porovnáme s príslušným kvantilom podľa vzťahu 5.27. Hladinu  $\alpha$  volíme  $\alpha = 0.05$  a túto hladinu porovnáme s testovanou štatistikou. V prípade, že je menšia ako  $\alpha$  nulovú hypotézu zamietame.

### 6.3.2 Závislosť hodnotenia v ankete od prospechu študenta

Znenie druhej vyslovenej hypotézy je: "Študenti s horším študijným priemerom sú v hodnotení vyučujúceho v ankete kritickejší."

Definovať kritiku hodnotenia je náročná úloha, keďže zanalyzovať každú študentskú odpoveď v ankete a prideliť jej číselného hodnotenie "na škále kritickosti" je vzhľadom k rozsahu dátového súboru pre hodnotenie otvorených otázok časovo nestihnuteľná úloha. Obmedzíme sa preto na hodnotenie uzavretých otázok, ktoré už boli sčasti popísané pre prvú stanovenú hypotézu. Kritické hodnotenie v našom ponímaní potom znamená, že čím vyššiu známku spomedzi hodnôt  $\{1, 2, 3, 4\}$  študent pridelí ako odpoveď na danú otázku, tým je jeho hodnotenie kritickejšie.

Študenti mimo odpovede na otázky týkajúce sa hodnotenia vyučujúcich, odpovedajú aj na otázky v kategórii hodnotenie predmetu. Z množiny všetkých otázok hodnotenia predmetu sme zvolili dve,

ktorých hodnotenie budeme brať do úvahy pri vyhodnocovaní stanovenej hypotézy 2.2.2.2. Otázku k hodnoteniu obtiažnosti predmetu a otázku týkajúcu sa prípravy študenta na predmet sme z hodnotenia vypustili, pretože sú hodnotené na inej škále ako otázky, ktoré berieme do úvahy pri vyhodnocovaní hypotézy.

- Predmet bol pre mňa prínosom.
- Študijné materiály boli kvalitné.

Študenti týmto otázkam opäť pridelujú známky ako v prípade hodnotenia vyučujúcich. Našou skúmanou veličinou je potom  $Y = \text{Udelené hodnotenie}$ , ktoré nadobúda hodnoty z množiny  $Z$ , kde  $Z = \{1, 2, 3, 4, 5\}$ .

Ako veličinu  $X$  sme stanovili  $X = \text{Priemer študenta}$ , ktorý môže nadobúdať hodnoty z množiny  $P = \{(1.0, 1.5), (1.5, 2.0), (2.0, 2.5), (2.5, 3.0)\}$ .

Označenie	Slovné vyjadrenie	Obor hodnôt	Typ náhodnej veličiny
$X$	Priemer študenta	$\{(1.0, 1.5), (1.5, 2.0), (2.0, 2.5), (2.5, 3.0)\}$	Diskrétna
$Y$	Udelené hodnotenie	$\{1, 2, 3, 4, 5\}$	Diskrétna

Tabulka 6.2: Skúmané náhodné veličiny závislosť hodnotenia v ankete od prospechu študenta

Závislosť hodnotenia v ankete od prospechu študenta budeme skúmať z dostupných dát za školský akademický rok 2017/2018. Stanovili sme potom nulovú hypotézu  $H_0$  voči alternatívnej hypotéze  $H_A$ .

- $H_0$ : Udelené hodnotenie a Priemer študenta sú nezávislé náhodné veličiny.
- $H_A$ : Udelené hodnotenie a Priemer študenta nie sú nezávislé náhodné veličiny.

### 6.3.2.1 Postup vyhodnotenia hypotézy o nezávislosti udeleného hodnotenia a priemeru študenta

Pre vyhodnotenie stanovenej hypotézy použijeme test nezávislosti v kontingenčnej tabuľke, ktorý sme popísali v predchádzajúcej kapitole 5.6.5.2. Kontingenčné tabuľky umožňujú pomocou testu nezávislosti dvoch veličín rozhodnúť, či spolu dve náhodné veličiny súvisia. Na základe nich je možné testovať veličiny, ktoré majú kvalitatívny charakter, diskkrétne veličiny alebo spojité kvantitatívne veličiny, ktorých hodnoty zlúčime do skupín.

Náhodná veličiny  $X = \text{Priemer študenta}$  je kvalitatívna a  $Y = \text{Udelené hodnotenie}$  je diskrétna kvantitatívna veličina. Vyššie sme si popísali aký obor hodnôt môžu obe náhodné veličiny nadobúdať. Zostavili sme si kontingenčnú tabuľku, ktorá je definovaná ako matica počtu výskytu dvojíc výberu  $(n_{i,j})$  [7].

Veličina $X$	Veličina $Y$					Celkom
	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	
$\langle 1.0, 1.5 \rangle$	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$n_{1,5}$	$\sum_{j=1}^5 n_{1,j}$
$\langle 1.5, 2.0 \rangle$	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$n_{2,5}$	$\sum_{j=1}^5 n_{2,j}$
$\langle 2.0, 2.5 \rangle$	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{3,4}$	$n_{3,5}$	$\sum_{j=1}^5 n_{3,j}$
$\langle 2.5, 3.0 \rangle$	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{4,4}$	$n_{4,5}$	$\sum_{j=1}^5 n_{4,j}$
Celkom	$\sum_{i=1}^4 n_{i,1}$	$\sum_{i=1}^4 n_{i,2}$	$\sum_{i=1}^4 n_{i,3}$	$\sum_{i=1}^4 n_{i,4}$	$\sum_{i=1}^4 n_{i,5}$	$\sum$

Tabulka 6.3: Kontingenčná tabuľka zostavená pre overenie závislosti veličín  $X$  a  $Y$ .

Na základe zostavenej kontingenčnej tabuľky pre stanovenú hypotézu  $H_0$ , vypočítame testovú štatistiku  $\chi^2$  podľa vzťahu 5.30. Vypočítanú p-hodnotu testovej štatistiky porovnáme s hladinou  $\alpha$  a na základe nej hypotézu zamietneme alebo nezamietneme.

### 6.3.3 Vyťaženosť vyučujúcich a ich vplyv na publikačnú činnosť

A na záver budeme testovať hypotézu: “Vyučujúci, ktorí sú viac vyťažení výukou, majú horšiu publikačnú činnosť.”

V hypotéze o závislosti medzi výučbou pracovníka a jeho publikačnou činnosťou predpokladáme fakt, že ak pracovník je priveľmi vyťažovaný výukou študentov, neostáva mu mnoho času pre jeho vedeckú činnosť. Preto sa snažíme stanovenou hypotézou otestovať fakt, že vysoký počet bodov za publikácie súvisí s nižšími hodnotami počtu odučených hodín pracovníka ako vyučujúceho a analogicky.

Výučba je jednou z mnohých činností vedecko-výskumných pracovníkov vysokých škôl a niektorí sa na nej podieľajú viac ako iní. Pripomeňme si, že pedagogický výkon vyučujúceho v predmete je určený na základe pomeru všetkých podielov vyučujúceho na výučbe predmetu, za prednášky a cvičenia a celkového počtu všetkých prednášok a cvičení predmetu 4.2.1 a predstavujú započítateľné hodiny vyučujúceho za jeho výučbu za semester. Započítateľné hodiny vyučujúceho sú potom číslo  $zh$ , kde  $zh \in \langle 0, \infty \rangle$ , ktoré budú skúmanou veličinou  $X = \text{Započítateľné hodiny vyučujúceho}$ , kde  $X$  je spojitá veličina.

Druhou skúmanou veličinou  $Y$  v rámci tretej hypotézy sú opäť publikačné výsledky vyučujúceho.

Označenie	Slovné vyjadrenie	Obor hodnôt	Typ náhodnej veličiny
$X$	Započítateľné hodiny	$\langle 0, \infty \rangle$	Spojité
$Y$	Publikačná činnosť	$\langle 0, \infty \rangle$	Spojité

Tabuľka 6.4: Skúmané náhodné veličiny vyťaženia vyučujúcich a ich vplyvu na publikačnú činnosť

Závislosť vyťaženia vyučujúcich a ich vplyvu na publikačnú činnosť budeme skúmať z dostupných dát za roky 2017 a 2018. Dostupné dáta o publikačnej činnosti sme upravili rovnakým spôsobom ako to bolo v prípade prvej hypotézy. Dáta o započítateľných hodinách vyučujúceho za tieto dva roky spriemerujeme, pretože nepredpokladáme u jednotlivých pracovníkov výraznú odchýlku v ich činnosti za dva po sebe idúce roky. Z dostupných dát neodstránime hodnoty  $zh = 0$ , a to z toho dôvodu, že očakávame, že ak bude človek veľa publikovať, bude to mať vplyv na jeho časovú vyťaženosť, a tak sa bude menej venovať výučbe.

Nižšie sa nachádza znenie stanovenej nulovej hypotézy  $H_0$  voči alternatívnej hypotéze  $H_A$ , pričom postup overenia je zhodný s postupom pre overenie prvej hypotézy 6.3.1.

- $H_0$ : *Vyťaženosť vyučujúcich nekoreluje s publikačnou činnosťou.*
- $H_A$ : *Vyťaženosť vyučujúcich koreluje s publikačnou činnosťou.*



# Kapitola 7

## Implementácia a testovanie

V predchádzajúcej kapitole 6 sme navrhli konceptuálny databázový model, stanovili hypotézy a navrhli riešenia, ktorými ich budeme overovať. V nasledujúcej kapitole nadviažeme na tú predchádzajúcu a popíšeme spôsob uloženia dátových zdrojov ako aj predstavíme algoritmizovaný postup overania hypotéz vhodnými štatistickými metódami, ktorými sme overovali platnosť hypotéz. Na záver kapitoly rozoberieme testovanie, ktoré je potrebné pre overenie správnosti našich výsledkov.

### 7.1 Implementácia

V rámci sekcie implementácia popíšeme implementovanú perzistentnú vrstvu a algoritmizáciu štatistických metód testovania hypotéz, ktoré sme využili v tejto práci s cieľom vyhodnotiť stanovené štatistické hypotézy.

#### 7.1.1 Databáza

Podľa konceptuálneho modelu popísaného v predchádzajúcej kapitole 6.1.1, sme využitím voľne dostupného relačného databázového systému PostgreSQL<sup>1</sup>, vytvorili relačnú databázu pre uloženie dátových zdrojov prostredníctvom skriptu. Relačná databáza nám umožňuje ukladať poskytnuté dáta do tabuliek, ktoré boli implementované presne podľa navrhnutého konceptuálneho schématu.

##### 7.1.1.1 Perzistentná vrstva

Perzistentná vrstva bola implementovaná v programovacom jazyku Java za použitia Java Persistence Api<sup>2</sup>, skratene JPA. JPA umožňuje objektovo relačné mapovanie ORM, pomocou knižnice Hibernate<sup>3</sup> a JPA anotácií.

Vrstva je rozdelená na dve časti *model* a *repository*. Jednotlivé tabuľky databázy sú implementované v triedach v časti *model* a za použitia anotácií JPA je vykonané mapovanie tried na databázové tabuľky. Časť *repository* obsahuje prístup do databázových tabuliek pomocou návrhového vzoru Data Access Object, skratene DAO, ktorý je implementovaný medzi aplikáciou a databázou.

Perzistentná vrstva databázovej aplikácie bola použitá najmä pre uľahčenie ukladania dát do databázy. Dátové zdroje boli pre účely práce poskytnuté formou .sql a .xml súborov a očakáva sa, že táto vrstva bude využitá aj po odovzdaní práce. Rôzny formát poskytnutých zdrojov nás viedol k myšlienke uloženia dát touto formou.

---

<sup>1</sup><https://www.postgresql.org/>

<sup>2</sup><https://docs.oracle.com/javaee/6/tutorial/doc/bnbpz.html>

<sup>3</sup><https://hibernate.org/>

### 7.1.1.2 Algoritmizácia postupu vyhodnocovania

Vyhodnocovanie štatistických hypotéz sme sa rozhodli implementovať v programovacom jazyku Python<sup>4</sup>, s použitím knižníc poskytujúcich funkcie pre štatistické výpočty.

Popis implementácie postupu vyhodnocovania môžeme rozdeliť na tri časti, a to skript pre pripojenie sa k databáze a tri ďalšie, ktoré boli implementované pre samotné vyhodnocovanie. Pripojenie k databáze je veľmi jednoduché a je implementované s použitím databázového adaptéru Psycopg<sup>5</sup>. Psycopg poskytuje funkciu `connect()`, ktorá vytvorí databázovú session a vráti objekt, ktorý reprezentuje databázové pripojenie, `connection`.

Jednotlivé skripty s implementáciou algoritmizácie postupu majú podobný charakter. Prostredníctvom funkcie `connection.cursor()` nám Python umožní zadávať PostgreSQL príkazy v databázovej session. Implementovanými queries sme získali dáta potrebné pre vyhodnocovanie hypotéz. Získané dáta sme uložili do dvojdimenzionálnej dátovej štruktúry DataFrame, ktorú poskytuje knižnica Pythonu Pandas<sup>6</sup>. Tá poskytuje funkciu `describe()`, ktorú sme využili pre popísanie charakteristík súborov a funkciu `cov()`, ktorej výsledkom je kovariančná matica. Pre výpočet Pearsonovho a Spearmanovho korelačného koeficientu sme využili Scipy<sup>7</sup>, čo je knižnica používaná pre matematické a štatistické výpočty. Histogramy a bodové grafy poskytnutých dát sme vykreslili prostredníctvom Seaborn<sup>8</sup>, čo je knižnica pre vizualizáciu dát.

## 7.2 Testovanie správnosti výsledkov

V nasledujúcom texte oboznámime čitateľa so spôsobom testovania správnosti výsledkov, ktoré sme získali podľa navrhnutého postupu pre overenie hypotéz v predchádzajúcej kapitole.

Testovanie potom môžeme rozdeliť na dva kroky. V prvom kroku sme testovali správnosť dát využitých pre testovanie štatistických hypotéz a v druhom kroku sme overili správnosť výsledkov.

Pripomeňme si stanovené hypotézy pre účely popisu testovania správnosti výsledkov.

- *H01*  
Publikačné výsledky vyučujúceho, ako vedecko-výskumného pracovníka, nekorelujú s jeho hodnotením v študentskej ankete.
- *H02*  
Študenti s horším študijným priemerom sú v hodnotení vyučujúceho v ankete kritickejší.
- *H03*  
Vyučujúci, ktorí sú viac vyťažení výukou, majú horšiu publikačnú činnosť.

Test správnosti dát, ktoré sme použili pre výpočet, sme vykonali len pre dáta nutné pre vyhodnotenie stanovených hypotéz. Ako sme už spomínali, dáta sme pre výpočet získali z databázy, prostredníctvom SQL dotazov, ktoré sú popísané vyššie. Výsledky dotazov sme porovnali s pôvodnými poskytnutými dátovými zdrojmi, uloženými v .xml súbore, pričom sa tieto dáta zhodovali.

Podľa navrhnutého postupu pre testovanie hypotéz *H01* a *H03* sme skúmali závislosť medzi veličinami na základe bodového grafu, hodnoty kovariancie, ako aj hodnoty korelačného koeficientu. Všetky tieto metódy testovania závislosti mali zhodné výsledky. Postup sme algoritmizovali a výsledok z každého kroku sme získali využitím knižníc Pythonu, poskytujúcich funkcie pre výpočet štatistických metód analýzy dát. Tieto knižnice nachádzajú široké využitie v oblasti matematiky

---

<sup>4</sup><https://www.python.org/>

<sup>5</sup><https://pypi.org/project/psycopg2/>

<sup>6</sup><https://pandas.pydata.org/>

<sup>7</sup><https://www.scipy.org/>

<sup>8</sup><https://seaborn.pydata.org/>

a informatiky a predpokladáme, že správnosť ich výsledkov je overená častým využívaním týchto funkcií.

Správnosť výsledkov testovej štatistiky  $\chi^2$  vypočítanej v druhej hypotéze sme, najmä pre jeho jednoduchosť, overovali ručne podľa postupu vyhodnotenia hypotézy o nezávislosti priemeru študenta 6.3.2.1 na základe dát dostupných z databázy, pričom sme získali rovnaké výsledky. Pre výpočet testovej štatistiky sme ako tomu bolo v predchádzajúcich dvoch prípadoch, opäť využili funkcie poskytované knižnicou pythonu.





## Kapitola 8

# Vyhodnotenie výsledkov

V rámci obsahu predchádzajúcich dvoch kapitol sme stanovili hypotézy a navrhli riešenia 6 pre zamietnutie alebo nezamietnutie ich platnosti. Tiež sme popísali spôsob uloženia dátových zdrojov 6.1.1 a algoritmizovali postup overenia vhodnými štatistickými metódami 7. Ostáva nám teda posledný cieľ bakalárskej práce, a tým je vyhodnotenie štatistických hypotéz z poskytnutých dát.

### 8.1 Vyhodnotenie závislosti publikačnej činnosti a celkového hodnotenia vyučujúceho

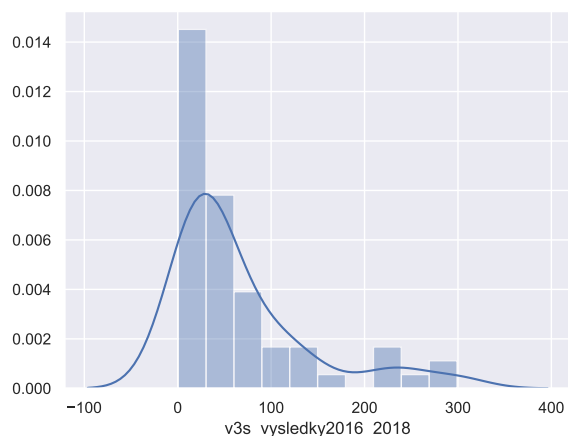
V rámci kapitoly 6 sme stanovili hypotézu "*Publikačná činnosť pracovníka koreluje s jeho hodnotením v študentskej ankete*", ktorou predpokladáme, že vedecko-výskumná činnosť pracovníka neovplyvňuje jeho schopnosti ako vyučujúceho.

Ako sme popísali v sekcii Vplyv publikačnej činnosti vyučujúceho na jeho hodnotenie v ankete 6.3.1, skúmali sme závislosť medzi veličinami  $X = \text{Publikačná činnosť}$  a  $Y = \text{Celkové hodnotenie}$ .

Podľa navrhovaného postupu overenia hypotéz sme si vykreslili histogramy dát publikačnej činnosti 8.1 a celkového hodnotenia vyučujúceho 8.2 po zjednotení dátových súborov pre roky 2016 - 2018, z ktorých môžeme vidieť, že publikačné výsledky vyučujúceho podliehajú exponenciálnemu rozdeleniu. Avšak, histogram celkového hodnotenia vyučujúceho tvarom pripomína približne Gaussovú krivku a môžeme ho prehlásiť za normálne rozdelenie 8.3.

Druhým bodom, v rámci prvého kroku bolo určiť základné charakteristiky dátových súborov. Hodnoty charakteristík publikačnej činnosti a celkového hodnotenia vyučujúceho sú určené po zjednotení výsledkov podľa osôb za roky 2016-2018 pre každú veličinu najskôr samostatne, a následne aj po zjednotení výsledkov publikácií a hodnotenia za roky 2016-2018 dohromady 8.3.

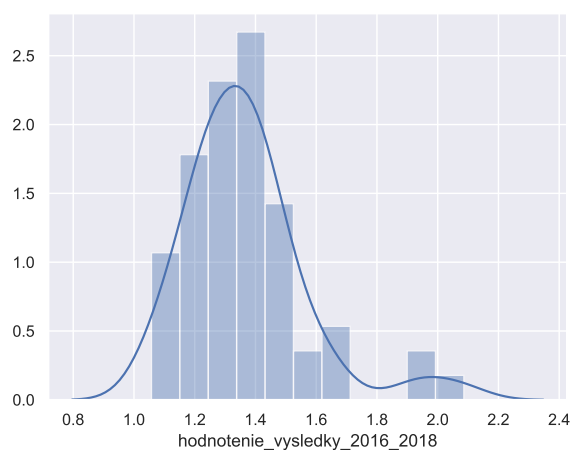
Tabuľka 8.1 popisuje charakter dát jednotlivo za roky 2016, 2017 a 2018 a následne ich sprimerované výsledky za roky 2016-2018. Ako môžeme vidieť, po zjednotení sa nám výrazne znížilo maximum spomedzi hodnôt a zvyšné charakteristiky ostali viac-menej zachované.



Obrázek 8.1: Histogram dát publikačnej činnosti vyučujúceho 2016-2018.

Charakteristika	Publikačné výsledky 2016	Publikačné výsledky 2017
Počet pozorovaní	294	294
Priemer	54.503	66.402
Štandardná odchýlka	90.267	102.896
Minimum	0.200	0.056
Maximum	824.156	764.444
Medián	14.652	31.057
	Publikačné výsledky 2018	Publikačné výsledky 2016-2018
Počet pozorovaní	294	294
Priemer	65.102	62.002
Štandardná odchýlka	96.066	76.222
Minimum	0.020	0.519
Maximum	779.75	488.284
Medián	32.150	34.65

Tabulka 8.1: Charakteristika dát publikačnej činnosti 2016-2018.



Obrázek 8.2: Histogram dát celkového hodnotenia 2016-2018.

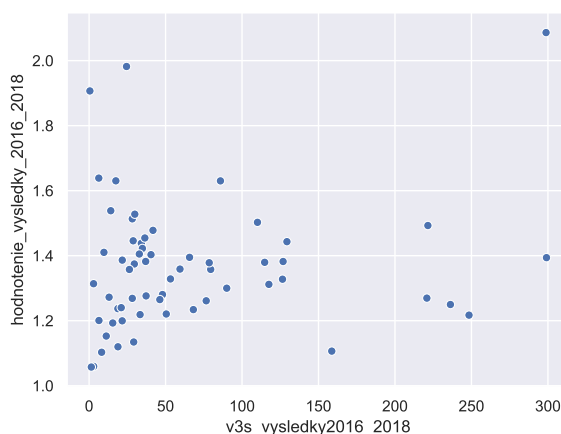
## 8.1. VYHODNOTENIE ZÁVISLOSTI PUBLIKAČNEJ ČINNOSTI A CELKOVÉHO HODNOTENIA VYUČUJÚCEHO

Charakteristiky hodnotenia vyučujúceho sú určené po zjednotení výsledkov podľa osôb za roky 2016-2018 a pre každý semester za akademické roky 2015/2016, 2016/2017 a 2017/2018 zvlášť. Charakter dát sa výrazne nemení ani v jednom zo sledovaných semestrov a ani v prípade spriemerovaných hodnôt za jednotlivé semestre rokov 2016-2018.

Charakteristika	Celkové hodnotenie B152	Celkové hodnotenie B161	Celkové hodnotenie B162
Počet pozorovaní	157	157	157
Priemer	1.349	1.363	1.360
Štandardná odchýlka	0.282	0.264	0.318
Minimum	1.000	1.000	1.000
Maximum	2.667	2.447	2.778
Medián	1.303	1.323	1.279
	Celkové hodnotenie B171	Celkové hodnotenie B172	Celkové hodnotenie B181
Počet pozorovaní	157	157	157
Priemer	1.334	1.348	1.338
Štandardná odchýlka	0.318	0.251	0.273
Minimum	1.000	1.000	1.000
Maximum	2.529	2.155	2.500
Medián	1.296	1.310	1.295
	Celkové hodnotenie 2016-2018		
Počet pozorovaní	157		
Priemer	1.349		
Štandardná odchýlka	0.197		
Minimum	1.0371		
Maximum	2.086		
Medián	1.329		

Tabulka 8.2: Charakteristika dát celkového hodnotenia 2016-2018

Druhým krokom postupu bolo vykreslenie bodového grafu, ktorý zobrazuje závislosť publikačnej činnosti a celkového hodnotenia. Graf svedčí o kladnej korelácii medzi veličinami.



Obrázek 8.3: Bodový graf dát publikačnej činnosti a celkového hodnotenia 2016-2018.

Súbor dát publikačnej činnosti a celkového hodnotenia sa po prieniku osôb zmenšil na 60 hodnôt a výrazne kleslo maximum publikačnej činnosti.

Charakteristika	Publikácie 2016-2018	Celkové hodnotenie 2016-2018
Počet pozorovaní	60	60
Priemer	66.254	1.365
Štandardná odchýlka	73.843	0.199
Minimum	0.519	1.057
Maximum	299.065	2.086
Medián	35.649	1.358

Tabulka 8.3: Charakteristika dát zjednotenia publikačnej činnosti a celkového hodnotenia 2016-2018.

Hodnota kovariancie veličín  $X$  a  $Y$   $cov(X, Y) = 2.347$  ukazuje na kladnú kovarianciu medzi veličinami, a teda vyššie hodnoty  $X$  súvisia s vyššími hodnotami  $Y$ .

Podľa bodového grafu sú splnené podmienky pre výpočet Pearsonovho korelačného koeficientu a jeho výsledná hodnota  $cor(X, Y) = 1.6$  hovorí o slabšej kladnej korelácii medzi pozorovanými veličinami.

Na základe  $p$  – hodnoty  $= 0.223$  nulovú hypotézu zamietame v prospech alternatívnej hypotézy "Publikačná činnosť pracovníka koreluje s jeho hodnotením v študentskej ankete", čím sme nepotvrdili našu pôvodnú stanovenú hypotézu. Vysoké hodnoty publikačných výsledkov dokonca súvisia s vyššími hodnotami hodnotenia vyučujúceho v študentskej ankete. To môže plynúť z faktu, že mu neostáva toľko čas na prípravu kvalitných študijných materiálov, prípadne zodpovedania dotazov jednotlivých študentov.

## 8.2 Vyhodnotenie závislosti hodnotenia v ankete od prospechu študenta

Stanovenú hypotézu "Študenti s horším študijným priemerom sú v hodnotení ankety kritickejší" sme overovali podľa postupu popísanom v kapitole Návrh riešenia 6.3.3.

Podľa postupu sme si zostavili kontingenčnú tabuľku náhodných veličín  $X =$  Priemer študenta a  $Y =$  Udelené hodnotenie.

Veličina $X$	Veličina $Y$					Celkom
	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	
$\langle 1.0, 1.5 \rangle$	36110	14105	4070	1450	2540	$\sum_{j=1}^5 n_{1,j}$
$\langle 1.5, 2.0 \rangle$	27980	12270	3360	1100	2365	$\sum_{j=1}^5 n_{2,j}$
$\langle 2.0, 2.5 \rangle$	15890	7500	2195	640	1630	$\sum_{j=1}^5 n_{3,j}$
$\langle 2.5, 3.0 \rangle$	1315	270	95	20	475	$\sum_{j=1}^5 n_{4,j}$
Celkom	$\sum_{i=1}^4 n_{i,1}$	$\sum_{i=1}^4 n_{i,2}$	$\sum_{i=1}^4 n_{i,3}$	$\sum_{i=1}^4 n_{i,4}$	$\sum_{i=1}^4 n_{i,5}$	$\Sigma$

Tabulka 8.4: Kontingenčná tabuľka zostavená pre overenie závislostí veličín Priemer študenta a Udeleného hodnotenia.

Na základe zostavenej tabuľky sme vypočítali testovú štatistiku  $\chi^2 = 1629.075$  a  $p$ -hodnotu, kde  $p$  – hodnota  $\leq 10^{-30}$  a porovnali ju s hladinou  $\alpha = 0.05$ .  $P$ -hodnota je menšia, čo znamená, že nulovú hypotézu zamietame v prospech alternatívnej hypotézy  $H_A$ : Udelené hodnotenia a priemer nie sú nezávislé náhodné veličiny a preukázali sme tak závislosť prospechu študenta a jeho pridelenom hodnotení v študentskej ankete.

### 8.3 Vyhodnotenie závislosti vyťaženia vyučujúcich a ich publikačnej činnosti

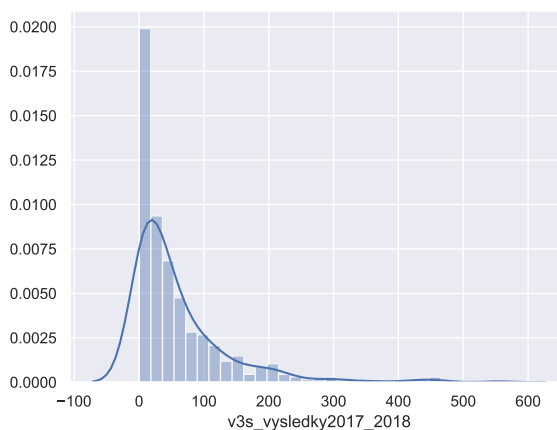
V kapitole 6 sme formulovali hypotézu "Vyučujúci, ktorí sú viac vyťažení výukou, majú horšiu publikačnú činnosť." a popísali postup overenia v časti Vyťaženie vyučujúcich a ich vplyv na publikačnú činnosť 6.3.3.

Stanovenou hypotézou sme skúmali závislosť medzi veličinou  $X = \text{Započítateľné hodiny}$  a veličinou  $Y = \text{Publikačná činnosť}$ , pričom sme si sformulovali nulovú hypotézu  $H_0$  voči alternatívnej hypotéze  $H_A$  nasledovne.

- $H_0$ : Vyťaženie vyučujúcich nekoreluje s publikačnou činnosťou.
- $H_A$ : Vyťaženie vyučujúcich koreluje s publikačnou činnosťou.

V procese dosiahnutia cieľa práce, vyhodnotenie výsledkov práce, sme postupovali podľa navrhnutého postupu, ktorý sme rozdelili do piatich krokov.

Prvým krokom bolo testovanie normality dátového súboru. Vykreslením histogramu pre dostupné dáta započítateľných hodín 8.5 vyučujúceho a publikačných výsledkov 8.4 vyučujúceho sme zistili, že dáta nemajú normálne rozdelenie a histogramy svojím tvarom u oboch sledovaných veličín pripomínajú exponenciálne rozdelenie. Druhým bodom prvého kroku bolo tiež vypočítanie charakteristiky náhodných veličín  $X, Y$ .



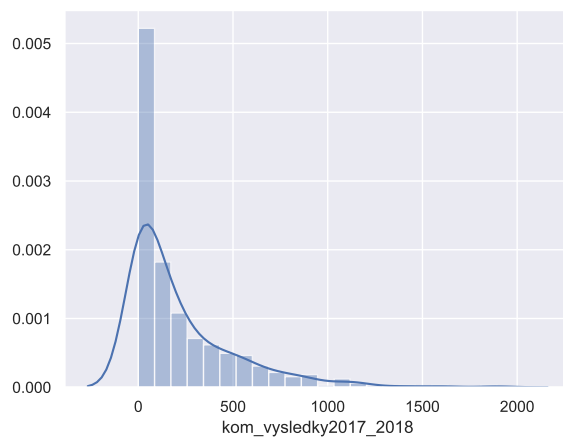
Obrázek 8.4: Histogram dát publikačnej činnosti vyučujúceho 2017-2018.

Podľa tabuľky Charakteristiky publikačnej činnosti 8.5 vidíme, že spriemerovaním hodnôt publikačnej činnosti za roky 2017 a 2018 sa nám znížilo maximum, štandardná odchýlka a mierne poskočil medián.

Charakteristika vyťaženia je popísaná v tabuľke 8.6, ktorá sa však ani po spriemerovaní hodnôt výrazne nezmenila.

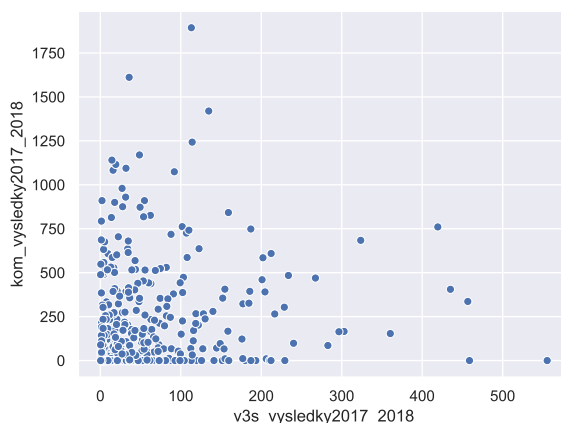
Charakteristika	Publikačné výsledky 2017	Publikačné výsledky 2018
Počet pozorovaní	379	379
Priemer	58.978	59.889
Štandardná odchýlka	95.115	90.705
Minimum	0.056	0.020
Maximum	764.444	779.750
Medián	22.310	26.740
<b>Publikačné výsledky 2017-2018</b>		
Počet pozorovaní		379
Priemer		59.434
Štandardná odchýlka		78.363
Minimum		0.278
Maximum		555.359
Medián		33.050

Tabulka 8.5: Charakteristiky publikačnej činnosti 2017-2018.



Obrázek 8.5: Histogram dát vyťažnosti vyučujúceho 2017-2018.

### 8.3. VYHODNOTENIE ZÁVISLOSTI VYŤAŽENOSTI VYUČUJÚCICH A ICH PUBLIKAČNEJ ČINNOSTI



Obrázek 8.6: Bodový dát publikačnej činnosti a vyťaženia vyučujúceho 2017-2018.

Charakteristika	Vyťaženosť 2017	Vyťaženosť 2018	Vyťaženosť 2017-2018
Počet pozorovaní	1413	1413	1413
Priemer	144.671	146.184	145.427
Štandardná odchýlka	267.301	274.948	268.550
Minimum	0.000	0.000	0.000
Maximum	1918.240	2120.960	1986.160
Medián	0.000	0.000	0.000

Tabulka 8.6: Charakteristiky dát vyťaženia 2017-2018.

Spriemerované výsledky vyťaženia a publikačnej činnosti pracovníkov za roky 2017 až 2018 sme zjednotili do jednej tabuľky podľa osôb, ktorým výsledky patrili. Po zjednotení sa náš dátový súbor, z ktorého budeme vyhodnocovať, obmedzil na 376 výsledkov a nasledujúca tabuľka 8.7 popisuje ako sa zmenila charakteristika dátového súboru. Vidíme, že hodnoty publikačných výsledkov sa výraznejšie nezmenili. Avšak, priemerná hodnota vyťaženia vyučujúceho a medián vyťaženia vyučujúcich sa zmenil na výrazne vyššie hodnoty.

Charakteristika	Publikačné výsledky 2017-2018	Vyťaženosť 2017-2018
Počet pozorovaní	376	376
Priemer	59.889	221.778
Štandardná odchýlka	78.510	291.595
Minimum	0.278	0.000
Maximum	555.359	1893.678
Medián	33.938	107.520

Tabulka 8.7: Zjednotenie dát publikačnej činnosti a vyťaženia za roky 2017-2018.

V druhom kroku sme si vykreslili bodový graf 8.6, ktorý zobrazuje vzájomný vzťah publikačnej činnosti a vyťaženia vyučujúceho. Z grafu môžeme usúdiť, že u veličín by mohla byť slabšia kladná korelácia a teda, že zvyšujúcimi sa hodnotami publikačnej činnosti sa zvyšujú hodnoty započítateľných hodín a analogicky. Keďže sme stanovenou hypotézou očakávali zápornú koreláciu medzi veličinami je tento výsledok naozaj prekvapivý. Podľa grafu, však vykreslené hodnoty nepripomínajú elipsu, a tak nie je splnený predpoklad pre výpočet Pearsonovho korelačného koeficientu.

Následne sme popísali vzájomný vzťah veličín  $X$  a  $Y$  kovarianciou. Na základe výslednej hodnoty  $cov(X, Y) = 2185.874$ , usudzujeme, že vyššie hodnoty publikácií sú zviazané s vyššími hodnotami vyťaženia a analogicky.

Podmienky pre výpočet Pearsonovho korelačného koeficientu splnené nie sú a tak sme vypočítali Spearmanov korelačný koeficient, ktorého hodnota pre veličiny  $X, Y$ , z dostupných dát, je  $cor(X, Y) = 0.159$ . Kladné znamienko korelácie nám hovorí, rovnako ako u hodnoty kovariancie, že so zvyšovaním sa hodnôt publikácií sa zvyšujú hodnoty započítateľných hodín. Podľa Evansovej príručky interpretujeme výslednú hodnotu ako veľmi slabú koreláciu.

Na záver nám ostáva porovnať výslednú  $p$ -hodnotu s hladinou  $\alpha = 0.05$ . Výsledná  $p$ -hodnota = 0.00193 je menšia ako hladina  $\alpha$ , a preto sme zamietli nulovú hypotézu v prospech alternatívnej hypotézy a naším záverom zo zistených výsledkov je, že existuje závislosť medzi publikačnou činnosťou a vyťaženosťou vyučujúcich, kde vyššie hodnoty publikačnej činnosti súvisia s vyššími hodnotami vyťaženosť vyučujúcich. Naša pôvodná stanovená hypotéza očakávala zápornú koreláciu medzi týmito veličinami, čo sa nepotvrdilo. Hodnota korelácie medzi týmito hodnotami je však veľmi nízka a je vhodné ďalej skúmať, prečo medzi nimi existuje závislosť.



# Kapitola 9

## Záver

Práca bola rozdelená podľa pokynov pre vypracovanie práce. Pokyny pre vypracovanie sme potom určili za ciele bakalárskej práce a postupovali sme po jednotlivých krokoch k ich dosiahnutiu.

Prvým krokom bola analýza poskytnutých dátových zdrojov hodnotenia vedecko-výskumnej činnosti, študijných výsledkov a študentského hodnotenia výuky. V kapitole Hodnotenie výuky na vysokých školách 2 sme popísali spôsob hodnotenia kvality výuky na vysokých školách, ktorou ministerstvo sleduje kvalitu vzdelávacej činnosti poskytovanú jednotlivými vysokoškolskými inštitúciami. Toto hodnotenie môžeme deliť na vnútorné a vonkajšie hodnotenie, a to na základe toho, či hodnotenia vypracúva škola, hodnotiacia agentúra alebo Akreditačný úrad. Ďalej sme predstavili študentskú anketu hodnotenia výuky, ktorá sa používa na ČVUT a popísali, aké dáta sme z nej schopný získať. Popísali sme tiež poskytnuté dátové zdroje z aplikácií KOMETA a V3S. V kapitole 3 sme čitateľa oboznámili so spôsobom, akým sa hodnotí vedecko-výskumná činnosť na vysokých školách v ČR a popísali kritéria hodnotenia vedecko-výskumnej činnosti na ČVUT FEL. Podľa týchto kritérií sú vedecko-výskumní pracovníci školy ohodnotení bodmi, ktoré nám poskytujú kvantitatívne vyjadrenie ich činnosti vo vedeckom výskume. Kapitola 4 popisovala študijný informačný systém KOS, slúžiaci k podpore študijných agend na ČVUT, aplikáciu KOMETA pre rozdeľovanie finančných prostriedkov na ČVUT a definovala pojem "*vyťaženosť vyučujúcich*". Na základe tohto pojmu sme kvantifikovali počet hodín, ktoré odzrkadľujú koľko sa pracovníci školy venujú výuke študentov za jeden akademický rok.

Pred samotnou definíciou hypotéz korelácie vedecko-výskumných výsledkov vyučujúcich a ich hodnotenia predmetov a študijných výsledkov študentov sme kapitolou 5 predstavili štatistické metódy, ktoré sme použili v rámci práce. Následne sme na základe týchto metód navrhli postup riešenia pre každú stanovenú hypotézu a navrhli sme si konceptuálny model pre uloženie dátových zdrojov.

Stanovené hypotézy, *H01 - "Publikačné výsledky vyučujúceho, ako vedecko-výskumného pracovníka, nekorelujú s jeho hodnotením v študentskej ankete"*, *H02 - "Študenti s horším študijným priemerom sú v hodnotení vyučujúceho v ankete kritickejší"* a *H03 - "Vyučujúci, ktorí sú viac vyťažení výukou, majú horšiu publikačnú činnosť"*, sme na základe navrhnutého postupu overili za použitia knižníc Pythonu, poskytujúcich funkcie pre štatistickú analýzu. V predchádzajúcej kapitole Vyhodnotenie výsledkov 8 sme zamietli, alebo nezamietli ich platnosť.

Na základe dostupných dát sme dospeli k záverom, že publikačná činnosť vyučujúcuho kladne koreluje s jeho hodnotením v študentskej ankete, čím sme zamietli pôvodnú hypotézu *H01*. Toto zistenie nás vedie k zamysleniu sa nad dôvodmi, pre ktoré sa mohla ukázať korelácia medzi nimi v kladnom smere. Uvažujeme, že vedecko-výskumnému pracovníkovi potom neostáva priveľa času na prípravu prednášok a cvičení a analogicky v role vyučujúceho. V ďalšej práci by bolo vhodné sa zamerať na ďalšie aspekty činnosti vedecko-výskumných pracovníkov, akými sú napríklad počet doktorandov, ktorých vedie, aplikačné výsledky a podobne. Vzťah by bolo vhodné preskúmať aj zo

strany pracovníka ako vyučujúceho, čiže vziať v úvahu náročnosť predmetov, ktoré vyučuje, koľko predmetov vyučuje a mnohé ďalšie.

Taktiež sme zistili, že existuje závislosť medzi hodnotením študenta v študentskej ankete a jeho študijným priemerom. Hodnotenie otázok zo študentskej ankety sme brali v úvahu súhrne a hypotézu sme nevyhodnocovali jednotlivo pre každú otázku. Preto by bolo v ďalšej práci vhodné vyhodnotiť závislosť medzi hodnotením v ankete a študijným priemerom študenta zvlášť pre každú otázku a zároveň vziať v úvahu textové otázky.

Ako poslednú sme overovali hypotézu  $H03$ , ktorou sme chceli preukázať závislosť publikačnej a pedagogickej činnosti pracovníkov školy. Dospeli sme k záveru, že publikačná činnosť slabo koreluje s vyťaženosťou pracovníka. A to dokonca opačným smerom, ako sme predpokladali. Publikačné výsledky pracovníka, podľa zistení, súvisia s vyššími hodnotami počtu hodín výuky. Čiže nám to dáva námet pre ďalší výskum závislosti vedeckej a pedagogickej činnosti pracovníkov vysokých škôl.

# Literatura

- [1] ANDĚL, J. *Statistické metody*. Matfyzpress, 2003.
- [2] DUPAČ, V. – HUŠKOVÁ, M. *Pravděpodobnost a matematická statistika*. Karolinum, 2005.
- [3] EVANS, J. D. *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co, 1996.
- [4] HARARI, Y. N. *Sapiens*. Bazarforlag AS, 2016.
- [5] HUDECOVÁ, Š. *Matematická statistika. Matematická sekce: Matematicko-fyzikální fakulta Univerzita Karlova v Praze*. 2012.
- [6] LIKEŠ, J. – MACHEK, J. *Matematická statistika*. Státní nakladatelství technické literatury, 1988.
- [7] MRKVIČKA, T. – PETRÁŠKOVÁ, V. *Úvod do teorie pravděpodobnosti*. Jihočeská univerzita, 2008.
- [8] PAVLÍK, T. – DUŠEK, L. *Biostatistika*. 1. vydání. Brno: Akademické nakladatelství cerm, sro, 2012. 132 stran.
- [9] RADEK, T. et al. *Marketingový výzkum: postupy, metody, trendy*. Grada Publishing as, 2017.
- [10] REIF, J. *Metody matematické statistiky*. Západočeská univerzita, 2000.

*LITERATURA*

---

## Příloha A

# Obsah přiloženého CD

Zložka	Popis
Persistence_Layer	Obsahuje perzistentní vrstvu realizovanou pomocí relační databáze a soubor <b>README.TXT</b> s popisem pro spuštění.
SQL_Scripts	Zložka obsahuje pg_dump soubor, který je zálohou PostgreSQL a soubor se sérií příkazů pro sestavení relační databáze.
Statistical_Analysis	Obsahuje skripty v programovacím jazyku python, které byly napsány pro vyhodnocování hypotéz statistickými metodami a soubor <b>README.TXT</b> s popisem pro spuštění.
Text	Zložka obsahuje PDF soubor bakalářské práce.