Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra matematiky

Bakalářská práce

Vstupní testy – jednotlivé příklady

Plzeň, 2019

Lenka Chladová

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a výhradně s užitím citovaných pramenů.

V Plzni dne 20. 5. 2019

Lenka Chladová

Poděkování

Děkuji vedoucímu své bakalářské práce panu Mgr. Michalu Frieslovi Ph.D. za poskytnutí dat, odborné vedení, vstřícný přístup a trpělivost během psaní této práce.

Abstrakt

Tato bakalářská práce se zabývá statistickým zpracováním dat ze vstupních testů z matematiky, které píší studenti prvních ročníků některých oborů Západočeské univerzity v Plzni. Práce se zaměřuje na souvislost mezi odpověďmi studentů na jednotlivé příklady a souvislost odpovědí s dostupnými informacemi o studentech v letech 2006 až 2018. Je provedeno základní statistické zpracování dat a následně zformulováno a testováno několik statistických hypotéz.

Klíčová slova: vstupní testy, analýza rozptylu, testování hypotéz

Abstract

This bachelor thesis deals with statistical processing of data from entrance tests in mathematics, which are written by students of the first years of some fields of study of the University of West Bohemia in Pilsen. The thesis focuses on the relationship among students responses to individual examples and the infuence of responses on available information about students in the years 2006 to 2018. Basic statistical data processing is performed and several statistical hypotheses are formulated and tested.

Keywords: entrance tests, analysis of variance, testing hypothesis

Obsah

1 Úvod

Cílem této bakalářské práce je statistické zpracování dat získaných ze vstupních testů z matematiky, které od roku 2006 absolvují studenti prvních ročníků na většině fakult Západočeské univerzity v Plzni. Tento test slouží k zjištění matematických schopností nově příchozích studentů.

Hlavní náplní práce je formulace a testování statistických hypotéz se zaměřením na souvislosti mezi odpověďmi studentů na jednotlivé příklady a souvislost odpovědí s dostupnými informacemi o studentech. Zpracována byla data z let 2006 až 2018. První čtyři roky měl sběr dat jiný charakter, než tomu bylo v pozdějších letech, a proto je většina kapitol rozdělena na dvě části. První část obsahuje data z let 2006 až 2009 a je nazývána jako „stará data“, druhá část se potom zabývá daty od roku 2010 do roku 2018 a je nazvaná „nová data“. Celkem je práce rozdělena do šesti kapitol. Druhá kapitola se zabývá popisem dat a důvody pro jejich rozdělení. V kapitole třetí je uvedeno, jak byla data zpracována a základní popis dostupných informací o studentech. Kapitola čtyři je věnována souvislosti mezi příklady. Jsou zde provedeny testy nezávislosti v kontingenčních tabulkách a vyšetřeny koeficienty korelace. Pátá kapitola se zabývá souvislostí odpovědí na příklady a různých faktorů. K tomu je použita analýza rozptylu a mnohonásobné porovnávání. Konečně šestá kapitola obsahuje shrnutí výsledků a závěr.

2 Data

Vedoucím práce bylo poskytnuto 13 souborů formátu *.csv obsahujících informace o vstupních testech z matematiky z let 2006 až 2018. Tento test píší každý rok studenti prvních ročníků některých oborů na Západočeské univerzitě v Plzni a v některých letech i studenti Přírodovědecké fakulty Jihočeské univerzity v Českých Budějovicích. Ukázka dat v tabulce 1 je oproti původním souborům převrácená a obsahuje pouze některé, v práci nejvíce používané, informace.

Poskytnuté soubory obsahují celkem 22248 záznamů. Ne vždy jsou však dostupné všechny hodnoty a informace. V následujících kapitolách budou studenti, kteří test absolvovali, roztřízeni podle různých dostupných kritérií.

kat	KMA	KMA	KMA	KMA	KMA	KMA
predm	M1	M1	M1	M1	M1	M1
fak	FAV	FAV	FAV	FAV	FAV	FAV
forma	P	P	P	P	P	P
obor	479	2239	479	2239	2239	479
odp	A4123322421	B1144141233	A4143312441	B2243121133	B0242131414	A5144520441
mat	a	a	a	a	a	b
ss	9	9	1	1	3	1
body	8	8	8	7	4	3
pr1	1	1	1	0	0	0
pr10	1	1	1	1	1	1
opak	0	0	1	0	1	0
roku	0	0	1	0	1	0
pohl	M	M	M	M	M	M
rok	2018	2018	2017	2018	2017	2018
kraj	Plzensky	Jihocesky	Plzensky	Plzensky	Karlovarsky	Plzensky

Tabulka 1: Ukázka informací o vstupních testech

2.1 Roky 2006 až 2009

Počty studentů v letech 2006 až 2009 uvádí tabulka 2. Ve zmíněném období máme o studentech menší množství informací než v pozdějších letech. V těchto souborech se vyskytuje 10 sloupců s body získanými v jednotlivých příkladech. Dále sloupec s celkovým počtem bodů a sloupec s variantou testu a odpověďmi na jednotlivé příklady. O studentovi se dozvídáme, na jaké fakultě studuje a v jakém semestru test psal. Předmět, ve kterém byl test absolvován, známe v letech 2006 a 2007 pouze u studentů elektrotechnické fakulty. Od roku 2008 je tento údaj zaznamenán i pro ostatní fakulty s výjimkou fakulty filozofické. Zda studuje kombinovanou nebo prezenční formu studia, víme u studentů od roku 2008.

V roce 2008 se v souboru objevuje 42 studentů ze Zemědělské fakulty a 142 studentů Přírodovědecké fakulty Jihočeské univerzity v Českých Budějovicích. Zde také chybí informace o fakultě a katedře.

Rok	2006	2007	2008	2009
počet studentů	1482	1585	2222	2041

Tabulka 2: Počty studentů v letech 2006 až 2009

2.2 Roky 2010 až 2018

Od roku 2010 probíhal sběr dat jiným způsobem a datové soubory obsahují větší množství informací o studentovi. Získáváme údaje jako pohlaví, počet předchozích studií a počet studovaných let na ZČU. Známý je i studijní program, obor a ročník, který student v době absolvování testu studuje. Dozvídáme se také, jaký typ střední školy navštěvoval a v jakém to bylo kraji. O studentově středoškolském studiu máme navíc i záznam o tom, v jakém roce skládal maturitní zkoušku a zda maturoval z matematiky.

Počínaje rokem 2011 obsahují soubory výše zmíněné informace i o studentech Filozofické fakulty. Není to však pravidlem, stále pro spoustu studentů této fakulty chybí záznamy o ročníku, oboru, programu a formě studia. Často není zaznamenáno ani pohlaví, kraj, typ střední školy nebo rok maturity. Tyto informace postrádáme i u studentů Přírodovědecké fakulty Jihočeské univerzity. U těchto studentů navíc neznáme ani předmět a katedru. Přírodovědecká fakulta se objevuje v souborech z let 2010 až 2013, dále v roce 2015 a 2018. V roce 2015 se navíc objevuje 25 studentů Vyšší odborné školy, u kterých známe opět pouze nejzákladnější informace jako je typ střední školy, kraj, rok maturitní zkoušky a maturita z matematiky.

Ve všech letech 2010 až 2018 jsou u všech studentů uvedeny počty celkových bodů z testu, body získané v jednotlivých příkladech, varianty testu a odpovědi na příklady. Celkové počty studentů v letech 2010 až 2018 zachycuje tabulka 3.

Rok	2010	2011	2012	2013	2014	2015	2016	2017	2018
počet studentů	2331	2263	2008	1883	1327	1480	1221	1241	1164

Tabulka 3: Počty studentů v letech 2010 až 2018

3 Základní charakteristiky

V této kapitole budou popsána vstupní data a jejich základní zpracování. K tomu byly použity softwary Matlab a Excel. Výpočty byly prováděny pouze v programu Matlab. V Excelu byly přidány chybějící sloupce do starších souborů. Navíc v každém dostupném souboru byl přidán sloupec R s údajem, z jakého roku data pochází. V Matlabu byly nejprve všechny soubory načteny do jedné tabulky a potom dále zpracovávány. Veškeré statistické testy byly prováděny na hladině významnosti $\alpha = 5$ %.

Zkoumané soubory obsahují informace z let 2006 až 2018. Test je v každém roce stejný, pouze v letech 2006 a 2007 se lišil osmý příklad. Místo počítání procent zde byl příklad na aritmetickou posloupnost. V prvních dvou letech byly také dvě správné možnosti u příkladu 2. Zadání testu ukazují obrázky 1 a 2. V tabulce 4 jsou uvedeny popisy všech deseti příkladů, které jsou průřezem celého středoškolského studia. Více informací o vstupním testu z matematiky je v [1].

Číslo příkladu	**Popis**	**Správná odpověď**
1	Úprava výrazu	4
2	Porovnávání zlomků	1
3	Analytické vyjádření kružnice	2
4	Řešení kvadratické nerovnice	3
5	Logaritmus	3
6	Graf funkce $sin(x)$	1
7	Funkce a $sin(x)$ a $cos(x)$	2
8	Procenta	4
9	Absolutní hodnota komplexního čísla	3
10	Vzájemná poloha přímek (obecná rovnice)	1

Tabulka 4: Popis příkladů a správné odpovědi (varianta A)

Jméno a příjmení: . . . . **Fakulta:** . . . .

Název střední školy: . . . . **Město (škola):** . . . .

Typ SŠ: (zakroužkujte)

1-Gym	2-Stroj	3-Elek	4-Ekon
5-Zdra	6-OdUč	7-Zem	8-Hotel
9-Inf	10-ObA	11-Stav	

Rok ukončení střední školy: . . . .

Maturita z matematiky | a-ANO | b-NE |

Na vypracování testu je 20 minut.
Alespoň jedna z odpovědí je vždy správná.

Pokud budete chtít pomoci s matematickými a informatickými předměty, je pro vás připraveno „Student Support Centre (SSC)", více na http://ssc.fav.zcu.cz/.

Obrázek 1: Vstupní test z matematiky (první strana)

A

1. Pro všechna $a > 0$ se výraz $\frac{1}{a} a^{\frac{-1}{2}} \sqrt{a^5}$ rovná

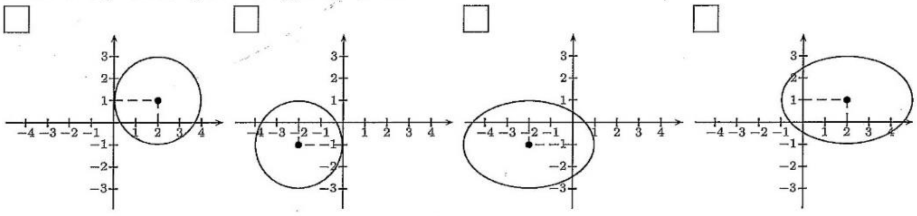
☐ -1 ☐ $\sqrt{a}$ ☐ 1 ☐ a ☐ všechny předchozí odpovědi jsou chybné

2. Mezi čísly $\frac{4}{3}$ a $\frac{6}{5}$ platí vztah

☐ $\frac{4}{3} > \frac{6}{5}$ ☐ $\frac{4}{3} = \frac{6}{5}$ ☐ $\frac{4}{3} \leq \frac{6}{5}$ ☐ $\frac{4}{3} < \frac{6}{5}$ ☐ všechny předchozí odpovědi jsou chybné

3. Rovnice $x^2 + y^2 + 4x + 2y + 1 = 0$ vyjadřuje křivku

☐ ☐ ☐ ☐



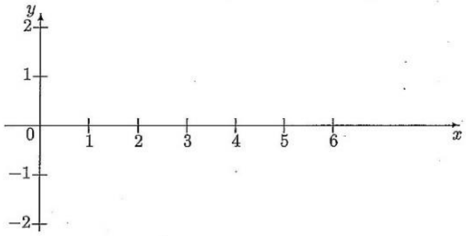
4. Počet celých čísel, která vyhovují nerovnici $x^2 - 7 \leq 0$, je

☐ -5 ☐ 7 ☐ 5 ☐ 3 ☐ všechny předchozí odpovědi jsou chybné

5. Číslo $\log_{10} 5$ je z intervalu

☐ $(-\infty, -1)$ ☐ $\langle -1, 0)$ ☐ $(0, 1)$ ☐ $\langle 1, 10)$ ☐ $(10, \infty)$

6. Do daného souřadnicového systému nakreslete graf funkce $y = \sin x$ na intervalu $\langle 0, 2\pi)$.



7. Je-li pro $\alpha \in \mathbb{R}$ $\sin^2 \alpha = 1$, potom se $\cos^2 \alpha$ rovná

☐ $\frac{1}{2}$ ☐ 0 ☐ 1 ☐ -1 ☐ všechny předchozí odpovědi jsou chybné

8. Po slevě stálo zboží 150 Kč. Sleva byla 25%. Kolik stálo zboží před slevou?

☐ 175 ☐ 187,5 ☐ 125 ☐ 200 ☐ všechny předchozí odpovědi jsou chybné

9. Absolutní hodnota komplexního čísla $z = -3i$ je

☐ -9 ☐ $3i$ ☐ 3 ☐ $\sqrt{3}$ ☐ všechny předchozí odpovědi jsou chybné

10. Přímky o rovnicích $2x - y = 1$, $mx - 2y = 1$ jsou rovnoběžné, pokud

☐ $m = 4$ ☐ $m = -4$ ☐ $m = 1$ ☐ $m = -2$ ☐ všechny předchozí odpovědi jsou chybné

Obrázek 2: Vstupní test z matematiky (druhá strana)

3.1 Varianta testu, odpovědi a body

Sloupec s názvem *odp* obsahuje jedenáct znaků. Prvním z nich je varianta testu. Vyskytuje se zde označení A a B, v roce 2010 se ve 40 případech u Přírodovědecké fakulty vyskytla i varianta C. U každé varianty se však jedná pouze o změnu pořadí příkladů. Toto bylo v práci zohledněno a sloupeček *odp* byl upraven. V případě, že se jednalo o variantu A, byly odpovědi ponechány v původním tvaru. U varianty B byla zaměněna první a druhá polovina odpovědí. V případě nejméně časté varianty C bylo pořadí příkladů vzhledem k variantě A následující: 8, 9, 10, 1, 2, 3, 4, 5, 6, 7.

Dalších deset znaků ve sloupci označuje odpovědi na jednotlivé příklady. Hodnoty 1, 2, 3, 4 a 5 označují zvolenou možnost odpovědi, hodnota 0 znamená, že student v příkladu neodpověděl. Mezi roky 2006 až 2009 se objevuje také hodnota 9, která znamená, že student z příkladu bod nezískal, ale už nevíme, zda odpověděl špatně nebo vůbec.

U příkladů 3 a 6 se v souborech objevuje více různých odpovědí, než připouští test. U příkladu 3 jsou v testu pouze čtyři možnosti odpovědi. Hodnota 5 byla nahrazena hodnotou 0 a to ve 181 případech. V příkladu 6 nejsou na výběr žádné možnosti. Zde má student nakreslit graf funkce $sin(x)$. Hodnota 1 ve sloupci s odpověďmi znamená, že graf nakreslil správně, hodnota 2 špatně a 0 znamená, že příklad vynechal. Odpovědi 3, 4 a 5 jsou tedy považovány za špatně nakreslený graf a byla jim přiřazena hodnota 2. Tato úprava nastala celkem v 15 případech. V tabulce 4 jsou v posledním sloupci zaznamenány správné odpovědi k příkladům.

Každý soubor obsahuje deset sloupců s body získanými v jednotlivých příkladech. Tyto sloupce nabývají pouze hodnot 0 a 1. Jeden bod získá student za správnou odpověď, v opačném případě nezískává žádné body.

3.2 Fakulta

Bylo nutné upravit zkratky názvů fakult. Zejména Fakulta designu a umění Ladislava Sutnara vystupovala v souborech ve třech různých označeních (FUD, FDU, UUD). Název pro všechny byl sjednocen jako FDU. Také Přírodovědecká fakulta Jihočeské univerzity měla dvě označení (PrF, PrF JCU). Pro všechny případy bylo ponecháno označení PrF. Počty studentů jednotlivých fakult ukazuje tabulka 5. Informace o fakultě, na které student studuje, máme ve všech třinácti letech a chybí pouze u tří studentů.

Název	**Zkratka**	**Počet studentů**
Fakulta aplikovaných věd	FAV	4494
Fakulta designu a umění Ladislava Sutnara	FDU	131
Fakulta ekonomická	FEK	6078
Fakulta elektrotechnická	FEL	4099
Fakulta filozofická	FF	1217
Fakulta pedagogická	FPE	1201
Fakulta strojní	FST	3378
Fakulta zdravotnických studií	FZS	274
Přírodovědecká fakulta Jihočeské univerzity	PrF	1306
Vyšší odborná škola	VOS	25
Zemědělská fakulta Jihočeské univerzity	ZF JCU	42
Celkem		**22245**

Tabulka 5: Fakulty a počty studentů

3.3 Obor

Informaci o studovaném oboru poskytují pouze soubory z roku 2010 a novější. Ve 1781 případech nebyl tento údaj vyplněn. Převážně se tak dělo u studentů z Českých Budějovic. Reálně tedy pracujeme s 13137 záznamy, ve kterých se objevuje celkem 81 různých databázových identifikátorů oborů. Podle těchto čísel byly názvy oborů dohledány na univerzitním portále [2]. Nejpočetněji zastoupené obory prezentuje tabulka 6. Obory s četností menší než 200 zde nejsou uvedeny.

Některé obory jsou zastoupeny více různými identifikátory. Jiné číslo je přiřazeno oboru v prezenční formě než tomu stejnému oboru ve formě kombinované. Vstupní test z matematiky tak psali studenti 62 různých oborů.

Databázový identifikátor	Název oboru	Fakulta	Forma	Počet studentů
484	Podniková ekonomika a management	FEK	P	1779
2398	Management obchodních činností	FEK	P	1427
2421	Strojní inženýrství	FST	P	1375
479	Informatika	FAV	P	1074
2145	Elektrotechnika a energetika	FEL	P	745
2147	Komerční elektrotechnika	FEL	P	643
2978	Systémy projektového řízení	FEK	P	586
2146	Elektronika a telekomunikace	FEL	P	490
3170	Stavitelství	FAV	P	385
2148	Technická ekologie	FEL	P	293
3102	Informační systémy	FAV	P	286
2133	Aplikovaná elektrotechnika	FEL	P	286
2436	Matematická studia	FPE	P	282
518	Strojírenství	FST	P	257
475	Kybernetika a řídicí technika	FAV	P	217
2402	Informační management	FEK	P	214
3063	Radiologický asistent	FZS	P	205
2433	Biologie se zaměřením na vzdělávání	FPE	P	203

Tabulka 6: Počty studentů v oborech

3.4 Forma studia

Na Západočeské univerzitě v Plzni mají studenti možnost v některých oborech studovat v prezenční nebo kombinované formě studia. Vstupního testu z matematiky se v letech 2008 až 2018, kdy je tato informace v souborech zaznamenána, zúčastnili studenti z 18 různých kombinovaných oborů. V těchto letech bylo studentů kombinovaného studia 1003, studentů prezenčního studia 15954 a pro 2224 případů záznam chybí.

3.5 Předmět

Informace, v jakém předmětu student test psal, je zaznamenána ve všech souborech. Chybí celkem u 4063 studentů většinou v prvních sledovaných letech. Tabulka 7 zachycuje nejpočetněji zastoupené z celkem devatenácti různých předmětů. Nejčastěji se zde objevují předměty z katedry matematiky Fakulty aplikovaných věd.

Zkratka	Předmět	Počet sudentů
KMA/ZM1	Základy matematiky 1	4977
KMA/M1	Matematika 1	3086
KMA/M1E	Matematika 1	2088
KMA/M1S	Matematika 1	1529
KMA/MS1	Matematika 1	1478
KMT/MMM1	Metody matematického modelování 1	969
KMA/MA1	Matematická analýza 1	936
KMA/ME1	Matematika 1	851
KMA/ZME1	Základy matematiky 1	742
KSS/ZZD	Základy zpracování kvantitativních dat	402

Tabulka 7: Nejpočetněji zastoupené předměty

3.6 Předchozí studium na ZČU

V poskytnutých datových souborech se od roku 2010 objevují sloupce s názvy *opak* a *roku*. První zmíněný sloupec znamená počet předchozích studií na ZČU a druhý udává počet předchozích studovaných let na ZČU.

Informace o počtu předchozích studií chybí u 1831 studentů. Z tabulky 8 vyplývá, že největší počet studentů psalo test z matematiky při svém prvním studiu na Západočeské univerzitě.

Počet studií	Počet studentů
0	11350
1	1429
2	238
3	42
4	18
5	6
6	4

Tabulka 8: Opakování studia a počty studentů

U počtu studovaných let na ZČU chybí tento údaj opět u 1831 studentů. Tabulka 9 zachycuje počty studentů a roky strávené na univerzitě před absolvováním testu z matematiky.

Počet studovaných let	Počet studentů
0	10765
1	1567
2	429
3	192
4	52
5	44
6	22
7	8
8	3
9	2
10	2
11	1

Tabulka 9: Počet studovaných let a počty studentů

3.7 Střední škola

Informace o typu střední školy jsou poprvé dostupné až u studentů, kteří psali vstupní test v roce 2011. Tabulka 10 zachycuje, kolik studentů přichází na univerzitu z jakého typu střední školy. První sloupec tabulky 10 udává číslo, pod kterým je daná střední škola zachycena v datových souborech ve sloupci s názvem *ss*, a poslední řádek zaznamenává celkový počet studentů, pro které známe v letech 2011 až 2018 typ střední školy. Tento údaj nebyl vyplněn u 860 studentů.

Číslo	Typ střední školy	Počet studentů
1	Gymnázium	3945
2	Strojní	1156
3	Elektrotechnický	2400
4	Ekonomický	964
5	Zdravotní	201
6	Odborné učiliště	192
7	Zemědělský	137
8	Hotelový	201
9	Informační	799
10	Obchodní akademie	1297
11	Stavební	435
Celkem		**11727**

Tabulka 10: Typy středních škol a počty studentů

3.8 Kraj

V poskytnutých souborech byly některé kraje značeny různě, bylo proto nutné názvy sjednotit. Pro studenty z Prahy bylo používáno označení *Praha* a *Hl. m. Praha*, ponechán byl kratší z názvů. Podobně tomu bylo u zahraničních studentů, u kterých se vyskytovala označení *Zahranici* a *Zahranicni*. Poslední úprava proběhla u Plzeňského kraje, odkud pocházel největší počet studentů. Tento kraj měl v souborech názvy *Plzensky, Mimo Plzen* a *Plzen* pro větší rozlišení studentů přímo z města Plzeň a z ostatních měst Plzeňského kraje. V této práci se však budeme zabývat jen kraji jako takovými, a proto byl název u všech tří možností sjednocen na *Plzensky*.

Informace o kraji, ve kterém student navštěvoval střední školu se poprvé objevuje v souboru z roku 2010. Do roku 2018 chybí tento záznam ve 2047 případech. Počty studentů z jednotlivých krajů zachycuje tabulka 11.

Kraj	**Počet studentů**
Jihočeský	1624
Jihomoravský	25
Karlovarský	1496
Kralovehradecký	59
Liberecký	80
Moravskoslezský	12
Olomoucký	9
Pardubický	25
Plzeňský	7368
Praha	237
Středočeský	688
Ústecký	972
Vysočina	65
Zlínský	15
Zahraničí	196
Celkem	**12871**

Tabulka 11: Kraje a počty studentů

3.9 Maturita z matematiky

Od roku 2011 máme o studentovi informaci, zda skládal maturitní zkoušku z matematiky. V letech 2012 a dále se v souborech ve sloupci *mat* objevují údaje *a* – maturoval z matematiky a *b*, což znamená opak. V roce 2011, kdy se ještě rozlišovaly různé úrovně maturitních zkoušek právě z předmětu matematika, byly zaznamenávány údaje *a*, *b*, *c*, *d*, *e*, *s*. Písmeno *c* znamenalo, že z matematiky nematuroval, ostatní písmena značí různé úrovně zkoušky. Značení bylo tedy sjednoceno na *a*/*b* (maturoval/nematuroval).

Maturitu z matematiky skládalo 6799 studentů. Naopak 5675 z ní nematurovalo a u 113 studentů údaj chybí.

3.10 Rok maturity

Záznam o tom, v jakém roce student skládal maturitní zkoušku je dostupný v souborech z roku 2010 a novějších. Vyskytují se téměř všechny roky od 1973 až po 2018. Tabulka 12 zaznamenává posledních třináct let. Ostatní roky jsou výrazně méně četné, a proto zde nebudou uvedeny. U 496 studentů informace o roku maturity chybí.

Rok maturity	Počet studentů
2018	868
2017	1053
2016	1002
2015	1277
2014	1181
2013	1673
2012	1880
2011	1902
2010	2133
2009	601
2008	249
2007	125
2006	115

Tabulka 12: Rok maturity a počty studentů

3.11 Pohlaví

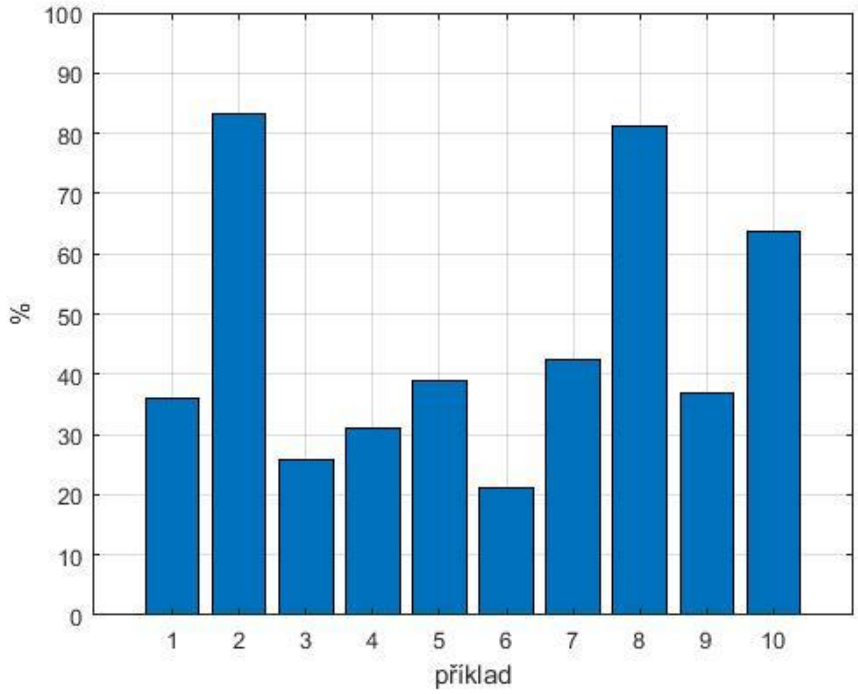
Informaci o pohlaví studenta nemáme ze všech dostupných souborů. Pouze mezi roky 2010 a 2018, kdy se testu zúčastnilo celkem 8445 mužů a 4642 žen. Záznam o pohlaví chybí u 1831 studentů z tohoto období.

4 Jednotlivé příklady

Následující podkapitoly obsahují zhodnocení úspěšnosti jednotlivých příkladů a zkoumání závislosti dvojic příkladů z pohledu dosažených bodů a studenty zvolených odpovědí.

4.1 Úspěšnost

Úspěšnost studentů byla vypočtena jako procentuální podíl počtu studentů, kteří v daném příkladu získali jeden bod, k počtu všech studentů, kteří test absolvovali. Obrázek 3 zachycuje celkovou úspěšnost jednotlivých příkladů. Je patrné, že největší úspěšnost mají příklady 2 a 8 s úspěšností 83,1 % a 81,2 %. Jako nejméně úspěšné se ukázaly příklady číslo 3 (úspěšnost 25,7 %), 4 (úspěšnost 31,0 %) a nejméně úspěšný příklad 6 (úspěšnost 21,2 %), jehož předmětem je nakreslit funkci $y = sin(x)$ do soustavy souřadnic. Úspěšnosti všech příkladů jsou číselně zaznamenány v tabulce 13.

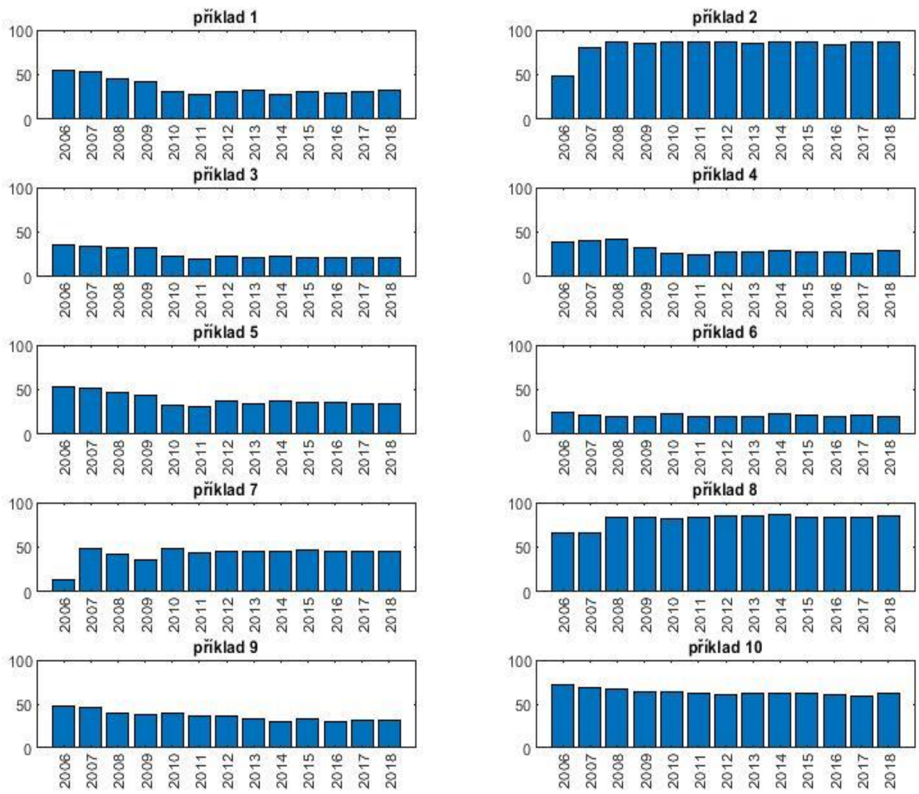


Obrázek 3: Úspěšnost jednotlivých příkladů

Úspěšnost příkladů byla zkoumána i z pohledu roků. Tabulka 13 zachycuje vypočtené nejvyšší a nejnižší úspěšnosti zaokrouhlené na jedno desetinné místo. Obsahuje také informaci o tom, v jakém roce bylo těchto hodnot dosaženo. Rok 2006 figuruje v šesti případech největší úspěšnosti, ale zároveň byla v tomto roce i nejmenší úspěšnost v porovnávání zlomků, u funkcí $sin(x)$ a $cos(x)$ a počítání s procenty. Obrázek 4 potom zachycuje vypočtené úspěšnosti ve všech letech u jednotlivých příkladů.

Příklad	Popis	Nejmenší Úspěšnost	Rok	Největší Úspěšnost	Rok	Celková úspěšnost
1	Výraz	28,5 %	2011	54,2 %	2006	36,0 %
2	Zlomky	48,2 %	2006	87,1 %	2011	83,1 %
3	Kružnice	20,0 %	2011	36,0 %	2006	25,7 %
4	Nerovnice	25,3 %	2011	41,8 %	2008	31,0 %
5	Logaritmus	31,6 %	2011	52,8 %	2006	39,0 %
6	Sinus	19,9 %	2018	24,2 %	2006	21,2 %
7	Sinus a kosinus	12,9 %	2006	49,0 %	2010	42,4 %
8	Procenta	65,1 %	2006	85,8 %	2014	81,2 %
9	Komplexní č.	29,8 %	2014	46,9 %	2006	36,8 %
10	Přímky	58,3 %	2017	72,20 %	2006	63,7 %

Tabulka 13: Úspěšnost jednotlivých příkladů



Obrázek 4: Úspěšnost příkladů v letech

4.2 Závislost bodů

Závislost mezi jednotlivými příklady byla zkoumána za využití kontingenčních tabulek a χ^2 testu nezávislosti v kontingenční tabulce. Více o testu nezávisloti například v [6] na straně 69. Nechť X_i je diskrétní veličina značící počet bodů z i-tého příkladu. Nulovou a alternativní hypotézu pro každou dvojici příkladů formulujeme následovně:

H_0 : veličiny X_i a X_j jsou nezávislé,

H_A : mezi veličinami X_i a X_j existuje závislost,

kde $i = 1, 2, \dots, j, \dots, 10.$

Realizace X_i nabývají hodnot 1 nebo 0. Náhodné výběry realizací náhodné veličiny X_i představuje vždy jeden ze sloupců *pr1*, *pr2*, …, *pr10* v datových souborech. Pro porovnání příkladu s příkladem tedy dostáváme kontingenční tabulky se dvěma řádky a dvěma sloupečky. Kritická hodnota pro všechny tabulky je 3,84. Pro výpočet byla v software Matlab využita funkce *crosstab*. Více o funkci lze nalézt v [3]. Tato funkce vrací, mimo jiné, asymptotickou p-hodnotu pro test nezávislosti v kontingenční tabulce.

Po sestavení kontingenčních tabulek a provedení testů nezávislosti zamítáme nulovou hypotézu H_0 o nezávislosti, na základě vypočtených p-hodnot, pro všechny dvojice příkladů ve všechny letech dohromady. Nejvyšší p-hodnota byla vypočtena z kontingenční tabulky pro dvojici příkladů 2 (porovnávání zlomků) a 9 (absolutní hodnota komplexního čísla) a měla hodnotu 0,0012. Ostatní hodnoty jsou výrazně menší, a proto zde nebudou uvedeny.

Nyní se zaměříme na výsledky příkladů v jednotlivých letech. Tabulka 14 zaznamenává nejvyšší vypočtené p-hodnoty. Hodnoty menší než 0,400 zde nejsou uvedeny. K nezamítnutí nulové hypotézy došlo celkem v padesáti případech. Ve většině případů se p-hodnota vyšší než 5 % týká příkladů 8 a 9, jak je patrné z tabulky 14.

příklad i	příklad j	p-hodnota	Rok
1	8	0,987	2018
2	9	0,949	2016
3	8	0,862	2014
2	9	0,860	2018
2	9	0,704	2014
8	9	0,559	2018
8	9	0,515	2013
8	9	0,495	2014
8	9	0,491	2012
1	2	0,417	2009

Tabulka 14: Nejvyšší p-hodnoty pro příklady v letech

V dalším kroku byl vypočten výběrový korelační koeficient r, který popisuje míru statistické lineární závislosti dvourozměrné náhodné veličiny (X_i, X_j). U veličin X_i a X_j předpokládáme, že jejich realizace nejsou konstantní a jejich směrodatné odchylky jsou tedy kladné. Pomocí vypočtených Následně byl proveden test významnosti výběrového korelačního koeficientu. Testována byla nulová hypotéza, kterou lze formulovat následovně:

$$H_0: \rho = 0,$$

tedy že veličiny X_i a X_j jsou nezávislé, proti oboustranné alternativě:

$$H_1: \rho \neq 0,$$

kde $\rho = \rho(X_i, X_j)$ je korelační koeficient.

Výše zmíněný test významnosti r předpokládá dvourozměrné normální rozdělení pravděpodobnosti. Více o korelačním koeficientu lze nalézt v [5] na straně 61 a v [6] na straně 82.

Jako veličiny X_i a X_j byli postupně voleny body z příkladů, jejich směrodatné odchylky jsou uvedeny v tabulce 15. Naše veličiny X_i a X_j nesplňují podmínku normality, i přes to však byl test použit.

Příklad	Směrodatná odchylka
1	0,480
2	0,375
3	0,437
4	0,463
5	0,488
6	0,409
7	0,494
8	0,391
9	0,482
10	0,481

Tabulka 15: Směrodatné odchylky pro body z příkladů

Pro všechny roky dohromady vycházely hodnoty výběrového korelačního koeficientu malé kladné. Nejvyšší hodnota korelace 0,218 je mezi šestým (vykreslení funkce $sin(x)$) a sedmým (hodnoty funkcí $sin(x)$ a $cos(x)$) příkladem. Pro výpočet koeficientů byla použita funkce *corrcoef* v software Matlab, více o funkci v [4]. Funkce poskytuje p-hodnoty pro test výše zmíněné nulové hypotézy. Tyto p-hodnoty hodnoty vyšly mezi všemi příklady velmi malé. Na 5% hladině významnosti tedy nulovou hypotézu zamítáme a můžeme tvrdit, že korelace jsou významné. Tabulka 16 obsahuje příklad výpočtu korelačních koeficientů pro kombinace prvního (úprava výrazu) příkladu se všemi ostatními. Tato tabulka také obsahuje 95% intervaly spolehlivosti pro vypočtené koeficienty a p-hodnoty. Jelikož je tato hodnota ve všech případech velice malá, nulovou hypotézu zamítáme.

Příklad	Popis	Koeficient	Spodní hranice	Horní hranice	p-hodnota
2	Zlomky	0,061	0,047	0,074	0,000
3	Kružnice	0,199	0,187	0,212	0,000
4	Nerovnice	0,196	0,183	0,209	0,000
5	Logaritmus	0,201	0,188	0,213	0,000
6	Sinus	0,200	0,187	0,212	0,000
7	Sinus a kosinus	0,172	0,159	0,185	0,000
8	Procenta	0,045	0,032	0,058	0,000
9	Komplexní č.	0,112	0,099	0,125	0,000
10	Přímky	0,134	0,121	0,147	0,000

Tabulka 16: Koeficienty korelace pro první příklad

Koeficient korelace byl počítán i pro jednotlivé roky. Tabulka 17 zachycuje kombinace příkladů s nejvyššími vypočtenými p-hodnotami. Jak je patrné i z koeficientů korelace, které jsou v těchto případech velice blízké nule, v jednotlivých letech se objevují dvojice příkladů, pro jejichž úspěšnosti nezamítáme na hladině významnosti $\propto = 5\ \%$ nulovou hypotézu. Většinou se to týká opět příkladů 8 a 9.

Příklad	**Příklad**	**Rok**	**Koeficient**	**Spodní hranice**	**Horní hranice**	***p*-hodnota**
1	8	2018	-0.0005	-0,058	0,057	0,987
2	9	2016	-0,002	-0,058	0,054	0,949
3	8	2014	-0,005	-0,059	0,049	0,862
2	9	2018	0,005	-0,052	0,063	0,860
2	9	2014	0,010	-0,043	0,064	0,705
8	9	2018	0,017	-0,040	0,075	0,560
8	9	2013	0,015	-0,030	0,060	0,515
8	9	2014	0,019	-0,035	0,073	0,496
8	9	2012	0,015	-0,028	0,059	0,491
1	2	2009	0,018	-0,025	0,061	0,417

Tabulka 17: Koeficienty korelace s největší p-hodnotou

4.3 Závislost odpovědí

Jak již bylo zmíněno výše, odpovědi na většinu příkladů nabývají hodnot 1, 2, 3, 4, 5 a hodnota 0 znamená nevyplněnou odpověď. Ve starších souborech se setkáváme i s hodnotou 9, která značí špatnou nebo žádnou odpověď. Následující výpočty byly proto rozděleny na dvě části.

V následujících dvou podkapitolách jsou opět sestaveny kontingenční tabulky a testována nulová hypotéza o nezávislosti náhodných veličin, tentokrát pro odpovědi na příklady. Postup je analogický jako v předchozí kapitole 4.2. Jako veličiny X_i a X_j zde vystupují odpovědi na příklady.

4.3.1 Stará data

Zde budou zkoumány roky 2006 až 2009. V těchto letech se totiž ve sloupci s odpověďmi u každého příkladu objevuje pouze hodnota správné odpovědi nebo hodnota 9. Četnosti odpovědí jsou zaznamenány v tabulce 18, kde řádky *správně* značí počty studentů, kteří v příkladu získali jeden bod.

Pro všechny příklady dostáváme kontingenční tabulky se dvěma řádky a dvěma sloupci. Pro všechny kombinace příkladů vyšly p-hodnoty pro test nulové hypotézy o nezávislosti veličin velmi malé. Největší z nich má hodnotu 0,018 pro příklady 2 a 9. Pro všechny dvojice příkladů tedy H_0 zamítáme.

Odpověď	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
správně	3476	5644	2456	2799	3538
9	3854	1686	4874	4531	3792
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
správně	1559	2620	5579	3092	4958
9	5771	4710	1751	4238	2372

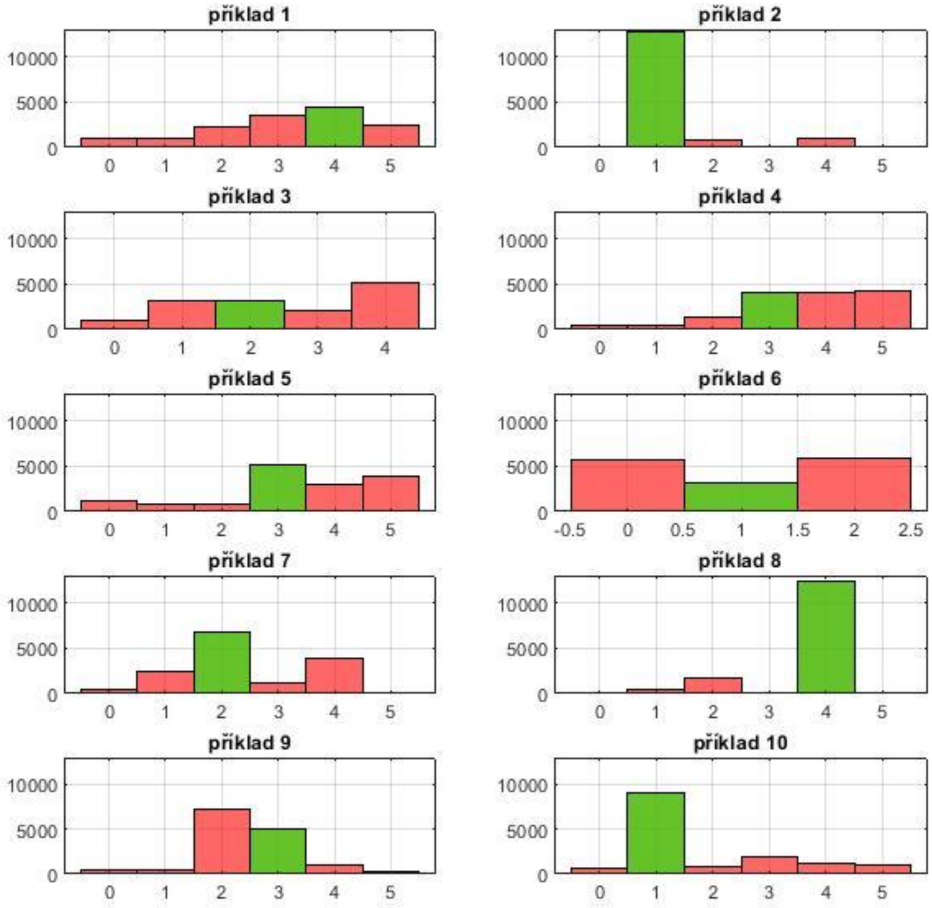
Tabulka 18: Četnosti odpovědí pro stará data

4.3.1 Nová data

V letech 2010 až 2018 už odpovědi nabývají více hodnot. Místo označení 9 je v datových souborech přesná odpověď, kterou student do testu vyplnil. V případě, že student na příklad neodpověděl, byla zaznamenána hodnota nula. I zde se ještě objevilo 233 studentů, kteří měli zaznamenanou odpověď 9. Jednalo se o studenty z Českých Budějovic z roku 2010. Tyto případy nebyly v dalších výpočtech uvažovány. Četnosti odpovědí pro jednotlivé příklady zachycuje obrázek 5 a přesné počty studentů, kteří zvolili jednotlivé odpovědi v příkladech jsou zaznamenány v tabulce 19. Zeleně vyznačené četnosti jsou četnosti správných odpovědí. Červené hodnoty v tabulce 19 znamenají, že četnost špatné odpovědi převýšila počet studentů, kteří na příklad odpověděli správně.

Odpověď	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
0	982	54	1065	391	1219
1	1048	12657	3230	535	728
2	2203	859	3214	1424	762
3	3560	89	2081	4045	5066
4	4475	1014	5095	4047	3011
5	2417	12	-	4243	3899
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
0	5644	506	40	479	634
1	3085	2370	437	545	9071
2	5956	6701	1698	7204	819
3	-	1152	37	5026	1867
4	-	3846	12314	1086	1218
5	-	110	159	346	1076

Tabulka 19: Přesné četnosti odpovědí pro nová data



Obrázek 5: Četnosti odpovědí pro nová data

Pro odpovědi získáváme různě velké kontingenční tabulky. Největší mají šest řádků a šest sloupců, nejmenší má pět řádků a tři sloupce (tabulka pro příklady 3 a 6). K výpočtu již nebyla používána funkce *crosstab*. Jak samotné kontingenční tabulky, tak tabulky s očekávanými četnostmi byly vytvořeny vlastním programem. Pořadí hodnot v řádku i sloupci je 0, 1, 2, 3, 4, 5 (resp. 0, 1, 2, 3, 4 u příkladu 3 a 0, 1, 2 u příkladu 6). Pokud v tabulce očekávaných četností počet hodnot menších než pět přesáhl 20 % celkového počtu hodnot tabulky, bylo nutné sloučit některé sloupce, konkrétně:

- Příklady 1 a 2: spojeny sloupce 1 a 6
- Příklady 2 a 4: spojeny sloupce 1, 2 a 3
- Příklady 2 a 5: spojeny sloupce 1, 2 a 3
- Příklady 2 a 7: spojeny sloupce 1, 4 a 6
- Příklady 2 a 8: spojeny sloupce 1, 2, 3, 4 a 6
- Příklady 2 a 9: spojeny sloupce 1, 2 a 6

- Příklady 2 a 10: spojeny sloupce 1 a 3, 5 a 6
- Příklady 7 a 8: spojeny sloupce 1 a 4
- Příklady 8 a 9: spojeny sloupce 1, 2 a 6
- Příklady 8 a 10: spojeny sloupce 1 a 3, 5 a 6.

Ve výše uvedených dvojicích příkladů jsou hodnoty příslušející prvnímu příkladu z dvojice v řádku a hodnoty druhého ve sloupci kontingenční tabulky.

V kontingenčních tabulkách opět testujeme nulovou hypotézu o nezávislosti dvou veličin, která je stejná jako v podkapitole 4.2. Zde jako veličiny X_i a X_j vystupují odpovědi na příklady. P-hodnoty pro všechny dvojice příkladů vyšly velmi malé, největší z nich měla hodnotu 0,00003 pro dvojici příkladů 2 (porovnávání zlomků) a 9 (absolutní hodnota komplexního čísla). U všech porovnávaných dvojic příkladů tedy dochází k zamítnutí nulové hypotézy. Můžeme se tak domnívat, že odpovědi na jednotlivé příklady jsou na 5% hladině významnosti navzájem závislé.

4.4 Shrnutí

Ve všech letech se ukázaly jako nejúspěšnější příklady 2 (porovnávání zlomků) a 8 (počítání procent). Nejméně úspěšné potom příklady 3 (analytické vyjádření kružnice), 4 (nerovnice) a 6 (graf funkce $sin(x)$).

Za využití kontingenčních tabulek, testu nezávislosti v kontingenčních tabulkách a koeficientů korelace bylo zjištěno, že ve všech letech dohromady jsou všechny dvojice příkladů závislé z pohledu získaných bodů.

Stejné postupy byly aplikovány na body z příkladů rozdělené po jednotlivých letech. Na základě testu nezávislosti v kontingenčních tabulkách pro celkem padesát různých dvojic příkladů nebyla zamítnuta hypotéza o nezávislosti. Většinou se to týkalo příkladů 8 (počítání procent) a 9 (absolutní hodnota komplexního čísla). Nezávislost příkladů 8 a 9 na některých jiných příkladech potvrzují i vypočtené koeficienty korelace, které právě pro kombinace některých příkladů s osmým a devátým příkladem vycházely velmi blízké nule.

Pokud se na příklady díváme z hlediska odpovědí, znamená to pouze rozšíření hodnoty 0 v případě získaných bodů. Pro stará data jsou vlastně odpovědi a body analogické. Jediný rozdíl je ve značení. V případě bodů jsou hodnoty označeny 0 a 1, v případě odpovědí máme hodnotu 9 a správnou odpověď (1, 2, 3, 4 nebo 5, u každého příkladu však pouze jednu). Pro nová data už můžeme mluvit o onom rozšíření, protože zisk 0 bodů se zde rozpadá na nesprávnou odpověď, kterou student v testu zvolil nebo hodnotu 0, pokud neodpověděl vůbec.

Z pohledu odpovědí tedy vyšly v kontingenčních tabulkách všechny dvojice příkladů závislé jak v nových, tak ve starých datech.

5 Příklady a různé faktory

Pro zkoumání vlivů různých faktorů na bodech z jednotlivých příkladů byla využita analýza rozptylu (anglicky Analysis of variance, zkráceně ANOVA). Stručný výklad ANOVY lze nalézt v [7] na straně 14 nebo podrobnější v [6] na straně 130.

Analýza rozptylu je statistický nástroj pro zkoumání závislostí mezi vysvětlovanou a vysvětlující proměnnou. V našem případě považujeme za vysvětlované proměnné body z jednotlivých příkladů. Jako vysvětlující proměnné budeme postupně volit různé faktory. Případy, ve kterých nebyla zkoumaná skutečnost vyplněna, nebudeme v dalších výpočtech uvažovat.

5.1 Formulace ANOVY

Sledujeme faktor A na k úrovních. Nechť pro i-tou úroveň faktoru A existuje n_i (nezávislých) pozorování $y_{i1}, y_{i2}, \dots, y_{in_i}$, u kterých předpokládáme, že jsou náhodným výběrem normálního rozdělení $N(\mu_i, \sigma^2)$. Předpokládáme také nezávislost jednotlivých výběrů. Pozorování potom lze popsat následovně:

$$y_{ip} = \mu_i + \varepsilon_{ip}, \qquad i = 1,2,\dots,k, \qquad p = 1,2,\dots,n_i,$$

kde je

ε_{ip} náhodná složka s normálním rozdělením $N(0, \sigma^2)$.

Jako faktor A v této kapitole vystupují postupně typ střední školy, kraj, pohlaví studenta, fakultu, zda student skládal maturitní zkoušku z matematiky, rok maturitní zkoušky, předmět, rok testu, obor, formu studia, variantu testu, kolik let už student na ZČU studoval a počet předchozích studií na ZČU. Úrovně faktoru jsou hodnoty, kterých právě zkoumaná skutečnost může nabývat. Jako pozorování $y_{i1}, y_{i2}, \dots, y_{in_i}$ vystupují body z jednotlivých příkladů, tedy hodnoty 1 a 0. To znamená, že data, která jsou vyhodnocována v této práci, předpoklady analýzy rozptylu nesplňují. Je jisté, že veličina nabývající pouze hodnot 0 a 1, nesplňuje podmínku normality. I přes to však bude metoda ANOVA použita, protože není příliš citlivá na porušení právě tohoto předpokladu.

Pomocí analýzy rozptylu testujeme nulovou hypotézu H_0 o neúčinnosti faktoru, kterou lze zformulovat takto:

$$H_0\colon\ \mu_1 = \mu_2 = \ldots = \mu_i = \ldots = \mu_k,$$

proti alternativní hypotéze:

$$H_1\colon\ alespoň\ jedna\ z\ rovností\ v\ H_0\ je\ porušena.$$

Pokud dojde k zamítnutí nulové hypotézy, přijímáme H_1. Jinak řečeno, vysvětlovaná proměnná na faktoru závisí.

V případě zamítnutí nulové hypotézy je přirozené se ptát, které úrovně faktoru se od sebe liší. Zde přichází na řadu mnohonásobné porovnávání. Z důvodu nevyváženého třídění byla použita Scheffého metoda. Tato metoda testuje následující nulovou hypotézu:

$$H_0\colon\ \mu_i - \mu_j = 0,$$

proti alternativní hypotéze:

$$H_1\colon \mu_i - \mu_j \neq 0.$$

V následujících kapitolách bude popsáno, jaký vliv mají jednotlivé faktory na body z příkladů a také jaké úspěšnosti dosahují příklady z hlediska úrovní faktorů. Úspěšnost byla vypočtena stejně jako v kapitole 4.1, tedy jako podíl studentů, kteří v příkladě získali jeden bod k počtu všech studentů příslušejících k dané úrovni faktoru. U každého faktoru byl výpočet analýzy rozptylu rozdělen na stará data (roky 2006 až 2009) a nová data (roky 2010 až 2018). Ze starých dat byly do ANOVY zařazeny pouze faktory varianta testu, fakulta, forma studia a rok. V případě, že právě zkoumaný faktor nebyl vyplněn, nebude takový student do výpočtu zařazen. Pro výpočet analýzy rozptylu ve výpočetním prostředí Matlab použita funkce *anova1*, její podrobný popis lze nalézt v [8]. Pokud došlo k zamítnutí nulové hypotézy, bylo provedeno mnohonásobné porovnání pomocí funkce *multcompare*, jejíž popis je v [9].

5.2 Stará data

- **Varianta testu**

Faktor varianta testu nabývá dvou hodnot A nebo B, tedy počet úrovní faktoru $k = 2$. Jako pozorování $y_{11}, y_{12}, \dots, y_{1n_1}$ zde vystupují body z příkladů příslušející studentům, kteří psali test ve variantě A, pozorování $y_{21}, y_{22}, \dots, y_{2n_2}$ potom přísluší studentům, kteří patřili k variantě B.

Úspěšnosti příkladů v jednotlivých variantách testu zachycuje tabulka 20. Pro příklady 5 (logaritmus) a 6 (graf funkce $sin(x)$) vyšly p-hodnoty pro test nulové hypotézy o shodnosti rozptylu menší než 5 %. Konkrétně 0,006 a 0,044. U těchto příkladů tedy správnost studentovy odpovědi závisí na variantě testu, který psal a jejich úspěšnosti jsou v tabulce 20 zachyceny červeně. Pro ostatní příklady na faktoru varianta testu nezáleží.

Varianta	**Příklad 1**	**Příklad 2**	**Příklad 3**	**Příklad 4**	**Příklad 5**
A	46,62	77,63	33,52	38,75	46,68
B	48,25	76,35	33,49	37,60	49,92
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
A	22,21	35,96	76,29	43,14	67,93
B	20,29	35,52	75,93	41,19	67,34

Tabulka 20: Úspěšnosti příkladů v závislosti na variantě testu (v %)

- **Forma studia**

U faktoru forma studia a všech ostatních je postup analogický jako u faktoru varianta testu.

Nulovou hypotézu nezamítáme pro příklady 1 ($p = 0{,}884$), 2 ($p = 0{,}894$), 8 ($p = 0{,}714$) a 9 ($p = 0{,}058$). Pro ostatní příklady je faktor formy studia statisticky významný. Úspěšnosti příkladů v závislosti na formě studia jsou zachyceny v tabulce 21. Pro ty příklady, u kterých získané body závisí na formě studia, jsou úspěšnosti vyznačeny červeně.

Forma	**Příklad 1**	**Příklad 2**	**Příklad 3**	**Příklad 4**	**Příklad 5**
K	44,62	87,18	22,82	31,03	37,44
P	44,23	86,94	35,28	39,30	48,54
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
K	13,33	35,13	83,33	34,87	63,08
P	23,21	40,85	84,05	39,83	38,31

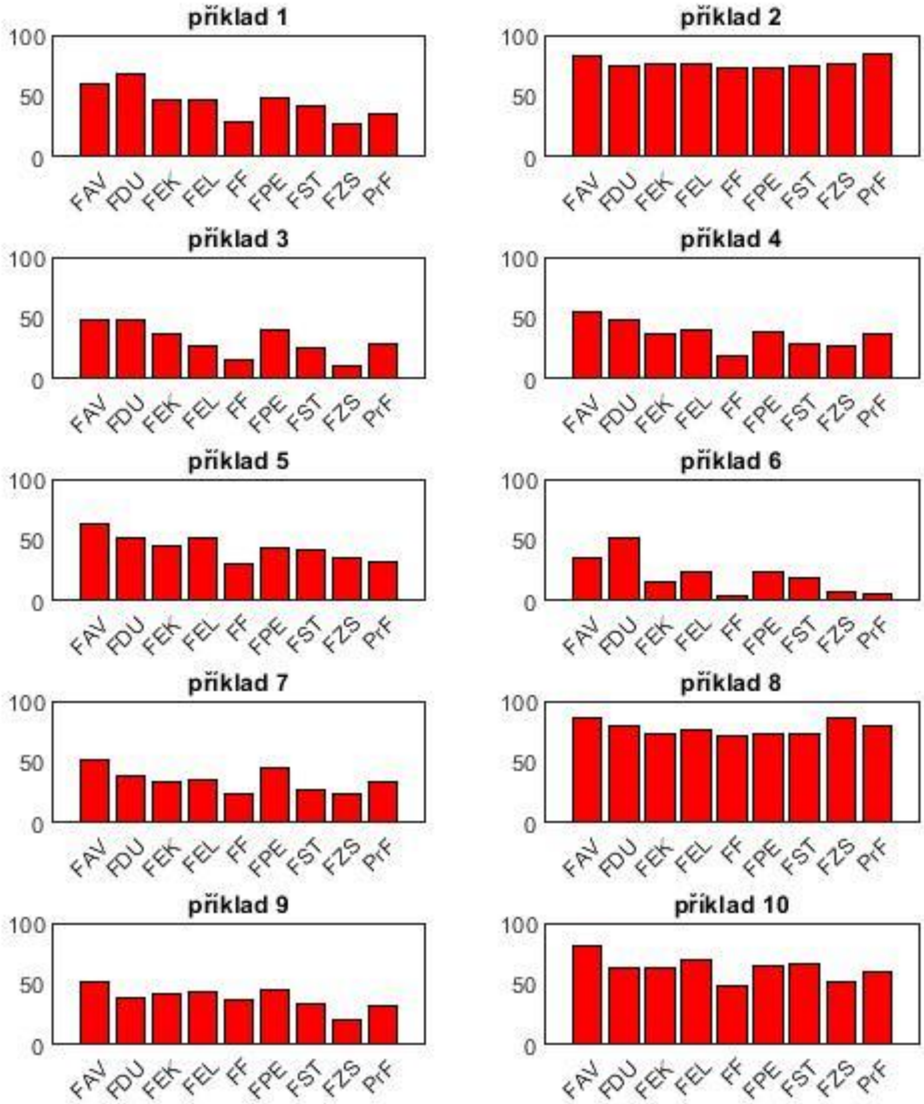
Tabulka 21: Úspěšnosti příkladů v závislosti na formě studia (v %)

- **Fakulta**

  Pro všechny příklady vyšly p-hodnoty velice malé, fakulta je tedy statisticky významný faktor. Tabulka 22 zaznamenává mnohonásobné porovnávání pro příklady 2 (porovnávání zlomků) a 8 (procenta). Sloupce s p-hodnotami přísluší testu nulové hypotézy o shodě středních hodnot v mnohonásobném porovnání pro příslušné úrovně faktoru. Pro dvojice fakult v tabulce 22 tedy nulovou hypotézu zamítáme a na hladině významnosti $\propto = 5$ % o těchto dvojicích řekneme, že se jejich střední hodnoty od sebe liší. U příkladů zaznamenaných v tabulce 22 vyšlo nejméně odlišných dvojic fakult. V ostatních příkladech se lišilo více fakult, většinou mezi nimi však převládala Fakulta aplikovaných věd, jejíž úspěšnost je téměř ve všech případech nejvyšší, jak je patrné i z obrázku 6. Jednou z výjimek je příklad 6 (graf funkce $sin(x)$), kde úspěšnost studentů Fakulty designu a umění Ladislava Sutnara převýšila úspěšnost studentů FAV o 15 %.

Příklad 2			**Příklad 8**		
Fakulta i	**Fakulta j**	**p-hodnota**	**Fakulta i**	**Fakulta j**	**p-hodnota**
FAV	FEL	0,007	FAV	FEL	0,000
FAV	FST	0,001	FAV	FST	0,000
FAV	FEK	0,034	FAV	FEK	0,000
-	-		FAV	FF	0,000

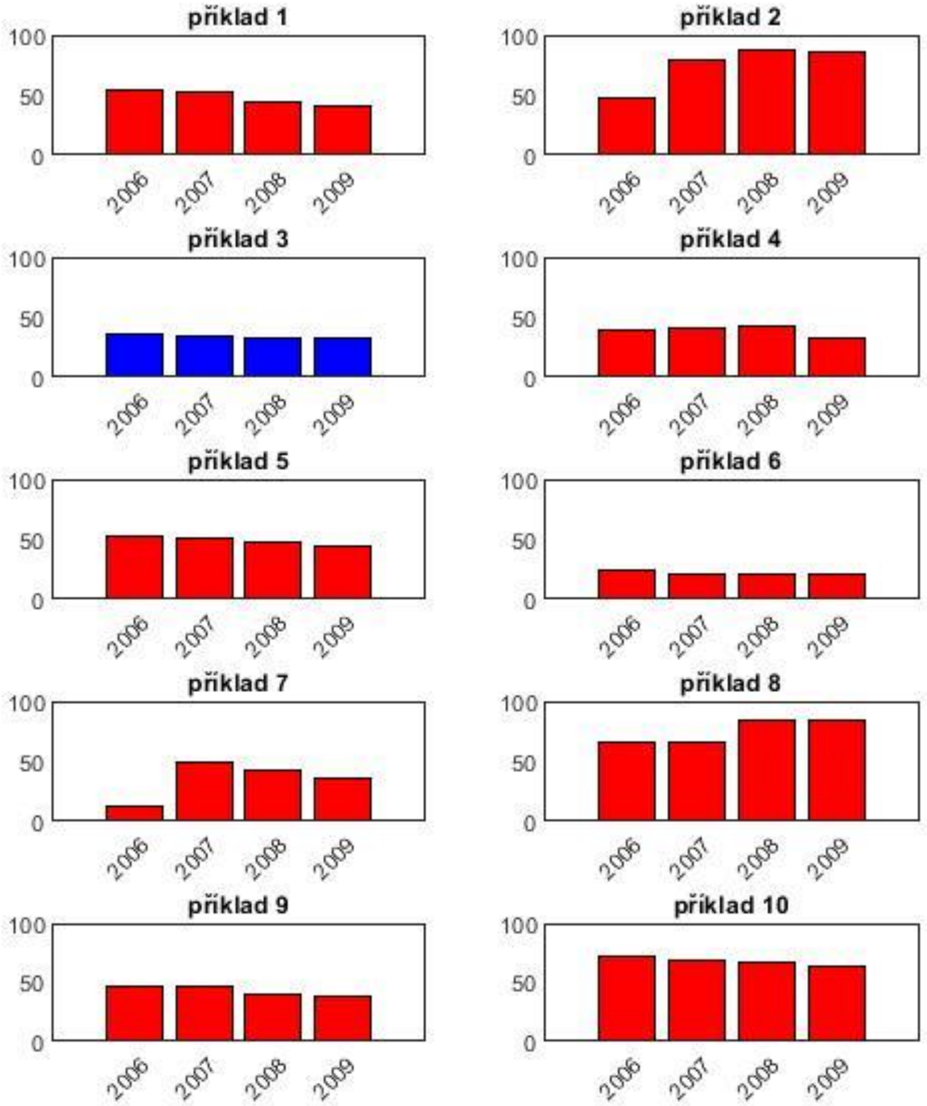
Tabulka 22: Mnohonásobné porovnávání pro příklady 2 a 8



Obrázek 6: Úspěšnosti příkladů pro faktor fakulta (v %)

- **Rok testu**

Na základě vypočtených p-hodnot pro analýzu rozptylu nezamítáme nulovou hypotézu pouze u příkadu 3 ($p = 0{,}078$). V ostatních případech H_0 zamítáme. Rok je tedy statisticky významným faktorem. V mnohonásobném porovnávání bylo zjištěno, že se od sebe neliší střední hodnoty získaných bodů z let 2006 a 2007 v téměř všech příkladech (výjimkou jsou příklady 2 a 7). Dále se od sebe neodlišují roky 2008 a 2009 v příkladech 1, 2, 6, 8, 9 a 10, jak je patrné z úspěšností jednotlivých příkladů, které jsou zachyceny na obrázku 7.



Obrázek 7: Úspěšnosti příkladů pro faktor rok testu (v %)

5.3.2 Nová data

- **Varianta testu**

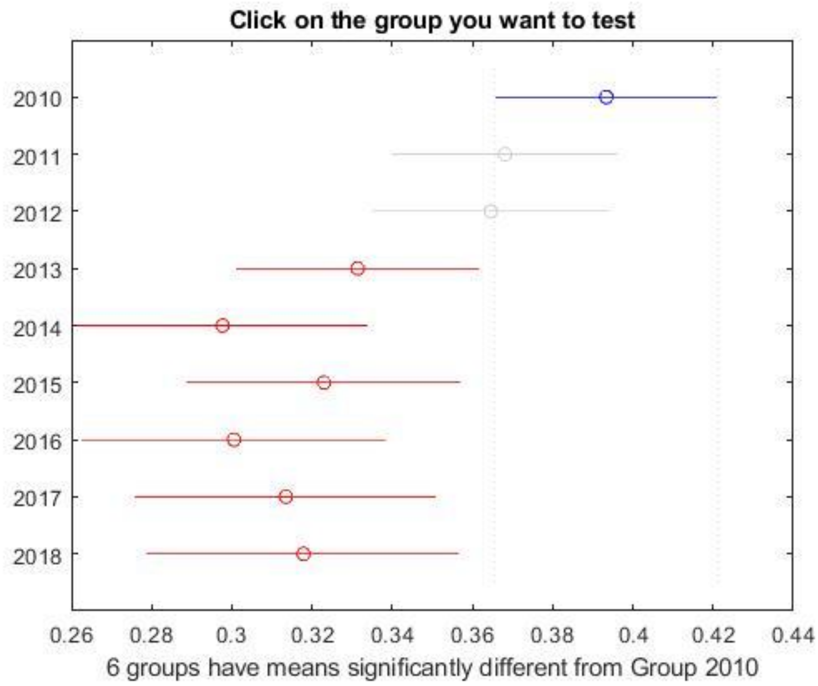
V datech od roku 2010 do roku 2018 získané body ve všech příkladech nezávisí na variantě testu. P-hodnoty pro nulovou hypotézu analýzy rozptylu vyšly ve všech příkladech větší než zvolená hladina významnosti. Nejmenší z nich má hodnotu 0,101 pro příklad 7 (funkce $sin(x)$ a $cos(x)$), ostatní jsou ještě vyšší a nulovou hypotézu tedy zamítáme pro všechny příklady. Tabulka 23 zachycuje úspěšosti jednotlich variant. Zde si můžeme všimnout, že se od sebe úspěšnosti variant A a B příliš neliší.

Varianta	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
A	29,99	85,74	21,45	27,23	33,85
B	30,92	86,50	22,17	27,70	35,01
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
A	21,36	45,58	83,41	34,32	61,95
B	20,85	45,75	84,05	34,04	61,55

Tabulka 23: Úspěšnosti příkladů pro faktor varianta testu (v %)

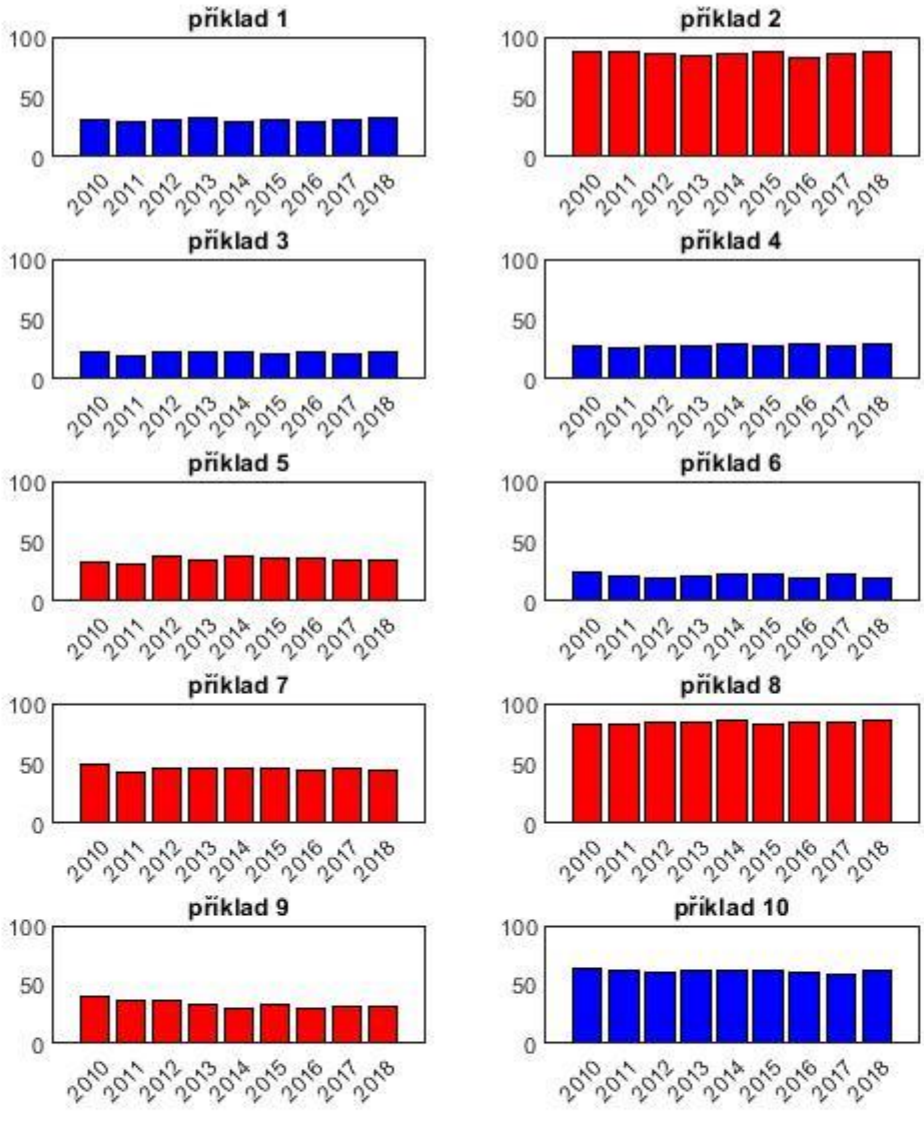
- **Rok testu**

  Rok, ve kterém studenti psali test z matematiky, je statisticky významný faktor pro polovinu příkladů. Obrázek 9 zachycuje úspěšnosti roků v jednotlivých příkladech. Ty příklady, u kterých byla zamítnuta nulová hypotéza o shodnosti středních hodnot, jsou v obrázku 9 vyznačeny červeně. U ostatní příkladů p-hodnota převyšují zvolenou hladinu významnosti, nezamítáme tedy u nich nulovou hypotézu a jsou v obrázku vyznačeny modře.



Obrázek 8: Mnohonásobné porovnání pro příklad 9

Mnohonásobné porovnání ukázalo, že se střední hodnoty počtu bodů liší pouze pro devět dvojic v příkladě 9 (absolutní hodnota komplexního čísla), jak ukazuje obrázek 8. Modře vyznačen je zde střední hodnota roku 2010 a červeně jsou označeny roky, jejichž střední hodnota je odlišná.

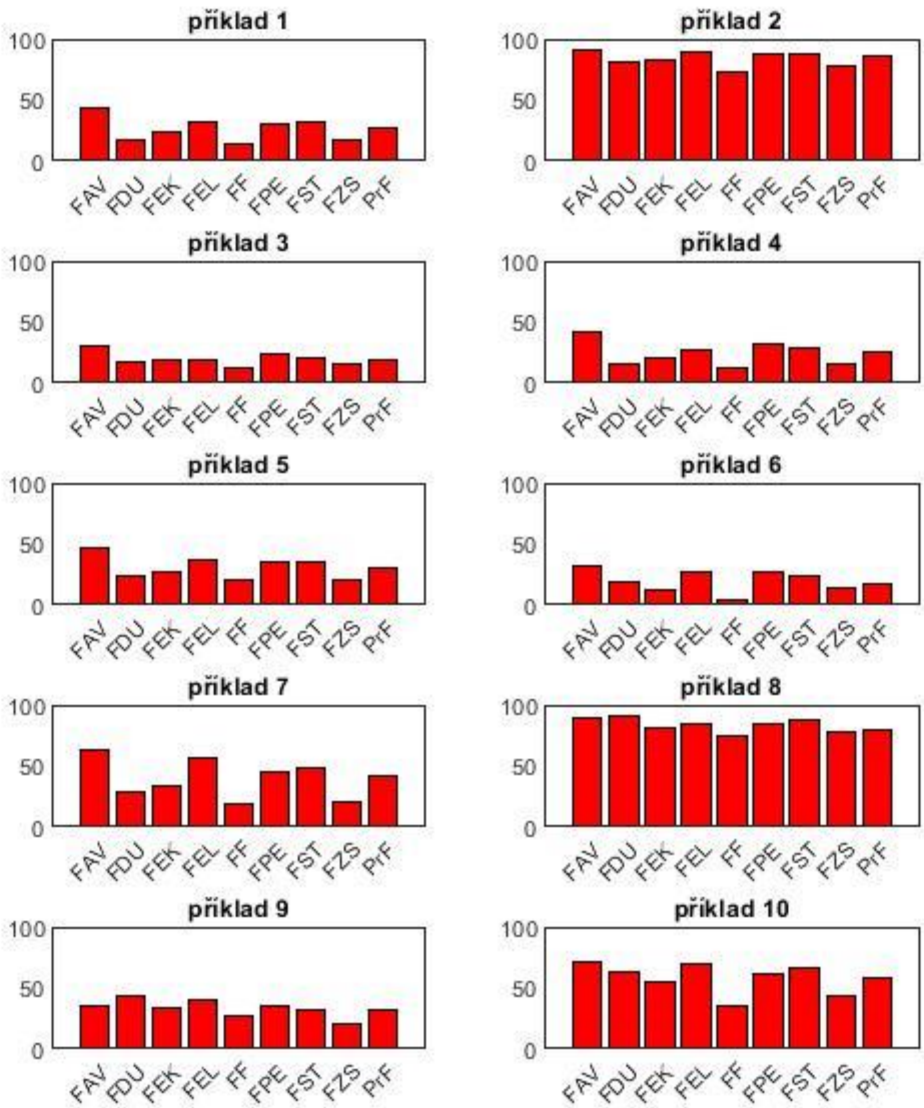


Obrázek 9: Úspěšnosti příkladů pro faktor rok testu (v %)

- **Fakulta**

  Faktor fakulta opět považujeme za statisticky významný pro všechny příklady. Všechny p-hodnoty pro analýzu rozptylu vyšly velice malé.

  Mnohonásobné porovnání ukázalo, že se ve všech příkladech od sebe navzájem liší větší množství dvojic fakult. Jak ukazuje obrázek 10, FAV dosahuje téměř ve všech příkladech největší úspěšnosti.

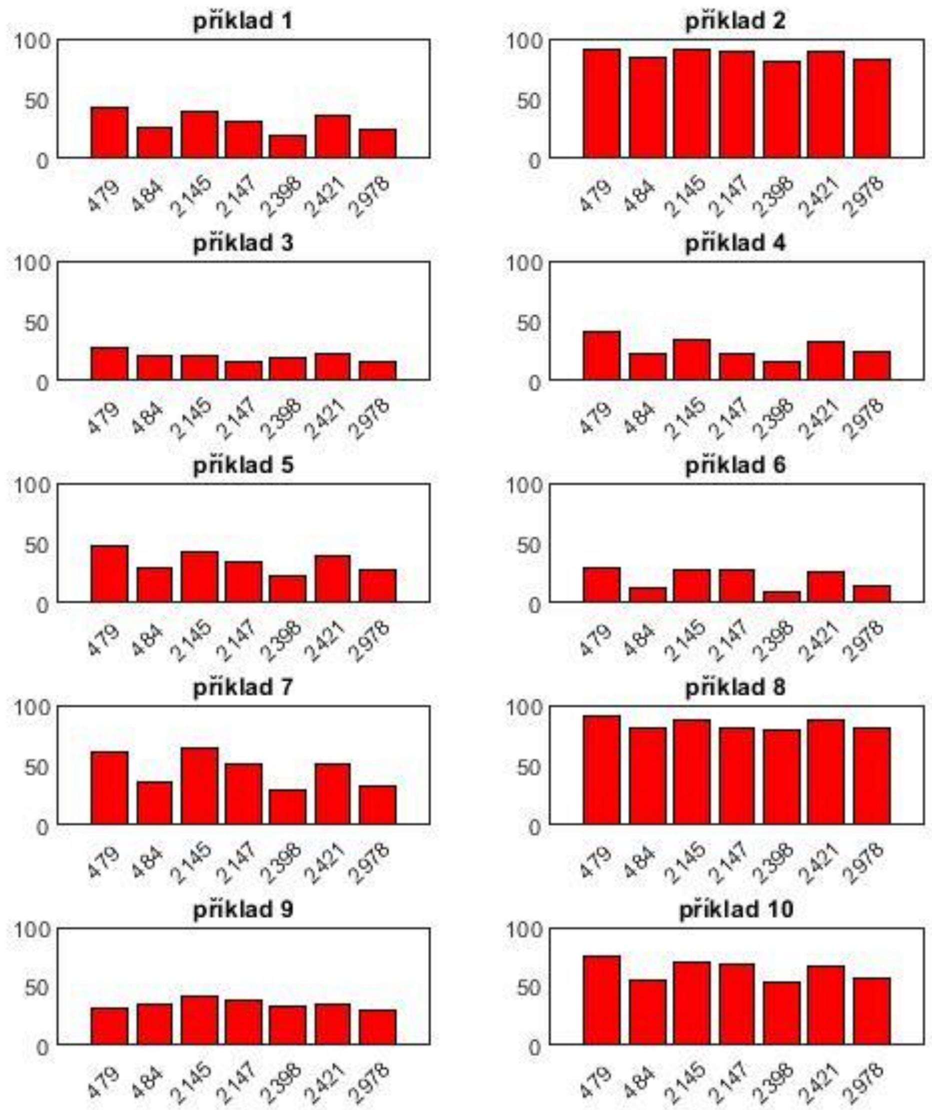


Obrázek 10: Úspěšnosti příkladů pro faktor fakulta (v %)

- **Obor**

  Studovaný obor je statisticky významný u všech příkladů. Všechny vypočtené *p*-hodnoty pro test nulové hypotézy o shodnosti rozptylů jsou velice malé. Obrázek 9 ukazuje úspěšnosti pro všechny předměty. Je zde zaznamenáno pouze sedm nejčastějších oborů. Tabulka 24 přiřazuje k databázovým identifikátorům, použitým v obrázku 11, názvy oborů.

  Mnohononásobné porovnání ukázalo, že se v příkladech 2,8 a 9 neliší žádné dvojice oborů. V ostatních příkladech nebylo mnoho odlišných středních hodnot, nejvíce v příkladu 7, kdy se lišilo 17 různých dvojic.



Obrázek 11: Úspěšnosti příkladů pro faktor obor (v %)

Databázový identifikátor	Název oboru
484	Podniková ekonomika a management
2398	Management obchodních činností
2421	Strojní inženýrství
479	Informatika
2145	Elektrotechnika a energetika
2147	Komerční elektrotechnika
2978	Systémy projektového řízení

Tabulka 24: Názvy nejčetnějších oborů

- **Forma studia**

  Ve formě studia se zabýváme tím, zda je studium kombinované nebo prezenční. Nulovou hypotézu o shodnosti rozptylů zamítáme u příkladu 9 (absolutní hodnota komplexního čísla), kde je vypočtená p-hodnota 0,000, a u příkladu 10 (vzájemná poloha přímek) s p-hodnotou 0,048. U ostatních příkladů H_0 nezamítáme.

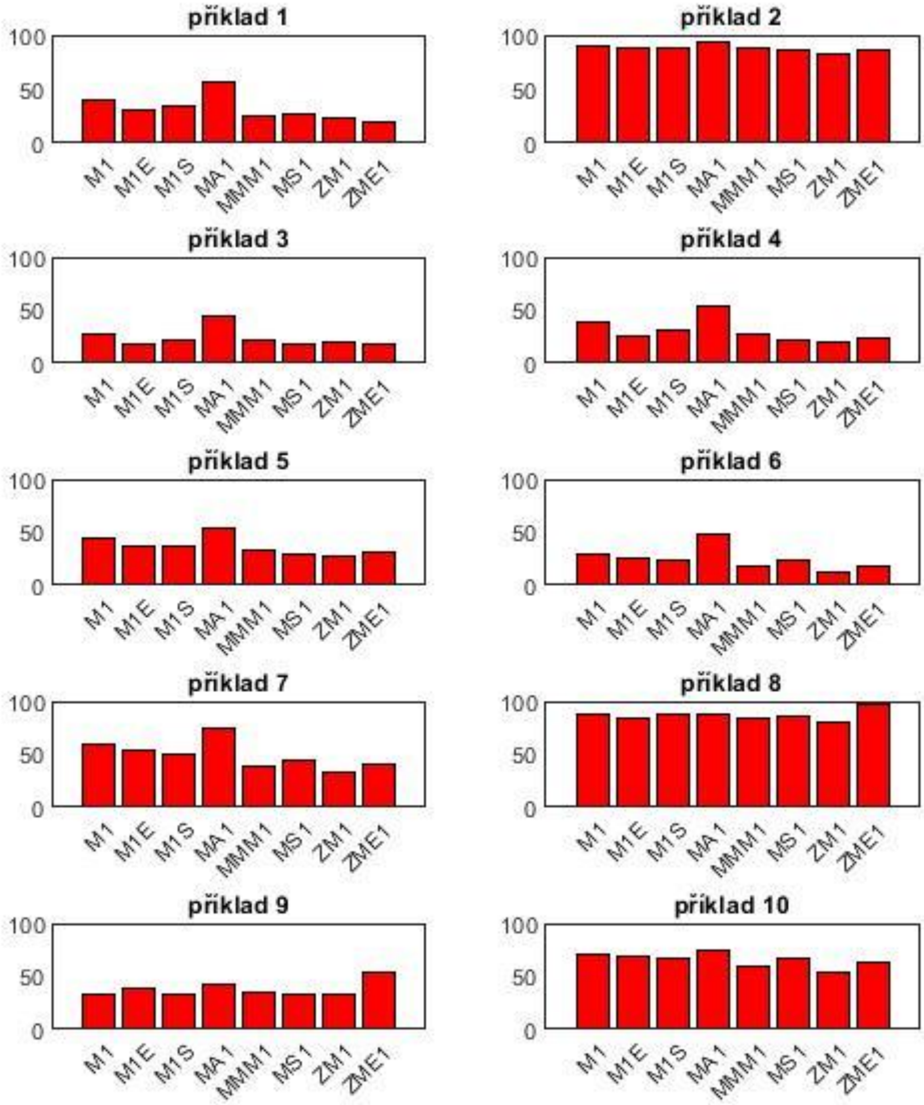
  Tabulka 25 zachycuje úspěšnosti pro jednotlivé příklady. Červeně vyznačené hodnoty jsou odhady u příkladů, u kterých je forma studia statisticky významným faktorem.

Forma	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
K	28,06	86,30	21,37	27,57	32,30
P	31,58	86,71	22,46	28,37	35,60
-	**Příklad 6**	**Příklad 7**	**Příklad 8**	**Příklad 9**	**Příklad 10**
K	22,84	46,00	85,64	25,29	59,54
P	22,11	47,20	84,45	35,27	63,49

Tabulka 25: Úspěšnosti příkladů pro faktor forma studia (v %)

- **Předmět**

  Předmět, ve kterém studenti vstupní test z matematiky psali, je statisticky významným faktorem pro všechny příklady. Všechny vypočtené p-hodnoty jsou velice malé. Pro všechny příklady bylo provedeno mnohonásobné porovnávání. U příkladu 9 jsou 4 dvojice předmětů, které se navzájem odlišují svými středními hodnotami. To je také nejmenší množství odlišných předmětů. U ostatních příkladů se liší větší množství dvojic. Ve dvojicích příkladů s odlišnými středními hodnotami ve všech příkladech vystupuje předmět MA1. Například v příkladech 1, 3, 4, 6 nebo 7 jsou studenti tohoto předmětu výrazně úspěšnější než ostatní studenti. Naopak u příkladů 2, 8 nebo 10 nejsou rozdíly tolik patrné. Obrázek 12 zachycuje osm nejpočetněji zastoupených předmětů.



Obrázek 12: Úspěšnosti příkladů pro faktor předmět (v %)

- **Předchozí studium na ZČU**

  Zde zkoumáme faktor počtu předchozích studií na ZČU a faktor počtu předchozích studovaných let na ZČU. Na základě hodnot vypočtených analýzou rozptylu, které obsahuje tabulka 26, je první z faktorů statisticky významný pro příklady 1, 3, 4, 6, 7, 8 a 10. Pro druhý faktor zamítáme nulovou hypotézu o shodnosti rozptylu u stejných příkladů, jako pro faktor předchozích studií, s výjimkou příkladu 8. Tabulky 27 a 28 zachycují úspěšnosti pro nejčastěji zastoupené úrovně obou faktorů.

Příklad	Popis	Předchozí studia	Předchozí roky
1	Výraz	0,001	0,000
2	Zlomky	0,305	0,138
3	Kružnice	0,029	0,002
4	Nerovnice	0,000	0,000
5	Logaritmus	0,123	0,105
6	Sinus	0,000	0,000
7	Sinus a kosinus	0,001	0,000
8	Procenta	0,001	0,134
9	Komplexní č.	0,145	0,673
10	Přímky	0,002	0,002

Tabulka 26: Výsledky *p*-hodnot pro předchozí studia a předchozí studované roky

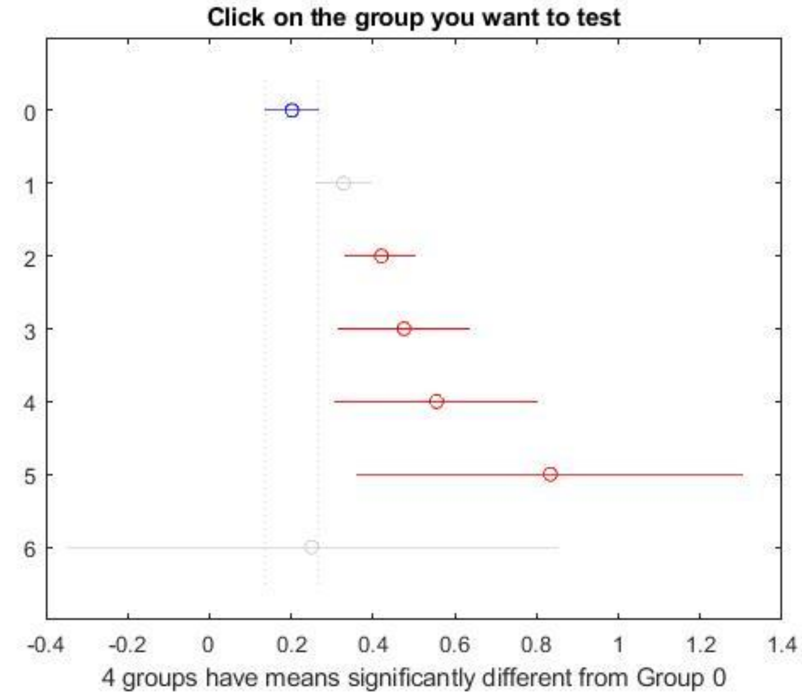
Předchozí studia	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
0	31,24	86,45	22,77	27,60	35,47
1	30,72	88,10	19,10	31,56	34,08
2	37,82	89,08	23,11	36,55	36,55
3	47,62	92,86	28,57	47,62	47,62
-	Příklad 6	Příklad 7	Příklad 8	Příklad 9	Příklad 10
0	20,15	46,56	84,00	35,22	62,71
1	32,75	49,69	87,05	32,12	65,78
2	42,02	55,04	90,34	34,03	73,11
3	47,62	64,29	90,48	35,71	78,57

Tabulka 27: Úspěšnosti příkladů pro faktor opakovaného studia (v %)

Předchozí studované roky	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
0	31,22	86,41	22,66	27,45	35,44
1	28,46	86,85	19,02	29,74	32,87
2	38,69	89,98	24,01	37,53	39,86
3	39,06	92,71	27,08	34,38	39,06
-	Příklad 6	Příklad 7	Příklad 8	Příklad 9	Příklad 10
0	19,44	46,02	83,99	35,25	62,69
1	30,82	49,90	86,73	32,10	63,88
2	40,09	56,64	85,78	35,43	71,56
3	38,54	57,29	88,54	33,85	69,79

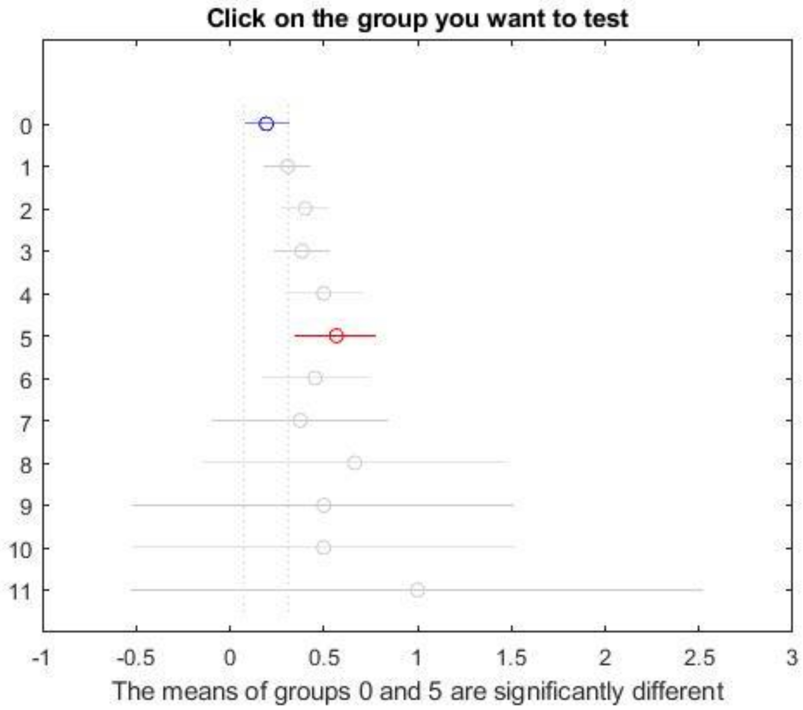
Tabulka 28: Úspěšnosti příkladů pro faktor předchozí studované roky (v %)

Mnohonásobné porovnání v případě počtu předchozí studií ukázalo, že zde není mnoho odlišností. Pouze pro příklad 6 se vyskytlo 5 dvojic u kterých došlo k zamítnutí nulové hypotézy, jak ukazuje obrázek 13.



Obrázek 13: Mnohonásobné porovnání pro příklad 6 (počet předchozích studií)

Mnohonásobné porovnání pro předchozí studované roky dopadlo podobně jako u předchozího faktoru v tom smyslu, že ani zde nebylo zjištěno příliš mnoho odlišností. U příkladu 6 se opět liší 5 dvojic středních hodnot, rozdíly zde však nejsou tolik patrné, jako u předchozího faktoru, jak ukazuje obrázek 14.



Obrázek 14: Mnohonásobné porovnání pro příklad 6 (počet předchozích studovaných let)

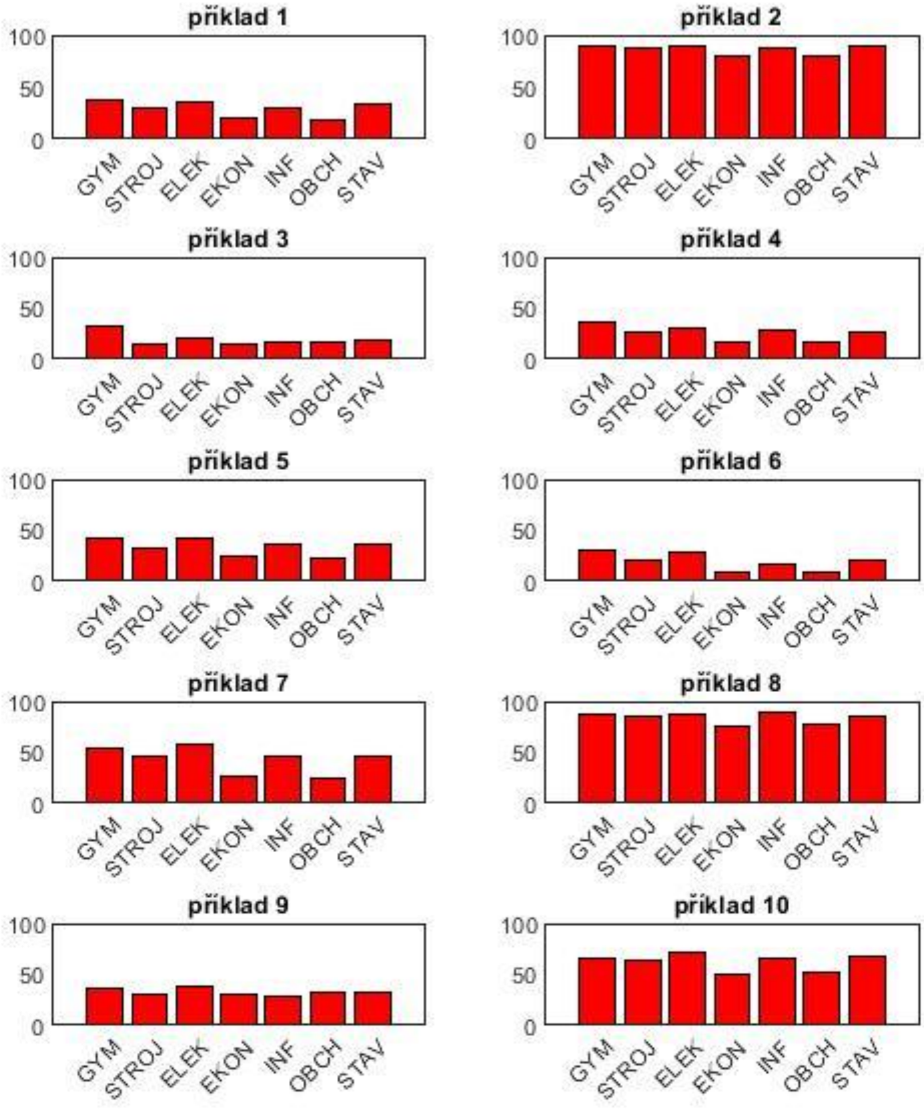
- **Střední škola**

  Střední škola, kterou student absolvoval před nástupem na vysokou školu je významným faktorem pro všechny příklady. P-hodnoty vycházejí velice malé. Úspěšnosti nejčastěji zastoupených středních škol v jednotlivých příkladech jsou zachyceny v obrázku 15, který ukazuje, že pozorované průměry gymnazistů jsou vyšší než průměry ostatních středních ve většině příkladů.

  Mnohonásobné porovnání ukázalo, že se ve většině příkladů je více dvojic středních škol, jejichž střední hodnoty se od sebe odlišují. Pro příklad 9 vyšly pouze dvě dvojice typů středních škol, u nichž nulovou hypotézu zamítáme, a to:

  Elektrotechnický a zemědělský ($p = 0{,}047$)

  Elektrotechnický a informatika ($p = 0{,}011$).



Obrázek 15: Úspěšnosti příkladů pro faktor střední škola (v %)

- **Kraj**

Kraj je statisticky významným faktorem pro všechny příklady s výjimkou příkladu 8. Poslední sloupec v tabulce 29 uvádí vypočtené p-hodnoty. V mnohonásobném porovnání bylo zjištěno, že se v jednotlivých příkladech odlišují pouze malé počty dvojic krajů. Lišili se střední hodnoty následujících krajů:

Příklad 1: Jihočeský a Karlovarský ($p = 0{,}004$)

Příklad 5: Jihočeský a Karlovarský ($p = 0{,}001$)

Příklad 6: Jihočeský a Karlovarský ($p = 0{,}008$)

Příklad 7: Jihočeský a Karlovarský ($p = 0{,}000$)

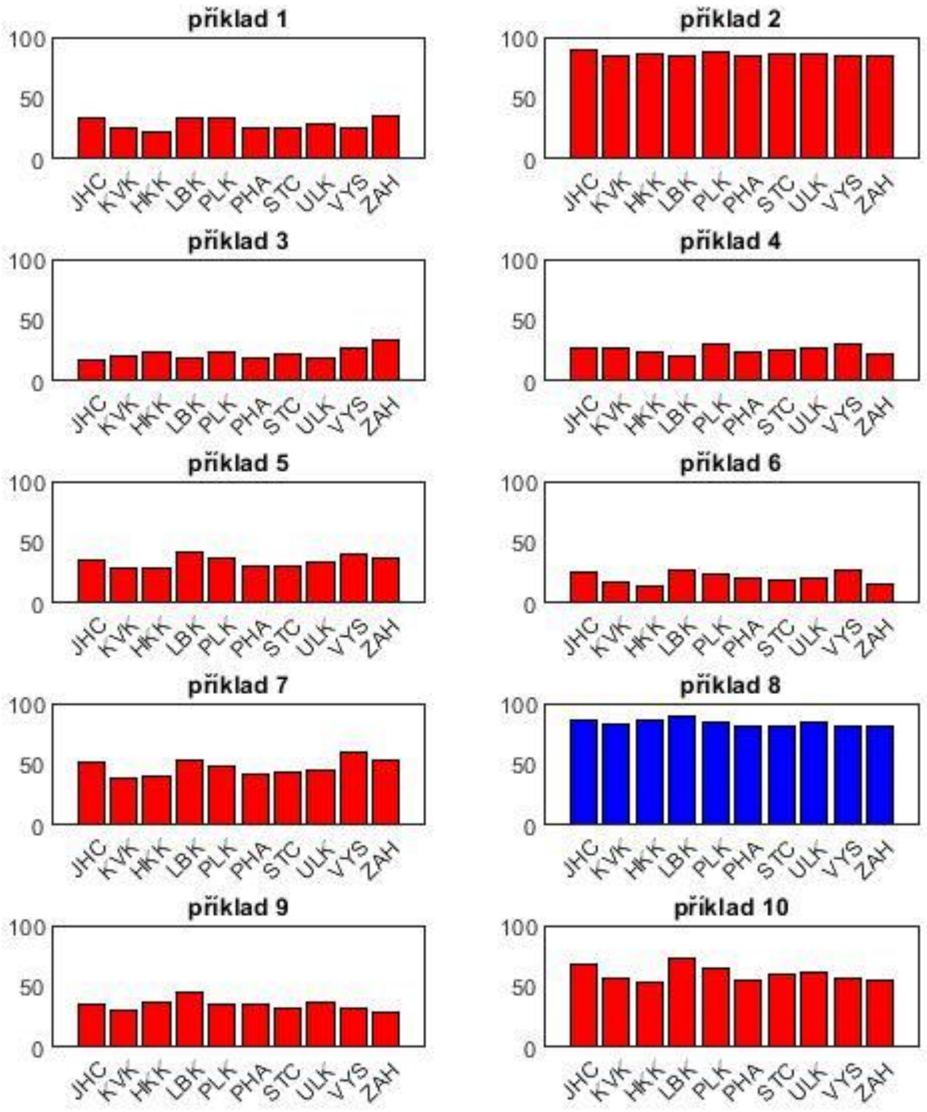
Příklad 10: Jihočeský a Karlovarský ($p = 0{,}000$)

Odlišnosti existují ještě mezi některými moravskými kraji, ty zde nebudou uvedeny, protože jsou velmi málo četné.

Obrázek 16 zachycuje úspěšnosti studentů z jednotlivých krajů v příkladech. Pro přehlednější zobrazení byly do grafů zařazeny pouze více četné kraje a jejich celé názvy byly nahrazeny zkratkami. Významy použitých zkratek lze nalézt v tabulce 29. Málo početné moravské kraje byly zanedbány.

Kraj	**Zkratka**	**Příklad**	***p*-hodnota**
Hlavní město Praha	PHA	1	0,000
Středočeský kraj	STC	2	0,006
Jihočeský kraj	JHC	3	0,000
Plzeňský kraj	PLK	4	0,002
Karlovarský kraj	KVK	5	0,000
Ústecký kraj	ULK	6	0,000
Liberecký kraj	LBK	7	0,000
Královehradecký kraj	HKK	8	0,263
Vysočina	VYS	9	0,014
Zahraničí	ZAH	10	0,000

Tabulka 29: Význam zkratek krajů a p-hodnoty



Obrázek 16: Úspěšnosti příkladů pro faktor kraj (v %)

- **Maturita z matematiky**

  Maturita z matematiky je statisticky významným faktorem pro všechny příklady. P-hodnoty ve všech případech vyšly velice malé a nulovou hypotézu tedy zamítáme. Tabulka 30 zachycuje úspěšnosti jednotlivých příkladů z pohledu faktoru maturita z matematiky. V tabulce je patrné, že úspěšnosti se mezi úrovněmi faktoru výrazně liší.

Maturita z matematiky	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
a (ano)	40,71	90,72	27,83	36,06	45,07
b (ne)	18,11	80,42	14,15	17,71	22,54
-	Příklad 6	Příklad 7	Příklad 8	Příklad 9	Příklad 10
a (ano)	29,89	57,79	87,13	36,71	70,63
b (ne)	9,80	29,81	80,62	29,13	50,84

Tabulka 30: Úspěšnosti příkladů pro faktor maturita z matematiky (v %)

- **Rok maturity**

  Datum maturity se na základě p-hodnot ukázal jako statisticky významný faktor pouze pro příklady 5 (logaritmus), 6(graf funkce $sin(x)$) a 9 (absolutní hodnota komplexního čísla), u kterých vyšly p-hodnoty menší než 0,05. V ostatních příkladech vyšly p-hodnoty výrazně vyšší a nezamítáme pro ně tedy nulovou hypotézu o shodě středních hodnot. Úspěšnosti příkladů pro nejčetnější roky maturity jsou uvedeny v obrázku 17.

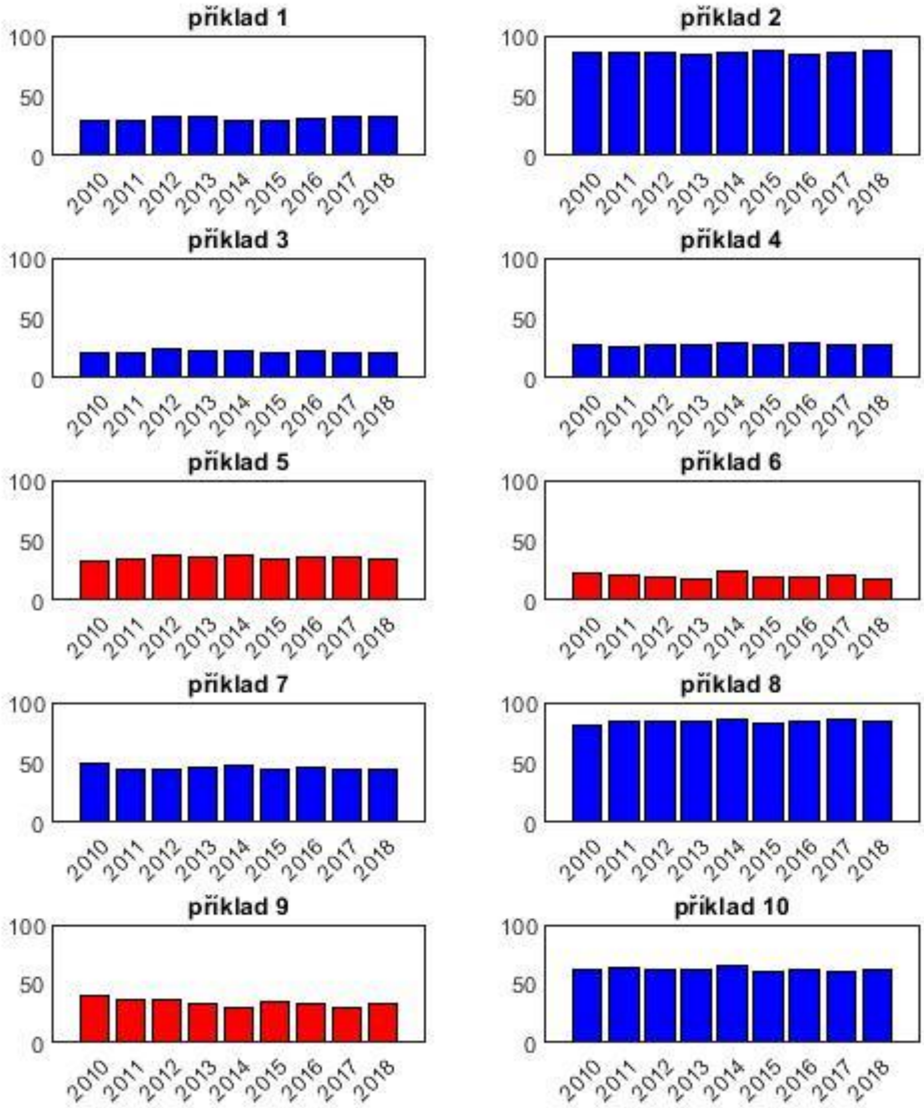
  V mnohonásobném porovnání nebyly zjištěny žádné odlišnosti středních hodnot.

- **Pohlaví**

  Faktor pohlaví se ukázal jako statisticky významný pro osm z deseti příkladů. Pouze u příkladu 3 (analytické vyjádření kružnice) a 9 (absolutní hodnota komplexního čísla) nemá pohlaví vliv na odpověď. P-hodnota analýzy roztylu pro příklad 3 vyšla 0,520, pro příklad 9 měla hodnotu 0,963. Ostatní výsledky p-hodnoty byly menší než zvolená hladina významnosti a zamítáme tedy u nich nulovou hypotézu. Tabulka 31 zaznamenává úspěšnosti příkladů pro faktor pohlaví.

Pohlaví	Příklad 1	Příklad 2	Příklad 3	Příklad 4	Příklad 5
M	33,59	89,22	22,59	32,39	39,21
Ž	27,45	82,10	22,10	20,92	28,52
-	Příklad 6	Příklad 7	Příklad 8	Příklad 9	Příklad 10
M	25,33	53,16	87,90	34,84	68,32
Ž	16,20	36,23	78,33	34,88	54,14

Tabulka 31: Úspěšnosti příkladů pro faktor pohlaví (v %)



Obrázek 17: Úspěšnosti příkladů pro faktor rok maturity (v %)

6 Závěr a shrnutí výsledků

Cílem této bakalářské práce bylo statistické zpracování dat získaných ze vstupních testů z matematiky se zaměřením na souvislost mezi odpověďmi a mezi odpověďmi a dostupnými informacemi o studentech. Z důvodu odlišností ve sběru dat byla data rozdělena na nová (roky 2010 až 2018) a stará (2006 až 2009).

V úspěšnostech jednotlivých příkladů se liší zejména roky 2006 a 2007. Je to patrné především u příkladů 2 a 8, které byly v těchto dvou letech zadávány jiným způsobem. Příklad 2 (porovnávání zlomků) měl dvě správné odpovědi, proto je jeho úspěšnost v prvních dvou letech nižší. Pokud totiž studen zvolil pouze jednu správnou odpověď, získal 0 bodů. Příklad 8 se nejprve zabýval aritmetickou posloupností, od roku 2008 je jeho předmětem počítání procent. Od této doby jsou úspěšnosti příkladu 8 vyrovnané a žádný rok se od ostatních výrazně neliší. Příklad 7 (funkce $sin(x)$ a $cos(x)$) dosahuje nejnižší úspěšnosti v roce 2006. V ostatních letech je úspěšnost téměř stejně velká. Naopak příklady 1 (úprava výrazu) nebo 5 (logaritmus) měly vyšší úspěšnost právě v letech 2006 až 2009 než v letech pozdějších.

Na základě testu nezávislosti v kontingenčních tabulkách a koeficientů korelace bylo zjištěno, že jsou příklady ve všech letech dohromady závislé z pohledu počtu získaných bodů v jednotlivých příkladech. Z pohledu odpovědí jsou příklady navzájem závislé v obou zkoumaných obdobích podle testu nezávislosti v kontingenčních tabulkách.

Ve starších datech byly pomocí jednofaktorové analýzy rozptylu zkoumány souvislosti odpovědí na příklady s faktory varianta testu, forma studia, fakulta a rok. Bylo zjištěno, že na hladině významnosti $\propto = 5\ \%$ správnost studentovy odpovědi nesouvisí s variantou testu, zatímco s faktorem fakulta a rok ano. Pro faktor forma studia byla souvislost prokázaná pro šest z deseti příkladů. Mezi studenty kombinovaného a prezenčního studia byly odlišnosti v úspěšnosti příkladů výrazné. V případě faktoru fakulta se od ostatních výrazněji lišila Fakulta aplikovaných věd. Pozorované průměry studentů FAV byly ve většině příkladů nejvyšší. Naopak nejmenší průměry byly u studentů Filozofické fakulty. Pro faktor rok testu byly zjištěny odlišnosti převážně pro roky 2006 a 2007.

V novějších datech bylo zkoumaných faktorů více. Zde byla na hladině významnosti $\propto = 5\ \%$ zamítnuta nulová hypotéza o shodnosti středních hodnot pro faktory fakulta, obor studia, předmět, střední škola, kraj, maturita z matematiky a pohlaví. Studenti, kteří složili maturitní zkoušku z matematiky byli ve všech příkladech výrazně úspěšnější než ti, kteří z matematiky nematurovali. Tento rozdíl je nejvíce patrný v příkladu 6, jehož předmětem je načrtnout graf funkce $sin(x)$ do soustavy souřadnic. Zde byli studenti s maturitou z matematiky dokonce o 20 % úspěšnější. Faktor pohlaví je také statisticky významný pro všechny příklady s výjimkou příkladů 3 a 9. Zde bylo zjištěno, že muži dosahují větší úspěšnosti v příkladech než ženy a u příkladu 7 byly muži úspěšnější dokonce o 17 % než ženy. Mezi studenty některých středních škol byly 20% rozdíly zejména u příkladů 6 a 7. Největších úspěšností dosahovali studenti gymnázií, naopak studenti středních škol ekonomického nebo zdravotnického zaměření dosahovali nejmenších úspěšností. Rozdíly mezi kraji nebyly tolik markantní. Opakované studium na ZČU a počet předchozích let strávených na ZČU se také ukázaly jako statisticky významné faktory na zvolené hladině významnosti. To však neplatí pro příklady 2, 5 a 9, u kterých k zamítnutí nulové hypotézy nedošlo. V příkladu 6, kde byl rozdíl mezi studenty v prvním ročníku, respektive studenti bez předchozích studií na ZČU, od úspěšností ostatních studentů téměř 20 %. U ostatních příkladů jsou rozdíly úspěšností mezi počty předchozích studií i předchozích studovaných let výrazně menší. Faktor rok testu byl statisticky významný pouze pro polovinu příkladů a rozdíly mezi úspěšnostmi úrovní tohoto faktoru nebyly příliš výrazné. Pro faktory varianta testu, forma studia a rok maturity na zvolené hladině významnosti nezamítáme nulovou hypotézu a můžeme se tedy domnívat, že správnost studentových odpovědí v příkladech na těchto faktorech nezávisí.

Dále by šlo podrobněji zkoumat vliv více faktorů najednou, například užitím dvou nebo více faktorové analýzy rozptylu. Jedna z možných otázek je, proč pohlaví studenta ovlivňuje správnost jeho odpovědi? Jedna z možností je, že zde bude patrná interakce faktoru pohlaví s faktorem fakulty. Je totiž možné, že například na fakultě ekonomické bude větší podíl žen a na Fakultě aplikovaných věd to pravděpodobně bude naopak. Dalším předmětem zkoumání by mohlo být, jak závisí přesné odpovědi na příklady na dostupných informacích o studentech. Bylo by také možné provést analýzu pro všechny roky dohromady, pro informace o studentech dostupné již od roku 2006, a nerozdělovat vše na nová a stará data.

7 Literatura

[1] *Statistici na KMA* [online]. [cit. 2019-05-14]. Dostupné z: https://stat.kma.zcu.cz/auth/

[2] *Portál ZČU* [online]. [cit. 2019-05-14]. Dostupné z: https://portal.zcu.cz/portal/studium/prohlizeni.html

[3] Documentation - crosstab. *MathWorks* [online]. [cit. 2019-05-18]. Dostupné z: https://www.mathworks.com/help/stats/crosstab.html?s_tid=doc_ta

[4] Documentation - corrcoef. *MathWorks* [online]. [cit. 2019-05-18]. Dostupné z: https://www.mathworks.com/help/matlab/ref/corrcoef.html?s_tid=doc_ta

[5] REIF, Jiří a Zdeněk KOBEDA. *Úvod do pravděpodobnosti a spolehlivosti*. 2. vydání-dotisk. Plzeň: Západočeská univerzita v Plzni, 2006. ISBN 80-7043-333-7.

[6] REIF, Jiří. *Metody matematické statistiky*. 2. upravené vydání. Plzeň: Západočeská univerzita v Plzni, 2004. ISBN 80-7043-302-7.

[7] HINDLS, Richard, Jara KAŇOKOVÁ a Ilja NOVÁK. *Statistické metody: Statistika B*. Praha: Vysoká škola ekonomická v Praze, 1995. ISBN 80-7079-354-6.

[8] Documentation - anova1. *MathWorks* [online]. [cit. 2019-05-18]. Dostupné z: https://www.mathworks.com/help/stats/anova1.html?s_tid=doc_ta

[9] Documentation - multcompare. *MathWorks* [online]. [cit. 2019-05-18]. Dostupné z: https://www.mathworks.com/help/stats/multcompare.html?s_tid=doc_ta