

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra matematiky

Bakalářská práce

**Příjmové nerovnosti ve vybraných evropských
zemích**

nahradit listem zadání

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne 24. května 2021

.....
vlastnoruční podpis

Poděkování

Ráda bych poděkovala vedoucímu své bakalářské práce Doc. RNDr. Petru Stehlíkovi, Ph. D. za cenné rady, ochotu, trpělivost, vstřícnost a veškerý čas, který mi při tvorbě práce věnoval.

Abstrakt

Tato bakalářská práce se zabývá měřením nerovností v příjmech. Nejprve jsou popsány metody měření příjmových nerovností: Giniho index a Lorenzova křivka, příjmová mezera a ANOVA. V další části je zpracováno a analyzováno přes 4 miliony dat z České republiky, Dánska, Finska, Německa, Norska, Slovenska a Švédska z let 2015 až 2019. Rozdělujeme je do skupin podle země, pohlaví, pracovní pozice a oborů zaměstnání. Zkoumáme nerovnosti v rámci těchto skupin i mezi nimi. V poslední části práce popisujeme, jak tyto faktory ovlivňují výši platu.

Klíčová slova: příjmové nerovnosti, Giniho index, Lorenzova křivka, příjmová mezera, ANOVA

Abstract

This bachelor thesis is focused on measuring income inequalities. First we describe measures of income inequality: Gini index, Lorenz curve, pay gap and ANOVA. In the next section we process and analyse over 4 million data from the Czech Republic, Denmark, Finland, Germany, Norway, Slovakia and Sweden from 2015 to 2019. We divide them into groups by country, gender, working position, and field of employment. We explore inequalities within these groups and among them. In the last section we describe the effect of these factors on salary.

Keywords: income inequalities, Gini index, Lorenz curve, pay gap, ANOVA

Obsah

Úvod	7
1 Měření příjmových nerovností	8
1.1 Giniho index a Lorenzova křivka	8
1.2 Příjmová mezera	9
1.3 ANOVA	9
1.3.1 Jednofaktorová ANOVA	10
1.3.2 Vícefaktorová ANOVA	11
1.4 Kolmogorovův-Smirnovův test	12
1.5 F-test shody rozptylů	12
2 Úprava a popis dat	13
2.1 Úprava dat	13
2.2 Popisná statistika	14
3 Zpracování dat - výpočet indexů nerovnosti	16
3.1 Giniho index a Lorenzova křivka	16
3.1.1 Nerovnosti v zemích	16
3.1.2 Nerovnosti u mužů a žen	17
3.2 Příjmová mezera	19
3.2.1 Příjmová mezera mezi muži a ženami	19
3.2.2 Příjmová mezera mezera mezi muži a ženami na jednotlivých úrovních a v oborech	20
3.2.3 Porovnání České republiky s ostatními zeměmi	25
4 Zpracování dat - ANOVA	27
4.1 Jednofaktorová ANOVA - pohlaví	27
4.2 Jednofaktorová ANOVA - země	28
4.3 Jednofaktorová ANOVA - úroveň	30
4.4 Trojfaktorová ANOVA	31
4.5 Jednofaktorová ANOVA - obor	33
4.6 Čtyřfaktorová ANOVA	34
Závěr	37
Přílohy	40
Příloha A - počty záznamů	40
Příloha B - popisná statistika	42
Příloha C - časový vývoj příjmové mezery mezi muži a ženami na jednotlivých úrovních a oborech	44
Příloha D - soubory na CD	51

Úvod

Tato práce se zabývá analýzou příjmových nerovností. Způsob rozdělení celkového bohatství ve společnosti je aktuálním a často diskutovaným tématem, protože měří chudobu a blahobyt ve společnosti. Touto problematikou se zabývají především sociologové, ale je zkoumána i v matematice nebo ekonomii, kde se příjmové nerovnosti měří.

Poslední dobou je jednou z častých otázek nerovnost mezi příjmy mužů a žen. Hovoří se o tzv. gender pay gap (příjmové mezeře mezi pohlavími), která udává, jak se platy mužů a žen liší. Nižší platy žen jsou často vysvětlovány tím, že ženy pracují na jiných pozicích, které jsou hůře placené. Například podle některých studií je ve většině vědeckých disciplín zaměstnáno méně žen než mužů, na jednu ženu připadají dva muži, [1]. Menší počet žen na některých pozicích bývá vysvětlován tím, že ženy se starají o rodinu více než muži a kariéra je pro ně méně důležitá než pro muže. Existují ale studie, podle kterých to tak není, [2]. Příjmová mezera proto bývá „očistěna“. To znamená, že jsou porovnáváni muži a ženy na srovnatelných pozicích, se srovnatelným vzděláním, ve srovnatelném věku apod., [3]. Tím je možné získat bližší údaj o skutečném rozdílu mezi platy mužů a žen.

V práci je zpracovááno přes 4 miliony dat z České republiky, Dánska, Finska, Německa, Norska, Slovenska a Švédska z let 2015 až 2019. Česká republika byla zvolena, protože pro nás bude pochopitelně nejzajímavější. Pro srovnání bylo vybráno Slovensko, se kterým má společnou historii. Jako další sousední země bylo vybráno Německo. Data byla doplněna skandinávskými zeměmi, které patří mezi nejvyspělejší státy s nejmenšími nerovnostmi.

Cílem je změřit příjmové nerovnosti v zemích a nerovnosti mezi muži a ženami a popsat jejich krátkodobý časový vývoj. Získaná data obsahují i informaci o pracovní pozici a oboru zaměstnání, a tak se pokusíme zjistit, zda muži a ženy na stejné pozici pobírají stejný plat a které faktory mají vliv na velikost platu.

V první části práce jsou popsány různé způsoby měření nerovnosti: Giniho index, kterým se měří nerovnosti ve skupině, již zmíněná příjmová mezera, která kvantifikuje nerovnosti mezi skupinami, a analýza rozptylu (ANOVA), která nám pomůže určit, jaký vliv mají různé faktory na velikost platu.

Druhá část se věnuje popisu vstupních dat, jejich zpracování, odstranění poškozených a nekoherentních údajů a výpočtu základních charakteristik platů jako je například průměr, medián nebo rozptyl. Součástí této kapitoly jsou i vizualizace některých základních charakteristik.

Ve třetí kapitole je vypočten Giniho index, abychom změřili nerovnosti v jednotlivých zemích. Je vypočtena také příjmová mezera mezi muži a ženami. Nejprve je určena ze všech dat a poté je rozdělujeme, abychom zjistili, zda muž a žena, kteří pracují na stejné pozici a ve stejném oboru, pobírají stejný plat. Nakonec je vypočtena i příjmová mezera mezi Českou republikou a ostatními zeměmi.

Ve čtvrté kapitole pomocí anovy zkoumáme, jakým způsobem země, pohlaví, pracovní pozice a obor zaměstnání ovlivňují velikost platu.

Poslední část je závěr, kde jsou shrnuty všechny výsledky.

1 Měření příjmových nerovností

Nerovnosti v příjmech je možné měřit buď ve skupině nebo mezi skupinami. Skupiny mohou tvořit například různé země, muži a ženy nebo regiony. Kromě příjmových nerovností lze měřit i jiné, zde se ale budeme zabývat pouze příjmovými.

Rozdíl mezi rovným a nerovným rozdělením příjmů lze poznat snadno. Pokud ale máme více skupin s nerovnoměrně rozdělenými příjmy, může být problém určit, které rozdělení je více nerovné. [4]

Existuje mnoho indexů sloužících k měření nerovností ve společnosti. Měly by splňovat přinejmenším to, že jsou rovny nule, pokud mají všechny osoby stejné příjmy, a nabývají kladných hodnot, když se příjem alespoň dvou jedinců liší. [4]

Dalším požadavkem je, že hodnota indexu by měla zůstat stejná, pokud všechny příjmy vynásobíme stejnou konstantou. Jinými slovy to znamená, že nezáleží na inflaci a měně. [4]

Při měření příjmových nerovností by také mělo být zohledněno, že více záleží na relativním rozdělení příjmů než na absolutním. Například pokud bychom měli dvě osoby, které vydělávají 50 a 100 korun, je mezi nimi velká nerovnost, protože druhá osoba vydělává dvojnásobek příjmu první. Kdybychom jim ale oběma přidali 10 000 korun, je už relativně malá. [4]

Další požadovanou vlastností při měření nerovností v příjmech je, že se index zvětší, pokud přesuneme část příjmů ve skupině od chudšího jedince k bohatšímu a zároveň příjmy ostatních zůstanou stejné. Zvětší se tím nerovnost ve společnosti, protože bude větší rozdíl mezi chudšími a bohatšími. [4]

Pro analýzu příjmů je nevhodnější použít Theilův index nebo Giniho index, který je nejběžnější, [4]. V této práci byl zvolen Giniho index z důvodu porovnání vypočtených hodnot s údaji zveřejněnými Světovou bankou, která používá také Giniho index.

1.1 Giniho index a Lorenzova křivka

Jednou z možností měření nerovností ve skupině je Giniho index (Giniho koeficient). Uvažujme, že máme populaci velikosti n a každá osoba má roční příjem x_i , kde $i = 1, 2, \dots, n$. Giniho index je definován jako [4]

$$G = \frac{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2\mu}. \quad (1)$$

Jedná se o podíl průměrné absolutní odchylky mezi příjmy všech osob a dvojnásobku průměru všech příjmů μ .

Pro výpočty je vhodnější použít vzorec [4]

$$G = \frac{2}{\mu n^2} \sum_{i=1}^n i x_i - \frac{n+1}{n}. \quad (2)$$

Giniho index nabývá hodnot z intervalu $(0, 1)$. Pokud je roven nule, je rozdělení příjmů rovné, a pokud se rovná jedné, vydělává jedna osoba celkový příjem celé populace a ostatní nemají žádný.

Giniho index souvisí s Lorenzovou křivkou. Předpokládáme, že máme příjmy seřazené tak, že $x_1 \leq x_2 \leq \dots \leq x_n$. Nejprve u každého vypočteme procento populace, které má stejný nebo menší příjem, a označíme ho p (jedná se o p -tý percentil). Dále vypočteme procento celkového příjmu, které p -tá část populace vydělává, a označíme ho $L(p)$. Vykreslením p na osu x a $L(p)$ na osu y získáme Lorenzovu křivku. Můžeme například zjistit, že 10 % nejchudších vydělává 5 % celkového příjmu, což by na Lorenzově křivce odpovídalo bodu $[0,1; 0,05]$. [4]

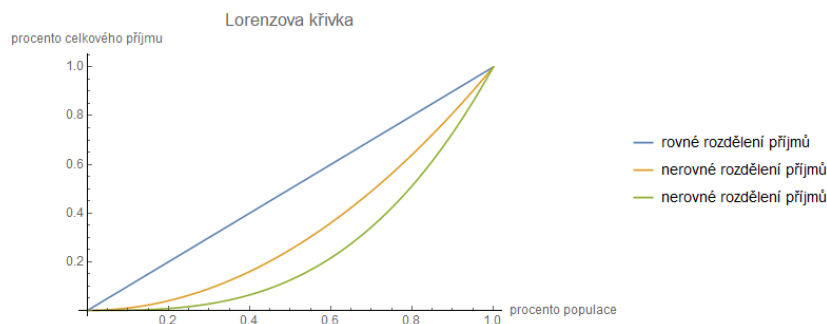
Příklad Lorenzovy křivky je na obrázku 1. Pokud má tvar přímky, je celkový příjem rozdělen rovnoměrně. V případě, že se Lorenzova křivka nachází pod touto přímkou, je rozdělení příjmů nerovné.

Pomocí Lorenzovy křivky může být definován i Giniho index jako [5]

$$G = 2 \int_0^1 [p - L(p)] dp. \quad (3)$$

Jinými slovy je roven dvojnásobku obsahu plochy mezi Lorenzovou křivkou a přímkou, která popisuje rovné rozdělení celkového příjmu. To znamená, že popisuje, jak se rozdělení příjmů liší od rovného.

Uvažujme situaci, kdy máme dvě Lorenzovy křivky popisující nerovná rozdělení příjmů a jedna z nich se celá nachází pod druhou. Potom má tato křivka větší Giniho index a je zde i vyšší nerovnost než v populaci, kterou popisuje druhá křivka. Na obrázku 1 by zelená křivka měla větší Giniho index než oranžová a rozdělení příjmů by bylo více nerovné.



Obrázek 1: Příklad rovného a nerovného rozdělení příjmů

1.2 Příjmová mezera

Příjmová mezera (anglicky pay gap) slouží k měření nerovností mezi skupinami. Jak již bylo zmíněno v úvodu, obvykle se počítá mezi muži a ženami (gender pay gap).

Je definována jako

$$GPG = \frac{Q_{\bar{z},50}}{Q_{m,50}}, \quad (4)$$

kde $Q_{m,50}$ je medián příjmu mužů a $Q_{\bar{z},50}$ medián příjmu žen. Také je možné místo mediánů počítat s průměrnými příjmy, [6].

Příjmová mezera nabývá kladných hodnot nebo je vyjádřena v procentech. Výsledkem je, kolika procentům mediánu příjmu mužů je roven medián příjmu žen, [6].

Pokud je hodnota příjmové mezery menší než jedna, pak je medián příjmu žen menší než medián příjmu mužů. Kdyby byla příjmová mezera rovna jedné, znamenalo by to, že jsou mediány příjmu mužů a žen shodné. Hodnoty větší než jedna nabývá příjmová mezera v případě, že je medián příjmu žen větší než medián příjmu mužů.

Existuje více způsobů výpočtu příjmové mezery. Eurostat, který zveřejňuje příjmovou mezera v evropských zemích, ji počítá jako rozdíl mezi průměrnou hodinovou mzdou mužů a žen vydělený průměrnou hodinovou mzdou mužů, [7]. Lze se opět setkat i s výpočtem pomocí průměru nebo mediánu měsíční mzdy, [3]. V těchto případech je výsledkem, o kolik procent mají ženy menší příjmy než muži. V této práci budeme ale používat pouze výpočet podle (4).

Příjmovou mezera je možné vypočítat i mezi jinými skupinami, než jsou muži a ženy, například mezi různými zeměmi. Je definována jako

$$PG = \frac{Q_{1,50}}{Q_{2,50}}, \quad (5)$$

kde $Q_{1,50}$ je medián příjmu první skupiny a $Q_{2,50}$ medián příjmu druhé skupiny.

Výsledkem je podobně jako u GPG , kolika procentům mediánu příjmů druhé skupiny odpovídá medián příjmu první skupiny. Stejně jako v případě mužů a žen, nabývá PG kladných hodnot.

Obdobně můžeme vypočítat i příjmovou mezera pro x -tý percentil, abychom zjistili, zda se v závislosti na percentilu mění. To lze podle vzorce [8]

$$PG_x = \frac{Q_{1,x}}{Q_{2,x}}, \quad (6)$$

kde $Q_{1,x}$ je x -tý percentil příjmu první skupiny, $Q_{2,x}$ x -tý percentil příjmu druhé skupiny a x je z intervalu $(0, 1)$.

Podobně jako v předchozích případech může příjmová mezera pro x -tý percentil nabývat kladných hodnot.

1.3 ANOVA

ANOVA (Analysis of Variance, analýza rozptylu) se používá, pokud chceme zjistit, jak na nějakou veličinu působí jeden nebo více faktorů. Například když zjišťujeme, zda je plat funkcí země, pohlaví nebo pracovní pozice. Faktory označme A, B, \dots . Každý z nich může nabývat konečného počtu hodnot, kterým se říká úrovně. V kapitolách 1.3.1 a 1.3.2 budeme vycházet z [9].

1.3.1 Jednofaktorová ANOVA

Pokud zkoumáme vliv jednoho faktoru A , jedná se o jednofaktorovou anovu (One-Way ANOVA, jednoduché třídění). Faktor A může nabývat I úrovní a předpokládáme, že $I \geq 2$.

Pro úroveň i faktoru A máme k dispozici n_i pozorování $y_{i1}, y_{i2}, \dots, y_{in}$ ($i = 1, \dots, I$). Předpokladem anovy je, že pozorování jsou nezávislá a jedná se o náhodný výběr z normálního rozdělení s konstantním rozptylem $N(\mu_i, \sigma^2)$. Pozorování potom mohou být vyjádřena jako

$$y_{ip} = \mu_i + e_{ip} \quad (i = 1, 2, \dots, I; p = 1, 2, \dots, n_i), \quad (7)$$

kde $e_{ip} \sim N(0, \sigma^2)$ jsou náhodné odchylky.

U analýzy rozptylu testujeme nulovou hypotézu o shodě středních hodnot

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_I.$$

Pokud vypočteme vážený průměr hodnot μ_i jako

$$\mu = \sum_{i=1}^I \mu_i \frac{n_i}{n},$$

kde $n = n_1 + n_2 + \dots + n_I$ je celkový počet naměřených pozorování, můžeme nulovou hypotézu také zapsat jako [9]

$$H_0 : \mu_i = \mu.$$

Když vypočteme odchylky μ_i od μ jako

$$\alpha_i = \mu_i - \mu, \quad (8)$$

můžeme pozorování zapsat ve tvaru

$$y_{ip} = \mu + \alpha_i + e_{ip} \quad (i = 1, 2, \dots, I; p = 1, 2, \dots, n_i) \quad (9)$$

a hypotéza H_0 je ekvivalentní s

$$H_0^{(\alpha)} : \alpha_1 = \alpha_2 = \dots = \alpha_I.$$

Tento model se nazývá model s pevnými efekty, protože μ_i a α_i jsou konstantní.

Výběrové průměry vypočteme jako

$$\bar{y}_{i.} = \frac{1}{n_i} \sum_{p=1}^{n_i} y_{ip} \quad (i = 1, 2, \dots, I) \quad (10)$$

a celkový průměr podle vzorce

$$\bar{y} = \frac{1}{n} \sum_{i=1}^I \sum_{p=1}^{n_i} y_{ip}. \quad (11)$$

Celková variabilita je definována jako

$$S_T = \sum_{i=1}^I \sum_{p=1}^{n_i} (y_{ip} - \bar{y})^2. \quad (12)$$

Meziskupinová variabilita, která je vysvětlována působením faktoru A , se vypočte jako

$$S_A = \sum_{i=1}^I (y_{i.} - \bar{y})^2. \quad (13)$$

Nakonec definujeme vnitroskupinovou variabilitu jako

$$S_e = \sum_{i=1}^I \sum_{p=1}^{n_i} (y_{ip} - \bar{y}_{i.})^2. \quad (14)$$

Ta je vysvětlována působením náhodných odchylek e a nazývá se také reziduální součet čtverců.

Veličiny S_T/σ^2 , S_A/σ^2 a S_e/σ^2 mají, pokud platí hypotéza H_0 , rozdělení χ^2 se stupni volnosti

$$f_T = n - 1, \quad (15)$$

$$f_A = I - 1, \quad (16)$$

a

$$f_e = n - I. \quad (17)$$

Veličina F_A se vypočte podle vzorce

$$F_A = \frac{S_A/f_A}{S_e/f_e}. \quad (18)$$

Pokud platí hypotéza H_0 , má F_A rozdělení F se stupni volnosti f_A a f_e . Hypotézu H_0 zamítáme na hladině významnosti α , pokud $F_A > F_{1-\alpha}(f_A, f_e)$. Výsledky anovy zapisujeme do tabulky (viz tabulka 1).

zdroj variability	součet čtverců	počet stupňů volnosti	průměrný čtverec	F
faktor A	S_A	$f_A = I - 1$	S_A/f_A	F_A
reziduální	S_e	$f_e = n - I$	S_e/f_e	-
celkem	S_T	$f_T = n - 1$	-	-

Tabulka 1: ANOVA - jednoduché třídění, dle [9]

Pokud byla zamítnuta hypotéza H_0 o shodě středních hodnot, tak nás většinou zajímá, mezi kterými úrovněmi faktoru je významný rozdíl. Tomu se říká mnohonásobné porovnávání. Lze k němu využít několika metod, například LSD metodu, Scheffého metodu nebo Tukeyovu metodu. V této práci bude využita Tukeyova metoda.

U Tukeyovy metody zjišťujeme platnost nulové hypotézy $\mu_i = \mu_j$. Nejprve je potřeba sestavit tabulku rozdílů průměrů (viz tabulka 2). Hypotézu, že $\mu_i = \mu_j$, zamítáme na hladině významnosti α , pokud

$$\bar{y}_i - \bar{y}_j \geq sq_{I,n-I}(\alpha) \sqrt{\frac{1}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (19)$$

kde $q_{I,n-I}$ je tzv. studentizované rozpětí a $s = \sqrt{S_e/f_e}$, [10].

	2	3	...	I
1	$\bar{y}_1 - \bar{y}_2$	$\bar{y}_1 - \bar{y}_3$...	$\bar{y}_1 - \bar{y}_I$
2	-	$\bar{y}_2 - \bar{y}_3$...	$\bar{y}_2 - \bar{y}_I$
...
I-1	-	-	-	$\bar{y}_{I-1} - \bar{y}_I$

Tabulka 2: Rozdíly průměrů, dle [10]

Při použití této metody platí, že pravděpodobnost, že alespoň jednou zamítneme nulovou hypotézu, pokud platí, je nejvýše α . Pokud $n_1 = n_2 = \dots = n_I$, je tato pravděpodobnost α . [10]

1.3.2 Vícefaktorová ANOVA

U dvoufaktorové anovy (dvojného třídění, Two-Way ANOVA) zjišťujeme, zda je nějaká veličina funkcí dvou faktorů A a B . Někdy se hovoří i o tzv. blokových experimentech, kdy faktor A považujeme za hlavní, měníme je a zkoumáme jeho vliv, a faktor B za rušivý. Pozorování jsou rozdělena do bloků, uvnitř kterých je faktor B stejný a mění se v nich faktor A .

Pokud se vliv faktorů A a B sčítá, jedná se o dvojný třídění bez interakcí. Když je jejich společné působení složitější, používá se model s interakcemi, který je nad rámec této práce.

U dvojného třídění s pevnými efekty bez interakcí máme model ve tvaru

$$y_{ijp} = \mu + \alpha_i + \beta_j + e_{ijp} \quad (i = 1, \dots, I; j = 1, \dots, J; p = 1, \dots, n_{ij}), \quad (20)$$

kde I je počet úrovní faktoru A , J je počet úrovní faktoru B (uvažujeme, že $I \geq 2$ a $J \geq 2$), y_{ijp} je počet pozorování pro i -tou úroveň faktoru A a j -tou úroveň faktoru B a n_{ij} je počet pozorování pro tuto kombinaci, μ je neznámá konstanta, α_i je pevný efekt i -té úrovně faktoru A , β_j je pevný efekt j -té úrovně faktoru B a e_{ijp} jsou nezávislé náhodné odchylky z normálního rozdělení $N(0, \sigma^2)$.

Dalšími požadavky je, aby $\sum \alpha_i = 0$ a $\sum \beta_j = 0$. Nazývají se reparametrizační podmínky.

Při různých četnostech n_{ij} je možné využít toho, že ANOVA je speciálním případem regresní analýzy nebo v některých případech tzv. vyvážené neúplné bloky. Více informací o řešení lze nalézt v [9].

U trojfaktorové anovy (trojného třídění) s pevnými efekty bez interakcí používáme model ve tvaru

$$y_{ijkp} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijkp} \quad (i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K; p = 1, \dots, n_{ijk}), \quad (21)$$

kde α_i , β_j a γ_k jsou efekty faktorů A , B a C a $e_{ijkp} \sim N(0, \sigma^2)$. Opět platí i reparametrizační podmínky $\sum \alpha_i = \sum \beta_j = \sum \gamma_k = 0$.

Více informací o řešení lze nalézt v [9]. Ve speciálním případě lze využít tzv. latinské čtverce a obdobně u čtyřfaktorové anovy řecko-latinské čtverce.

1.4 Kolmogorovův-Smirnovův test

Kolmogorovův-Smirnovův test (Kolmogorovův test) zkoumá platnost nulové hypotézy, že data pochází z určitého rozdělení pravděpodobnosti. Následující kapitola vychází z [9].

Předpokládáme, že máme uspořádaný náhodný výběr $x_{(i)}$, kde $i = 1, 2, \dots, n$. Test je založen na porovnání distribuční funkce zkoumaného rozdělení a empirické (výběrové) distribuční funkce (EDF), která je definována jako

$$F_n(x) = \frac{\text{počet hodnot ve výběru} \leq x}{n}. \quad (22)$$

Testovací statistika D je definována jako

$$D = \sup\{|F_n(x) - F(x)| : x \in R\}. \quad (23)$$

Jedná se o největší vzdálenost mezi hodnotami empirické distribuční funkce a distribuční funkce rozdělení, které testujeme. Testovou statistiku lze také vyjádřit jako

$$D = \frac{1}{2n} + \max \left\{ \left| \frac{i - 0,5}{n} - F(x_{(i)}) \right| : i = 1, 2, \dots, n \right\}. \quad (24)$$

Hypotézu H_0 , že náhodný výběr pochází z daného rozdělení zamítáme, pokud statistika D nabývá velkých hodnot. Pro velká n se používá přibližná kritická hodnota $\sqrt{-\frac{1}{2n} \ln(\alpha/2)}$.

Pokud neznáme parametry hypotetického rozdělení, je potřeba je odhadnout. Například pokud testujeme nulovou hypotézu, že data pochází z normálního rozdělení, ale neznáme přesné parametry μ a σ^2 , odhadujeme je pomocí výběrového průměru a výběrového rozptylu. Test se potom nazývá Lillieforsův. Testovací statistika je stejná, ale má jiné kritické hodnoty. Pro velké n je to přibližně $0,89/\sqrt{n}$.

1.5 F-test shody rozptylů

F-test shody rozptylů používáme, pokud máme dva nezávislé náhodné výběry z rozdělení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Rozsahy výběrů jsou n_1 a n_2 .

Testujeme platnost nulové hypotézy, že $\sigma_1^2 = \sigma_2^2$. Alternativní hypotéza je, že $\sigma_1^2 \neq \sigma_2^2$. Poměr výběrového rozptylu prvního výběru ku výběrovému rozptylu druhého výběru se řídí F rozdělením se stupni volnosti $n_1 - 1$ a $n_2 - 1$, a proto nulovou hypotézu zamítáme, pokud [9]

$$\frac{s_1^2}{s_2^2} > F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$$

nebo

$$\frac{s_2^2}{s_1^2} > F_{1-\frac{\alpha}{2}}(n_2 - 1, n_1 - 1).$$

2 Úprava a popis dat

Cílem práce je analýza příjmů v České republice a dalších vybraných evropských zemích. Konkrétně se jedná o Slovensko, Německo, Dánsko, Norsko, Švédsko a Finsko.

Data byla získána od firmy, která analyzuje příjmy soukromých firem z celého světa. Poskytla data z výše zmíněných zemí z let 2015 až 2019. Bohužel neznáme přesný způsob výběru dat, ale pravděpodobně se nejedná o náhodný výběr ze všech záznamů, protože byly vybírány podle toho, zda byly údaje kompletní.

Data byla poskytnuta pouze za podmínky, že firma nebude jmenována, aby nebylo poškozeno její jméno, protože se jedná o citlivá data. Z toho důvodu také data nemohla být přiložena na CD. Jejich ukázka je na obrázku 2.

POHLAVÍ	UROVEN	REGION	UVAZE	ROČNÍ PLAT	ROČNÍ PLAT NA PLNY	ROČNÍ PLAT + FIXNÍ	OBOR
muz	Operation	Ústecký kraj	1	306000	306000	306000	Production
zena	Supervisors	Praha	1	420000	420000	420000	Administration/Support/Service
zena	Supervisors	Praha	1	408000	408000	408000	Finance and Accounting

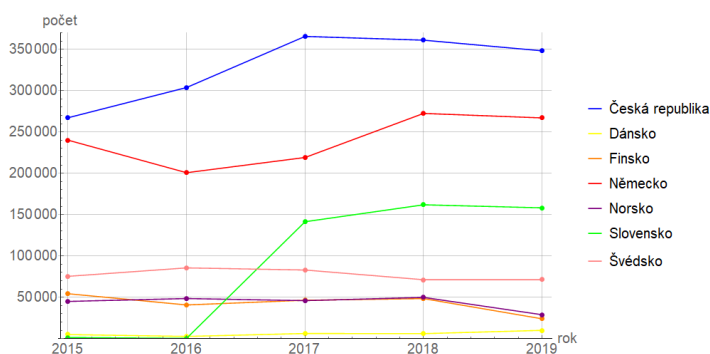
Obrázek 2: Ukázka vstupních dat

Data obsahují informaci o pohlaví zaměstnance, úvazku, ročním platu, ročním platu přepočteném na celý úvazek a oboru zaměstnání. Dále je u každého uvedena úroveň (pracovní pozice): řadový pracovník (operation), vedoucí pracovník (supervisor), manažer (manager) nebo řídicí pracovník (executive). Také je zde součet roční mzdy a fixních plateb (například třináctý plat nebo jiné bonusy). V některých datech je také uveden region. Všechny mzdy jsou v měně dané země a v názvu souboru je uvedeno datum, ke kterému se mají mzdy převádět.

Celkem je tedy k dispozici třicet pět souborů (sedm zemí po pěti letech) a v každém sedm, případně osm (pokud je uveden i region) údajů o zaměstnancích. Zpracování dat bylo provedeno v programu Wolfram Mathematica.

2.1 Úprava dat

Data obsahovala nekompletní a chybné záznamy, a proto je bylo potřeba upravit. Ve sloupci pohlaví byly odstraněny řádky, které obsahovaly něco jiného než muž nebo žena. Stejným způsobem bylo postupováno i u úrovní. V případě úvazku byly vybrány jen hodnoty z intervalu (0, 1). Ve sloupci roční plat převedený na celý úvazek bylo u každé země zvoleno minimum, aby se nezapočítávaly extrémně nízké chybné hodnoty. Nakonec byly odstraněny řádky, kde byl prázdný obor a region. Kromě odstranění chybných dat byly také provedeny drobné úpravy v chybných regionech (například Plzeň byla nahrazena Plzeňským krajem apod.). Počty dat v jednotlivých krocích jsou uvedeny v příloze A v tabulkách 44 až 50. Na obrázku 3 je znázorněn počet záznamů po všech úpravách v jednotlivých letech.



Obrázek 3: Graf počtu záznamů ve všech zemích po všech úpravách v závislosti na čase

Po všech úpravách je nejvíce záznamů z České republiky (v roce 2019 jich je 348 252) a Německa (v roce 2019 267 068), naopak nejméně ze skandinávských zemí (v roce 2019 je v Dánsku 10 032 dat, ve Finsku 24 013, v Norsku 28 947 a ve Švédsku 71 382). Na první pohled si můžeme všimnout, že slovenských dat je poměrně málo v letech 2015 a 2016 (1 405 a 362), v dalších letech už podstatně více (např. 158 272 v posledním roce). Z toho důvodu můžeme očekávat, že se výsledky na Slovensku v letech 2015 a 2016 budou lišit od dalších let.

Dále byly mzdy převedeny z původních měn na české koruny, aby bylo možné mezi sebou porovnávat různé státy. V původních datech bylo v názvu souboru uvedeno datum, ke kterému se platy mají převádět, a tak byly zvoleny kurzy z těchto dnů, [11]. U Německa datum chybělo, a proto byl zvolen měsíční kumulovaný průměr kurzu za leden až prosinec, [12].

2.2 Popisná statistika

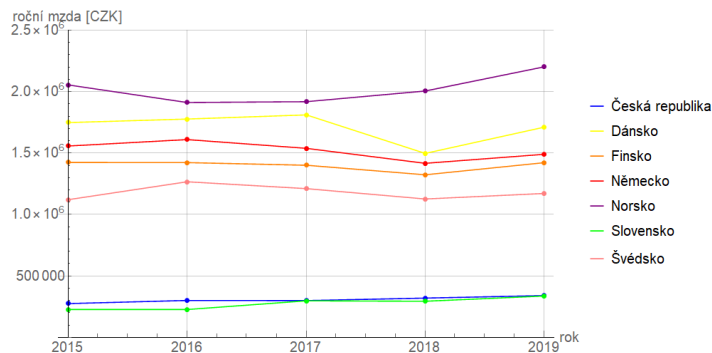
Po úpravě dat byly vypočteny základní charakteristiky ročních platů: počet, minimum, maximum, průměr, medián, rozptyl, směrodatná odchylka, dolní a horní kvartil, šikmost a špičatost. Výsledky dat z roku 2019 jsou uvedeny v tabulce 3, ostatní v příloze B v tabulkách 51 až 57.

	Česká republika	Dánsko	Finsko	Německo	Norsko	Slovensko	Švédsko
počet	348	10	24	267	29	158	71
minimum	145	386	343	258	243	108	337
maximum	20 624	1 7824	19 325	127 632	23 737	11 108	32 958
průměr	432	1 824	1 572	1 678	2 326	411	1 307
medián	340	1711	1 422	1 490	2 202	336	1 171
rozptyl	$9,93 \times 10^7$	$8,54 \times 10^8$	$5,96 \times 10^8$	$7,81 \times 10^8$	$7,39 \times 10^8$	$7,93 \times 10^7$	$4,40 \times 10^8$
směrodatná odchylka	315	924	772	884	860	282	663
dolní kvartil	264	1 307	1 158	1 055	1 750	256	886
horní kvartil	480	2 206	1 780	2 134	2 696	469	1 549
šikmost	7,34	3,15	6,41	12,63	3,03	6,27	6,63
špičatost	198,89	31,90	94,18	1 563,21	33,61	109,02	150,64

Tabulka 3: Popisná statistika pro data z r. 2019, počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²

Z posledního roku 2019 máme k dispozici 348 252 českých, 10 032 dánských, 24 013 finských, 267 068 německých, 28 947 norských, 158 272 slovenských a 71 382 švédských dat.

Na obrázku 4 je zobrazen vývoj mediánu platů v čase. Nejmenší medián má Česká republika a Slovensko, ale v čase roste. V roce 2015 byl v České republice 274 560 CZK a v roce 2019 339 600 CZK. Na Slovensku se za stejnou dobu zvýšil z 225 997 CZK na 335 889 CZK. Ostatní země mají oproti České republice a Slovensku medián vyšší (v posledním roce je v Dánsku 1 710 590 CZK, ve Finsku 1 421 780 CZK, v Německu 1 490 110 CZK, v Norsku 2 202 040 CZK a ve Švédsku 1 171 200 CZK). V Norsku klesl mezi lety 2015 a 2016, v Dánsku mezi lety 2017 a 2018 a v Německu, Finsku a Švédsku klesal od roku 2016 do roku 2018. V ostatních případech i v těchto zemích rostl.



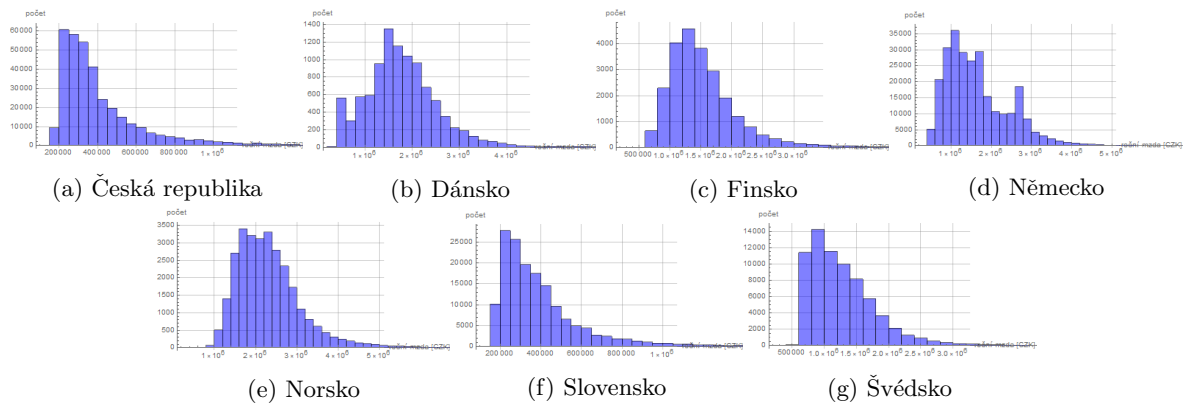
Obrázek 4: Vývoj mediánu roční mzdy v čase - všechny země

Abychom měli představu, jaké rozdělení mají roční platy, byly vytvořeny histogramy. Na obrázku 5 jsou zobrazeny histogramy dat z roku 2019. Je vidět, že data jsou vychýlená vlevo od průměru. Tomu odpovídají i vypočtené kladné koeficienty šikmosti (Česká republika 7,24, Dánsko 31,9, Finsko 94,18, Německo 1563,21, Norsko 33,61, Slovensko 109,02 a Švédsko 150,64). V záznamech je tedy více zaměstnanců s platem nižším než průměr a méně těch s vyšším.

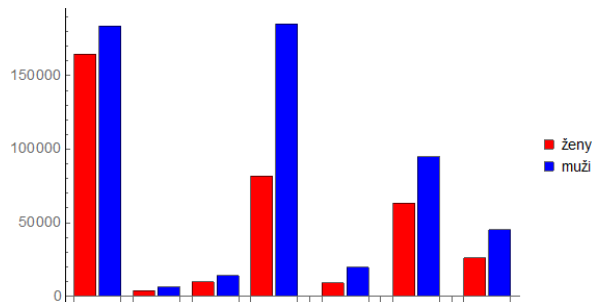
Protože je cílem práce analýza příjmových nerovností podle pohlaví, bylo zjištěno, kolik je v datech mužů a žen. Na obrázku 6 jsou zobrazeny počty dat z roku 2019. Žen je ve všech zemích méně než mužů, někdy se jedná i přibližně jen o polovinu oproti mužům.

U dat rozdělených na muže a ženy byly v každé zemi vypočteny mediány platů a znázorněn jejich časový vývoj podobně jako pro celé země (obrázek 7). Na první pohled je vidět, že téměř ve všech případech byl medián platů žen menší než medián platu mužů. Výjimkou je Slovensko v roce 2016, kde to může být opět způsobeno výrazně menším počtem dat.

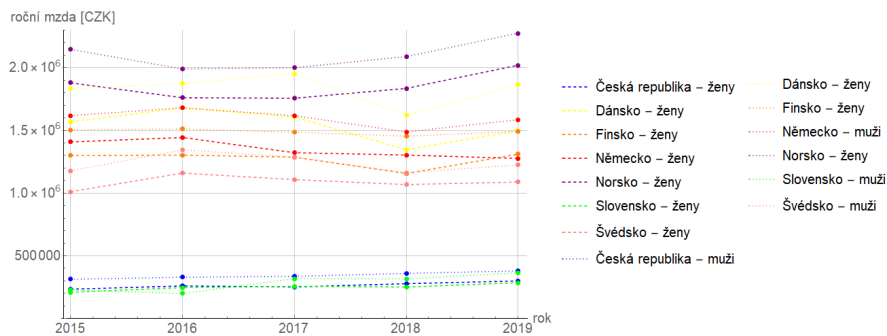
Pro data rozdělená podle pohlaví byly také sestaveny histogramy (viz obrázek 8), ve kterých je opět vidět, že je v záznamech více mužů než žen. Mzdy žen jsou ve všech zemích kromě Švédska více



Obrázek 5: Histogramy ročních mezd v r. 2019

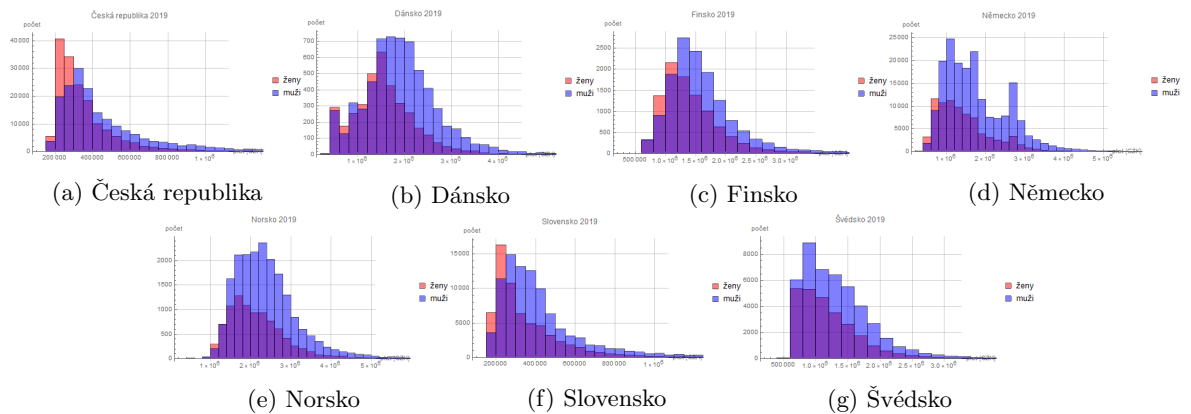


Obrázek 6: Počty záznamů v r. 2019 podle pohlaví, zleva Česká republika, Dánsko, Finsko, Německo, Norsko, Slovensko, Švédsko



Obrázek 7: Vývoj mediánu roční mzdy v čase, rozděleno podle pohlaví

vychýlené doleva než mzdy mužů.



Obrázek 8: Histogramy ročních mezd v r. 2019

3 Zpracování dat - výpočet indexů nerovnosti

3.1 Giniho index a Lorenzova křivka

Cílem práce je analýza příjmových nerovností. Při zpracování dat jsme se nejprve zaměřili na nerovnosti uvnitř skupin. Byly analyzovány pomocí Giniho indexu a Lorenzovy křivky, které byly popsány v kapitole 1.1.

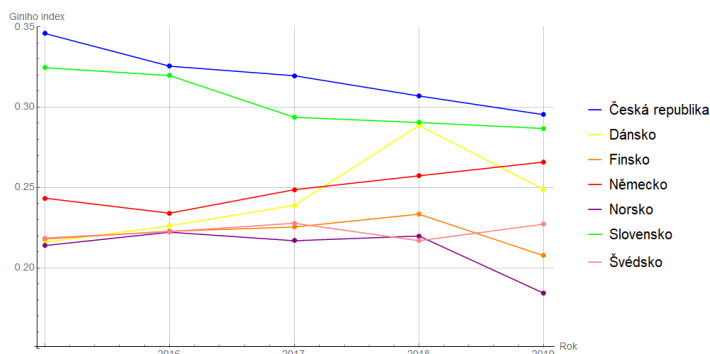
3.1.1 Nerovnosti v zemích

Nejdříve jsme se zabývali nerovnostmi v jednotlivých zemích. V posledním roce je v České republice Giniho index 29,5 % a na Slovensku 28,7 %, a proto jsou v těchto zemích nerovnosti největší. V Německu je index 26,6 %, v Dánsku 24,9 %, ve Švédsku 22,7 %, ve Finsku 20,8 % a v Norsku 18,4 %.

Giniho index v ostatních letech je v tabulce 1.1 a na obrázku 9 je zobrazen jeho časový vývoj ve všech zemích. V České republice a na Slovensku klesá, v Německu roste. Můžeme si všimnout, že v roce 2015 měly všechny skandinávské země a Německo Giniho indexy zhruba podobné (Dánsko 21,6 %, Finsko 21,8 %, Německo 24,3 %, Norsko 21,4 % a Švédsko 21,9 %), ale v roce 2019 se index v Dánsku a Německu blíží indexu v Česku a Slovensku. V Dánsku se navíc v roce 2018 hodnota Giniho indexu liší od ostatních let (28,9 %). Jinými slovy nerovnosti v České republice a na Slovensku se v čase zmenšují a v Dánsku a Německu se zvětšují.

	2015	2016	2017	2018	2019
Česká republika	0,346	0,326	0,320	0,307	0,295
Dánsko	0,216	0,226	0,239	0,289	0,249
Finsko	0,218	0,223	0,225	0,233	0,208
Německo	0,243	0,234	0,249	0,257	0,266
Norsko	0,214	0,222	0,217	0,220	0,184
Slovensko	0,325	0,320	0,294	0,290	0,287
Švédsko	0,219	0,223	0,228	0,217	0,227

Tabulka 4: Giniho indexy v zemích

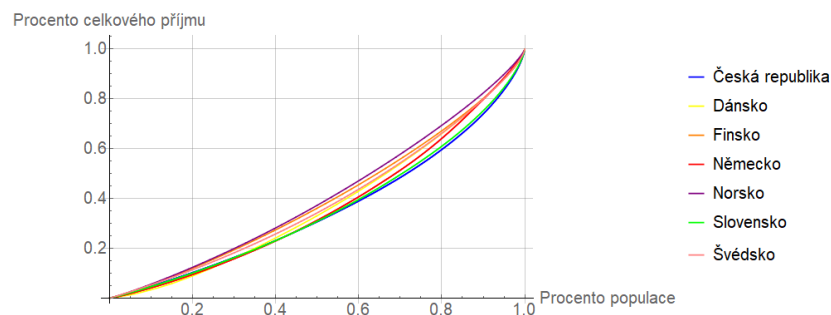


Obrázek 9: Vývoj Giniho indexu v čase

Byly vykresleny Lorenzovy křivky pro rok 2019, které znázorňují rozdělení celkového příjmu ve společnosti (viz obrázek 10). Nejblíže přímce, která by znamenala rovné rozdělení, má Norsko s Giniho indexem 18,4 %. Naopak nejvíce nerovné rozdělení má Česká republika s indexem 29,5 % a indexy ostatních států se nachází mezi nimi.

Vypočtené hodnoty Giniho indexu je možné porovnat s těmi, které uvádí Světová banka (viz tabulka 5), [13]. Zatím byly uvedeny pouze do roku 2016 nebo 2017. Podle těchto údajů má nejvyšší Giniho index Německo (v posledním uvedeném roce 31,9 %) a nejnižší Česká republika (24,9 %) a Slovensko (25,2 %). Poslední uvedená hodnota v Dánsku je 28,7 %, ve Finsku 27,1 %, v Norsku 27 % a ve Švédsku 28,8 %.

U České republiky a Slovenska jsou vypočtené hodnoty vyšší, než uvádí Světová banka. Tudíž podle Světové banky jsou zde nerovnosti menší než podle našich dat. U ostatních zemí naopak Světová banka uvádí větší indexy, a tak jsou podle ní nerovnosti větší, než jsme vypočetli. Rozdíl může být způsoben povahou dat, o jejichž detailním původu nemáme bohužel informace (jejich struktura je více popsána v kapitole 3.2.2). Například nemáme k dispozici data osob s jinými příjmy, než je zaměstnání, třeba z podnikání nebo brigády.



Obrázek 10: Lorenzovy křivky v roce 2019 pro jednotlivé země

	2015	2016	2017
Česká republika	0,259	0,254	0,249
Dánsko	0,282	0,282	0,287
Finsko	0,271	0,271	-
Německo	0,317	0,319	-
Norsko	0,275	0,285	0,270
Slovensko	0,265	0,252	-
Švédsko	0,292	0,296	0,288

Tabulka 5: Giniho indexy uvedené Světovou bankou (chybějící údaje Světová banka zatím ne-zveřejnila), [13]

3.1.2 Nerovnosti u mužů a žen

Všechna data byla rozdělena na muže a ženy a v obou skupinách byly vypočteny Giniho indexy. Například v České republice byl v posledním roce Giniho index pro muže 31,2 % a pro ženy 24,4 %, ve Finsku 21,2 % pro muže a 19,3 % pro ženy, na Slovensku 28,7 % pro muže a 27,2 % pro ženy a ve Švédsku 22,9 % pro muže a 21,9 % pro ženy. V těchto zemích mají muži větší Giniho index než ženy. To znamená, že mají mezi sebou větší příjmové nerovnosti než ženy. Pouze v Německu byly příjmové nerovnosti mezi muži menší než mezi ženami, Giniho index pro muže byl 25,9 % a pro ženy 26,5 %. V Dánsku a Norsku jsou indexy pro obě pohlaví přibližně stejné (v Dánsku 24,1 % a 24,2 % a v Norsku 18,3 % a 18,1 %), a tak jsou příjmové nerovnosti u obou pohlaví také téměř stejné. Hodnoty indexů v předchozích letech jsou v tabulkách 6 a 7.

	2015	2016	2017	2018	2019
Česká republika	0,357	0,334	0,332	0,322	0,312
Dánsko	0,220	0,245	0,250	0,282	0,241
Finsko	0,221	0,229	0,232	0,234	0,212
Německo	0,242	0,230	0,243	0,254	0,259
Norsko	0,217	0,228	0,220	0,222	0,183
Slovensko	0,326	0,370	0,289	0,290	0,287
Švédsko	0,215	0,223	0,228	0,220	0,229

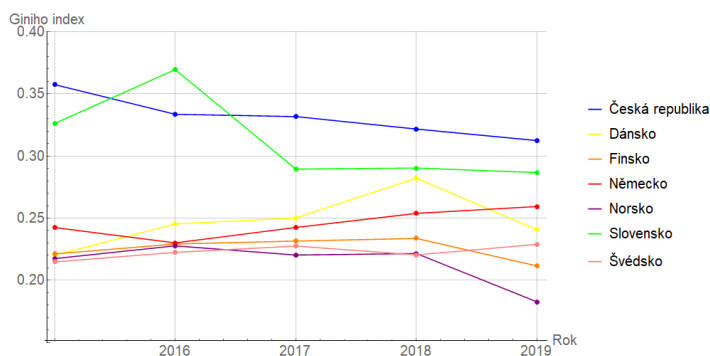
Tabulka 6: Giniho indexy pro muže

	2015	2016	2017	2018	2019
Česká republika	0,297	0,287	0,272	0,254	0,244
Dánsko	0,191	0,186	0,197	0,283	0,242
Finsko	0,201	0,201	0,204	0,213	0,193
Německo	0,232	0,233	0,251	0,257	0,265
Norsko	0,198	0,202	0,202	0,209	0,181
Slovensko	0,310	0,270	0,288	0,277	0,272
Švédsko	0,217	0,213	0,220	0,206	0,219

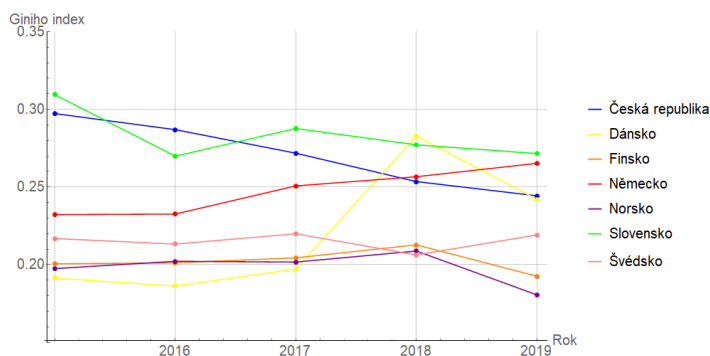
Tabulka 7: Giniho indexy pro ženy

Největší rozdíly mezi indexy mužů a žen jsou v České republice (v roce 2018 a 2019 byl rozdíl 6,8 %), Dánsku mezi lety 2016 a 2017 (5,1 % a 5,3 %) a na Slovensku v roce 2016 (10 %). Slovensko v roce 2016 se může lišit z důvodu menšího počtu záznamů, který byl zjištěn v kapitole 2.2.

Časový vývoj Giniho indexů je znázorněn na obrázcích 11 a 12. U mužů lze pozorovat podobný časový vývoj jako u celých zemí (muži tvoří ve všech datech větší část než ženy, jak bylo znázorněno v kapitole 2.2 na obrázku 6). Největší Giniho index je opět v České republice a na Slovensku. V čase



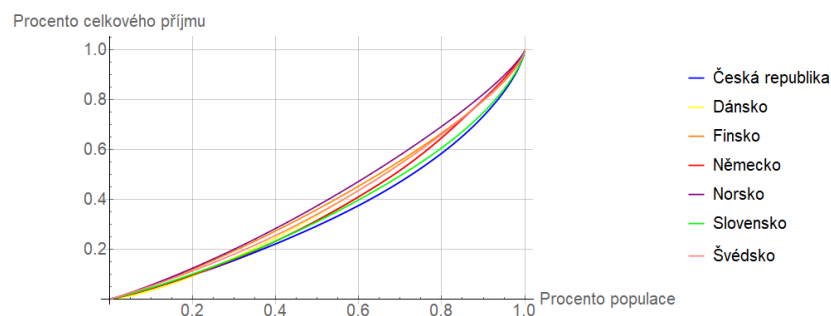
Obrázek 11: Vývoj Giniho indexu v čase - muži



Obrázek 12: Vývoj Giniho indexu v čase - ženy

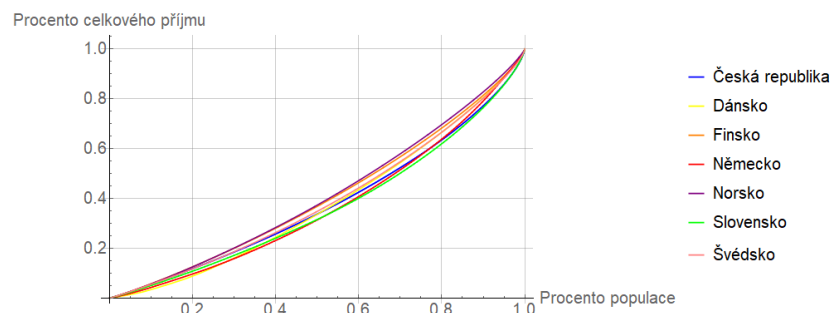
zde jeho hodnota klesá (kromě již zmíněného Slovenska v roce 2016, které se liší). V České republice klesl z počátečních 35,7 % na 31,2 % a na Slovensku z 32,6 % na 28,7 %. To znamená, že příjmové nerovnosti mezi muži se zde v čase zmenšují a zároveň se přibližují hodnotám v ostatních zemích. V Německu, Dánsku a Švédsku se hodnota Giniho indexu naopak zvýšila a zvětšují se tedy i příjmové nerovnosti mezi muži. V Německu se z počátečních 24,2 % zvýšila na 25,9 %, v Dánsku z 22 % na 24,1% a ve Švédsku z 21,5 % na 22,9 %. Kromě toho v Dánsku došlo ke zvětšení na 28,2 % v roce 2018 a v dalším roce opět k poklesu podobně jako u Giniho indexu celé země. Nejnižší Giniho indexy, a tedy i nejmenší příjmové nerovnosti mezi muži jsou opět v Norsku a Finsku, kde poslední rok dokonce došlo k poklesu. V Norsku je poslední hodnota 18,3 % a ve Finsku 21,2 %.

U žen jsou časové vývoje Giniho indexu v jednotlivých zemích také podobné jako u celých zemí, ale rozdíly mezi indexy v různých zemích jsou menší. Německo a Dánsko mělo v posledních dvou letech vyšší Giniho index (Německo 25,7 % a 26,5 % a Dánsko 28,3 % a 24,2 %) než Česká republika (25,4 % a 24,4 %), tj. ženy v Německu a Dánsku měly mezi sebou v posledních dvou letech větší příjmové nerovnosti než ženy v České republice. V Dánsku se znovu liší rok 2018, kdy byl Giniho index ze všech zemí nejvyšší (28,3 %). Dalším rozdílem oproti mužům a oběma pohlavím dohromady je, že nejvyšší Giniho index, a tedy i největší příjmové nerovnosti mezi ženami má po celou dobu převážně Slovensko (v posledním roce 27,2 %).



Obrázek 13: Lorenzovy křivky pro muže v roce 2019

Byly vykresleny také Lorenzovy křivky pro rok 2019, abychom znázornili, jak je celkový příjem mužů a žen rozdělen (viz obrázky 13 a 14). U mužů je situace opět podobná jako u celých zemí.



Obrázek 14: Vývoj Lorenzovy křivky pro ženy v roce 2019

Nejblíže rovnému rozdělení má Norsko (Giniho index je 18,3 %) a nejdále Česká republika (31,2 %). V případě žen se Lorenzovy křivky trochu liší od celých zemí. Nejblíže rovnému rozdělení má kromě Norska také Finsko (Giniho indexy jsou 18,1 % a 19,3 %). Naopak největší nerovnosti jsou podle křivky v Německu a na Slovensku (26,5 % a 27,2 %).

3.2 Příjmová mezera

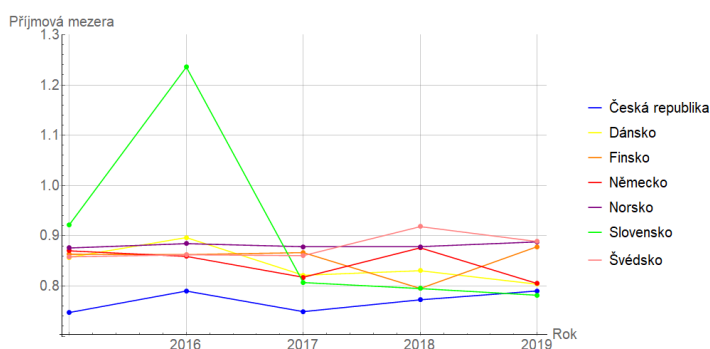
Dalším vypočteným ukazatelem je příjmová mezera, která slouží k analýze nerovností mezi skupinami. Byla popsána v kapitole 1.2.

3.2.1 Příjmová mezera mezi muži a ženami

Nejdříve byla vypočtena příjmová mezera mezi muži a ženami jako poměr mediánu roční mzdy žen a mediánu roční mzdy mužů. Popisuje, kolika procentům mediánu mzdy mužů je roven medián mzdy žen.

	2015	2016	2017	2018	2019
Česká republika	0,747	0,790	0,749	0,773	0,790
Dánsko	0,857	0,897	0,822	0,831	0,804
Finsko	0,863	0,863	0,867	0,795	0,879
Německo	0,870	0,859	0,818	0,876	0,805
Norsko	0,876	0,885	0,878	0,879	0,888
Slovensko	0,922	1,236	0,807	0,795	0,782
Švédsko	0,858	0,862	0,861	0,918	0,889

Tabulka 8: Příjmová mezera mezi pohlavími



Obrázek 15: Časový vývoj příjmové mezery mezi pohlavími

V České republice dosahovaly v roce 2019 ženy na 79 % platu mužů a na Slovensku 78,2 %. Jedná se o nejnižší hodnoty ze všech zemí. Naopak nejvyšších hodnot nabývá příjmová mezera v Norsku, Finsku a Švédsku (88,8 %, 87,9 % a 88,9 %). V Dánsku dosahují ženy na 80,4 % platu mužů a v Německu na 87 %.

Výsledky z předchozích let jsou uvedeny v tabulce 8 a na obrázku 15 je znázorněn vývoj příjmové mezery v čase. U Dánska a Německa si můžeme všimnout, že příjmová mezera v čase poklesla a ženy tedy dosahují na menší procento platu mužů než před pěti lety. V roce 2015 dosahovaly v Dánsku na 85,7 % a v Německu na 87 %. V roce 2019 už jen na 80,4 % v Dánsku a na 80,5 % v Německu.

Poklesu z 92,2 % na 78,2 % si lze také všimnout u Slovenska (pokud znovu neuvažujeme rok 2016, ze kterého je málo dat).

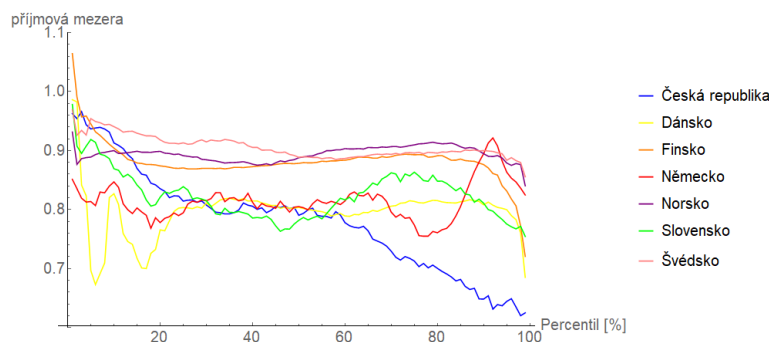
Příjmovou mezeru v evropských zemích zveřejňuje Eurostat, ale používá jiný způsob výpočtu, u kterého je výsledkem, o kolik procent mají ženy nižší mzdu než muži [14]. Vyšší příjmová mezera v tomto případě znamená na rozdíl od našich výpočtů větší rozdíly mezi muži a ženami. Z toho důvodu nemůžeme porovnat konkrétní hodnoty.

Podle Eurostatu je příjmová mezera v České republice 18,9 %, v Dánsku 14 %, ve Finsku 16,6 %, v Německu 19,2 %, v Norsku 13,2 %, na Slovensku 18,4 %, ve Švédsku 11,8 % a průměr v Evropské unii je 14,1 %, [14]. Zajímavé je, že Finsko je jako jediná skandinávská země nad průměrem Evropské unie, ale podle našich dat je to jedna ze zemí, kde jsou nerovnosti mezi muži a ženami nejmenší.

Dále byla vypočtena i příjmová mezera pro percentily v roce 2019, abychom věděli, zda se na různých percentilech liší. Výsledky vybraných percentilů jsou v tabulce 9 a graf závislosti příjmové mezery na percentilu na obrázku 16. Z něj lze na první pohled vidět, že v České republice se příjmová mezera se zvyšujícím percentilem zmenšuje. Na prvním percentilu si ženy vydělávají téměř stejně jako muži (příjmová mezera je 96,3 %), ale na devadesátém devátém už dosahují už jen na 62,5 % platu mužů.

percentil	Česká rep.	Dánsko	Finsko	Německo	Norsko	Slovensko	Švédsko
1	0,963	0,986	1,065	0,851	0,932	0,978	0,969
10	0,913	0,827	0,904	0,847	0,900	0,869	0,941
25	0,814	0,803	0,871	0,795	0,891	0,839	0,913
50	0,790	0,804	0,879	0,805	0,888	0,782	0,889
75	0,713	0,814	0,893	0,774	0,910	0,863	0,896
90	0,648	0,808	0,876	0,895	0,895	0,811	0,900
99	0,625	0,685	0,721	0,825	0,840	0,754	0,856

Tabulka 9: Příjmová mezera mezi muži a ženami pro vybrané percentily



Obrázek 16: Příjmová mezera mezi muži a ženami pro percentily

3.2.2 Příjmová mezera mezi muži a ženami na jednotlivých úrovních a v oborech

Doted' byla příjmová mezera počítána pro všechna data dohromady, ale bylo by lepší porovnávat mezi sebou muže a ženy, kteří pracují ve stejném oboru a na stejné pozici. Jinak se může stát, že u žen bude většina záznamů například o uklízečkách a u mužů to budou například právníci ve vedoucích pozicích. Uklízečky by dosahovaly na malé procento platu právníků, a tak by příjmová mezera byla nízká.

Abychom si udělali lepší představu, o jakých zaměstnancích a firmách máme záznamy, bylo vypočteno, kolik procent zaměstnanců je na jednotlivých úrovních (pracovních pozicích). V tabulce 10 jsou uvedeny výsledky pro rok 2019.

V České republice tvoří 49,7 % dat z roku 2019 řadoví pracovníci, 43,87 % vedoucí pracovníci, 6,2 % manažerů a 0,23 % řídicích pracovníků. Na Slovensku je 51,77 % řadových zaměstnanců, 43,3 % vedoucích, 4,72 % manažerů a 0,21 % řídicích pracovníků. V obou zemích tedy převládají řadoví pracovníci a je zde poměrně velké procento vedoucích.

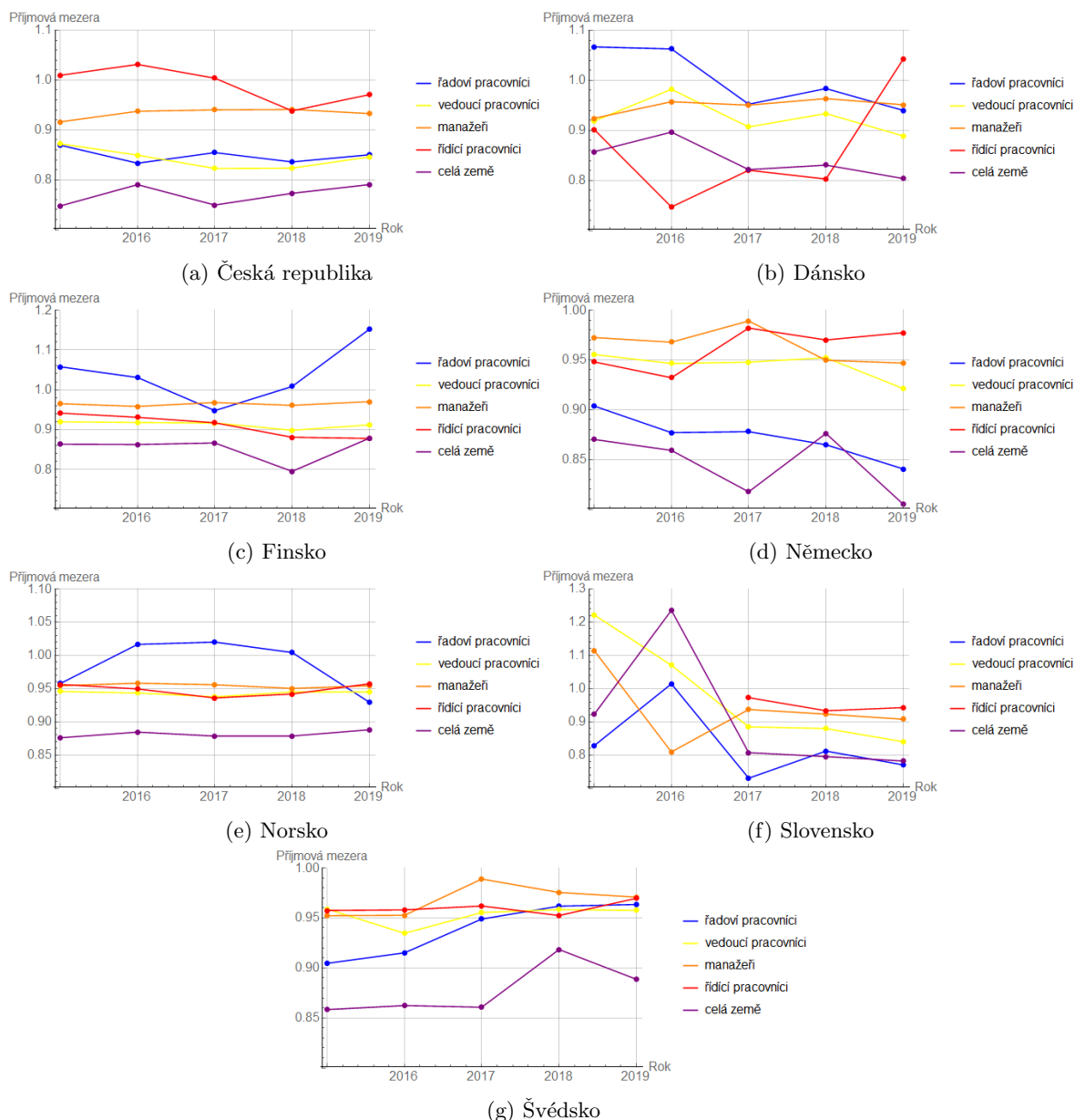
V Dánsku, Finsku, Německu a Švédsku tvoří většinu záznamů z roku 2019 vedoucí pracovníci. V Dánsku 55,97 %, ve Finsku 65,22 %, v Německu 50,11 % a ve Švédsku 59,45 %. Oproti České republice a Slovensku je zde menší procento řadových zaměstnanců a větší procento manažerů a řídicích pracovníků. Přesné počty procent jsou uvedeny v tabulce 10.

V Norsku tvoří většinu záznamů manažerů (53,38 %). Vedoucích je 37,77 % a řídicích pracovníků 6,74 %. Je zde poměrně málo řadových zaměstnanců (2,11 %).

	řadoví prac.	vedoucí prac.	manažeři	řídící prac.
Česká republika	49,70	43,87	6,20	0,23
Dánsko	18,41	55,97	23,66	1,95
Finsko	4,08	65,22	27,95	2,75
Německo	24,70	50,11	23,25	1,95
Norsko	2,11	37,77	53,38	6,74
Slovensko	51,77	43,30	4,72	0,21
Švédsko	15,96	59,45	22,96	1,63

Tabulka 10: Procenta osob na jednotlivých úrovních (pracovních pozicích) v roce 2019

Data byla rozdělena podle pracovních pozic. V každé ze skupin byla vypočtena příjmová mezera mezi muži a ženami, abychom zjistili, zda se na jednotlivých úrovních liší. Na obrázku 17 jsou zobrazeny jejich časové vývoje a v tabulce 11 výsledky z roku 2019.



Obrázek 17: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a v zemích celkově

Můžeme si všimnout, že pokud data rozdělíme, je ve většině zemí příjmová mezera na jednotlivých úrovních vyšší, než když byla počítána ze všech záznamů dohromady. Někde je i větší než jedna (např. v roce 2019 u řadových pracovníků ve Finsku je 115,2 % nebo u řídicích pracovníků v Dánsku je 104,3 %). V České republice, Německu, Švédsku a na Slovensku (opět kromě let 2015 a 2016, kdy bylo málo slovenských dat) lze vidět vyšší hodnoty příjmové mezery spíše u řídicích pracovníků

	řadoví prac.	vedoucí	manažeři	řídící prac.	celá země
Česká republika	0,850	0,846	0,933	0,971	0,790
Dánsko	0,940	0,889	0,951	1,043	0,804
Finsko	1,152	0,912	0,970	0,878	0,879
Německo	0,841	0,921	0,947	0,977	0,805
Norsko	0,929	0,945	0,954	0,957	0,888
Slovensko	0,770	0,840	0,908	0,943	0,782
Švédsko	0,963	0,958	0,971	0,970	0,889

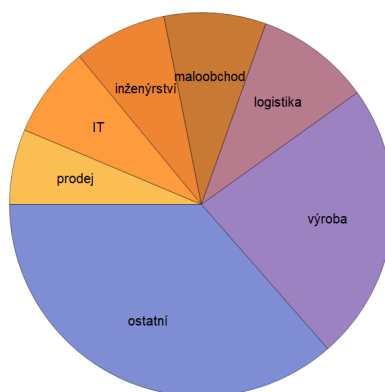
Tabulka 11: Příjmová mezera mezi muži a ženami na jednotlivých úrovních a v celých zemích v roce 2019

a manažerů a v ostatních skandinávských zemích spíše u řadových pracovníků.

Data obsahují také informaci o oboru zaměstnání. Celkem máme k dispozici ze všech zemí a let 4 158 520 záznamů ze 140 různých oborů. V tabulce 12 jsou počty údajů z každého oboru a jejich procento z celkového počtu záznamů. Do tabulky byly vybrány pouze ty, které tvoří alespoň 1 % všech dat. K další práci byly použity pouze obory, které tvoří alespoň 5 % všech dat: výroba (23,46 %), logistika (9,64 %), maloobchod (8,59 %), inženýrství (7,83 %), IT (7,69 %) a prodej (6,34 %). Struktura dat je zobrazena na obrázku 18.

Obor	Počet záznamů	Procento z celk. počtu
lidské zdroje	73 274	1,76
marketing	83 720	2,01
bankovníctví	102 497	2,46
výzkum a vývoj	103 834	2,50
projektový a programový management	108 866	2,57
zajištění kvality	115 616	2,78
zákaznický servis	141 320	3,40
administrativa	147 928	3,56
finančníctví a účetnictví	197 465	4,75
prodej	263 817	6,34
IT	319 659	7,69
inženýrství	325 841	7,84
maloobchod	357 138	8,59
logistika	400 986	9,64
výroba	975 676	23,46

Tabulka 12: Počty oborů zaměstnání ve všech datech a procenta z celkového počtu záznamů (byly vybrány pouze obory, které tvořily alespoň 1 % záznamů z roku 2019)

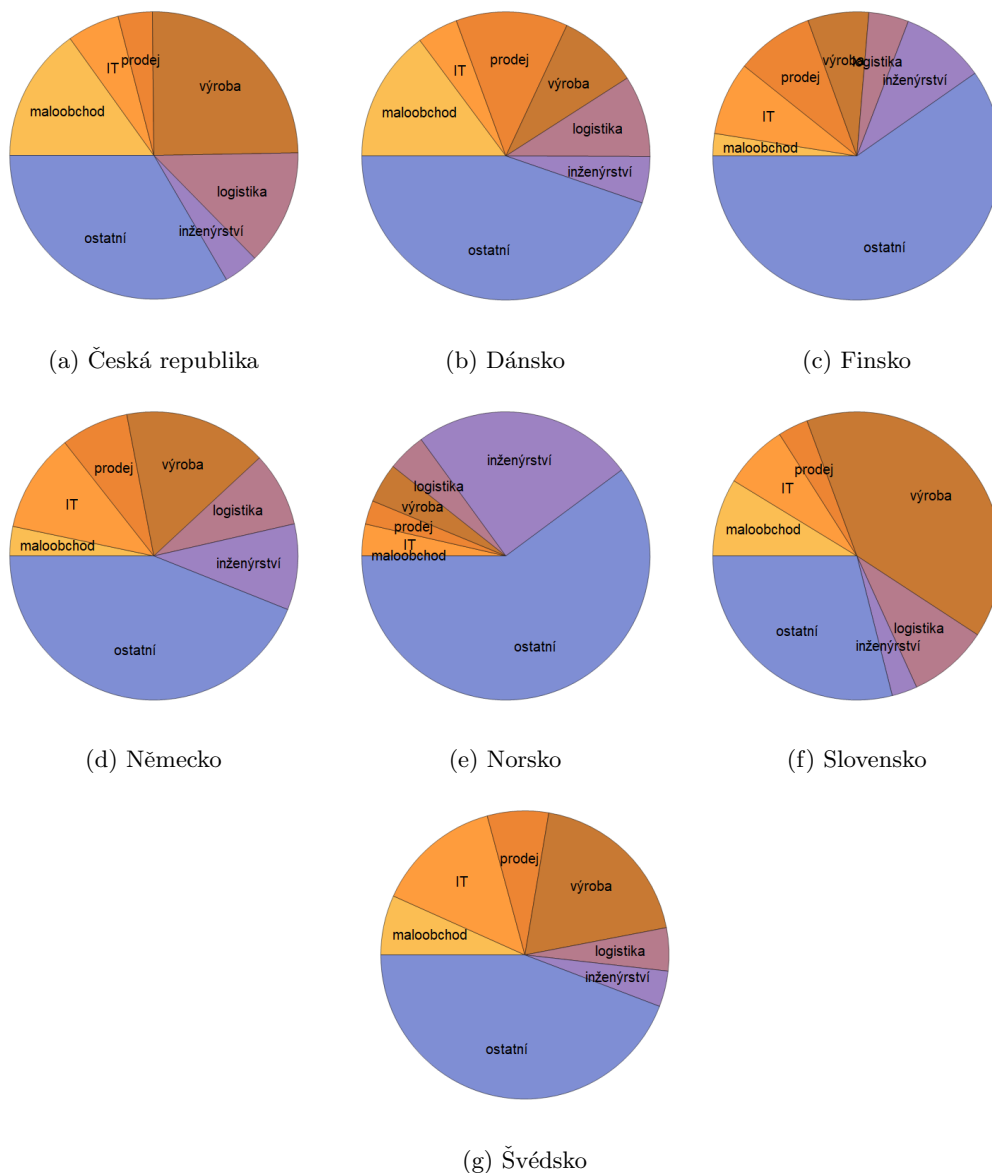


Obrázek 18: Struktura všech dat

Na obrázku 19 je zobrazena struktura dat v jednotlivých zemích v roce 2019. V České republice a na Slovensku je nejvíce zastoupena výroba. V České republice tvoří v roce 2019 25,9 % všech záznamů a na Slovensku 41,01 %. V Německu a Švédsku tvoří také poměrně velkou část dat výroba (17,83 % a 20,11 %). V Norsku je zase nejčastější inženýrství, které tvoří 33,05 % dat.

Data byla dále rozdělena i podle oborů. Celkem jsou tedy rozdělena podle států, pozic a oborů, aby bylo možné spočítat příjmovou mezera například mezi muži a ženami z České republiky, kteří jsou zaměstnáni v oboru inženýrství na manažerské pozici. V následujících tabulkách jsou výsledky z posledního roku. Pokud někde chybí, tak v dané zemi a daném roce nebyli žádní muži nebo ženy.

Výsledky pro Českou republiku jsou uvedeny v tabulce 13. Příjmová mezera vypočtená mezi všemi



Obrázek 19: Struktura dat v roce 2019 v jednotlivých zemích

muži a ženami v České republice byla 79 %. S výjimkou řadových pracovníků v oboru výroba, kde ženy dosahují na 78 % platu mužů a vedoucích pracovníků v oboru prodej, kde dosahují na 75,6 %, je příjmová mezera vždy větší (tj. ženy dosahují na větší procento platu mužů) než v celé zemi.

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	1,071	0,962	1,219	0,780	0,896	0,986
vedoucí prac.	0,947	0,854	0,756	0,926	0,939	0,901
manažeri	0,987	0,914	0,856	0,930	0,946	0,946
řídící prac.	1,200	-	0,900	-	0,758	-

Tabulka 13: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání v České republice v roce 2019

Výsledky příjmové mezery v Dánsku jsou uvedeny v tabulce 14. V celé zemi v tomto roce dosahovaly ženy na 80,4 % platu mužů. Pokud data rozdělíme podle pozic a oborů, je příjmová mezera vždy větší nebo stejná. To znamená, že ženy dosahují na stejné nebo vyšší procento platu mužů. U řadových pracovníků ve všech oborech je příjmová mezera dokonce větší než 100 %, a tak si ženy na těchto pozicích vydělají více než muži.

Výsledky z Finska jsou v tabulce 15. V celé zemi byla příjmová mezera 87,9 %. Kromě řídicích pracovníků v oboru výroba (86,7 %) je všude příjmová mezera vyšší (tj. ženy dosahují na větší část platu mužů) než v celé zemi. V některých skupinách je příjmová mezera i 100 % nebo dokonce vyšší.

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	1,000	-	1,141	1,052	1,029	-
vedoucí prac.	0,897	0,987	0,880	0,901	0,893	0,919
manažeři	-	0,891	0,975	0,949	0,981	0,960
řídící prac.	-	-	0,804	-	0,982	-

Tabulka 14: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání v Dánsku v roce 2019

Jde o řadové pracovníky v oborech maloobchod (100 %), IT (123,8 %), výroba (105 %) a logistika (120,1 %), vedoucí pracovníky v oboru maloobchod (125,6 %) a řídící pracovníky v logistice (124,6 %).

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	1,000	1,238	0,932	1,056	1,201	0,937
vedoucí prac.	1,256	0,949	0,873	0,924	0,897	0,967
manažeři	0,982	0,958	0,927	0,971	0,985	0,951
řídící prac.	-	0,904	0,931	0,867	1,246	-

Tabulka 15: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání ve Finsku v roce 2019

Příjmová mezera v Německu v různých oborech a pozicích je uvedena v tabulce 16. V celé zemi byla 80,5 %. Na první pohled si můžeme všimnout, že ženy dosahují na poměrně malé procento platu mužů, pokud se jedná o řídící pracovníky v maloobchodu (57,2 %) a řadové zaměstnance v prodeji (58,2 %). V ostatních skupinách je příjmová mezera větší než v celé zemi. Ženy si vydělávají stejně jako muži, když jsou zaměstnány v oboru IT jako manažerky nebo na řídící pozici. Jako řadové pracovníce v maloobchodu si vydělávají dokonce 112,6 % platu mužů, jako řadové pracovníce v inženýrství 103 % a jako řídící pracovníce v inženýrství 108,6 %.

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	1,126	0,968	0,582	0,859	0,967	1,030
vedoucí prac.	0,964	0,981	0,828	1,036	0,946	0,950
manažeři	0,995	1,000	0,950	0,938	0,930	0,900
řídící prac.	0,572	1,000	0,921	0,891	0,876	1,086

Tabulka 16: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání v Německu v roce 2019

Výsledky z Norska jsou uvedeny v tabulce 17, kde si můžeme všimnout, že v oboru maloobchod nebyli žádní muži nebo ženy, a proto nemáme výsledky. V celé zemi byla v roce 2019 příjmová mezera 88,8 %. Na menší část platu mužů zde dosahovaly pouze ženy zaměstnané jako řadové pracovníce v IT (75 %) a jako řídící pracovníce v prodeji (80,8 %). Ve většině skupin je tedy příjmová mezera opět vyšší, než když byla počítána z celé země dohromady. Kromě dvou zmíněných skupin je příjmová mezera všude alespoň 90 % a u několika je i vyšší než 100 %.

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	-	0,750	0,993	0,936	1,067	0,952
vedoucí prac.	-	0,977	0,900	1,010	0,961	0,959
manažeři	-	0,959	0,901	1,063	0,926	0,971
řídící prac.	-	0,980	0,808	-	1,368	0,978

Tabulka 17: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání v Norsku v roce 2019

V tabulce 18 jsou uvedeny výsledky příjmové mezery na Slovensku. Nejprve si můžeme všimnout, že chybí výsledky řídících pracovníků, protože ve zvolených oborech nebyli na této pozici žádní muži nebo ženy. V celé zemi ženy pobíraly 78,2 % platu mužů. Ve všech skupinách je tedy příjmová mezera vyšší. Řadové pracovníce v logistice dosahují na 100 % platu mužů a řadové pracovníce v inženýrství dokonce na 106,4 %.

Výsledky ze Švédska jsou v tabulce 19. V celé zemi byla příjmová mezera 88,9 %. Příjmová mezera je tedy opět ve většině skupin vyšší kromě vedoucích v maloobchodě (87,4 %) a v prodeji (88,5 %). Je zde i několik skupin, kde je příjmová mezera větší než 100 %.

Časové vývoje příjmové mezery u dat rozdělených podle pracovní pozice a oboru zaměstnání jsou znázorněny v příloze C na obrázcích 27 až 33. Celkově lze ale říci, že pokud data takto rozdělíme, dosahují ženy ve většině případů na vyšší procento platu mužů než u nerozdělených dat.

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	0,910	0,965	0,840	0,746	1,000	1,064
vedoucí prac.	0,829	0,835	0,818	0,872	0,907	0,857
manažeři	0,930	0,883	0,865	0,962	0,889	0,997
řídící prac.	-	-	-	-	-	-

Tabulka 18: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání na Slovensku v roce 2019

	maloobchod	IT	prodej	výroba	logistika	inženýrství
řadoví prac.	0,996	0,753	1,229	0,985	1,041	1,095
vedoucí prac.	0,874	0,967	0,885	0,979	0,951	1,016
manažeři	1,069	0,988	0,935	0,983	0,945	0,938
řídící prac.	0,957	1,294	0,9853	1,125	0,957	-

Tabulka 19: Příjmová mezera mezi muži a ženami na jednotlivých pozicích ve vybraných oborech zaměstnání ve Švédsku v roce 2019

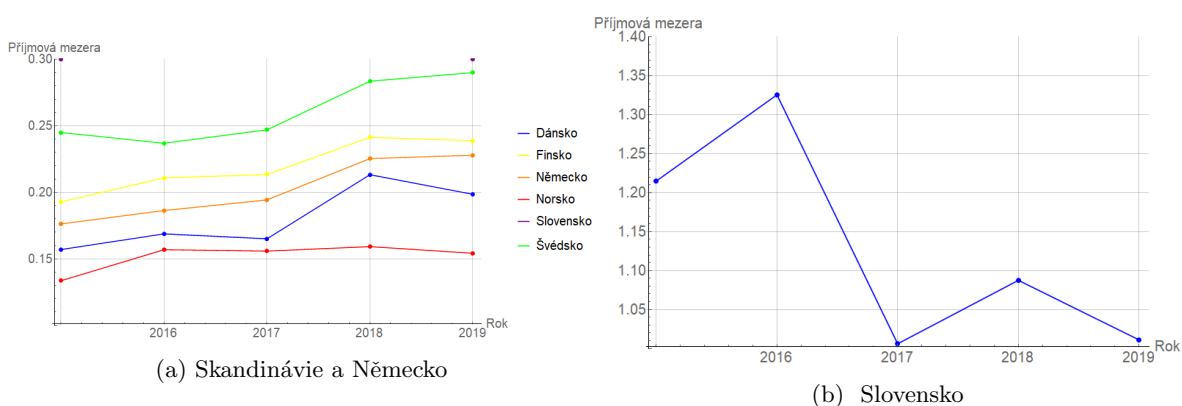
3.2.3 Porovnání České republiky s ostatními zeměmi

Byla vypočtena také příjmová mezera mezi Českou republikou a ostatními státy jako podíl mediánu roční mzdy České republiky a mediánu dané druhé země. Výsledkem je, kolika procentům mediánu ročního platu vybrané země je roven medián ročního platu České republiky.

Jediná země, která má medián platu nižší než Česká republika, je Slovensko, a tak je zde příjmová mezera větší než 1 (respektive 100 %). Například v roce 2019 byla 101,1 %. Z ostatních zemí má medián mzdy České republiky nejbližší ke Švédsku. V posledním roce dosahovaly české platy na 29 % švédského platu. Nejnižší je příjmová mezera mezi Českou republikou a Norskem (15,4 %). Dále Česká republika dosahuje na 19,9 % dánského platu, 23,9 % finského a 22,8 % německého. Výsledky z ostatních let jsou uvedeny v tabulce 20 a na obrázku 20 je znázorněn časový vývoj příjmové mezery.

	2015	2016	2017	2018	2019
Dánsko	0,157	0,169	0,165	0,213	0,199
Finsko	0,193	0,211	0,213	0,241	0,239
Německo	0,176	0,186	0,194	0,225	0,228
Norsko	0,134	0,157	0,156	0,159	0,154
Slovensko	1,215	1,326	1,006	1,088	1,011
Švédsko	0,245	0,237	0,247	0,284	0,290

Tabulka 20: Příjmová mezera mezi Českou republikou a ostatními zeměmi vypočtená jako podíl mediánu mzdy České republiky a mediánu mzdy dané druhé země



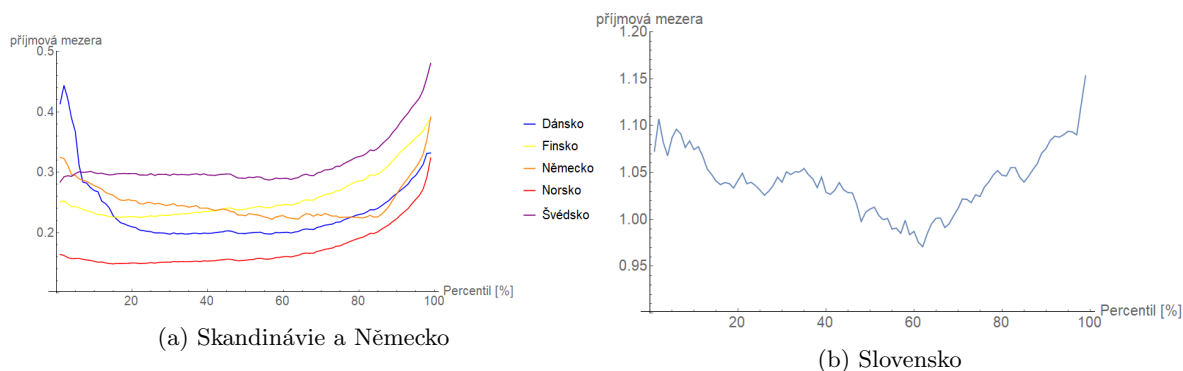
Obrázek 20: Časový vývoj příjmové mezery mezi Českou republikou a ostatními zeměmi

Opět byla vypočtena i příjmová mezera pro percentily v roce 2019. Výsledky vybraných percentilů jsou v tabulce 21 a závislost příjmové mezery na percentilu je znázorněna na obrázku 21. U skandinávských zemí si na první pohled můžeme všimnout, že přibližně od sedmdesátého percentilu příjmová mezera roste, zatímco u nižších zůstává přibližně stejná nebo klesá. Také si můžeme všimnout, že u nižších percentilů je příjmová mezera mezi Dánskem a Českou republikou poměrně vysoká. Například na prvním percentilu dosahuje český plat na 41,3 % dánského. U Německa příjmová mezera klesá přibližně do devadesátého percentilu a od něj roste. U Slovenska klesá zhruba do

šedesátého na hodnotu menší než jedna a potom zase roste. Jinými slovy přibližně na šedesátém percentilu si Slováci vydělávají více než Češi a na ostatních mají větší plat Češi.

percentil	Dánsko	Finsko	Německo	Norsko	Slovensko	Švédsko
1	0,413	0,251	0,325	0,164	1,077	0,284
10	0,270	0,232	0,278	0,153	1,074	0,299
25	0,202	0,228	0,250	0,151	1,031	0,298
50	0,199	0,239	0,228	0,154	1,011	0,290
75	0,218	0,270	0,225	0,178	1,024	0,310
90	0,265	0,326	0,264	0,223	1,075	0,373
99	0,332	0,390	0,391	0,324	1,153	0,480

Tabulka 21: Příjmová mezera mezi Českou republikou a ostatními zeměmi - vybrané percentily



Obrázek 21: Příjmová mezera pro percentily v roce 2019 mezi Českou republikou a ostatními státy

4 Zpracování dat - ANOVA

Dalším krokem zpracování dat byla ANOVA (analýza rozptylu), která byla představena v kapitole 1.3. Pokusíme se zjistit, zda je roční plat funkcí dalších veličin jako je země, pohlaví, pozice nebo obor zaměstnání a jak ho tyto faktory ovlivňují. Při zpracování analýzy rozptylu byly použity pouze záznamy z roku 2019.

4.1 Jednofaktorová ANOVA - pohlaví

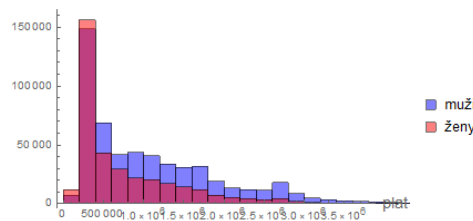
Prvním zkoumaným faktorem bylo pohlaví. Uvažujeme model platu

$$y_i = \mu + \alpha_i + e_i \quad (i \in \{1, 2\}), \quad (25)$$

kde μ je průměrný plat vypočtený ze všech záznamů, α_i jsou odchylky průměrných platů mužů a žen od μ a e_i jsou náhodné odchylky.

Nejprve bylo zjištěno, zda data splňují požadavky anovy. Pozorování z obou pohlaví mají být nezávislá a pocházet z normálního rozdělení s konstantním rozptylem, [9].

Data byla rozdělena na muže a ženy (viz obrázek 22) a pomocí Kolmogorovova-Smirnovova testu (viz kapitola 1.4) byla u obou skupin otestována nulová hypotéza, zda platy pochází z normálního rozdělení pravděpodobnosti.



Obrázek 22: Histogram platů rozdělených podle pohlaví

Počet mužů je 549 470. Výběrový průměr (odhad střední hodnoty) jejich platů je 1 105 940 a výběrová směrodatná odchylka je 941 287. Byla vypočtena distribuční funkce normálního rozdělení a testovací statistika, jejíž výsledek je 0,9975. Hladina významnosti byla zvolena 5 %. Kritická hranice je 0,0012, a proto zamítáme nulovou hypotézu, že platy mužů pochází z normálního rozdělení. Tomu odpovídá i p-hodnota rovná nule.

Podobně byl test proveden i pro platy žen. Jejich počet je 358 496, výběrový průměr je 760 343 a výběrová směrodatná odchylka je 708 334. Testové kritérium vyšlo 0,9974 a kritická hranice 0,0015, a tak byla nulová hypotéza o normalitě dat také zamítnuta na hladině významnosti 5 %. I v tomto případě byla p-hodnota rovna nule.

Dále byla pomocí F-testu (viz kapitola 1.5) otestována hypotéza, zda mají platy mužů a žen shodné rozptyly. Výběrový rozptyl mužů je $8,86 \times 10^{11}$ a výběrový rozptyl žen je $5,02 \times 10^{11}$. Testovací statistika je 1,7659. P-hodnota vyšla nula, a proto byla zamítnuta hypotéza H_0 o shodě rozptylů.

Předpoklady nebyly splněny, a tak mohou být výsledky analýzy rozptylu zkrácené. Neměl by to ale být velký problém z důvodu velkého množství dat. Celkově máme k dispozici 907 966 záznamů.

Zkoumali jsme platnost nulové hypotézy, že pohlaví nemá vliv na plat (střední hodnoty platů mužů a žen jsou shodné). Výsledky anovy jsou uvedeny v tabulce 22. Testové kritérium F je 35 289. To znamená, že meziskupinová variabilita je velká ve srovnání s vnitroskupinovou. P-hodnota testu je rovna 0, a proto zamítáme nulovou hypotézu, že pohlaví nemá vliv na velikost platu. Protože pohlaví má jen dvě úrovně, není potřeba zjišťovat, mezi kterými je významný statistický rozdíl.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
pohlaví	1	$2,59 \times 10^{16}$	$2,59 \times 10^{16}$	35 289	0
reziduální	907 964	$6,67 \times 10^{17}$	$7,34 \times 10^{11}$	-	-
celkem	907 965	$6,93 \times 10^{17}$	-	-	-

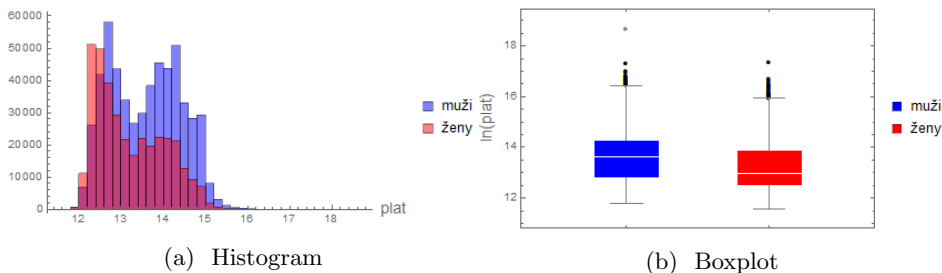
Tabulka 22: Jednofaktorová ANOVA - pohlaví

V tabulce 23 jsou průměrné platy obou pohlaví μ_i a odchylky celkového průměru α_i . Průměr všech platů μ je 969 488. Průměrný plat mužů je 1 105 940 a průměrný plat žen je 760 343. Jejich odchylky od váženého průměru α_i jsou 136 454 u mužů a -209 145 u žen. To znamená, že muži mají roční plat v průměru o 136 454 CZK vyšší než průměrný plat a ženy ho mají v průměru o 209 145 CZK menší. Asymetrie je dána větším počtem dat u mužů.

země	μ_i	$\hat{\alpha}_i$
všechny záznamy	969 488	-
muži	1 105 940	136 454
ženy	760 343	-209 145

Tabulka 23: Průměry platů u mužů a žen podle modelu (25)

Místo platů můžeme u anovy použít také jejich přirozené logaritmy, které jsou méně vychýlené doleva, a tak se více blíží normálnímu rozdělení (viz obrázek 23(a)), které je jedním z předpokladů anovy. Na obrázku 23(b) jsou vykreslené boxploty přirozených logaritmů platu mužů a žen.



Obrázek 23: Vizualizace přirozených logaritmů platů

Když používáme logaritmy platů, je model ve tvaru

$$\ln(y_i) = \mu + \hat{\alpha}_i + e_i \quad (i \in \{1, 2\}), \quad (26)$$

kde μ je průměr všech přirozených logaritmů platů, $\hat{\alpha}_i$ jsou odchylky průměrů logaritmů platů mužů a žen od μ a e_i jsou náhodné odchylky. Plat lze potom zapsat v multiplikačním tvaru jako

$$y_i = e^\mu e^{\hat{\alpha}_i} e^{e_i}. \quad (27)$$

Nejprve bylo opět zkoumáno, zda přirozené logaritmy platů splňují požadavky anovy. Pomocí Kolmogorova-Smirnova testu bylo otestováno, jestli pochází z normálního rozdělení.

U mužů je průměr 13,59 a směrodatná odchylka 0,82. Testovací statistika je 0,9916 a kritická hodnota 0,0012, a proto byla na hladině významnosti 5 % zamítnuta nulová hypotéza, že data pochází z normálního rozdělení. Odpovídá tomu i p-hodnota testu, která byla 0.

Průměr u žen je 13,20 a směrodatná odchylka 0,79. Testovací statistika je 0,9831, kritická hranice 0,0015 a p-hodnota testu 0. Nulová hypotéza byla proto také zamítnuta na hladině významnosti 5 %.

Dále byl použit F-test, abychom zjistili, zda jsou rozptyly u mužů a žen shodné. Výběrový rozptyl je 0,6799 u mužů a 0,6319 u žen. Testovací statistika vyšla 1,0760 a p-hodnota testu $7,38 \times 10^{-128}$. Nulová hypotéza byla také zamítnuta, protože p-hodnota je malá.

Předpoklady tedy nejsou splněny podobně jako u platů. Výsledky anovy jsou v tabulce 24. Testové kritérium F je 49 028, tudíž znovu vyšlo, že je velká variabilita mezi skupinami oproti vnitroskupinové. Tomu odpovídá i p-hodnota testu rovna 0, a proto zamítáme hypotézu, že pohlaví nemá vliv na plat.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
pohlaví	1	32 404	32 404	49 028	0
reziduální	907 964	600 094	0,66	-	-
celkem	907 965	632 497	-	-	-

Tabulka 24: Jednofaktorová ANOVA - pohlaví (s přirozenými logaritmy platů)

V tabulce 25 jsou průměry logaritmů platů obou pohlaví μ_i a odchylky $\hat{\alpha}_i$. Průměr μ všech záznamů je 13,44. To odpovídá platu 684 191 CZK. Průměrná hodnota u mužů je 13,59 a u žen 13,20. Po přepočtení podle modelu (27) vychází, že ženy mají plat 541 508 CZK a muži 796 979 CZK. Odchylky od průměru $\hat{\alpha}_i$ jsou 0,1526 u mužů a -0,2339 u žen. Podle modelu (27) to znamená, že plat žen je 79,15 % ze 684 191 CZK a plat mužů 116,45 % ze 684 191 CZK.

4.2 Jednofaktorová ANOVA - země

Dalším faktorem zkoumaným faktorem byla země. Model platu máme ve tvaru

$$y_i = \mu + \alpha_i + e_i \quad (i = 1, 2, \dots, 7), \quad (28)$$

země	μ_i	$\hat{\alpha}_i$	e^{μ_i}	$e^{\hat{\alpha}_i}$
všechny záznamy	14,436	-	684 191	-
muži	13,5886	0,1526	796 979	1,1645
ženy	13,2021	-0,2339	541 508	0,7915

Tabulka 25: Průměry přirozených logaritmu platů u mužů a žen podle modelů (26) a (27)

kde μ je průměrný plat vypočtený ze všech záznamů, α_i jsou odchylky průměrných platů v zemích od μ a e_i jsou náhodné odchylky.

Začali jsme opět kontrolou předpokladů. Platnost nulové hypotézy, že data rozdělená podle zemí pochází z normálního rozdělení, byla zkoumána pomocí Kolmogorovova-Smirnovova testu. U všech zemí je p-hodnota testu 0, a tak byly nulové hypotézy zamítnuty.

Dále jsme testovali, zda jsou rozptyly platů všech zemí shodné ¹. Tato hypotéza byla zamítnuta, protože p-hodnota testu byla 0.

Předpoklady anovy tedy znovu nejsou splněny, a tak mohou být její výsledky zkreslené. Z důvodu velkého množství dat by to ale neměl být problém.

Výsledky anovy jsou v tabulce 26. Testové kritérium F je 165 023, což značí velkou meziskupinovou variabilitu oproti vnitroskupinové. Nulovou hypotézu, že země nemá vliv na velikost platu, zamítáme, protože p-hodnota testu je 0.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
země	6	$3,61 \times 10^{17}$	$6,02 \times 10^{16}$	165 023	0
reziduální	907 959	$3,31 \times 10^{17}$	$3,65 \times 10^{11}$	-	-
celkem	907 965	$6,93 \times 10^{17}$	-	-	-

Tabulka 26: Jednofaktorová ANOVA - země

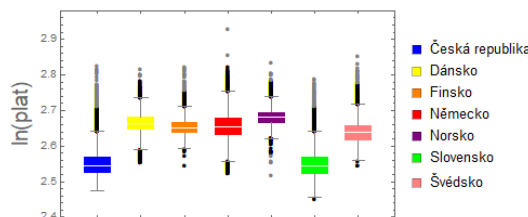
Podle výsledků anovy má země na velikost platu vliv. To vede k otázce, mezi kterými zeměmi je statisticky významný rozdíl. Pomocí Tukeyovy metody (viz kapitola 1.3) bylo zjištěno, že se jedná o všechny dvojice zemí.

Průměrné platy μ_i a odchylky od celkového průměru α_i jsou v tabulce 27. Průměrný plat všech zaměstnanců je 969 488 CZK. Pokud je zaměstnanec z České republiky, tak má plat průměrně o 537 595 CZK menší než průměr. Podobně zaměstnanci ze Slovenska ho mají o 558 071 CZK menší. Dánští zaměstnanci jej mají průměrně o 854 264 CZK větší než průměr, finští o 602 758 CZK větší, němečtí o 708 144 CZK větší, norští o 1 356 220 CZK větší a švédští o 337 906 CZK větší.

země	μ_i	α_i
všechny	969 488	-
Česká republika	431 893	-537 595
Dánsko	1 823 750	854 264
Finsko	1 572 250	602 758
Německo	1 677 630	708 144
Norsko	2 325 710	1 356 220
Slovensko	411 418	-558 071
Švédsko	1 307 390	337 906

Tabulka 27: Průměrné platy zemí podle modelu (28)

Analýzu rozptylu je opět možné udělat i pomocí přirozených logaritmu platů. Na obrázku 24 jsou zobrazeny jejich boxploty.



Obrázek 24: Boxplot přirozených logaritmu platů v zemích

¹Protože F test lze použít jen pokud testujeme shodu dvou rozptylů, bylo využito testu, který automaticky používal program Wolfram Mathematica. Stejně je to i v dalších případech, kde testujeme shodu více než dvou rozptylů.

Potom používáme model

$$\ln(y_i) = \mu + \hat{\alpha}_i + e_i \quad (i = 1, 2, \dots, 7), \quad (29)$$

kde μ je průměr všech přirozených logaritmů platů, $\hat{\alpha}_i$ jsou odchylky průměrů logaritmů platů v jednotlivých zemích od μ a e_i jsou náhodné odchylky. Platy potom můžeme vyjádřit jako

$$y_i = e^\mu e^{\hat{\alpha}_i} e^{e_i}. \quad (30)$$

Nejprve byla opět otestována platnost předpokladů. Pomocí Kolmogorovova-Smirnovova testu jsme otestovali, jestli data rozdělená podle zemí pochází z normálního rozdělení. U všech zemí je p-hodnota testu 0, a tak zamítáme nulovou hypotézu, že data pochází z normálního rozdělení. Také bylo otestováno, zda jsou shodné rozptyly logaritmů platů všech zemí. P-hodnota je rovna nule, a tak byla tato hypotéza zamítnuta. Požadavky nejsou splněny, a proto mohou být výsledky anovy opět zkreslené. Kvůli velkému počtu dat by se ale nemělo jednat o problém.

Výsledky analýzy rozptylu jsou v tabulce 28. Testové kritérium F je 340 721. P-hodnota je 0, a proto zamítáme nulovou hypotézu, že země nemá vliv na logaritmus platu.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
země	6	437 976	72 996	340 721	0
reziduální	907 959	194 521	0,21	-	-
celkem	907 965	632 497	-	-	-

Tabulka 28: Jednofaktorová ANOVA - země (s přirozenými logaritmy platů)

Průměrné hodnoty v zemích μ_i a odchylky od průměru všech dat $\hat{\alpha}_i$ jsou v tabulce 29. Průměr všech zaměstnanců je 13,4360, což odpovídá 684 191 CZK. V České republice je podle modelu (29) roční plat 375 424 CZK, v Dánsku 1 629 050 CZK, ve Finsku 1 462 840 CZK, v Německu 1 497 810 CZK, v Norsku 2 200 850 CZK, na Slovensku 360 691 CZK a ve Švédsku 1 203 560 CZK.

země	μ_i	$\hat{\alpha}_i$	e^{μ_i}	$e^{\hat{\alpha}_i}$
všechny záznamy	13,4360	-	684 191	-
Česká republika	12,8358	-0.6002	375 424	0,5487
Dánsko	14,3035	0.8675	1 629 050	2,3810
Finsko	14,1959	0.7599	1 462 840	2,1381
Německo	14,2195	0.7835	1 497 810	2,1892
Norsko	14,6044	1.1683	2 200 850	3,2167
Slovensko	12,7958	-0.6402	360 691	0,5272
Švédsko	14,0008	0.5648	1 203 560	1,7591

Tabulka 29: Průměry přirozených logaritmů platů v zemích podle modelů (29) a (30)

4.3 Jednofaktorová ANOVA - úroveň

Dále jsme provedli analýzu rozptylu, kde jsme zkoumali vliv pracovní pozice (úrovně). Uvažovali jsme model

$$y_i = \mu + \alpha_i + e_i \quad (i = 1, 2, \dots, 4), \quad (31)$$

kde μ je průměrný plat vypočtený ze všech záznamů, α_i jsou odchylky průměrných platů na jednotlivých úrovních od μ a e_i jsou náhodné odchylky.

Začali jsme kontrolou splnění předpokladů. Pro každou pozici vyšla p-hodnota Kolmogorovova-Smirnovova testu 0, a proto zamítáme nulovou hypotézu, že data mají normální rozdělení. U testu shody rozptylů byla p-hodnota také 0, a tak zamítáme i hypotézu, že platy na všech pozicích mají stejný rozptyl. Předpoklady tedy nebyly splněny ani u pracovních pozic, a tak může dojít ke zkreslení výsledků, které by ale nemělo být příliš velké z důvodu velkého množství dat.

Výsledky anovy jsou v tabulce 30. Testovací statistika F je 399 115, což opět značí velkou variabilitu mezi skupinami. Nulovou hypotézu, že pracovní pozice nemá vliv na velikost platu, zamítáme, protože p-hodnota testu je 0.

Otázkou je, mezi kterými pozicemi je statisticky významný rozdíl. Pomocí Tukeyovy metody bylo zjištěno, že podobně jako v předchozím případě jde o všechny možné dvojice pozic.

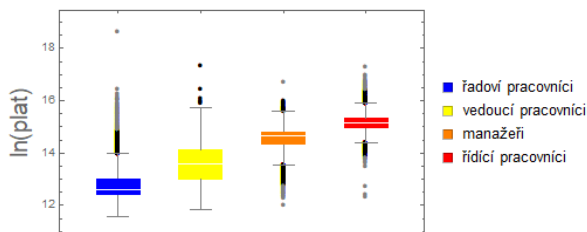
Průměrné platy všech skupin μ_i a odchylky od průměru všech záznamů α_i jsou v tabulce 31. Průměrný roční plat je 969 488 CZK. Řadoví pracovníci si vydělávají v průměru o 539 102 CZK ročně méně, vedoucí pracovníci o 28 389 CZK méně, manažeři o 1 221 010 CZK více a řídicí pracovníci o 3 100 770 CZK více.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
úroveň	3	$3,94 \times 10^{17}$	$1,31 \times 10^{17}$	399 115	0
reziduální	907 962	$2,99 \times 10^{17}$	$3,29 \times 10^{11}$	-	-
celkem	907 965	$6,93 \times 10^{17}$	-	-	-

Tabulka 30: Jednofaktorová ANOVA - úroveň

pozice	μ_i	α_i
všechny záznamy	969 488	-
řadoví prac.	430 386	-539 102
vedoucí prac.	941 099	-28 389
manažeri	2 190 490	1 221 010
řídící prac.	4 070 260	3 100 770

Tabulka 31: Průměrné platy na jednotlivých úrovních podle modelu (31)



Obrázek 25: Boxplot přirozených logaritmu platů na různých pozicích

Znovu můžeme anovu udělat i s přirozenými logaritmy platů. Na obrázku 25 jsou zobrazené jejich boxploty.

Při použití logaritmu má model tvar

$$\ln(y_i) = \mu + \hat{\alpha}_i + e_i \quad (i = 1, 2, \dots, 4), \quad (32)$$

kde μ je průměr všech přirozených logaritmu platů, $\hat{\alpha}_i$ jsou odchylky průměrů logaritmu platů na jednotlivých úrovních od μ a e_i jsou náhodné odchylky. Platy lze vyjádřit jako

$$y_i = e^\mu e^{\hat{\alpha}_i} e^{e_i}. \quad (33)$$

Data rozdělená na řadové pracovníky, vedoucí, manažery a řídící pracovníky byla nejdříve otestována, jestli pochází z normálního rozdělení s konstantním rozptylem. P-hodnoty všech testů byly opět rovny nule, a proto mohou být výsledky anovy zkreslené. Protože máme k dispozici mnoho dat, nemělo by toto zkreslení být velké.

Výsledky anovy jsou v tabulce 32. Testové kritérium je 329 716, tudíž je zde znovu velká variabilita mezi skupinami oproti vnitroskupinové. Nulovou hypotézu, že pozice nemá vliv na logaritmus platu, zamítáme, protože p-hodnota testu je 0.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
úroveň	3	329 782	109 927	329 716	0
reziduální	907 962	302 715	0,3334	-	-
celkem	907 965	632 497	-	-	-

Tabulka 32: Jednofaktorová ANOVA - úroveň (s přirozenými logaritmy platů)

Průměrné hodnoty na jednotlivých pozicích μ_i a odchylky od průměru všech dat $\hat{\alpha}_i$ jsou uvedeny v tabulce 33. Jsou přepočteny i na platy podle modelu (33). U řadových zaměstnanců vychází roční plat 360 591 CZK, u vedoucích 776 513 CZK, u manažerů 2 019 730 a u řídících pracovníků 3 808 550 CZK.

4.4 Trojfaktorová ANOVA

Se všemi doposud uvedenými faktory byla provedena trojfaktorová ANOVA bez interakcí, kde jsme zkoumali vliv pohlaví, země a úrovně na výši platu. (Lze udělat i anovu s interakcemi, ale ta je nad rámec této práce.)

Nejprve jsme uvažovali model ve tvaru

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk} \quad (i \in \{1, 2\}; j = 1, \dots, 7; k = 1, \dots, 4), \quad (34)$$

pozice	μ_i	$\hat{\alpha}_i$	e^{μ_i}	$e^{\hat{\alpha}_i}$
všechny záznamy	13,4360	-	684 191	-
řadoví prac.	12,7955	-0,6405	360 591	0,5270
vedoucí prac.	13,5626	0,1266	776 513	1,1349
manažeři	14,5185	1,0825	2 019 730	2,9520
řídící prac.	15,1528	1,7168	3 808 550	5,5665

Tabulka 33: Průměrné hodnoty přirozených logaritmu platů na jednotlivých úrovních podle modelů (32) a (33)

kde μ je průměrný plat vypočtený ze všech záznamů, α_i jsou odchylky průměrných platů mužů a žen od μ , β_j jsou odchylky průměrných platů v jednotlivých zemích od μ , γ_k jsou odchylky průměrných platů na jednotlivých úrovních od μ a e_{ijk} jsou náhodné odchylky.

Výsledek testu je v tabulce 34. Testovací statistika je 174 291 u pohlaví, 384 150 u země a 423 864 u úrovně (pozice). P-hodnota je u všech faktorů 0, a proto zamítáme hypotézu, že nemají vliv na pohlaví.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
pohlaví	1	$2,59 \times 10^{16}$	$2,59 \times 10^{16}$	174 291	0
země	6	$3,43 \times 10^{17}$	$5,71 \times 10^{16}$	384 150	0
úroveň	3	$1,89 \times 10^{17}$	$6,30 \times 10^{16}$	423 864	0
reziduální	907 955	$1,35 \times 10^{17}$	$1,45 \times 10^{11}$	-	-
celkem	907 965	$6,93 \times 10^{17}$	-	-	-

Tabulka 34: Trojfaktorová ANOVA

Dále jsme použili model s přirozenými logaritmy platů

$$\ln(y_{ijk}) = \mu + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + e_{ijk} \quad (i \in \{1, 2\}; j = 1, \dots, 7; k = 1, \dots, 4), \quad (35)$$

kde μ je průměrný logaritmus platu vypočtený ze všech záznamů, $\hat{\alpha}_i$ jsou odchylky průměrných logaritmu platů mužů a žen od μ , $\hat{\beta}_j$ jsou odchylky průměrných logaritmu platů v jednotlivých zemích od μ , $\hat{\gamma}_k$ jsou odchylky průměrných logaritmu platů na jednotlivých úrovních od μ a e_{ijk} jsou náhodné odchylky. Platy lze vyjádřit jako

$$y_{ijk} = e^\mu e^{\hat{\alpha}_i} e^{\hat{\beta}_j} e^{\hat{\gamma}_k} e^{e_{ijk}}. \quad (36)$$

Výsledek trojfaktorové anovy počítané s přirozenými logaritmy platů je v tabulce 35. Testovací statistika u pohlaví je 423 995, u země 904 957 a u úrovně 504 809. U všech faktorů je p-hodnota rovna 0, a proto zamítáme hypotézu, že země, pohlaví a úroveň nemají vliv na velikost platu.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
pohlaví	1	32 404	32 404	423 995	0
země	6	414 965	69 161	904 957	0
úroveň	3	115 739	38 580	504 809	0
reziduální	907 955	69 390	0,08	-	-
celkem	907 965	632 497	-	-	-

Tabulka 35: Trojfaktorová ANOVA (s přirozenými logaritmy platů)

Můžeme také použít redukováný model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \quad (i = 1; j = 1, \dots, 6; k = 1, \dots, 3), \quad (37)$$

kde jako μ zahrnujeme bazové hodnoty žena, Švédsko a vedoucí pracovník, α_i je efekt pohlaví, β_j je efekt země, γ_k je efekt úrovně a e_{ijk} jsou náhodné odchylky.

Model s logaritmy je potom ve tvaru

$$\ln(y_{ijk}) = \mu + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + e_{ijk} \quad (i = 1; j = 1, \dots, 6; k = 1, \dots, 3). \quad (38)$$

Podobně jako v předchozím případě používáme pro μ bazové hodnoty žena, Švédsko a vedoucí pracovník, $\hat{\alpha}_i$ je efekt pohlaví, $\hat{\beta}_j$ je efekt země, $\hat{\gamma}_k$ je efekt úrovně a e_{ijk} jsou náhodné odchylky. Plat potom můžeme vyjádřit v multiplikatívním tvaru

$$y_{ijk} = e^\mu e^{\hat{\alpha}_i} e^{\hat{\beta}_j} e^{\hat{\gamma}_k} e^{e_{ijk}}. \quad (39)$$

V tabulce 36 jsou uvedeny bodové a intervalové odhady platů vypočtené podle redukováného modelu. Pro bazové hodnoty žena, Švédsko a vedoucí pracovník vychází odhad ročního platu $1,03 \times 10^6$

CZK. Muž by měl plat o 13 % vyšší. Když tedy započítáme vliv země a pracovní pozice, je rozdíl mezi muži a ženami menší, než vyšlo u jednofaktorové anovy. Pokud by žena na vedoucí pozici byla z České republiky, byl by její plat 42,1 % báze a na Slovensku 40,9 %. V Dánsku by byl o 36 % vyšší než u základních hodnot, ve Finsku o 10,3 % vyšší, v Německu o 28,4 % vyšší a v Norsku o 32,9 % vyšší. Kdyby žena ze Švédska pracovala jako řadový zaměstnanec, tak pobírá 60,5 % báze, jako manažerka 184,3 % a na řídicí pozici 303,7 %.

		log-bodový ($\hat{\alpha}_i, \hat{\beta}_j, \hat{\gamma}_k$)	bodový ($e^{\hat{\alpha}_i}, e^{\hat{\beta}_j}, e^{\hat{\gamma}_k}$)	log interval spolehlivosti	interval spolehlivosti
báze	μ	13,845	$1,03 \times 10^6$	(13,843; 13,848)	($1,028 \times 10^6; 1,033 \times 10^6$)
pohlaví	muž	0,122	1,13	(0,121; 0,123)	(1,128; 1,131)
země	CZE	-0,864	0,421	(-0,867; -0,862)	(0,42; 0,422)
	DEN	0,308	1,36	(0,302; 0,313)	(1,352; 1,368)
	FIN	0,098	1,103	(0,094; 0,102)	(1,098; 1,107)
	GER	0,25	1,284	(0,248; 0,252)	(1,281; 1,287)
	NOR	0,285	1,329	(0,281; 0,289)	(1,324; 1,335)
	SVK	-0,893	0,409	(-0,896; -0,891)	(0,408; 0,41)
úroveň	řídicí prac.	1,111	3,037	(1,105; 1,116)	(3,02; 3,053)
	manažeri	0,611	1,843	(0,61; 0,613)	(1,84; 1,846)
	řadový prac.	-0,503	0,605	(-0,504; -0,502)	(0,604; 0,605)

Tabulka 36: Trojfaktorová ANOVA (s přirozenými logaritmy platů) podle modelů (38) a (39), odhad parametrů a 95% intervaly spolehlivosti

4.5 Jednofaktorová ANOVA - obor

Posledním faktorem je obor zaměstnání, ale z důvodu velkého množství různých oborů byly vybrány pouze ty, které tvoří alespoň 5 % všech záznamů, jak bylo popsáno v kapitole 3.2.2. Jedná se o výrobu, logistiku, maloobchod, inženýrství, IT a prodej,

Používáme model

$$y_i = \mu + \alpha_i + e_i \quad (i = 1, 2, \dots, 6), \quad (40)$$

kde μ je průměrný plat vypočtený ze všech záznamů s vybranými obory, α_i jsou odchylky průměrných platů v jednotlivých oborech od μ a e_i jsou náhodné odchylky.

Předpoklady anovy znovu nebyly splněny. P-hodnoty všech Kolmogorovových-Smirnovových testů, kterými jsme zjišťovali, jestli platy z jednotlivých oborů pochází z normálního rozdělení, vyšly 0. P-hodnota testu shody rozptylů byla také 0, a proto zamítáme hypotézu, že jsou rozptyly shodné. Z důvodu velkého množství dat by ale nesplnění předpokladů znovu neměl být problém.

Výsledky anovy jsou v tabulce 37. Testovací statistika F je 45 590,2 a p-hodnota testu 0. Zamítáme nulovou hypotézu, že obor nemá vliv na velikost platu.

Pomocí Tukeyovy metody bylo zjištěno, že významné rozdíly ve střední hodnotě jsou mezi všemi dvojicemi oborů.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
obor	5	$1,04 \times 10^{17}$	$2,08 \times 10^{16}$	45 590,2	0
reziduální	596 306	$2,72 \times 10^{17}$	$4,57 \times 10^{11}$	-	-
celkem	596 311	$3,76 \times 10^{17}$	-	-	-

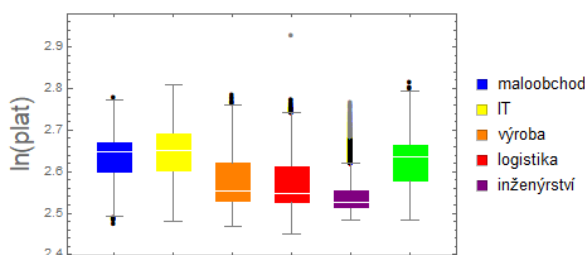
Tabulka 37: Jednofaktorová ANOVA - obor zaměstnání

Průměrné platy v jednotlivých oborech zaměstnání μ_i a jejich odchylky od průměru všech zaměstnanců pracujících ve vybraných oborech α_i jsou v tabulce 38. Průměrný roční plat ve vybraných oborech je 850 135 CZK. Pokud někdo pracuje v oboru inženýrství, tak má plat průměrně o 552 896 CZK vyšší než průměr, v oboru IT o 715 305 CZK vyšší, v logistice o 145 570 CZK nižší, v maloobchodu 465 891 CZK nižší a v prodeji o 382 917 CZK vyšší.

obor	μ_i	α_i
všechny záznamy	850 135	-
inženýrství	1 403 030	552 896
IT	1 565 440	715 305
logistika	704 565	-145 570
výroba	586 382	-263 753
maloobchod	384 243	-465 891
prodej	1 233 050	382 917

Tabulka 38: Průměrné platy ve vybraných oborech zaměstnání podle modelu (40)

Anovu můžeme znovu spočítat i s přirozenými logaritmy platů. Na obrázku 26 jsou vykresleny boxploty přirozených logaritmů platů v různých oborech zaměstnání.



Obrázek 26: Boxplot přirozených logaritmů platů ve vybraných oborech zaměstnání

Model potom zapisujeme jako

$$\ln(y_i) = \mu + \hat{\alpha}_i + e_i \quad (i = 1, 2, \dots, 6), \quad (41)$$

kde μ je průměr všech přirozených logaritmů platů z vybraných oborů, $\hat{\alpha}_i$ jsou odchylky průměru logaritmů platů v jednotlivých oborech od μ a e_i jsou náhodné odchylky. Platy můžeme vyjádřit jako

$$y_i = e^\mu e^{\hat{\alpha}_i} e^{e_i}. \quad (42)$$

Opět bylo otestováno, zda přirozené logaritmy platů rozdělené podle oborů pochází z normálního rozdělení a mají konstantní rozptyl. P-hodnoty všech testů jsou rovny nule, a tak nulovou hypotézu zamítáme, a proto mohou výsledky analýzy rozptylu být znovu zkreslené.

Výsledky anovy jsou v tabulce 39. Testové kritérium F je 45 590, což opět znamená velkou variabilitu mezi skupinami. Tomu odpovídá i p-hodnota rovná nule, a tak zamítáme nulovou hypotézu, že obor zaměstnání nemá vliv na plat.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
obor	5	135 674	27 135	59 204	0
reziduální	596 306	273 305	0,46	-	-
celkem	596 311	408 979	-	-	-

Tabulka 39: Jednofaktorová ANOVA - obor zaměstnání (s přirozenými logaritmy platů)

Průměrné hodnoty v různých oborech μ_i a jejich odchylky od průměru $\hat{\alpha}_i$ jsou v tabulce 40. Podle modelu 40 mají zaměstnanci v oboru inženýrství roční plat 1 173 370 CZK, v IT 1 261 150 CZK, v logistice 508 074 CZK, ve výrobě 451 433 CZK, v maloobchodu 327 133 CZK a v prodeji 948 254 CZK.

obor	μ_i	$\hat{\alpha}_i$	e^{μ_i}	$e^{\hat{\alpha}_i}$
všechny záznamy	13,2961	-	594 850	-
inženýrství	13,9754	0,6793	1 173 370	1,9725
IT	14,0475	0,7515	1 261 150	2,1201
logistika	13,1384	-0,1577	508 074	0,8541
výroba	13,0202	-0,2759	451 433	0,7589
maloobchod	12,6981	-0,5979	327 133	0,5499
prodej	13,7624	0,4663	948 254	1,5941

Tabulka 40: Průměrné hodnoty přirozených logaritmů platů ve vybraných oborech zaměstnání podle modelů (41) a (42)

4.6 Čtyřfaktorová ANOVA

Nakonec byl zkoumán vliv všech čtyř faktorů (země, pohlaví, úroveň, obor) pomocí modelu bez interakcí. Také by bylo možné použít model s interakcemi, který je nad rámec této práce. Výsledky jsou v tabulce 41.

Model bez interakcí je ve tvaru

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + e_{ijkl} \quad (i \in \{1, 2\}; j = 1, \dots, 7; k = 1, \dots, 4; l = 1, \dots, 6), \quad (43)$$

kde μ je průměrný plat vypočtený ze všech záznamů, α_i jsou odchylky průměrných platů mužů a žen od μ , β_j jsou odchylky průměrných platů v jednotlivých zemích od μ , γ_k jsou odchylky průměrných

platů na jednotlivých úrovních od μ , δ_l jsou odchylky průměrných platů v jednotlivých oborech od μ a e_{ijkl} jsou náhodné odchylky.

Testovací statistika F je 325 034 u země, 41 941,3 u pohlaví, 284 685 u úrovně a 5 174,82 u oboru. P-hodnoty všech faktorů jsou rovny 0, a proto zamítáme, že nemají vliv na plat.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
země	6	$2,12 \times 10^{17}$	$3,53 \times 10^{16}$	325 034	0
pohlaví	1	$4,55 \times 10^{15}$	$4,55 \times 10^{15}$	41 941,3	0
úroveň	3	$9,27 \times 10^{16}$	$3,09 \times 10^{16}$	284 685	0
obor	5	$2,81 \times 10^{15}$	$5,61 \times 10^{14}$	5 174,82	0
reziduální	596 296	$6,47 \times 10^{16}$	$1,08 \times 10^{11}$	-	-
celkem	596311	$3,76 \times 10^{17}$	-	-	-

Tabulka 41: Čtyřfaktorová ANOVA

Při výpočtu s přirozenými logaritmy používáme model

$$\ln(y_{ijkl}) = \mu + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l + e_{ijkl} \quad (i \in \{1, 2\}; j = 1, \dots, 7; k = 1, \dots, 4; l = 1, \dots, 6), \quad (44)$$

kde μ je průměrný plat vypočtený ze všech záznamů, $\hat{\alpha}_i$ jsou odchylky průměrných platů mužů a žen od μ , $\hat{\beta}_j$ jsou odchylky průměrných platů v jednotlivých zemích od μ , $\hat{\gamma}_k$ jsou odchylky průměrných platů na jednotlivých úrovních od μ , $\hat{\delta}_l$ jsou odchylky průměrných platů v jednotlivých oborech od μ a e_{ijkl} jsou náhodné odchylky. Plat se při použití tohoto modelu vypočte podle vzorce

$$y_{ijkl} = e^\mu e^{\hat{\alpha}_i} e^{\hat{\beta}_j} e^{\hat{\gamma}_k} e^{\hat{\delta}_l} e^{e_{ijkl}}. \quad (45)$$

Při výpočtu s přirozenými logaritmy je statistika F u země 763 095, u pohlaví 143 303, u úrovně 331 834 a u oboru 11 279. U všech faktorů je p-hodnota rovna 0, a tak zamítáme hypotézu, že země, pohlaví, úroveň a obor zaměstnání nemají vliv na plat. Všechny výsledky jsou uvedeny v tabulce 42.

zdroj variability	počet stupňů volnosti	součet čtverců	průměrný čtverec	F	p-hodnota
země	6	293 959	48 993	763 095	0
pohlaví	1	9 200	9 200	143 303	0
úroveň	3	63 914	21 305	331 834	0
obor	5	3 621	724	11 279	0
reziduální	596 296	38 284	0,06	-	-
celkem	596 311	408 979	-	-	-

Tabulka 42: Čtyřfaktorová ANOVA (s přirozenými logaritmy platů)

Můžeme také použít redukováný model

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + e_{ijkl} \quad (i = 1; j = 1, \dots, 6; k = 1, \dots, 3; l = 1, 2, \dots, 5), \quad (46)$$

kde jako μ zahrnujeme základní hodnoty žena, Švédsko, vedoucí pracovník a prodej, δ_j je efekt oboru zaměstnání a e_{ijkl} jsou náhodné odchylky.

Místo platů můžeme znovu použít i jejich logaritmy

$$\ln(y_{ijkl}) = \mu + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l + e_{ijkl} \quad (i = 1; j = 1, \dots, 6; k = 1, \dots, 3; l = 1, 2, \dots, 5), \quad (47)$$

kde základní hodnoty pro μ jsou žena, Švédsko, vedoucí pracovník a prodej, $\hat{\alpha}_i$ je efekt pohlaví, $\hat{\beta}_j$ je efekt země, $\hat{\gamma}_k$ je efekt úrovně, $\hat{\delta}_l$ je efekt oboru a e_{ijkl} jsou náhodné odchylky. Plat potom můžeme vyjádřit v multiplikatívním tvaru jako

$$y_{ijk} = e^\mu e^{\hat{\alpha}_i} e^{\hat{\beta}_j} e^{\hat{\gamma}_k} e^{\hat{\delta}_l} e^{e_{ijkl}}. \quad (48)$$

Bodové a intervalové odhady platů podle redukováného modelu jsou uvedeny v tabulce 43. Pro základní hodnoty žena, Švédsko, vedoucí pracovník a prodej vychází odhad ročního platu $1,029 \times 10^6$ CZK. Pokud by to byl muž, tak si vydělá o 12,1 % procent více. To znamená, že když plat ještě navíc očistíme o vliv oboru zaměstnání, je rozdíl mezi muži a ženami ještě menší, než vyšlo u trojfaktorové anovy.

		log-bodový	bodový	log interval spolehlivosti	interval spolehlivosti
		$(\hat{\alpha}_i, \hat{\beta}_j, \hat{\gamma}_k, \hat{\delta}_l)$	$(e^{\hat{\alpha}_i}, e^{\hat{\beta}_j}, e^{\hat{\gamma}_k}, e^{\hat{\delta}_l})$		
báze	μ	13,844	$1,029 \times 10^6$	(13,841; 13,847)	$(0,026 \times 10^6; 1,032 \times 10^6)$
pohlaví	muž	0,114	1,121	(0,112; 0,115)	(1,119; 1,122)
země	CZE	-0,881	0,414	(-0,884; -0,878)	(0,413; 0,415)
	DEN	0,312	1,366	(0,305; 0,319)	(1,357; 1,376)
	FIN	0,0975	1,102	(0,092; 0,103)	(1,096; 1,108)
	GER	0,229	1,257	(0,226; 0,231)	(1,253; 1,260)
	NOR	0,282	1,325	(0,277; 0,287)	(1,319; 1,332)
	SVK	-0,909	0,403	(-0,912; -0,906)	(0,402; 0,404)
úroveň	řídící prac.	1,001	2,721	(0,993; 1,010)	(2,698; 2,745)
	manažeri	0,547	1,728	(0,545; 0,549)	(1,724; 1,732)
	řadoví prac.	-0,393	0,675	(-0,395; -0,391)	(0,674; 0,676)
obor	inženýrství	0,046	1,047	(0,043; 0,049)	(1,044; 1,050)
	IT	0,155	1,167	(0,152; 0,157)	(1,164; 1,171)
	logistika	-0,068	0,934	(-0,071; -0,065)	(0,931; 0,937)
	výroba	-0,084	0,919	(-0,087; -0,0818)	(0,917; 0,921)
	maloobchod	-0,184	0,832	(-0,187; -0,181)	(0,829; 0,834)

Tabulka 43: Čtyřfaktorová ANOVA (s přirozenými logaritmy platů) podle modelů (47) a (48), odhad parametrů a 95% intervaly spolehlivosti

Závěr

Cílem práce byla analýza příjmových nerovností v České republice, Dánsku, Finsku, Německu, Norsku, Slovensku a Švédsku. Nejprve byly představeny různé způsoby měření nerovností v příjmech. Je možné je měřit buď ve skupině nebo mezi skupinami. Skupiny mohou být vytvořeny například rozdělením záznamů podle země nebo pohlaví.

K měření nerovností uvnitř skupiny slouží například Giniho index a lze je zobrazit pomocí Lorenzovy křivky. Pokud měříme nerovnosti mezi skupinami, můžeme využít například příjmovou mezeru, která popisuje, kolika procentům platu mužů je roven plat žen. Obvykle se používá v souvislosti s pohlavím, ale je možné ji využít i u jiných skupin. K výpočtu se obvykle využívají průměrné platy nebo jejich mediány, ale můžeme zkoumat i závislost příjmové mezery na percentilu.

Dále byla popsána analýza rozptylu, která se používá, pokud chceme zjistit, zda je nějaká veličina (v našem případě plat) funkcí jednoho nebo více faktorů (země, pohlaví, pracovní pozice nebo oboru). Také byla představena Tukeyho metoda, kterou používáme, pokud zjistíme, že daný faktor má na plat vliv, abychom určili, mezi kterými jeho úrovněmi je významný rozdíl.

V další části práce byla zpracována data. Nejdříve byly vypočteny základní charakteristiky platů jako například počet dat, průměr, medián, rozptyl nebo směrodatná odchylka a udělány základní vizualizace dat. Mohli jsme si tak udělat lepší představu o datových souborech. Nejvíce dat bylo z České republiky a Německa, ze skandinávských zemí bylo dat méně. Slovenských dat bylo poměrně málo v letech 2015 a 2016, ale později už více. Bylo vidět, že v České republice a na Slovensku jsou poměrně nízké platy (například medián) oproti ostatním zemím nebo že jsou platy ve všech zemích vychýlené doleva. Zároveň jsme upozornili na odlišnou strukturu dat z jednotlivých zemí.

Z důvodu analýzy dat podle pohlaví byly některé charakteristiky a vizualizace udělány i pro záznamy rozdělené na muže a ženy. Bylo zjištěno, že ve všech zemích je méně záznamů o ženách než o mužích. Platy žen jsou také oproti platům mužů více vychýlené doleva a mediány platů žen jsou nižší než mediány platů mužů.

Dále byl vypočten Giniho index ve všech zemích a byly vykresleny jejich časové vývoje a pro poslední rok také Lorenzovy křivky. Největší nerovnosti byly v České republice a na Slovensku, nejmenší v Norsku, Finsku a Švédsku. V České republice a na Slovensku se nerovnosti v čase zmenšují, v Dánsku a Německu zase spíše zvětšují.

Data byla dále rozdělena na muže a ženy a v těchto dvou skupinách bylo postupováno obdobně. U mužů byly výsledky i časové vývoje podobné jako v zemích celkově. U žen byly mezi zeměmi menší rozdíly. Největší nerovnosti mají ženy mezi sebou na Slovensku, v Německu, Dánsku a České republice, nejmenší v Norsku a Finsku.

V další části byla vypočtena příjmová mezera mezi muži a ženami, tj. na jaké procento platu mužů ženy dosahují. Vyšší hodnoty byly ve Švédsku, Norsku a Finsku, v ostatních zemích byly oproti nim nižší. Také byla znázorněna příjmová mezera v závislosti na percentilu. V České republice klesá, v ostatních zemích se příliš nemění.

Data byla dále rozdělena podle pracovních pozic (úrovní) a dále i podle oborů zaměstnání (z důvodu jejich velkého množství byly vybrány pouze nejčastější) a opět byla vypočtena příjmová mezera mezi muži a ženami. Ve většině případů byla příjmová mezera vyšší (tj. ženy dosahují na vyšší procento platu mužů) než u nerozdělených dat. Pokud tedy porovnáváme muže a ženy na stejné pozici pracující ve stejném oboru, tak jsou mezi nimi menší rozdíly, než když porovnáváme všechny dohromady. V několika případech byla dokonce příjmová mezera větší než jedna. To znamená, že ženy si na některých pozicích vydělávají více než muži. V posledním roce to v České republice byly například řadové pracovníce v oborech maloobchod a prodej.

Také byla vypočtena příjmová mezera mezi Českou republikou a ostatními zeměmi, tedy na kolik procent platu dané země Česká republika dosáhne (pouze v případě Slovenska to bylo naopak, na kolik procent českého platu dosáhnou na Slovensku). Byly zobrazeny i časové vývoje a závislost příjmové mezery na percentilu.

Dalším krokem byla analýza rozptylu (ANOVA), kde bylo zjištěno, že na výši platu má vliv pohlaví, země, pracovní pozice a obor.

Nejprve byla provedena jednofaktorová ANOVA pro pohlaví. Byla zamítnuta nulová hypotéza, že pohlaví nemá na plat vliv. Anovu jsme udělali také pro přirozené logaritmy platů, protože mají blíže normálnímu rozdělení, což je jeden z předpokladů anovy. Zde byla nulová hypotéza také zamítnuta.

Stejně bylo postupováno i u faktoru země a pracovní pozice. V obou případech byla zamítnuta nulová hypotéza, že daný faktor nemá na plat vliv. Pomocí Tukeyho metody byl zjištěn významný rozdíl mezi všemi dvojicemi zemí i pozic.

Vliv pohlaví, země a pozice byl zjišťován také pomocí trojfaktorové anovy, kde bylo opět zamítnuto, že plat není jejich funkcí. Použili jsme také redukovaný model, u kterého jsme zjistili, že pokud

plat očistíme o vliv země a pracovní pozice, jsou rozdíly mezi platy mužů a žen menší, než vyšly u jednofaktorové anovy. Další možností by byl model s interakcemi, který je ale nad rámec této práce.

Dalším zkoumaným faktorem byl obor zaměstnání, kde byly opět z důvodu jejich velkého množství vybrány pouze nejčastější. Opět bylo zamítnuto, že obor nemá na plat vliv a statisticky významný rozdíl byl znovu mezi všemi dvojicemi oborů.

Nakonec byla provedena čtyřfaktorová anova se všemi faktory, kde bylo opět zamítnuto, že nemají vliv na plat. Také jsme použili redukovaný model, kde jsme zjistili, že očištěním platu o vliv oboru jsou rozdíly mezi platy mužů a žen ještě menší než u trojfaktorové anovy. Podobně jako u trojfaktorové anovy by bylo možné použít i model s interakcemi, který je nad rámec této práce.

Reference

- [1] HUANG, Junming, Alexander J. GATES, Robert SINATRA a Albert-László BARABÁSI. Historical comparison of gender inequality in scientific careers across countries and disciplines. PNAS [online]. March 3, 2020, 117(9), 4609-4616 [cit. 2021-5-10]. Dostupné z: www.pnas.org/cgi/doi/10.1073/pnas.1914221117
- [2] ELY, Robin J., Pamela STONE a Colleen AMMERMAN. Rethink What You "Know" About High-Achieving Women. Harvard Business Review. December 2014.
- [3] Co je to Gender Pay Gap a jak se počítá? Rovná odměna [online]. [cit. 2021-03-21]. Dostupné z: <https://www.rovnaodmena.cz/www/img/uploads/76fef6e6.pdf>
- [4] ALLISON, Paul D. Measures of Inequality. American Sociological Association [online]. 1978, 43(6), 865-880 [cit. 2021-03-13]. Dostupné z: <https://www.jstor.org/stable/2094626>
- [5] FARRIS, Frank A. The Gini Index and Measures of Inequality. The American Mathematical Monthly [online]. 2010, 851-864 [cit. 2021-03-13]. Dostupné z: <https://doi.org/10.4169/000298910X523344>
- [6] Mzdy. Český statistický úřad [online]. [cit. 2021-03-21]. Dostupné z: <https://www.czso.cz/documents/10180/20568983/311110a04.pdf/aff2944d-1279-4f36-bde7-5501184bea29?version=1.0>
- [7] Gender pay gap in unadjusted form - NACE Rev. 2 activity (earn_grgpg2). Eurostat [online]. [cit. 2021-03-21]. Dostupné z: https://ec.europa.eu/eurostat/cache/metadata/en/earn_grgpg2_esms.htm
- [8] SUDOVÁ, Blanka. Analýza příjmových nerovností. Plzeň, 2016. Diplomová práce. Západočeská univerzita v Plzni.
- [9] REIF, Jiří. Metody matematické statistiky. 2. upr. vyd. V Plzni: Západočeská univerzita, Fakulta aplikovaných věd, 2004. ISBN 80-704-3302-7.
- [10] ANDĚL, Jiří. Statistické metody. 4., upr. vyd. Praha: Matfyzpress, 2007. ISBN isbn80-7378-003-8.
- [11] Kurzy devizového trhu. Česká národní banka [online]. [cit. 2020-12-29]. Dostupné z: <https://www.cnb.cz/cs/financni-trhy/devizovy-trh/kurzy-devizoveho-trhu/kurzy-devizoveho-trhu/>
- [12] Kurzy devizového trhu: Kurzy devizového trhu - měsíční kumulované průměry. Česká národní banka [online]. [cit. 2020-12-29]. Dostupné z: https://www.cnb.cz/cs/financni-trhy/devizovy-trh/kurzy-devizoveho-trhu/kurzy-devizoveho-trhu/prumerne_mena.html?mena=EUR
- [13] Gini index (World Bank estimate). World Bank [online]. [cit. 2021-03-14]. Dostupné z: https://data.worldbank.org/indicator/SI.POV.GINI?end=2017&most_reprotect\ discretionary{\char\hyphenchar\font}{-}{-}cent_value_desc=false&start=1992&view=chart
- [14] Gender pay gap statistics. Eurostat [online]. 2021 [cit. 2021-03-21]. Dostupné z: https://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics

Přílohy

Příloha A - počty záznamů

V tabulkách 44 až 50 jsou zobrazeny počty záznamů v jednotlivých letech, které jsme měli k dispozici na začátku a počty po odstranění chybných a neúplných záznamů. Postupně byly odebrány údaje o zaměstnancích, u kterých nebyla informace o pohlaví, úrovni, úvazku, roční mzdě, pozici a regionu. V některých souborech nebyl uveden region, a tak jsou příslušné buňky tabulek prázdné.

počet záznamů	2015	2016	2017	2018	2019
vše	268 000	303 887	369 114	361 686	348 458
s pohlavím	268 000	303 887	369 114	361 686	348 458
s úrovní	268 000	303 887	369 114	361 686	348 458
s úvazkem	268 000	303 887	369 114	361 686	348 458
s platem	267 999	303 887	369 114	361 686	348 458
s pozicí	267 999	303 887	369 114	361 686	348 458
s regionem	267 246	303 841	365 676	361 128	348 252

Tabulka 44: Počet záznamů za jednotlivé roky - Česká republika

počet záznamů	2015	2016	2017	2018	2019
vše	11 467	6 929	9 888	10 110	11 532
s pohlavím	11 467	6 929	9 888	10 110	11 532
s úrovní	11 467	6 929	9 888	10 110	11 532
s úvazkem	11 467	6 929	9 888	10 110	11 532
s platem	11 467	6 929	9 888	10 098	11 513
s pozicí	11 467	6 929	9 888	10 098	11 513
s regionem	4 952	2 712	6 226	6 075	10 032

Tabulka 45: Počet záznamů za jednotlivé roky - Dánsko

počet záznamů	2015	2016	2017	2018	2019
vše	54 360	40 784	46 465	52 734	27 641
s pohlavím	54 360	40 784	46 465	52 734	27 641
s úrovní	54 360	40 784	46 465	52 734	27 641
s úvazkem	54 360	40 784	46 465	52 734	27 641
s platem	54 360	40 784	46 465	52 732	27 641
s pozicí	54 360	40 784	46 465	52 732	27 641
s regionem	-	-	-	48 463	24 013

Tabulka 46: Počet záznamů za jednotlivé roky - Finsko

počet záznamů	2015	2016	2017	2018	2019
vše	307 433	434 891	342 813	365 351	397 934
s pohlavím	240 847	314 725	245 290	312 683	347 048
s úrovní	240 847	314 725	245 290	312 683	347 048
s úvazkem	240 847	233 730	219 178	272 513	271 555
s platem	240 166	200 782	219 176	272 508	267 068
s pozicí	240 166	200 782	219 176	272 508	267 068
s regionem	-	-	-	-	-

Tabulka 47: Počet záznamů za jednotlivé roky - Německo

počet záznamů	2015	2016	2017	2018	2019
vše	45 038	48 520	45 975	49 984	29 465
s pohlavím	45 038	48 520	45 975	49 984	29 465
s úrovní	45 038	48 520	45 975	49 984	29 465
s úvazkem	45 038	48 520	45 975	49 982	29 465
s platem	45 037	48 520	45 975	49 939	29 465
s pozicí	45 037	48 520	45 975	49 939	29 465
s regionem	-	-	-	-	28 947

Tabulka 48: Počet záznamů za jednotlivé roky - Norsko

počet záznamů	2015	2016	2017	2018	2019
vše	1 405	362	141 486	163 192	158 305
s pohlavím	1 405	362	141 486	163 192	158 305
s úrovní	1 405	362	141 486	163 192	158 305
s úvazkem	1 405	362	141 486	163 192	158 305
s platem	1 405	362	141 482	163 191	158 272
s pozicí	1 405	362	141 482	163 191	158 272
s regionem	1 405	362	141 482	161 936	158 272

Tabulka 49: Počet záznamů za jednotlivé roky - Slovensko

počet záznamů	2015	2016	2017	2018	2019
vše	75 380	85 623	83 039	71 310	98 784
s pohlavím	75 380	85 623	83 039	71 310	98 784
s úrovní	75 380	85 623	83 039	71 310	98 784
s úvazkem	75 380	85 623	83 039	71 310	98 784
s platem	75 380	85 623	83 039	71 296	98 774
s pozicí	75 380	85 623	83 039	71 296	98 774
s regionem	-	-	-	-	71 382

Tabulka 50: Počet záznamů za jednotlivé roky - Švédsko

Příloha B - popisná statistika

V tabulkách 51 až 57 jsou vypočteny základní charakteristiky polohy a variability ročních mezd pro všechna upravená data. Mzdy byly přepočteny na české koruny.

	2015	2016	2017	2018	2019
počet	267	304	366	361	348
minimum	102	108	120	138	145
maximum	10 625	13 201	18 266	49 669	20 624
průměr	362	381	379	409	432
medián	275	300	299	319	340
rozptyl	$9,19 \times 10^7$	$8,85 \times 10^7$	$8,72 \times 10^7$	$1,03 \times 10^8$	$9,93 \times 10^7$
směrodatná odchylka	303	297	295	321	315
dolní kvartil	193	215	216	240	264
horní kvartil	413	437	432	462	480
šikmost	5,00	5,29	6,06	16,63	7,34
špičatost	58,27	71,45	115,74	1688,05	198,89

Tabulka 51: Základní charakteristiky ročních platů převedených na plný úvazek - Česká republika (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	5,0	2,7	6,2	6,1	10,0
minimum	741	853	684	362	386
maximum	22 075	22 680	40 605	17 370	17 824
průměr	1 947	2 047	2 094	1 621	1 824
medián	1 748	1 777	1 811	1 498	1 711
rozptyl	$9,17 \times 10^8$	$1,29 \times 10^9$	$1,98 \times 10^9$	$9,46 \times 10^8$	$8,54 \times 10^8$
směrodatná odchylka	958	1 138	1 406	973	924
dolní kvartil	1 409	1 464	1 461	971	1 307
horní kvartil	2 202	2 311	2 333	2 016	2 206
šikmost	5,17	6,21	9,25	3,79	3,15
špičatost	66,20	78,25	159,27	41,26	31,90

Tabulka 52: Základní charakteristiky ročních platů převedených na plný úvazek - Dánsko (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	54	41	46	48	24
minimum	436	433	367	332	342
maximum	29 788	28 461	27 436	77 701	19 325
průměr	1 594	1 600	1 576	1 487	1 572
medián	1 424	1 422	1 401	1 323	1 422
rozptyl	$6,90 \times 10^8$	$7,51 \times 10^8$	$7,53 \times 10^8$	$8,05 \times 10^8$	$5,96 \times 10^8$
směrodatná odchylka	831	866	868	897	772
dolní kvartil	1 147	1 141	1 117	1030	1 156
horní kvartil	1 803	1 811	1 790	1 706	1 780
šikmost	7,40	7,89	7,85	18,19	6,41
špičatost	139,10	151,52	147,54	1 159,80	94,18

Tabulka 53: Základní charakteristiky ročních platů převedených na plný úvazek - Finsko (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	240	201	219	273	267
minimum	303	325	300	300	258
maximum	51 238	17 218	15 798	19 265	127 632
průměr	1 744	1 749	1 698	1 621	1 678
medián	1 557	1 610	1 539	1 417	1 490
rozptyl	$6,90 \times 10^8$	$5,92 \times 10^8$	$6,13 \times 10^8$	$6,34 \times 10^8$	$7,81 \times 10^8$
směrodatná odchylka	831	770	783	796	884
dolní kvartil	1 136	1 191	1 104	1 054	1 055
horní kvartil	2 271	2 163	2 230	1 996	2 134
šikmost	4,02	1,59	1,36	1,81	12,63
špičatost	111,63	10,93	8,41	12,55	1563,21

Tabulka 54: Základní charakteristiky ročních platů převedených na plný úvazek - Německo (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	45	49	46	50	29
minimum	374	597	335	286	243
maximum	26 860	25 702	24 688	28 444	23 737
průměr	2 259	2 078	2 076	2 184	2 326
medián	2054,84	1911,82	1918,53	2005,44	2202,04
rozptyl	$9,31 \times 10^8$	$8,63 \times 10^8$	$8,23 \times 10^8$	$1,02 \times 10^9$	$7,39 \times 10^8$
směrodatná odchylka	965	929	907	1 008	860
dolní kvartil	1 610	1 471	1 478	1 559	1 750
horní kvartil	2 684	2 498	2 472	2 581	2 696
šikmost	2,75	3,00	2,97	4,44	3,03
špičatost	26,95	31,45	30,17	67,06	33,61

Tabulka 55: Základní charakteristiky ročních platů převedených na plný úvazek - Norsko (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	1,4	0,3	141,5	161,9	158,3
minimum	120	157	111	110	108
maximum	4 134	2 434	12 502	11 056	11 108
průměr	303	320	365	365	411
medián	226	226	297	294	336
rozptyl	$5,88 \times 10^7$	$6,67 \times 10^7$	$6,95 \times 10^7$	$6,50 \times 10^7$	$7,93 \times 10^7$
směrodatná odchylka	242	258	264	255	282
dolní kvartil	173	179	222	221	256
horní kvartil	361	338	420	419	469
šikmost	5,56	4,20	7,40	6,54	6,27
špičatost	61,31	28,12	158,25	124,77	109,02

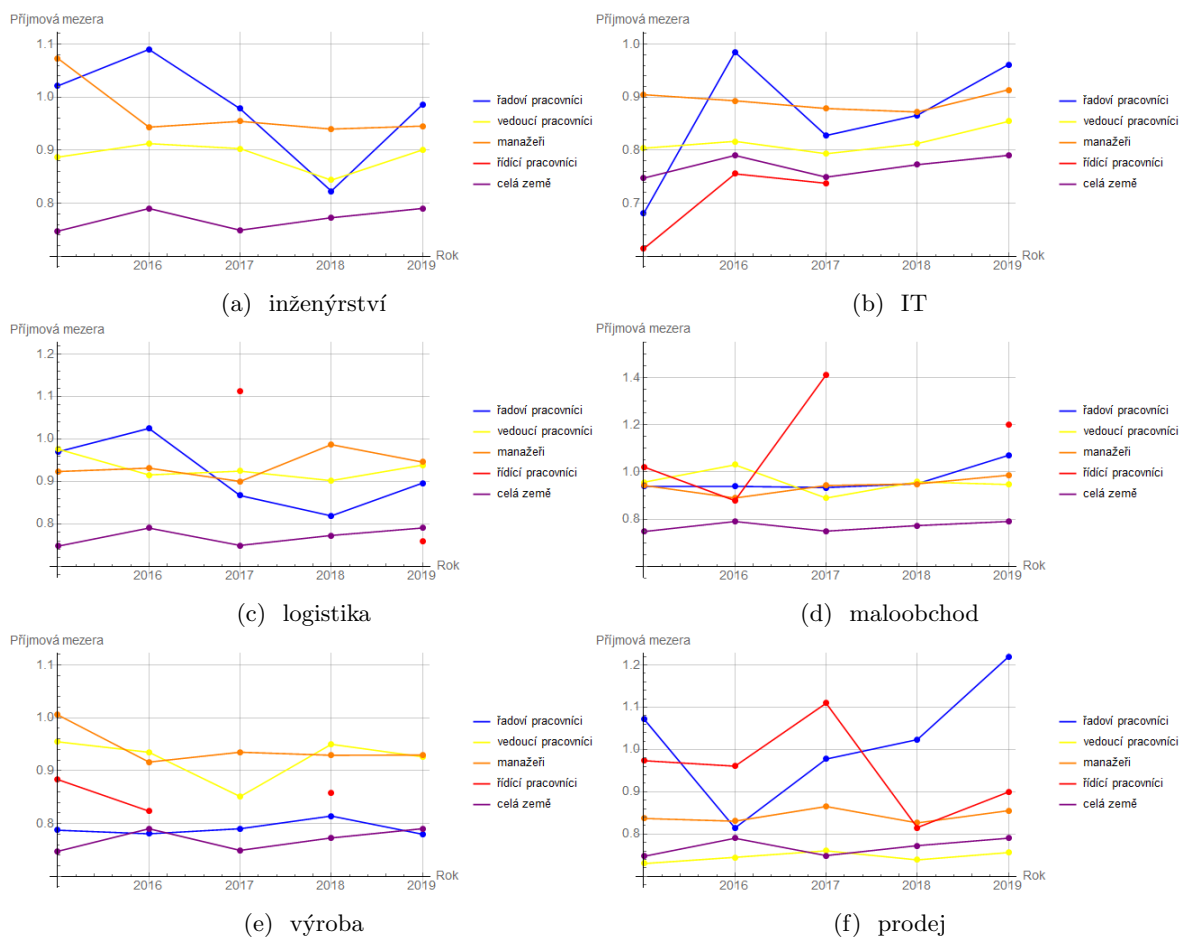
Tabulka 56: Základní charakteristiky ročních platů převedených na plný úvazek - Slovensko (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

	2015	2016	2017	2018	2019
počet	75	86	83	71	71
minimum	479	506	426	248	337
maximum	23 580	32 391	38 788	35 452	32 958
průměr	1 254	1 431	1 376	1 269	1 307
medián	1120,81	1266,44	1210,51	1126,29	1171,2
rozptyl	$3,66 \times 10^8$	$4,76 \times 10^8$	$5,20 \times 10^8$	$3,96 \times 10^8$	$4,40 \times 10^8$
směrodatná odchylka	605	690	721	629	663
dolní kvartil	868	992	937	887	886
horní kvartil	1 489	1 699	1 624	1 484	1 549
šikmost	6,38	5,91	8,15	7,76	6,63
špičatost	124,09	131,21	221,86	212,91	150,64

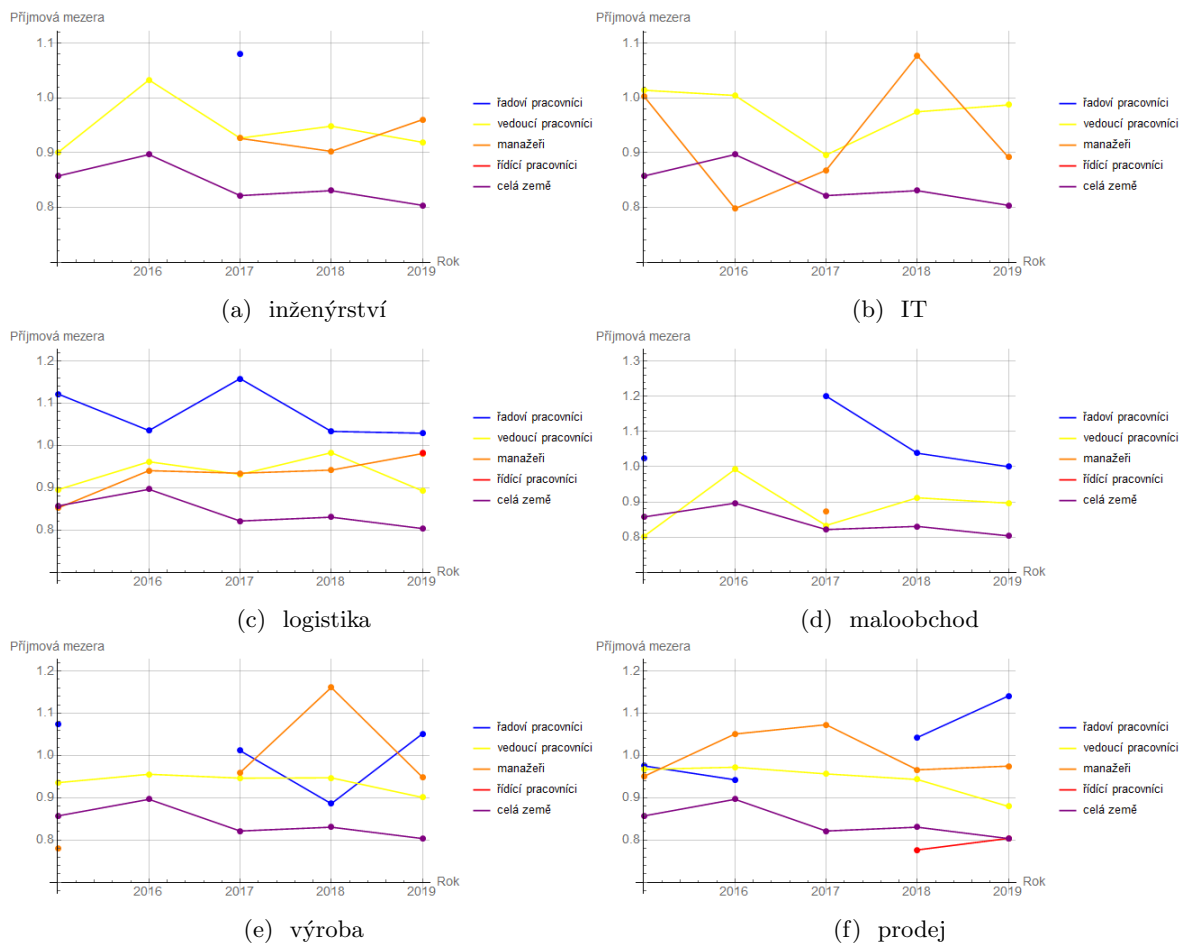
Tabulka 57: Základní charakteristiky ročních platů převedených na plný úvazek - Švédsko (počet je uvedený v tisících a ostatní údaje s výjimkou šikmosti a špičatosti jsou v tisících CZK, případně CZK²)

Příloha C - časový vývoj příjmové mezery mezi muži a ženami na jednotlivých úrovních a oborech

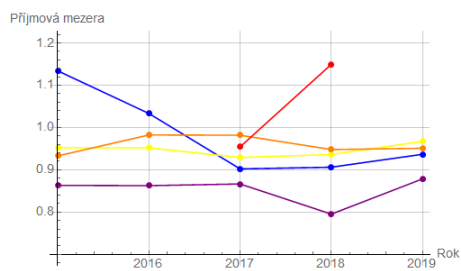
Na obrázcích 27 až 33 jsou znázorněny časové vývoje příjmové mezery pro data rozdělená podle úrovně a oboru zaměstnání. Byly vybrány pouze nejčastější obory: maloobchod, IT, výroba, logistika a zásobování a inženýrství. Pro srovnání jsou uvedeny také příjmové mezery v celých zemích, které byly vypočteny ze všech záznamů dohromady. Ve většině případů je příjmová mezera celé země nižší než příjmová mezera u rozdělených dat. Pokud někde příjmová mezera chybí, tak nebylo dost mužů nebo žen v daném oboru a na dané pozici.



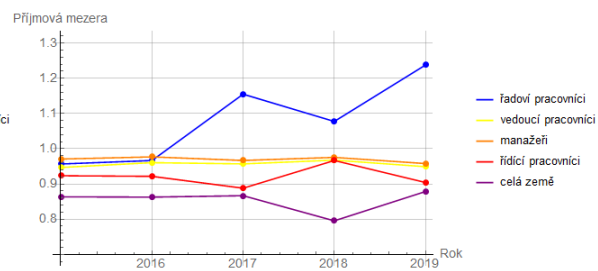
Obrázek 27: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech v České republice



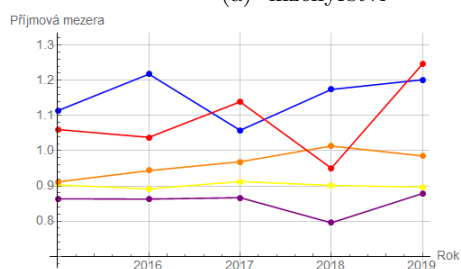
Obrázek 28: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech v Dánsku



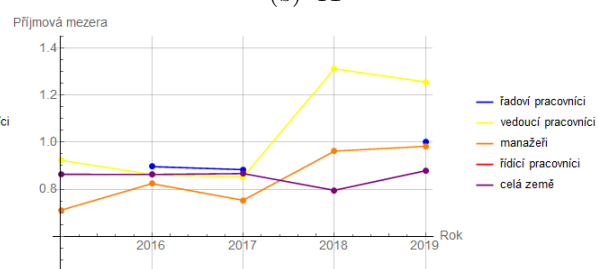
(a) inženýrství



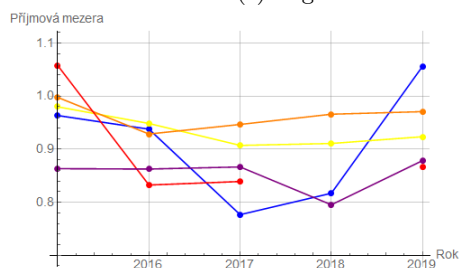
(b) IT



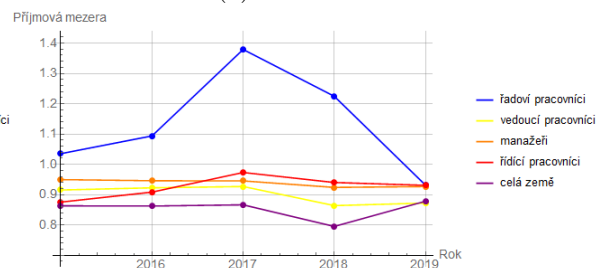
(c) logistika



(d) maloobchod

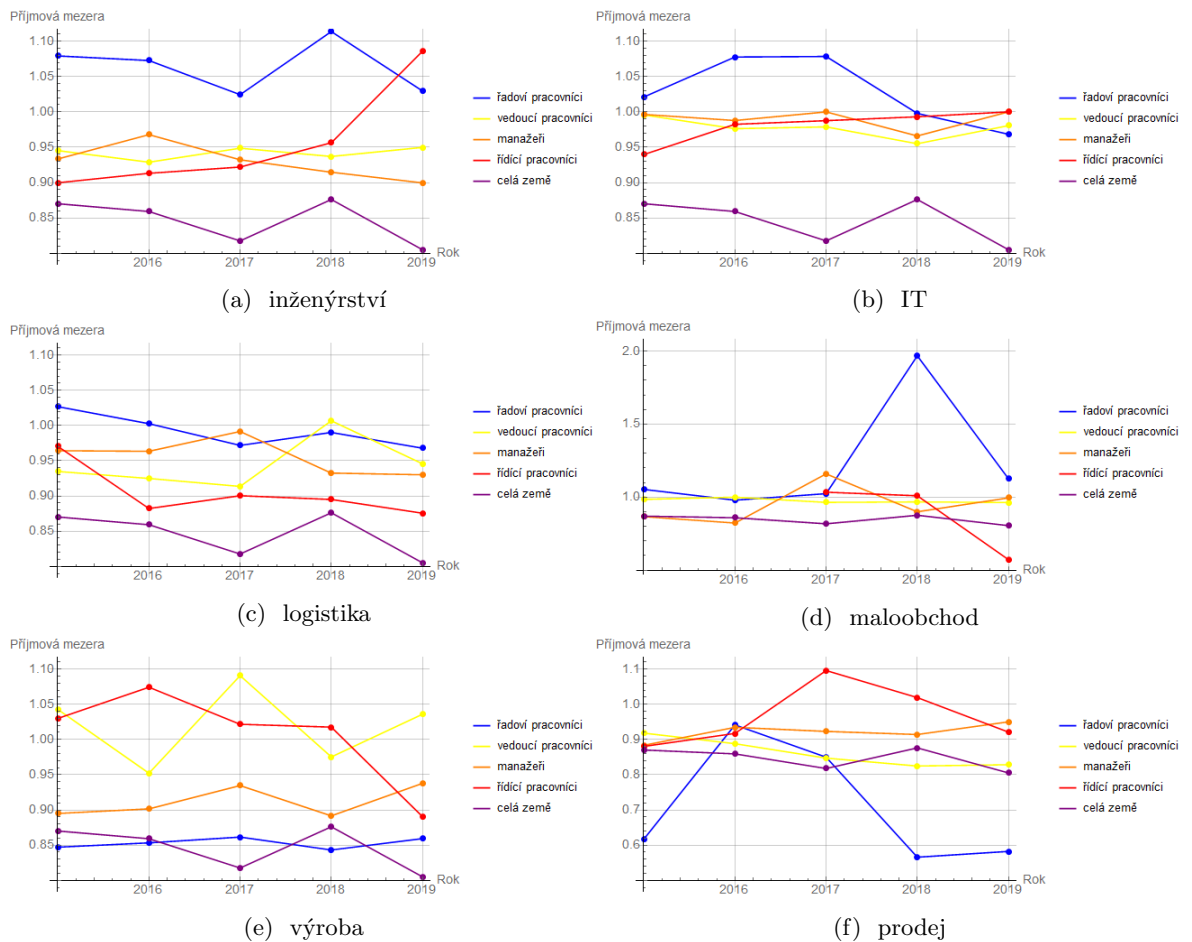


(e) výroba

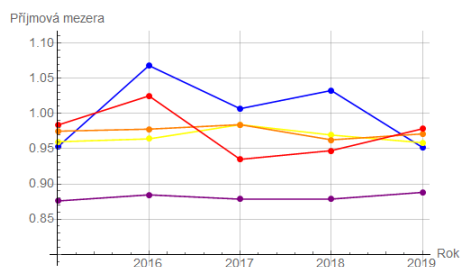


(f) prodej

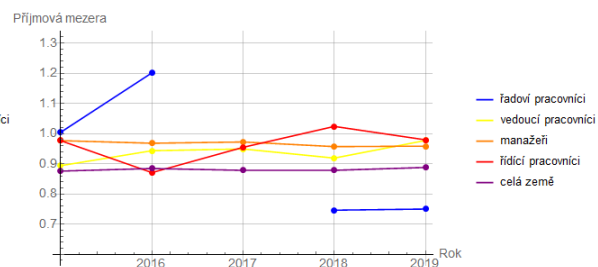
Obrázek 29: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech ve Finsku



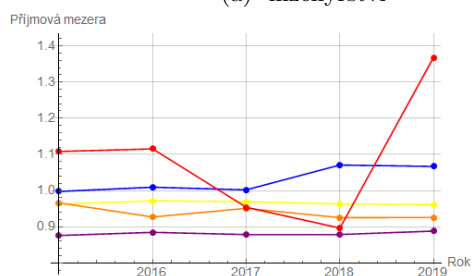
Obrázek 30: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech v Německu



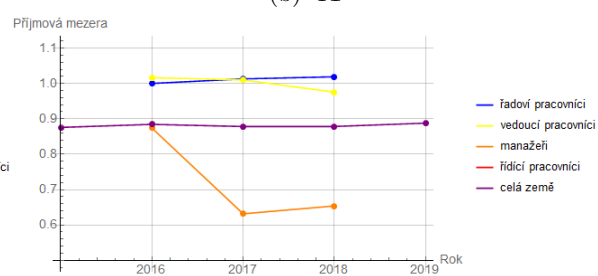
(a) inženýrství



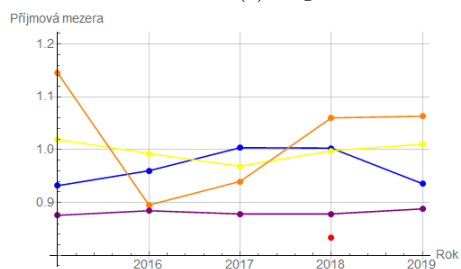
(b) IT



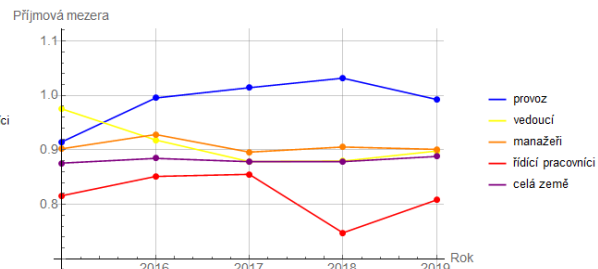
(c) logistika



(d) maloobchod

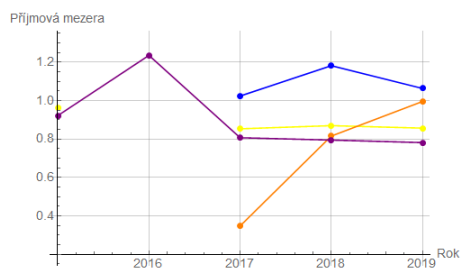


(e) výroba

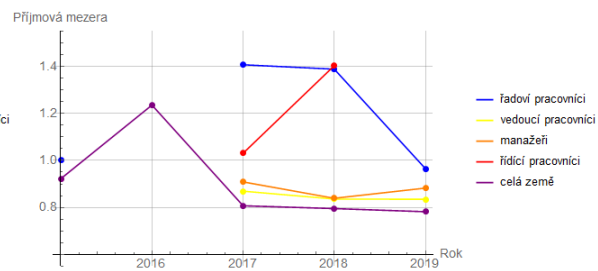


(f) prodej

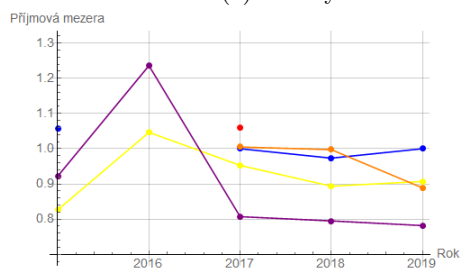
Obrázek 31: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech v Norsku



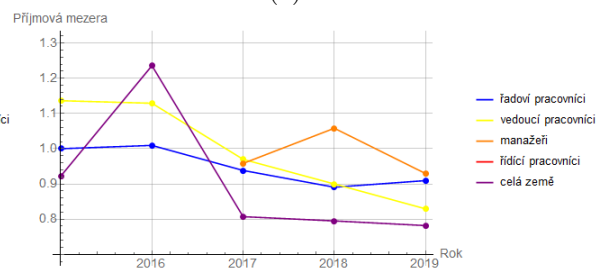
(a) inženýrství



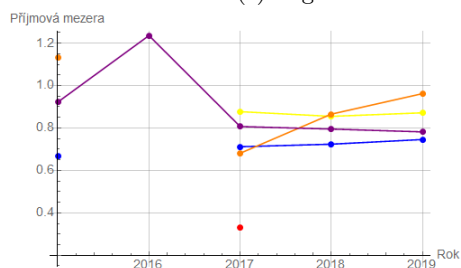
(b) IT



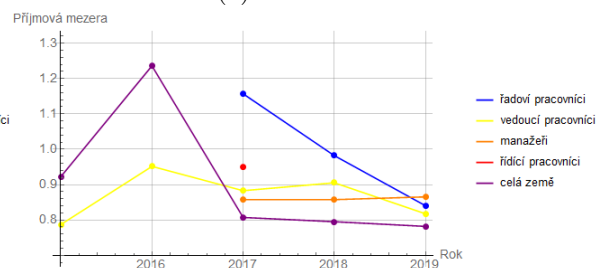
(c) logistika



(d) maloobchod

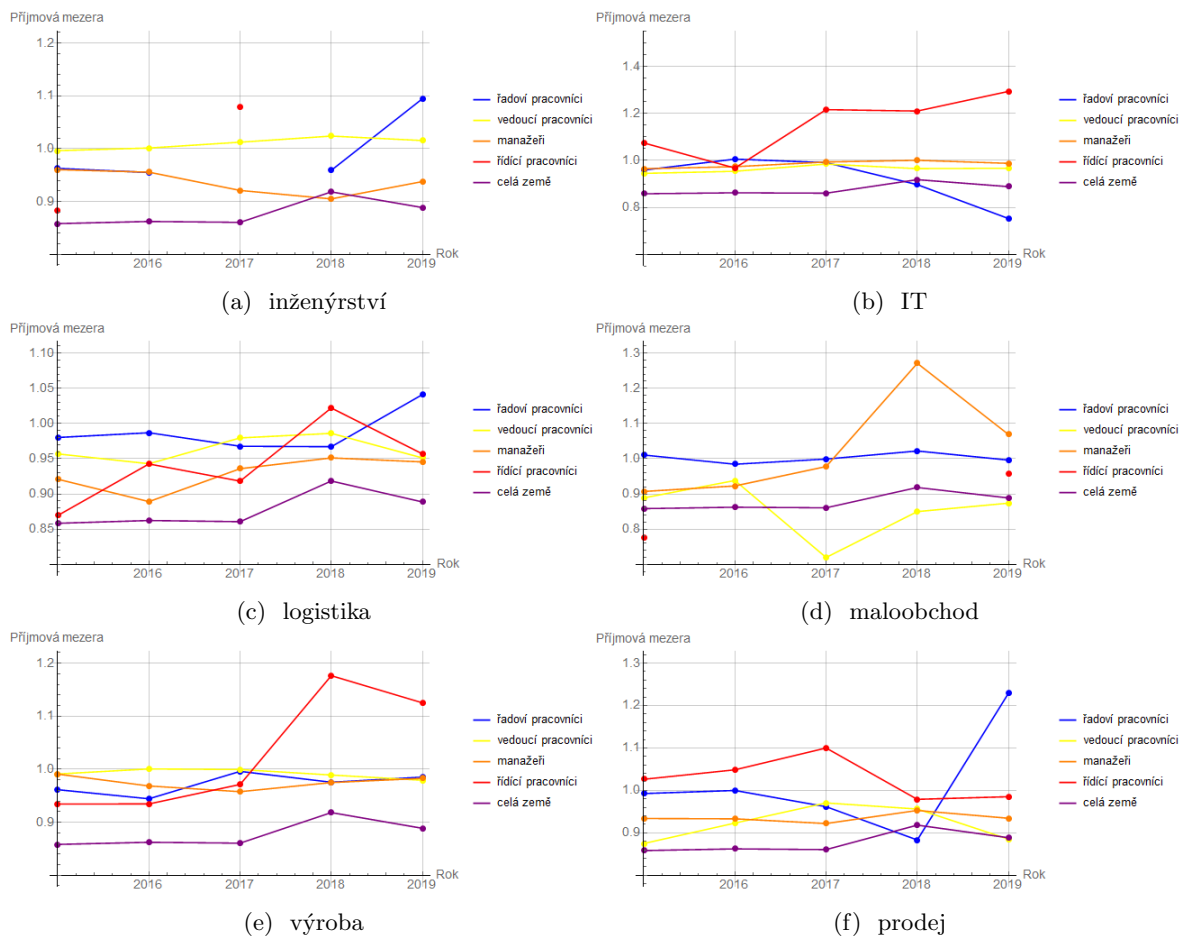


(e) výroba



(f) prodej

Obrázek 32: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech na Slovensku



Obrázek 33: Časový vývoj příjmové mezery mezi muži a ženami na jednotlivých pozicích a oborech ve Švédsku

Příloha D - soubory na CD

Vstupní data byla získána od firmy, která analyzuje platy soukromých firem. Z důvodu citlivosti dat nemohla být přiložena na CD. Pro ilustraci přikládáme všechny zdrojové kódy.

- **BP_Martina_Kůsová_2020_21.pdf:** text bakalářské práce
- **Lorenzova_krivka.nb:** příklad Lorenzovy křivky pro rovné a nerovné rozdělení
- **uprava_dat.nb:** soubor, kde jsou upraveny nebo vymazány chybné a neúplné záznamy a zjištěn počet dat po jednotlivých úpravách
- **popisna_statistika.nb:** výpočet a vizualizace základních charakteristik dat
- **Giniho_index.nb:** všechny výpočty a vizualizace Giniho indexu a Lorenzových křivek
- **obory.nb:** soubor, kde je popsáno, kolik záznamů je z jednotlivých úrovní a oborů a zjištěno, které obory se v datech vyskytují nejčastěji
- **prijmova_mezera.nb:** všechny výpočty příjmové mezery
- **anova_pohlavi.nb:** výpočet jednofaktorové anovy s faktorem pohlaví a kontrola splnění jejích předpokladů
- **anova_logaritmy_pohlavi.nb:** výpočet jednofaktorové anovy s faktorem pohlaví s použitím logaritmů platů a kontrola splnění jejích předpokladů
- **anova.nb:** soubor se všemi anovami kromě redukovaných modelů a s kontrolou splnění jejich předpokladů
- **anova_logaritmy.nb:** soubor se všemi anovami kromě redukovaných modelů s použitím logaritmů platů a s kontrolou splnění jejich předpokladů
- **anova_redukovane_modely.nb:** redukované modely trojfaktorové a čtyřfaktorové anovy s použitím logaritmů platů