

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce
Analýza trhu ocenění zaměstnanců

Zde bude oficiální zadání

Čestné prohlášení

Zde prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných zdrojů.

V Plzni dne _____

Lukáš Havlík

Poděkování

Velmi rád bych poděkoval Ing. Pavlu Novému, Ph.D. za nekonečnou trpělivost a velmi vstřícný přístup při vedení této práce. Dále Ing. Petru Chocholkovi za cenné rady a celkově za možnost tuto studii uskutečnit. Nakonec ještě celé své rodině, přítelkyni a všem přátelům, kteří mě po dobu studií nezměrně podporovali.

Abstrakt

Cílem této práce je sestavení nástrojů nutných k celému procesu analýzy trhu ocenění zaměstnanců. Ta poskytuje pomoc potencionálním zákazníkům k lepšímu přerozdělení finančních zdrojů k tomu určených. Pro takového klienta je vytvořen dotazník a manuál k jeho používání. Sebraná data se pak zpracují statistickými metodami a jsou nahrána do interního databázového systému. Na základě takto očištěných dat jsou pak určeny modely sloužící k lepšímu zachycení chyb v příštím sběru. Společnosti je na oplátku poskytnuta základní analýza trhu ocenění zaměstnanců.

Klíčová slova: Analýza trhu, ocenění zaměstnanců, modely trhu.

Abstract

The purpose of this project is to create the necessary tools to the employee valuation analysis process. This analysis allows us to help potential customers to improve reallocation of their resources. A questionnaire is created in this project, as well as a user manual. The collected data is processed using statistical methods and afterwards uploaded into an internal database system. Tolerance limits will be created based on the processed data, which help us to avoid mistakes and typos in subsequent data collection. In return, the client gets basic feedback of employee valuation market analysis.

Key words: Market analysis, employee valuation, market model.

Obsah

Úvod	1
1 Teoretická část – statistický základ.....	2
1.1 Základní pojmy	2
1.2 Beta rozdělení.....	3
1.3 Uspořádaný náhodný výběr	4
1.4 Popisná statistika	5
1.4.1 Charakteristiky polohy	5
1.4.2 Charakteristiky variability.....	6
1.5 Toleranční a predikční interval.....	7
1.6 Intervaly spolehlivosti pro kvantily spojitých rozdělení	7
1.7 Wilksovy toleranční meze	8
2 Praktická část – případová studie	11
2.1 Dotazníkové šetření	12
2.1.1 Vytvoření základního databázového schéma.....	12
2.1.2 Sestrojení dotazníku.....	13
2.1.3 Manuál k vyplnění dotazníku.....	16
2.2 Zpracování dat a statistická analýza	22
2.2.1 Data a jejich předzpracování	22
2.2.2 Statistická analýza	26
2.3 Základní analýza pro klienta.....	34
2.3.1 Ukázka protokolu pro firmu č. 117	35
2.3.2 Ukázka protokolu pro firmu č. 9	39
2.4 Rozšířené analýzy	40
2.5 Diskuze nad výsledky a vzniklé problémy	45
Závěr	49
Literatura a zdroje	50
Zdrojová data a soubory přiložené na CD.....	52

Seznam obrázků

Obrázek 1: Schéma celého procesu.	11
Obrázek 2: Relační schéma databázového systému.	13
Obrázek 3: Vytvoření formuláře pomocí VBA.....	14
Obrázek 4: Toolbox.....	14
Obrázek 5: Dotazník.....	15
Obrázek 6: Chybová hlášení.	16
Obrázek 7: Výběr pozice.	18
Obrázek 8: Název pracovní pozice.....	21
Obrázek 9: Počet zaměstnanců se stejnými podmínkami.....	21
Obrázek 10: Ukázka uložených dat.....	22
Obrázek 11: Počty dat na pozičních úrovních.....	22
Obrázek 12: Vyplnění ID společnosti.	23
Obrázek 13: Cleaning form.....	24
Obrázek 14: Převod měny.....	24
Obrázek 15: Wilksovy toleranční meze a ocenění zaměstnanců.....	25
Obrázek 16: Opravené hodnoty.	26
Obrázek 17: Porovnání agregátů.....	30
Obrázek 18: Poziční úrovně 5 až 12.....	30
Obrázek 19: Poziční úrovně 12 až 21.....	31
Obrázek 20: Intervaly spolehlivosti pro 90% kvantil.	32
Obrázek 21: Porovnání modelu trhu s Wilksovými mezemi.....	33
Obrázek 22: Porovnání intervalů spolehlivosti s Wilksovými mezemi.....	34
Obrázek 23: Basic Salary společnosti č. 117.	35
Obrázek 24: Base Salary společnosti č. 117.	36
Obrázek 25: Total Cash společnosti č. 117.	36
Obrázek 26: Porovnání mediánů pro BA.	37
Obrázek 27: Porovnání mediánů pro BS.....	37
Obrázek 28: Porovnání mediánů pro TC.....	38
Obrázek 29: Basic Salary pro firmu č. 9.	39
Obrázek 30: Porovnání mediánů pro BA a TC.....	40
Obrázek 31: Počet dat trhu Production.....	41
Obrázek 32: Model ocenění zaměstnanců trhu Production.....	42
Obrázek 33: Rozšířená analýza pro firmu.	43
Obrázek 34: Overflow.	46
Obrázek 35: Ukázka opraveného modelu trhu.	48

Seznam tabulek

Tabulka 1: Ukázka přiřazení pracovních pozic k referenčním úrovním.	19
Tabulka 2: Rodiny pracovních pozic.	20
Tabulka 3: Ukázka vytvoření intervalu spolehlivosti.	28
Tabulka 4: Nejnižší možné počty dat potřebné k výpočtu intervalů spolehlivosti. ...	29
Tabulka 5: Nejnižší možné počty dat potřebné k výpočtu (2).	29
Tabulka 6: Fixní přídavky a variabilní bonusy společnosti 117.	38
Tabulka 7: Porovnání zaměstnanců individuálně.	44

Úvod

Obecně si každá společnost chrání své citlivé údaje, zejména informace o mzdách a jiných ocenění svých zaměstnanců. Zároveň ale potřebuje vědět, jak dobře nebo špatně své zaměstnance oceňuje vůči konkurenci. Základní myšlenkou této práce je vytvořit věrohodný model trhu ocenění zaměstnanců a na jeho základě pak pomoci daným společnostem tuto problematiku řešit. Firmy se díky tomu mohou konkurenci na trhu přizpůsobit. Podílení se na této analýze je oboustranně prospěšné. Firmy poskytnou informace o ocenění jejich zaměstnanců, tím přispějí ke zvětšení databáze, tedy k většímu statistickému vzorku. Na druhé straně dostane společnost za vyplnění dotazníku základní analýzu obecného trhu.

Je nutné zaměstnance rozdělit do jednotlivých tříd tak, aby jejich pozice ve firmě odpovídala všem ostatním ve stejné třídě. Rozdělit se musí nejen zaměstnanci, ale i společnosti. Firmy jsou rozděleny do několika sektorů podle jejich specifikace. Statistický vzorek každé třídy zaměstnanců v každém sektoru musí být natolik dostačující, aby výsledky analýz měly co nejlepší výpovědní hodnotu a aby bylo vůbec možné je prezentovat jako model. S rostoucí velikostí statistického vzorku roste i možný počet pozičních tříd a sektorů. Je tedy nezbytné, zajistit správný kompromis mezi podrobností databáze a statistickou významností.

Při převzetí dat dochází k jejich kontrole, a to srovnáním s modelem vytvořeným na základě dat minulých. Informace, které těmto modelům nevyhovují, jsou firmám hlášeny a je požadováno potvrzení korektnosti. Na základě již sesbíraných dat je určen toleranční interval, v jakém by se měla roční ocenění zaměstnanců pohybovat příští rok. Jak je známo, v každé skupině zaměstnanců v každém odvětví se mzdy pohybují v jiném rozmezí.

Ve výsledku tato práce představuje kombinaci aplikace statistických metod ve vytvořených nástrojích a manuálu k použití těchto nástrojů pro poradenskou firmu, která se touto problematikou zabývá. Tento manuál, předvedení funkčnosti nástrojů na reálných datech a ukázka analýz pro klienty jsou předvedeny v případové studii.

V dnešní době existuje již mnoho společností, které poskytují podobné analýzy a účtují si za ně nemalé částky. V tomto projektu si ukážeme, jak je pomocí námi vytvořených nástrojů možné tyto analýzy uskutečnit. Součástí práce je praktické využití dostupných programů *Oracle SQL Developer* a *MS Office*, konkrétně *Excel*, *Visual Basic*, *Word* a *PowerPoint*, pomocí nichž jsme schopni projít celým procesem.

1 Teoretická část – statistický základ

Před sestrojením samotné analýzy ocenění trhu zaměstnanců je nezbytné nadefinovat teoretický základ. Bude se jednat o základní popisnou statistiku, která poskytne prvotní informace o našich datových výběrech. Pokročilejší statistický podklad nám pak umožní sestavit model požadovaného trhu, což je účel naší práce.

Předmětem statistiky je kvantitativní popis zkoumaných jevů a procesů, tedy požadujeme datové soubory prozkoumat, zpracovat a vyhodnotit nějaké závěry. Volba statistických metod aplikovaných při řešení konkrétních problémů v praxi závisí do značné míry na našich předchozích zkušenostech a znalostech o náhodných veličinách, jimiž se máme zabývat. Tedy o jejich rozděleních, případně některých parametrech jejich rozdělení. Pokud se snažíme o úspěšné řešení problémů v situacích, kdy naše dosavadní znalosti o rozdělení příslušných náhodných veličin nejsou příliš velké nebo pokud si nejsme jisti oprávněností předpokladů příslušnosti rozdělení, pak poslouží metody nezávislé na tvaru rozdělení. Takové metody jsou založeny na pořádkových statistikách a budeme je nazývat neparametrické. Konkrétně sestrojení a automatizace intervalů spolehlivosti pro kvantily spojitě náhodné veličiny a univerzálních tolerančních mezí je kromě základních statistických charakteristik naší hlavní náplní statistické části.

1.1 Základní pojmy

Konečná **populace (základní statistický soubor)** X rozsahu n je n -tice reálných nebo podmnožiny reálných čísel. Bez dalších omezení se prvky v souboru mohou běžně opakovat a pořadí prvků nepřikládáme prozatím žádný význam. Rozsah n je počet prvků, objektů souboru. Základní soubor může být i nekonečný.

Populační charakteristika je kvantitativní charakteristika sledované vlastnosti v populaci, např. průměrná mzda nebo variabilita bonusových ocenění. Běžně se tento parametr pokládá za reálné číslo, které je ale obecně neznámé. Neznámý parametr se obvykle značí Θ (specifičtěji např. μ pro střední hodnotu nebo σ pro směrodatnou odchylku).

Výběr prvků je konečná podmnožina populace. Zkoumané vlastnosti populace se tedy měří pouze u některých prvků (výběru) populace. Je vždy konečný rozsahu n . Pokud takový výběr kvalitně napodobuje strukturu základního souboru, pak tento výběr nazveme reprezentativní. Výběry se člení podle způsobu získávání prvků z populace. Úplné šetření nám poskytuje informaci o celé populaci, selektivní výběr je výběr zvolený pomocí dané selekce, dále záměrný výběr, atd.

Provedeme-li náhodný pokus, jehož výsledkem je hodnota x jednorozměrné náhodné veličiny X , která má distribuční funkci $F(x)$ a tento pokus opakujeme nezávisle n -krát, dostaneme hodnoty x_1, x_2, \dots, x_n . Tyto hodnoty lze považovat za realizace náhodné veličiny $X_i, i = 1, 2, \dots, n$. Protože n uvažovaných pokusů je n nezávislých opakování téhož pokusu, jsou náhodné veličiny X_1, X_2, \dots, X_n vzájemně nezávislé a všechny mají stejné rozdělení, jaké má veličina X . Celý proces nazýváme **náhodný výběr**. Lze ho provádět buď bez vracení prvků zpět do základního souboru, nebo s vracením.

Prostý náhodný výběr se provádí různými technikami losování, které musí zaručit, aby každý prvek populace měl stejnou možnost být zařazen do výběru. **Oblastní náhodný výběr** pak spočívá v tom, že populace je rozdělena do dílčích oblastí. Jednotlivé oblasti jsou uvnitř „sama sebe“ homogenní (sledované charakteristiky se příliš neliší) a naopak mezi sebou jsou oblasti heterogenní. Z každé oblasti se vybere vzorek metodou prostého náhodného výběru. [1], [2], [6], [9].

1.2 Beta rozdělení

Toto rozdělení je základem naší problematiky. Jedná se o rozdělení náhodné veličiny X s hustotou pravděpodobnosti

$$f_{\beta}(x) = \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}; 0 < x < 1; p, q > 0, \quad (1.2.1)$$

tedy rozdělení beta s parametry p, q . Distribuční funkce rozdělení beta vypadá následovně:

$$F_{\beta}(x) = \frac{1}{B(p, q)} \int_0^x t^{p-1} (1-t)^{q-1} dt; 0 < x < 1; p, q > 0. \quad (1.2.2)$$

Tento tvar je tzv. neúplná beta funkce $I_x(p, q) = 1 - I_{1-x}(q, p)$.

$$I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1} (1-t)^{q-1} dt, \quad (1.2.3)$$

kde pro přirozená čísla n je $\Gamma(n) = (n-1)!$ tzv. gama funkce.

Úplná beta funkce je pak dána předpisem:

$$I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 t^{p-1} (1-t)^{q-1} dt. \quad (1.2.4)$$

Dále také platí, že pomocí neúplné beta funkce lze vyjádřit distribuční funkci binomického rozdělení:

$$\sum_{t=0}^x \binom{n}{t} \pi^t (1-\pi)^{n-t} = \frac{1}{B(n-x, x+1)} \int_0^{1-\pi} z^{n-x-1} (1-z)^x dz \quad (1.2.5)$$

$$= I_{1-\pi}(n-x, x+1)$$

V následujících kapitolách bychom se bez neúplné funkce beta a samotného beta rozdělení neobešli. [1], [8].

1.3 Uspořádaný náhodný výběr

Mějme náhodný výběr $\{x_1, x_2, \dots, x_n\}$ z daného rozdělení a takový výběr seřadíme podle velikosti vzestupně. Tento uspořádaný náhodný výběr označíme $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde hodnotu $x_{(i)}$ nazveme i -tou **pořádkovou statistikou**. V následujícím předpokládáme případy, kdy výběr pochází z rozdělení majícího spojitou distribuční funkci $F(x)$ a hustotu pravděpodobnosti $f(x) = dF(x)/dx$. Platí tedy $P(x_i = x_{i-1}) = 0$ pro všechna i . Pro distribuční funkci $F_{(i)}(x)$ náhodné veličiny $X_{(i)}$ platí:

$$F_{(i)}(x) = P(X_i < x) = \sum_{j=i}^n \binom{n}{j} [F(x)]^j [1-F(x)]^{n-j}, \quad (1.3.1)$$

kde $-\infty < x < \infty$. Jelikož náhodné veličiny X_1, X_2, \dots, X_n jsou vzájemně nezávislé a všechny mají tutéž distribuční funkci $F(x)$, je pravděpodobnost toho, že právě j z n veličin X_1, X_2, \dots, X_n nabude hodnoty menší než x , rovna:

$$\binom{n}{j} [F(x)]^j [1-F(x)]^{n-j}, \quad j = 0, 1, 2, \dots, n.$$

Porovnáme-li (1.3.1) s (1.2.5), vidíme již podobnost s beta rozdělením a platí následující:

$$F_{(i)}(x) = I_{F(x)}(i, n-i+1) = \frac{n!}{(i-1)!(n-i)!} \int_0^{F(x)} t^{i-1} (1-t)^{n-i} dt, \quad (1.3.2)$$

kde $-\infty < x < \infty$. Hustota pravděpodobnosti $f_{(i)}(x) = dF_{(i)}(x)/dx$ statistiky $X_{(i)}$ vypadá takto:

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} f(x) [1-F(x)]^{n-i}, \quad (1.3.3)$$

kde opět $-\infty < x < \infty$. V případě náhodného výběru n z rovnoměrného rozdělení $R(1/2, 1/2)$ má statistika $X_{(i)}$ hustotu pravděpodobnosti:

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1} (1-x)^{n-i}, 0 < x < 1 \quad (1.3.4)$$

= 0, jinde.

Tedy statistika $X_{(i)}$ má rozdělení $B(i, n - i + 1)$.

Jestliže $\{x_1, x_2, \dots, x_n\}$ je náhodný výběr, potom $\{y_1, y_2, \dots, y_n\}$, kde $y_i = F(x_i)$ je náhodný výběr transformovaných náhodných proměnných. Dále pokud $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ je uspořádaný náhodný výběr náhodných proměnných, pak $\{y_{(1)}, y_{(2)}, \dots, y_{(n)}\}$, kde $y_{(1)} < y_{(2)} < \dots < y_{(n)}$ a $y_{(i)} = F(x_{(i)})$ je transformace, která při definovaných podmínkách na $F(x)$ dodržuje pořadí uspořádaného náhodného výběru. [1], [4], [8].

1.4 Popisná statistika

Zde se omezíme na pojmy týkající se jednorozměrných statistických souborů, jejichž prvky jsou reálná čísla. Kompletní seznam níže popsaných charakteristik a vzorců nalezneme v [2], [6] a [9].

1.4.1 Charakteristiky polohy

Aritmetický průměr souboru $X = x_1, x_2, \dots, x_n$ je hodnota $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Jsou-li v_i , kde i je stejného rozsahu n jako rozsah statistického souboru x_i , potom součet $\sum_{i=1}^n x_i v_i$ se nazývá **vážený průměr** \bar{x}_v hodnot x_1, x_2, \dots, x_n s vahami v_1, v_2, \dots, v_n . O vahách v_i se obvykle předpokládá, že jsou nezáporné a jejich součet je roven jedné. Není to ovšem podmínkou. Vážený průměr lze vyjádřit i následovně:

$$\bar{x}_v = \frac{\sum_{i=1}^n x_i v_i}{\sum_{i=1}^n v_i} \text{ nebo } \bar{x}_v = \frac{\sum_{i=1}^n \frac{x_i}{v_i}}{\sum_{i=1}^n \frac{1}{v_i}}.$$

Záleží vždy na tom, jaký charakter váhy mají a jak je

chceme použít. Vážený průměr může mít i význam v situacích, kde dostáváme několik průměrů z různých datových výběrů, které ale pocházejí ze stejného základního souboru. Vážený průměr je pak vyjádřen vztahem: $\bar{x}_v = \frac{\sum_{i=1}^n \bar{x}_i v_i}{\sum_{i=1}^n v_i}$ nebo

$$\text{obdobně } \bar{x}_v = \frac{\sum_{i=1}^n \frac{\bar{x}_i}{v_i}}{\sum_{i=1}^n \frac{1}{v_i}}.$$

Dalšími průměry jsou: **Geometrický průměr** $\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i}$,

harmonický průměr $\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$, **kvadratický průměr** $\bar{x}_K = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$, atd. Vztah

mezi jednotlivými průměry je $x_{(1)} \leq \bar{x}_H \leq \bar{x}_G \leq \bar{x} \leq \bar{x}_K \leq x_{(n)}$.

Jinou charakteristikou polohy souboru je **medián**, který se značí \tilde{x} . Předpokládejme, že hodnoty x_1, x_2, \dots, x_n jsou seřazeny podle velikosti. Pro n liché se definuje $\tilde{x} = x_m$, kde $m = (n + 1)/2$, pro n sudé je pak medián definován jako $\tilde{x} = (x_m + x_{m+1})/2$, kde $m = n/2$. Výhodou mediánu oproti průměru je jeho

robustnost (malá citlivost na odlehlá nebo chybná data). Někdy se při charakterizaci souboru udává také tzv. **modus**, což je hodnota s nejvyšší četností. Modus nemusí být vždy definován jednoznačně, tedy v souboru se nachází více hodnot se stejnou nejvyšší četností.

Pro datové soubory většího rozsahu se často určují tzv. kvartily (z anglického quartile). **Dolní kvartil** $x_{0,25}$ je laicky řečeno taková hodnota, že čtvrtina členů datového souboru je menší a tři čtvrtiny členů je větší než $x_{0,25}$. Analogicky definujeme **horní kvartil** $x_{0,75}$. Prostřední kvartil je roven mediánu. Jemnější členění dat je dělení na kvantily. 25% kvantil odpovídá dolnímu kvartilu, atd. Obecně, je-li $p \in (0,1)$, pak pro statistický soubor definujeme **100p% kvantil** x_p následovně. Uspořádáme soubor podle velikosti vzestupně a tím získáme pořádkové statistiky. Označíme je $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde platí $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Je-li číslo np celé, potom $x_p = [x_{np} + x_{(np+1)}]/2$. Nemá-li číslo np celé, definujeme $x_p = x_{(i)}$, kde i je nejmenší celé číslo vyšší než np . Nezapomínáme ani na **maximum** $x_{(n)}$ a **minimum** $x_{(1)}$.

1.4.2 Charakteristiky variability

Statistické soubory je vhodné charakterizovat nejen na základě průměru a kvantilů, ale také podle toho, jak se jednotlivá data liší od sebe navzájem nebo od zvolené charakteristiky polohy. **Rozptyl** (druhý centrální moment) statistického souboru x_1, x_2, \dots, x_n je charakterizován hodnotou $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$. Jedná se o průměrnou kvadratickou odchylku hodnot od aritmetického průměru. Výpočetní tvar rozptylu je $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$. Pro $n > 1$ se definuje též **výběrový rozptyl** $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{n-1} \sigma^2$. Výběrový rozptyl má stejný význam jako rozptyl, ale lepší statistické vlastnosti, proto je v rámci statistických metod používanější. Číslo $\sigma_n = \sqrt{\sigma_n^2}$ se nazývá **směrodatná odchylka**, **výběrová směrodatná odchylka** je pak definována číslem $s = \sqrt{s^2}$.

Variační rozpětí $R = x_{(n)} - x_{(1)}$ je nejjednodušší charakteristikou variability. Jeho problémem je ale závislost na extrémních hodnotách. Často používanou charakteristikou je potom **kvartilové rozpětí** $R_{(75;25)} = x_{0,75} - x_{0,25}$, které je robustní vůči odlehlým pozorováním. Obecně lze definovat **kvantilové rozpětí** $R_{(p,q)} = x_p - x_q$, kde $p, q \in (0,1)$ a $0 < q < 0,5 < p < 1$, ať už je toto rozpětí symetrické či nikoliv.

1.5 Toleranční a predikční interval

Mějme náhodný výběr $\{x_1, x_2, \dots, x_n\}$ (iid) z náhodné proměnné ξ s distribuční funkcí $F_\xi(x)$ a dvě statistiky $L(x_1, x_2, \dots, x_n)$ a $U(x_1, x_2, \dots, x_n)$. Pokud platí, že:

$$L(x_1, x_2, \dots, x_n) < U(x_1, x_2, \dots, x_n) \text{ a} \\ P_{x_1, x_2, \dots, x_n} (P_\xi \{L(x_1, x_2, \dots, x_n) \leq \xi \leq U(x_1, x_2, \dots, x_n)\} \geq \beta) \geq \gamma, \quad (1.5.1)$$

potom $\langle L, U \rangle$ budeme nazývat tolerančním intervalem (β, γ) . Tedy pravděpodobnost toho, že interval $\langle L, U \rangle$ pokryje více než $100\beta\%$ pravděpodobnosti výskytu náhodné proměnné ξ , je alespoň $100\gamma\%$.

Za výše uvedených podmínek budeme pod tzv. β predikčním intervalem rozumět obě statistiky, pokud platí:

$$E_{x_1, x_2, \dots, x_n} [P_\xi \{L(x_1, x_2, \dots, x_n) \leq \xi \leq U(x_1, x_2, \dots, x_n)\}] \geq \beta \quad (1.5.2)$$

U tolerančního intervalu požadujeme, aby pravděpodobnost jevu, že náhodná proměnná bude ležet v odhadnutých (spočtených) mezích s pravděpodobností β , bude alespoň γ . U predikčního intervalu požadujeme, aby střední hodnota pravděpodobnosti jevu, že náhodná proměnná bude ležet v odhadnutých (spočtených) mezích, bude alespoň β . [1], [3], [8].

1.6 Intervaly spolehlivosti pro kvantily spojitých rozdělání

Mějme uspořádaný výběr $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ z rozdělání, které má spojitou distribuční funkci $F(x)$ a hledáme intervaly spolehlivosti pro $100p\%$ kvantil rozdělání x_p , kde $p \in (0,1)$. Pro dvojici uspořádaných statistik $x_{(i)}, x_{(j)}$, kde $1 \leq i < j \leq n$, platí

$$P(x_{(i)} < x_p) = P(x_{(i)} < x_{(j)} < x_p) + P(x_{(i)} < x_p < x_{(j)}). \quad (1.6.1)$$

Z tohoto a ze vztahu (1.3.2) vyplývá, že

$$P(x_{(i)} < x_p < x_{(j)}) = P(x_{(i)} < x_p) - P(x_{(j)} < x_p) = \\ = I_{F(x_p)}(i, n - i + 1) - I_{F(x_p)}(j, n - j + 1). \quad (1.6.2)$$

Tato dvojice uspořádaných statistik $x_{(i)} < x_p < x_{(j)}$ je potom dvoustranným intervalem spolehlivosti pro kvantil x_p , přičemž koeficient spolehlivosti je vzhledem ke vztahu $F(x_p) = p$ roven číslu:

$$I_p(i, n - i + 1) - I_p(j, n - j + 1) = \\ = 1 - I_{1-p}(n - i + 1, i) - I_p(j, n - j + 1). \quad (1.6.3)$$

Pokud se $j = n - i + 1$, tak koeficient spolehlivosti je roven číslu:

$$\begin{aligned}
& I_p(i, n - i + 1) - I_p(n - i + 1, i) = \\
& = I_{1-p}(i, n - i + 1) - I_{1-p}(n - i + 1, i) = \\
& = 1 - I_{1-p}(n - i + 1, i) - I_p(n - i + 1, i).
\end{aligned} \tag{1.6.4}$$

Pro $i = 1$ a $j = n$ platí:

$$\begin{aligned}
P(x_{(1)} < x_p < x_{(n)}) & = 1 - I_{1-p}(n, 1) - I_p(n, 1) \\
& = 1 - (1 - p)^n - p^n.
\end{aligned} \tag{1.6.5}$$

Všechny vzorce a odvození jsou k nalezení v [1].

1.7 Wilksovy toleranční meze

V některých případech, jako zde, není vhodný předpoklad o rozdělení, ze kterého pochází pozorování náhodné proměnné, přijatelný. Potom je možné využít pořádkových statistik.

Mějme náhodný výběr $\{x_1, x_2, \dots, x_n\}$ rozsahu n náhodné proměnné X s distribuční funkcí $F(x)$ a hustotou $f(x)$. Je-li $F(x)$ spojitá a rostoucí, potom vstoupně seřazená pozorování $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ jsou pořádkové statistiky nad výběrem $\{x_1, x_2, \dots, x_n\}$ a $x_{(i)}$ je pak i -tá pořádková statistika. Dále předpokládáme, že jednotlivá pozorování jsou po dvou různá, tedy platí: $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

Pro libovolné $i < j \leq n$ nazveme $F_X(x_{(j)}) - F_X(x_{(i)})$ pravděpodobnostním pokrytím intervalu $\langle x_{(i)}, x_{(j)} \rangle$. Uspořádaný náhodný výběr $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ dělí celý prostor na $n + 1$ intervalů $J_i = \langle x_{(i-1)}, x_{(i)} \rangle$, kde $x_{(0)} = -\infty$ a $x_{(n+1)} = \infty$. Označme $c_i = P(X \in J_i)$ jako pravděpodobnostní pokrytí již zmíněného intervalu $J_i = \langle x_{(i-1)}, x_{(i)} \rangle$. Tato pokrytí jsou pak spojené vztahem $\sum_{i=1}^{n+1} c_i = 1$.

Součet W libovolných $r \leq n$ po dvou různých náhodných proměnných c_i má pak hustotu:

$$\begin{aligned}
f_W(x) & = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} x^{r-1}(1-x)^{n-r} = \\
& = \frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r} = r \binom{n}{r} x^{r-1}(1-x)^{n-r},
\end{aligned} \tag{1.7.1}$$

tedy beta rozdělení.

Pro jednorozměrnou populaci X hledáme toleranční meze ve tvaru $L(x_1, x_2, \dots, x_n) = x_{(k)}$ a $U(x_1, x_2, \dots, x_n) = x_{(n-k+1)}$. Potom:

$$\begin{aligned} \gamma &= P_{x_1, x_2, \dots, x_n} [P_{\xi} \{L(x_1, x_2, \dots, x_n) \leq \xi \leq U(x_1, x_2, \dots, x_n)\} \geq \beta] = \\ &= P_{x_1, x_2, \dots, x_n} [F_{\xi}(U) - F_{\xi}(L) \geq \beta] = \\ &= P_{x_1, x_2, \dots, x_n} [F_{\xi}(x_{(n-k+1)}) - F_{\xi}(x_{(k)}) \geq \beta] = \\ &= P \left[\sum_{j=1}^{n-k+1} c_j - \sum_{j=1}^k c_j \geq \beta \right] = P \left[\sum_{j=k+1}^{n-k+1} c_j \geq \beta \right] = P[S \geq \beta]. \end{aligned} \quad (1.7.1)$$

$S = \sum_{j=k+1}^{n-k+1} c_j$ je pak součet $n - 2k + 1$ elementárních pravděpodobnostních pokrytí.

Shrneme-li výše uvedené vztahy, potom:

$\gamma = P(S \geq \beta) = 1 - P(S < \beta) = 1 - F_S(\beta)$. Náhodná proměnná S má hustotu:

$$f_S(x) = \frac{\Gamma(n+1)}{\Gamma(n-2k+1)\Gamma(n-(n-2k+1)+1)} x^{n-2k+1-1} (1-x)^{n-(n-2k+1)}, \quad (1.7.2)$$

po úpravě:

$$f_S(x) = \frac{\Gamma(n+1)}{\Gamma(n-2k+1)\Gamma(2k)} x^{n-2k} (1-x)^{2k-1}, \quad (1.7.3)$$

tedy dostáváme, že:

$$\begin{aligned} 1 - F_S(\beta) &= 1 - \int_0^{\beta} \frac{\Gamma(n+1)}{\Gamma(n-2k+1)\Gamma(2k)} x^{n-2k} (1-x)^{2k-1} dx = \\ &= 1 - I(\beta; n - 2k + 1, 2k), \end{aligned} \quad (1.7.4)$$

kde $I(\beta; n - 2k + 1, 2k)$ je distribuční funkce beta rozdělení s parametry $p = n - 2k + 1$ a $q = 2k$. Nakonec platí, že:

$$\gamma = 1 - I(\beta; n - 2k + 1, 2k) \Leftrightarrow 1 - \gamma = I(\beta; n - 2k + 1, 2k).$$

Parametrickým řešením rovnice $\beta = I^{-1}(1 - \gamma; n - 2k + 1, 2k)$ nalezneme vztah mezi n, k pro dané β . Je-li vhodně zvolené γ dostatečně velké, potom k bude malé. Všechny vztahy mezi n, k, β a γ jsou tabelovány v [7]. Kompletní odvození ostatních vzorců a rovnic nalezneme v [3], [4], [5], [8].

Tento statistický podklad nám postačí k řešení požadovaných analýz této práce. Za poskytnutá data od firem je jim přislíbena základní analýza trhu ocenění zaměstnanců. Rozšířené analýzy selektivních výběrů jsou poté požadovány individuálně od jednotlivých klientů. Intervaly spolehlivosti pro kvantily nám poslouží jako model trhu za daný rok. Wilksovy toleranční meze potom mají význam při sběru dat v roce příštím. Zajistí nám označení extrémních hodnot

a prevenci proti chybám ve sběru. Tím při dlouhodobém používání z části docílíme toho, že naše výběry jsou „bezchybné“, stejně rozdělené a nezávislé (iid).

Pomocí statistiky nelze předpoklad nezávislosti náhodných výběrů a skutečnosti, že pochází ze stejného rozdělení, nikdy potvrdit. Tento výrok lze pouze zamítnout. Informaci o tomto předpokladu dostáváme z části mimo statistiku. Jelikož se obecně u sběru dat provádí oblastní výběr, tedy výběr, u něhož předpokládáme, že jednotlivé datové soubory jsou homogenní (rozdělení firem do sektorů a zaměstnanců do pozičních úrovní), potom předpokládáme, že tyto výběry jsou nezávislé a stejně rozdělené.

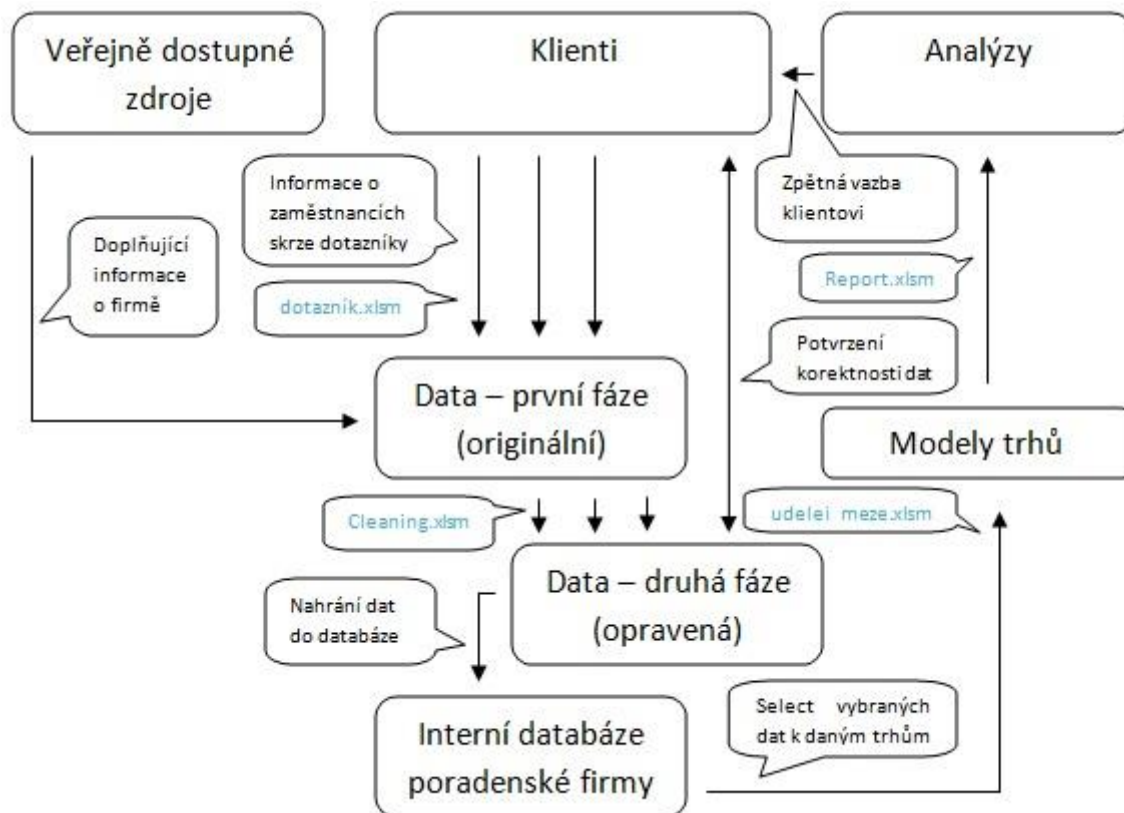
Mimo statistické rešerše byl proveden průzkum na téma ocenění zaměstnanců a k tomu podobných výrazů. Dostupná literatura [10], [11], [12], [13] a [14] se nejčastěji zabývá problémem správného zařazení zaměstnanců do jednotlivých skupin a tříd tak, aby byly zaměstnanci mezi sebou v dané třídě homogenní a mezi třídami heterogenní. Mezi tyto metody patří např. metody pořadí (bodovací metoda) nebo klasifikační metoda (metoda faktorového porovnávání). Dalšími tématy jsou správné odměňování zaměstnanců, a jak toto oceňování může pracovníky motivovat. Dále z pohledu vlastníka společnosti je hlavním problémem řízení lidských zdrojů a jejich ohodnocení. Tyto metody a celková problematika správného zařazení zaměstnanců je obsáhlým tématem. Předpokládáme, že tento krok je již vyřešen a soustředíme se tedy na statistickou část, která přichází hned poté. V praktické části této studie se budeme zabývat pravděpodobnostními modely a statistickou analýzou reálných dat. Použijeme k tomu metody popsané v teoretické části v kombinaci se znalostmi a zkušenostmi společnosti, která na této práci spolupracuje. Ta po dobu své působnosti má problematiku správného zařazení zaměstnanců již optimalizovanou. Data (informace o zaměstnancích), s kterými je v praktické části pracováno, jsou věrohodná a zaměstnanci, kteří se průzkumu účastnili, jsou již správně zařazení do jednotlivých kategorií (popsáno v 2.1.3).

V následující kapitole si ukážeme praktickou část této práce. Analytici poradenské firmy, která vytváří analýzy pro její klienty, jsou na základě manuálů a případových studií schopni sestavit požadované analýzy. Bude také ukázána funkčnost programových nástrojů na reálných datech.

2 Praktická část – případová studie

V praktické části budeme aplikovat výše uvedené metody na reálná data. Ukážeme si, jak vyčíslit základní statistiky, sestavit intervaly spolehlivosti pro kvantily a toleranční meze. Sestrojíme model trhu ocenění zaměstnanců a vytvoříme analýzy pro klienty, kteří se průzkumu zúčastnili. Uvažujeme, že máme k dispozici velké množství klientů, kteří se našeho průzkumu účastnili již v minulých letech. Na základě minulých dat máme již spočítané toleranční meze pro sběr dat v tomto roce (bude ukázáno v kapitole 2.2.1)

Nejdříve obecně nastíníme celý proces od začátku do konce. Vše začne sběrem dat od firem. Těm se pošle navrhnutý dotazník. Klienti dotazník vyplní a zašlou ho zpět. Po přijetí vyplněného dotazníku ho analytici vyhodnotí a data tzv. vyčistí. Dále se musí vyjasnit všechny nesrovnalosti a posléze se data nahrají do databáze. Po přijetí všech dotazníků od všech klientů se začnou data zpracovávat a vyhodnocovat. Určí se intervaly spolehlivosti pro kvantily rozdělení náhodných veličin a sestaví se model trhu. Vypracuje se analýza trhu a případná doporučení pro jednotlivé klienty. Po odevzdání vyhodnocených dat končí spolupráce se společnostmi pro daný rok. V poslední řadě se aktualizují makra a ostatní programové nástroje na rok příští. Na následujícím diagramu lze vidět celý proces.



Obrázek 1: Schéma celého procesu.

2.1 Dotazníkové šetření

Základním kamenem je sestavení dotazníku, který slouží ke sběru dat. Správné navržení dotazníku často zmenšuje dobu potřebnou k datovému zpracování. Mohou mít nejrůznější formy od papírových po mnoho typů elektronických dotazníků. Prvním úkolem je stanovit si, na co se v dotazníku budeme společností ptát. Jak jsme již naznačili v úvodu, je zapotřebí určit ten správný poměr počtu tříd (otázek v dotazníku a jejich možnosti) a velikosti statistického vzorku. V této práci je znázorněna pouze základní verze s nejpodstatnějšími informacemi, i přesto lze data rozdělit na více než tisíc možných variant trhu. Dotazníky velkých společností obsahují ještě mnohem více informací a jejich databáze jsou tím tolikrát větší.

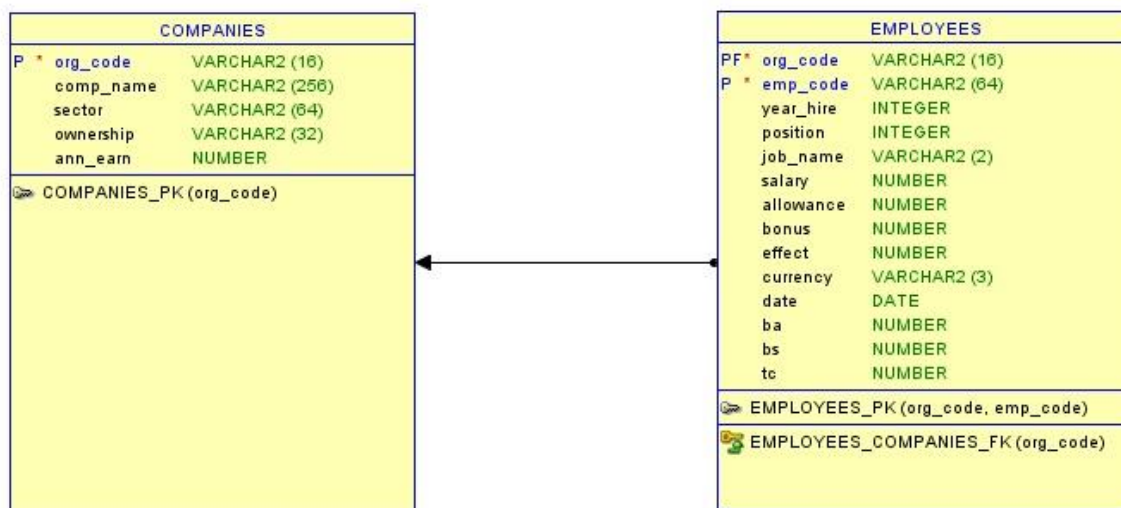
2.1.1 Vytvoření základního databázového schéma

Úkolem je navrhnout jednoduché databázové schéma pro účely našeho příkladu. Systém bude obsahovat pouze dvě základní tabulky, tabulku společností „COMPANIES“ a tabulku zaměstnanců „EMPLOYEES“.

Tabulka společností obsahuje informace týkající se jednotlivých firem, které se dotazníkového šetření zúčastnily. Unikátní identifikační označení společnosti „org_code“ je v této tabulce primárním klíčem. Dále v tabulce vystupují následující atributy. Jméno společnosti „comp_name“, sektor ve kterém firma vystupuje „sector“, informaci o vlastnictví firmy „ownership“ a celkové roční výnosy firmy „ann_earn“.

Druhá tabulka „EMPLOYEES“ nese informace o jednotlivých zaměstnancích. Primárním klíčem je kombinace atributů „org_code“ (cizí klíč k atributu „org_code“ tabulky „COMPANIES“) a „emp_code“, což je identifikační označení zaměstnance. Společně tvoří unikátní kód, podle kterého lze najít specifického zaměstnance dané firmy. Dále jsou v ní atributy nesoucí obecné informace, jako jsou rok nástupu do firmy „year_hire“, pozice ve firmě „position“ a kategorické zařazení zaměstnance „job_name“, zjednodušeně název pozice ve firmě, např. CEO, IT manažer, účetní, právní oddělení apod. Dalšími atributy jsou „salary“, což je roční základ mzdy, „allowance“, to jsou nejrůznější příspěvky, např. na mobilní telefon, jídlo, cestu atd. a atribut „bonus“, do kterého se vkládají informace o motivačních bonusech, provizích z prodeje a dalších. V neposlední řadě jsou v tabulce ještě počet zaměstnanců se stejnými podmínkami „effect“, měna „currency“ a datum „date“, který vyjadřuje datum sběru dat a tedy i datum, kdy bylo celkové ocenění zaměstnance převedeno na jednotnou měnu, pokud je tedy zrovna tento zaměstnanec placen v měně jiné.

Posledními atributy jsou „ba“, „bs“ a „tc“. To jsou již spočtené agregáty. Agregát *Basic Salary* „ba“ vyjadřuje základ mzdy zaměstnance. *Base Salary* „bs“ je ocenění zaměstnance včetně fixních přídavků (již zmíněno u atributu „allowance“). Poslední agregát *Total Cash* „tc“ vyjadřuje celkové roční ocenění zaměstnance včetně všech fixních i variabilních příspěvků. Na obrázku 2 je znázorněno relační schéma našeho interního informačního databázového systému.



Obrázek 2: Relační schéma databázového systému.

K sestavení schématu byl použit program *Oracle SQL Developer*. Veškerá manipulace s databází a případné ukázky příkazů *SQL* jsou v této práci vytvořeny právě pro výše zmíněný software.

2.1.2 Sestrojení dotazníku

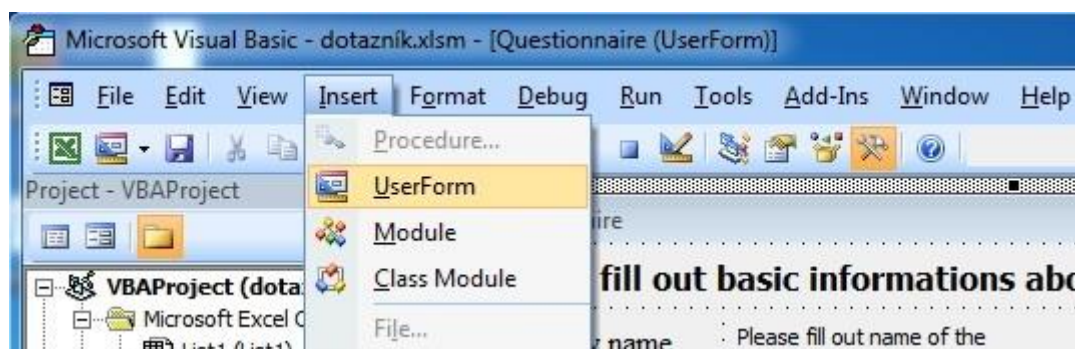
Dotazníky by měly být stručné, přehledné a srozumitelné. Ve studii jsme využili nástrojů *MS Excel* a *MS Visual Basic*. Námi vytvořený dotazník je vhodný jak pro nové klienty, tak i pro klienty, kteří dotazník již vyplňovali v minulosti. Pro nováčky je sestrojen dotazník v prostředí *MS Visual Basic*. Je přehledný a nelze v něm udělat mnoho chyb, to vše je ale na úkor času stráveného u vyplňování.

Pokud se klient, a tedy i potenciální zákazník, rozhodne našeho průzkumu zúčastnit, je mu takový dotazník zaslán či předán. Základní informace o firmě, které jsou později nahrány do databázové entity „*COMPANIES*“, jsou zjištěny při předání tohoto dotazníku klientovi. Zjistíme tedy v první řadě, jak se společnost jmenuje celým jménem a v jakém sektoru se projevuje (např. automobilový průmysl, těžba, potravinářství apod.). Dále kolik má společnost zaměstnanců, kdo ji vlastní a nakonec, jaké jsou její roční výnosy z prodeje produktů či služeb. Informací by mohlo být více, např. jak dlouho firma působí na trhu nebo jestli se jedná o akciovou společnost atd. Pro účely práce postačí informace základní. Firma má samozřejmě

právo na to, aby některé z informací neposkytla, tím ovšem přichází o možnost se účastnit více analýz.

Jedná-li se o novou společnost, je vhodné, abychom osobu odpovědnou za vyplnění dotazníku osobně seznámili s procesem vyplnění. Hlavně s možnými odpověďmi, jak mezi nimi vybrat tu správnou a obecně, co dělat, kdyby se klient dostal do úzkých. V následujících kapitolách je dotazník podrobně popsán.

Mějme k dispozici programy *MS Excel* a *MS Visual Basic*. V této práci se konkrétně jedná o anglickou verzi z roku 2007. Dotazník pro nové zákazníky je vytvořen pomocí formulářů již zmíněného programu *MS Visual Basic*.



Obrázek 3: Vytvoření formuláře pomocí VBA

Po vložení formuláře si již sestavíme dotazník podle vlastních představ. Pomocí okna *Toolbox* lze do formuláře přidat nejrůznější prvky, od nadpisů a popisků, přes textová a zaškrťovací políčka až po samostatná tlačítka. Jednotlivé prvky se dají pojmenovat podle libosti a dá se s nimi dále pracovat.



Obrázek 4: Toolbox.

Přidáváním jednotlivých prvků a jejich nastavením postupně sestojíme požadovaný dotazník. Forma námi vytvořeného dotazníku je vidět na obrázku 5. Jednotlivé otázky jsou popsány v následující kapitole. Odpovědi jsou ukládány do skrytého listu, který lze podle libosti i opatřit heslem, abychom k datům měli přístup jen my.

Pokud by se jednalo o již zaučeného klienta, tedy klienta, který by dotazník nevyplňoval poprvé, lze tento list nechat zobrazený a osoba odpovědná za vyplnění dotazníku může hodnoty zapisovat rovnou do tohoto listu. Ušetří se tím čas a my se již nemusíme strachovat, že budou vyplněny špatné odpovědi, protože je klient již seznámen s možnými odpověďmi z minula.

The image shows a software window titled "Questionnaire" with a close button in the top right corner. The main heading inside the window is "Please fill out basic information about employee". The form consists of several rows, each with a label, a description, and an input field:

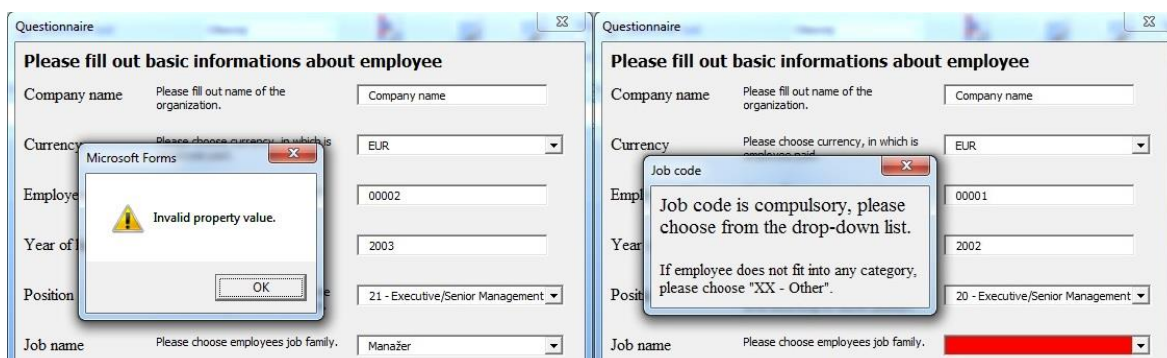
- Company name:** "Please fill out name of the organization." followed by a text input field.
- Currency:** "Please choose currency, in which is employee paid." followed by a dropdown menu.
- Employee ID:** "Please fill out unique identification code for every employee." followed by a text input field.
- Year of hire:** "Please fill out year, when the employee was hired." followed by a text input field.
- Position:** "Please choose employee reference level according to his/her position." followed by a dropdown menu.
- Job name:** "Please choose employees job family." followed by a dropdown menu.
- Salary:** "Please fill out basic annual salary." followed by a text input field.
- Allowances:** "Is the employee eligible? If yes, please fill out sum of annual allowances (meal, mobile, shift, transportation, temperature allowance etc.)" followed by two radio buttons labeled "Yes" and "No", and a text input field.
- Variable bonuses:** "Is the employee eligible? If yes, please fill out sum of annual bonuses (sales commision, profit share, bonus incentive etc.)" followed by two radio buttons labeled "Yes" and "No", and a text input field.

At the bottom of the form, there are two buttons: "Save" (with a dotted border) and "Multiple save".

Obrázek 5: Dotazník.

Na obrázku 5 vidíme již dotazník sloužící ke sběru informací potřebných pro naplnění databázové entity „EMPLOYEES“. Je sestrojen v anglickém jazyce, protože neuvažujeme, že poradenská firma, působí pouze na českém trhu a data získává pouze od českých společností. Tento dotazník je určen pro nové klienty. Je opatřen

proti zadání nechtěných nebo jinak nevyhovujících odpovědí. Např. pokud klient nevyplní povinnou informaci, tak je pole zvýrazněno červenou barvou a zobrazí se okno s chybovou hláškou.



Obrázek 6: Chybová hlášení.

2.1.3 Manuál k vyplnění dotazníku

Dotazník nelze klientovi poslat jen tak a předpokládat, že vyplní všechny informace přesně, jak by měly být vyplněny. V této kapitole rozebereme jednotlivé otázky a možné odpovědi. Některé otázky jsou jednoznačné a je jednoduché je vyplnit, jiné jsou složitější. Pro jistotu by měly být ale klientovi popsány všechny otázky. V následujících odstavcích jsou vysvětleny všechny otázky a jejich možné odpovědi. V závorkách se nachází anglické ekvivalenty.

Název společnosti (Company name): Klient vyplní celý název společnosti včetně zkratk s.r.o. nebo a.s. (inc., Ltd.) apod. Tato kolonka se vyplní pouze poprvé a poté se uloží do paměti a vždy načte, takže od prvního uložení názvu společnosti není již potřeba tuto informaci vyplňovat znovu.

Měna (Currency): V další otázce se zadává měna, ve které je zaměstnanec vyplácen. Odpovědi jsou vybrány z rozbalovacího seznamu. Seznam lze kdykoliv rozšířit či upravit. Pro naše účely jsou možné odpovědi nastaveny na CZK, USD a EUR, korunu českou, americký dollar a euro. Uvažujeme, že firma vyplácí všechny nebo alespoň většinu svých zaměstnanců v jedné měně, proto je tato informace také opakovaně načítána.

Identifikační kód zaměstnance (Employee ID): Zde se vyplní unikátní identifikační kód zaměstnance. Je možné použít značení, které používá firma nebo lze zadat jméno zaměstnance. Pokud firma nepoužívá žádné značení zaměstnanců a i přes to se u jmen zaměstnanců objevují shody, je možné použít obyčejný seznam od jedné po součet zaměstnanců.

Rok nástupu do firmy (Year of hire): Tato informace není povinná, není tedy třeba ji vyplňovat. Firma ovšem ztrácí možnost zúčastnit se analýz trhu, založeného na době setrvání ve firmě.

Poziční, referenční úroveň zaměstnance (Employee reference level): Poziční úroveň zaměstnance je nejtěžší otázkou dotazníku. Díky tomu, že různé firmy jsou vždy jinak strukturálně založeny a vedeny, je občas velmi těžké rozhodnout o správné poziční úrovni zaměstnance. Tyto úrovně lze chápat jako obdobu platových tříd státní sféry. Pozice zaměstnanců mohou být na jedné poziční úrovni různorodé. Všichni zaměstnanci jsou si ovšem rovni v základních parametrech, jako rozhodování, řešení problémů, tvořivost, míra zodpovědnosti, vliv na výsledky, iniciativa, plánování/organizování, argumentování, vztahy k ostatním pracovníkům a mnoho dalších, nebo v jejich kombinaci. V naší studii jsme pozice rozdělili na čtyři kategorie a všechny tyto kategorie jsou ještě děleny na pět dílčích úrovní.

Do první kategorie patří zaměstnanci, kteří ve firmách zastávají nejnižší pozice. Jsou to zaměstnanci zajišťující základní operativní výkony, produkci, montáž apod. (Clerical / Operations). Tito zaměstnanci nenesou žádnou odpovědnost za někoho jiného, pouze za své výkony. Není u nich potřeba, aby dosáhli vyššího než základního, popř. středoškolského vzdělání. Úrovně této kategorie jsou 5 až 9, kde úroveň 5 vyjadřuje nově zaměstnaného člověka, většinou ve zkušební době, pobírajícího nejnižší možný plat, zpravidla ve společnostech zabývajících se nějakou určitou výrobou či montáží. Úroveň 9 potom naopak obsahuje zaměstnance, kteří jsou již zkušení, ve firmě působí delší dobu a chystá se pro ně v nejbližší době povýšení na vyšší pozice. Záleží také na druhu práce. Mějme např. montážní firmu, která se zabývá sestrojováním komponentů do automobilů. Tato práce není obzvláště namáhavá ani nevyžaduje zvláštní vzdělání. Dělníky tohoto druhu bychom tedy zařadili do úrovní 5 až 7, podle platových rozhraní firmy, náročnosti, druhu práce apod. Oproti tomu v analytické firmě poskytující nějaké poradenské služby, je i pro základní prvotní pozici analytika potřeba určitého vzdělání, většinou nejméně středoškolského ukončeného maturitou. Nástupní plat bývá také vyšší než ve společnostech zabývajících se výrobou či montáží. Těmto zaměstnancům bychom přidělili úrovně 7 až 9.

Další tři kategorie jsou již profesionálové a manažerské pozice, nižší, střední a top management. První z těchto tří kategorií jsou mladší profesionálové a nižší management (Supervisors / Junior Professionals) s přiřazenými úrovněmi 10 až 14. Tyto pozice jsou obsazeny vedoucími určitých lokálních sekcí, jako např. kontrola. Jde o přímé nadřízené zaměstnanců, odpovídají za přidělování úkolů

zaměstnancům z předchozí kategorie, průběžně kontrolují jejich práci a získávají od nich informace. Měli by přicházet s iniciativními návrhy a předávat je středním manažerům. Úkolem je sladit zájmy a potřeby zaměstnanců vyplývajících z jejich pracovních činností se zájmy podniku jako celku. Vedle lidských a komunikačních dovedností jsou u nich důležité technické znalosti a dovednosti.

Funkce středního managementu (Seasoned Professionals / Middle Management) bývá vytvořena u větších podniků. Jsou to např. vedoucí závodů, vedoucí odborných úseků apod. Jejich úkolem je být prostředníkem mezi top manažery a základním managementem. Sbírat, třídit, zpracovávat a předávat informace od nižších článků řízení vyššímu a transformovat informace vyšších článků pro potřeby nižších. Střední manažeři samostatně rozhodují o metodách plnění úkolů. Musí disponovat dobrými sociálními a komunikačními dovednostmi. Mezi jejich činnosti patří péče o design, tvorba cen, propagace, reklama, podpora prodeje, distribuční cesty atd. Jsou zařazovány do úrovně 15 až 19.

Vrcholový management (Executive / Senior Management). Top manažeři přebírají místo vlastníků odpovědnost za úspěšný rozvoj podniku. Tvoří koncepce a koordinují všechny zájmy. Rozhodují o možných rozvojových záměrech, jako zásadní investice, zavedení nových produktů apod. Jejich úrovně jsou 20 až 24. Generální ředitel velké společnosti (Chief Executive Officer) je na úrovni 24. Na úrovni 20 je např. nezkušený ředitel nějakého z hlavních oblastí podniku. Jsou to ředitelé techničtí, výrobní, obchodní, personální atd.

Position	Please choose employee reference level according to his/her position.	07 - Clerical/Operations
Job name	Please choose employees job family.	08 - Clerical/Operations
Salary	Please fill out basic annual salary.	09 - Clerical/Operations
		10 - Supervisor/Junior Professionals
		11 - Supervisor/Junior Professionals
		12 - Supervisor/Junior Professionals
		13 - Supervisor/Junior Professionals
		14 - Supervisor/Junior Professionals
		15 - Seasoned Professionals/Middle

Obrázek 7: Výběr pozice.

Pokud by se jednalo o menší společnost, která nedisponuje velkým počtem zaměstnanců ani obsáhlým managementem, tedy pokud by se jednalo např. o podnik čítající 50 dělníků, 4 mistry, 2 vedoucí směny a jednoho ředitele, kde mistři i vedoucí směny jsou na stejné úrovni, tak by pro ně klient zvolil úroveň prostřední. Mistři by tedy byli zařazeni do kategorie Supervisor / Junior Professionals na úroveň 12, vedoucí směny do kategorie Seasoned Professionals / Middle Management na úroveň 17 a jediný ředitel společnosti pak na úroveň 21

kategorie Executive / Senior Management. Ředitel společnosti není zařazen do střední úrovně nejvyšší kategorie, protože velikost firmy tomu ještě neodpovídá.

Pro ještě lepší představu je v následující tabulce předvedena ukázka pozic příslušných k jednotlivým referenčním úrovním. Většina těchto ukázek je z rodiny pozic FA – Finance and accounting. Tato rodina pracovních pozic je v rozsahu pozičních úrovní 8 až 22, zbylé úrovně jsou potom prezentovány pozicemi z jiných oblastí výkonů zaměstnanců. Ukázky rodin pracovních oblastí jsou popsány v tabulce 2.

Poziční úroveň	Poziční kategorie	Rodina oblastí výkonů	Pracovní pozice a její popis
5	Clerical	XX	Uklízečka
6	Clerical	AS	Obsluha kopírovacího centra
7	Clerical	LS	Řidič
8	Clerical	FA	Sekretářka účetního
9	Clerical	FA	Sekretářka ředitele
10	Junior professionals	FA	Asistent finančního oddělení
11	Junior professionals	FA	Mzdový účetní
12	Junior professionals	FA	Daňový účetní
13	Junior professionals	FA	Finanční účetní
14	Junior professionals	FA	Finanční kontrolér
15	Seasoned professionals	FA	Finanční poradce
16	Seasoned professionals	FA	Vedoucí finanční kontroly
17	Seasoned professionals	FA	Vedoucí účetního oddělení
18	Seasoned professionals	FA	Finanční manažer (malá firma)
19	Seasoned professionals	FA	Finanční manažer (velká firma)
20	Executive	FA	CFO (malá firma)
21	Executive	FA	CFO (střední firma)
22	Executive	FA	CFO (velká firma)
23	Executive	EM	Vedoucí divize
24	Executive	EM	CEO

Tabulka 1: Ukázka přiřazení pracovních pozic k referenčním úrovním.

Rodiny oblastí výkonů zaměstnance, název pracovní pozice (Job name): Zde klient vybere název pozice, pod kterou je zaměstnanec veden. V následující tabulce je seznam námi zvoleného rozdělení pracovních pozic. Ten lze kdykoliv upravit či přidat nové rodiny názvů pozic. Jedná se o pouhou ukázkou, počet oblastí pracovních výkonů bude v každé zemi pravděpodobně jiný. Lze vytvořit nové pozice a nové rodiny, popř. odstranit nepotřebné.

Zkratka	Anglický název oddělení	Český název oddělení
AC	Actuarial	Pojistně-matematické
AE	Asset management	Správní
AS	Administration / Support	Administrativní / Podpora
CA	Corporate affairs	Podnikové záležitosti
CC	Call center	Centrum telefonních hovorů
CD	Credit / Credit cards	Kredit / Kreditní karty
CJ	Capital project development	Vývoj investičních projektů
CL	Claims	Nároky
CR	Clinical research	Klinický výzkum
CS	Customer service	Zákaznické služby
DE	Design	Design
EG	Engineering	Inženýrské
EM	Executive / Management	Manažerské
EN	Energy market and trading	Trh a obchodování s energiemi
FA	Finance and accounting	Finance a účetnictví
HI	Health insurance	Zdravotní pojištění
HR	Human resources	Lidské zdroje
HS	Hotel service	Hotelový servis
IT	Information technology	Informační technologie
LG	Legal	Právní
LS	Logistics / Supply Chain	Logistické / Dodavatelský řetězec
MA	Marine operations	Mořské operace
MD	Medical	Lékařské
ME	Merchandise operations	Operace se zbožím
MK	Marketing	Marketing
PC	Policy service	Politické
PD	Product development	Vývoj produktů
PE	Petroleum engineering	Ropné inženýrství
PH	Pharmaceutical research	Farmaceutický výzkum
PR	Production	Produkce
QA	Quality assurance	Zajištění kvality
RD	Research and development	Výzkum a vývoj
RO	Retail operations	Maloobchodní operace
SW	Software development	Vývoj softwaru
SL	Sales	Prodej, odbyt
SP	Strategic planning	Strategické plánování
TM	Technical maintenance	Technická údržba
XX	Other	Jiné, ostatní oddělení

Tabulka 2: Rodiny pracovních pozic.

Pokud klient není schopen zaměstnance zařadit, zvolí poslední možnost XX – Other. Jak jsme již zmínily, na základě komunikace s klienty lze v průběhu sběru dat doplnit další významná oddělení.

Job name	Please choose employees job family.	<input type="text"/> QA - Quality Assurance RD - Research and Development RO - Retail Operations SW - Software Development SL - Sales SP - Strategic Planning TM - Technical Maintenance XX - Other
Salary	Please fill out basic annual salary.	
Allowances	Is the employee eligible? If yes, please fill out sum of annual allowances (meal, mobile, shift, transportation, temperature allowance etc.)	

Obrázek 8: Název pracovní pozice.

Výplata, mzda (Basic salary): Zde se vyplní základ roční mzdy zaměstnance. Po domluvě s klientem se lze domluvit i na vyplňování měsíčních výplat, které se posléze vynásobí dvanácti.

Příspěvky (Allowances): To je souhrn všech ročních přídavků pro zaměstnance. Zde se sčítají přídavky na jídlo, transport do zaměstnání, mobilní telefon, přídavky za noční směny, za teplotní podmínky a další nejrůznější fixní příspěvky.

Bonusy (Bonuses): Jedná se o variabilní část mzdy. Vyplní se zde roční úhrn těchto bonusů. Patří mezi ně motivační bonusy, provize z prodeje, podíl na celkových tržbách apod.

Takto vyplněný dotazník se uloží pomocí tlačítka *Save* (Obrázek 5). Pokud by se ve společnosti nacházelo více zaměstnanců, kteří pracují za stejných podmínek (většinou se jedná o zaměstnance první kategorie), tedy mají stejnou roční výplatu, stejné přídavky a většinou žádné bonusy, lze použít tlačítko *Multiple save*. Po zmáčknutí tohoto tlačítka zadá klient počet zaměstnanců se stejnými podmínkami.

The image shows a Windows-style dialog box titled "Employee count". Inside the dialog, there is a text prompt: "Please fill out count of employees with same conditions." Below this prompt is a text input field containing the number "15". At the bottom of the dialog is a button labeled "Save".

Obrázek 9: Počet zaměstnanců se stejnými podmínkami.

Data se postupně ukládají do požadované formy na *List 2*. Jsou ve standardním formátu *MS Excel*. Jsou připraveny na čištění a následnou práci s nimi.

	A	B	C	D	E	F	G	H	I	J
1	COMPANY_NAME	CURRENCY	EMPLOYEE_ID	YEARS_IN_SERVICE	POSITION_LEVEL	JOB_NAME	SALARY	ALLOWANCES	BONUSES	EFFECT
2	Company name	EUR	4	2009	05 - Clerical/Operations	AE - Asset Management	15000	100		15
3	Company name	EUR	3	2001	12 - Supervisor/Junior Professionals	CD - Credit/Credit Card	28000	500	150	
4	Company name	EUR	2	2003	08 - Clerical/Operations	CA - Corporate Affairs	18500	200	0	
5	Company name	EUR	1	2009	07 - Clerical/Operations	CD - Credit/Credit Card	17250	200		

Obrázek 10: Ukázka uložených dat.

Pokud se jedná o již zkušeného klienta, tak lze data uložit přímo na *List 2* a není nezbytně nutné používat dotazník. Tito klienti jsou již seznámeni se všemi možnými variantami a vyplňovali dotazník již v předešlých letech.

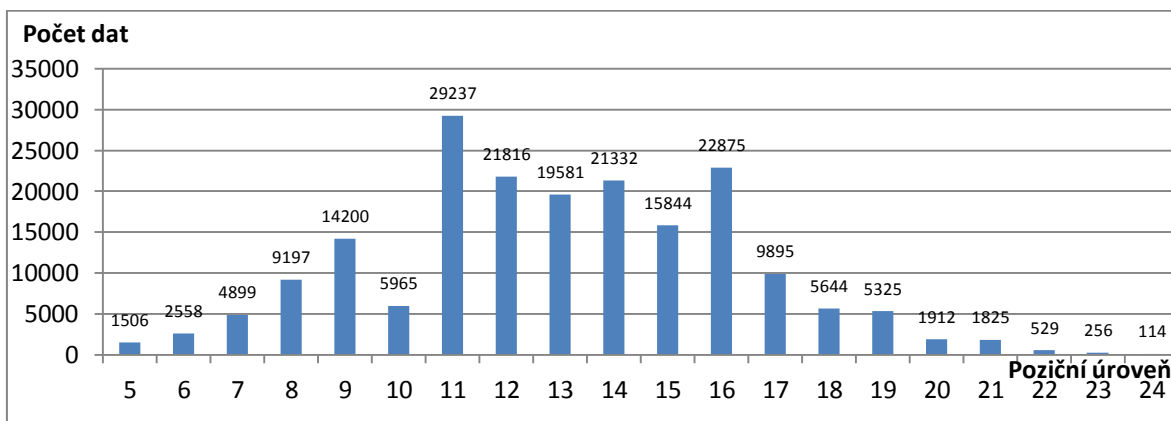
Tím je dotazník uzavřen a končí sběr dat u dané společnosti. Data se vyberou od co možná největšího počtu podniků. Všechny soubory s daty jsou následně podrobeny jejich očištění (úprava na požadovanou formu pro databázový systém, převod na jednotnou měnu apod.) a statistické analýze. Postup je ukázán v následující kapitole. [dotazník.xlsm]

2.2 Zpracování dat a statistická analýza

Mějme vybraná data za dané období. Většinou tím máme na mysli jeden kalendářní rok, resp. data by neměla být starší než jeden rok. Můžeme tedy uvažovat i kratší období, např. šest měsíců. Jelikož sběr dat je ale časově náročnou záležitostí, uvažujeme zpravidla datové období délky právě jednoho roku, pro jednoduchost to bude jeden kalendářní rok. Laicky řečeno, osoby odpovědné za sběr dat se v průběhu roku starají o to, aby klienti byli v úplném rozsahu seznámeni s našimi analýzami, postupy a produkty. Poslední měsíc v roce, když už je většinou obecně známo, jakou výši ocenění zaměstnanci obdrží, se vyplní dotazník. Předpokládáme tedy, že všechna data jsou sebrána k poslednímu dni v roce. Nyní přichází na řadu již několikrát zmíněné „čištění dat“.

2.2.1 Data a jejich předzpracování

Ve studii máme k dispozici reprezentativní výběr z populace reálného ocenění zaměstnanců. Naší analýzy se zúčastnilo 194 510 zaměstnanců z 341 společností.



Obrázek 11: Počty dat na pozičních úrovních.

Nástroje, které jsou v příloze této práce, umožňují provádět analýzu trhu obecně na jakákoliv data. Jako ukázkou uvažujeme výše zmíněné počty dat. Na jejich základě a aplikace statistických metod sestrojíme později model trhu.

Nyní jsme ve fázi, kdy jsou všechna data již v rukou analytiků. Ti musí data projít a zjistit, jestli v nich nejsou nějaké nechtěné chyby nebo překlepy. K čištění dat analytikovi poslouží další nástroj *Cleaning.xlsm*. Toto makro je uzpůsobeno tak, aby konvertovalo data do požadovaného formátu, který je nutný pro databázový systém. Dále zkontroluje duplicity a označí hodnoty podezřelé jako extrémní. Tyto hodnoty se pak zkoumají individuálně analytikem.

Prvním krokem je, si soubor *Cleaning.xlsm* zkopírovat do složky „XLSTART“, kde se nachází nainstalovaný software *MS Excel* (dále už jen Excel) a na pás karet si pro jednodušší manipulaci přidat makro *Company_ID* (obsaženo již v *dotazník.xlsm*). Máme k dispozici vyplněný dotazník. Makro *Company_ID* spustíme jako první. Otevře se nám formulář k vyplnění identifikačního kódu, který si určujeme sami (unikátní, primární klíč). Analytik vyplní ID společnosti a zmáčkne na tlačítko „Done“.



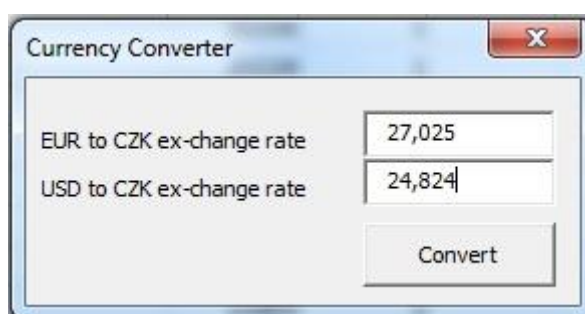
Obrázek 12: Vyplnění ID společnosti.

Po vyplnění kódu společnosti se data ze souboru *dotazník.xlsm* přesunou na nově vytvořený sešit v Excelu „*Sešit1*“. Nyní přichází na řadu již zmíněné makro *Cleaning.xlsm*. Tento nástroj je rozdělen na několik kroků. Obrázek 12 nastíní pohled na strukturu tohoto makra.



Obrázek 13: Cleaning form.

První tlačítko *Currency converter* je určené k převodu měny na měnu jednotnou pro databázi. Tou je CZK. V naší studii máme v dotazníku předurčeny tři měny. Pokud by se v dotaznících objevily, lze je převést vyplněním daného kurzu platnému k datu sběru.



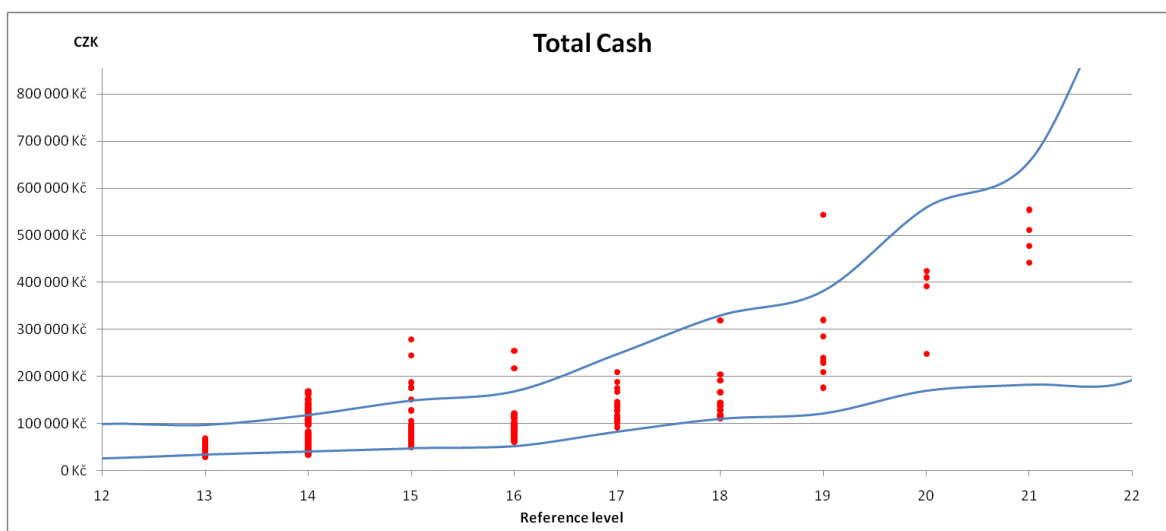
Obrázek 14: Převod měny.

Záznamy jsou převedeny na jednotnou měnu a na řadu přichází dalších pět kroků čistícího makra. Po kliknutí na druhé tlačítko *Formatting* se textová pole převedou na základní tvar (05 - Clerical/Operation na číslo 5, AE - Asset Management na AE), zkontrolují se duplicity ve jménu společnosti (duplicity v ID společnosti se kontrolovat nemusí, zadáváme je sami, tudíž v jednom vyplněném dotazníku je vždy právě jedno unikátní označení firmy) a spočítají se naše tři agregáty. Duplicitní hodnoty se v případě nálezu zabarví.

Kroky *Salary*, *Allowance* a *Bonus check* kontrolují hodnoty ve sloupcích s příslušnými částkami. Z dat minulého roku máme spočítané Wilksovy toleranční meze. Pravděpodobnost β stanovujeme na 80 %. To je úmyslné, protože chceme zachytit nechtěné chyby ve sběru. Hodnoty statistik $L(x_1, x_2, \dots, x_n)$

a $U(x_1, x_2, \dots, x_n)$ jsou nastaveny tak, že nová hodnota padne s pravděpodobností zhruba 10 % pod $L(x_1, x_2, \dots, x_n)$ a s 10% pravděpodobností nad $U(x_1, x_2, \dots, x_n)$. Hodnota nového ocenění zaměstnance tedy do našich mezí padne s pravděpodobností 80 %. Pravděpodobnost tohoto výroku je běžně vyšší než 99 %. K docílení podmínky $\gamma > 99 \%$ je zapotřebí data v rozsahu alespoň 150 a více. Ukázkou aplikace teorie tolerančních mezí v Excelu a jejich automatizace bude předvedena později.

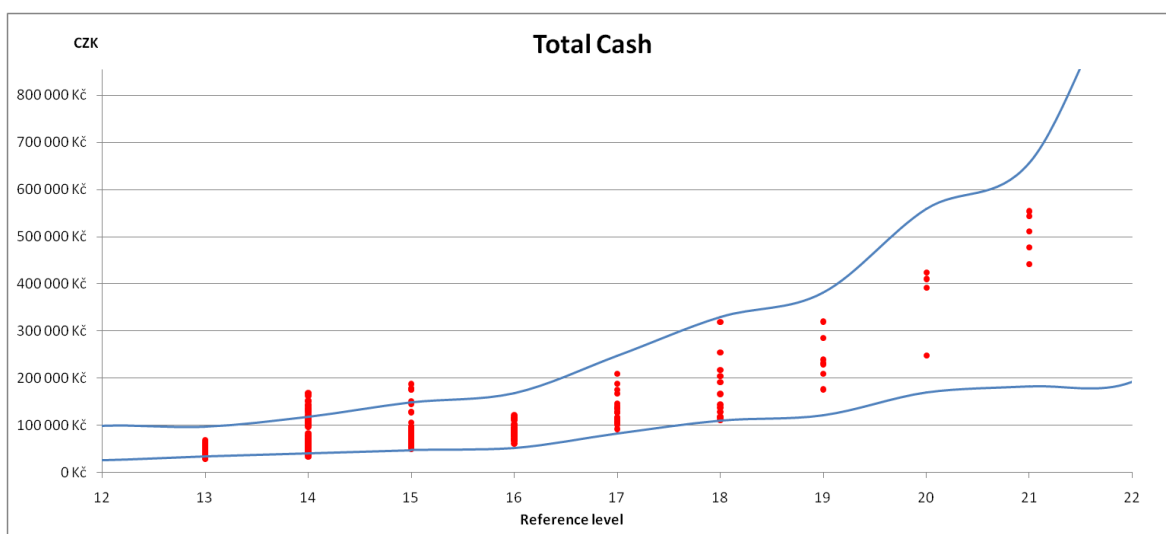
Pokud nová hodnota bude mimo tyto meze, je makrem označena jako podezřelá z extrému. Pro vizuální kontrolu se prozatím posledním tlačítkem *Charts and ranges* vytvoří grafy. Pomocí *Cleaning.xlsm* jsme soubor překontrolovali a momentálně máme v sešitě tři listy. List „DATA“, kde je základní soubor dat, která jsou připravena pro možné nahrání do databáze. Na listu „CHARTS“ jsou pak grafy s vyobrazením hodnot v porovnání s tolerančními mezemi.



Obrázek 15: Wilksovy toleranční meze a ocenění zaměstnanců.

Na obrázku je vidět, že by se měly nahlásit vysoké hodnoty na pozici 19, 16, 15 a 14. Je také vidět, že hodnoty na pozici 14 mají relativně malý rozptyl oproti ostatním úrovním, které toleranční meze překračují. Kdyby tedy hodnoty ocenění překročily toleranční meze pouze na úrovni 14, tak bychom nejspíše kontrolu korektnosti nevyžadovali. Je celkem věrohodné, že tyto hodnoty jsou opravdu správně. Zamýšlíme tak právě na základě menšího rozptylu. Jelikož ale potřebujeme určitě ověřit nejvyšší hodnotu úrovně 19, dvě vysoké hodnoty z 16 a minimálně dvě nejvyšší hodnoty z 15, tak se nahlásí i ostatní hodnoty mimo meze. Na posledním listu „PROBLEMS“ jsou již tyto problémy pro klienta vypsané. Pokud se na listu „DATA“ zabarví jakákoliv hodnota, tak se i automaticky zapíše problém.

Většinou není chyba ve špatném zadání ocenění, ale ve špatném zadání poziční úrovně zaměstnance. Vidíme, že pokud je výše ocenění (mimo meze) na pozici 19 správně, měla by se přehodnotit poziční úroveň tohoto zaměstnance. Dále by se mělo prozkoumat, jestli zaměstnanec není z jiné země (např. ze země, kde se nachází centrála firmy). Ocenění zaměstnanců v této zemi mohou být obecně vyšší. Může se také jednat o velmi kvalifikovaného experta nebo prodejce, kterému se povedlo tento rok prodat opravdu velké množství produktů, ze kterých má provizi. Možností je mnoho. Zpravidla se takto vysoká hodnota opraví, ale jsou i situace, kdy tato hodnota bude naprosto korektní. Po opravě by data měla vypadat následovně.



Obrázek 16: Opravené hodnoty.

Původní vysoké hodnoty z úrovně 16 a 19 byly správně, ale opravila se jejich úroveň. Posunuly se o dvě úrovně výše. Dvě nejvyšší hodnoty na úrovni 15 byly naopak mylně vyplněny. Jejich hodnota klesla o sto tisíc Kč. Nyní považujeme data za korektní. Rozptyly dat jednotlivých úrovní jsou již menší a až na výjimky jsou všechny hodnoty uvnitř našich tolerančních mezí.

Tento proces opakujeme se všemi 341 dotazníky. Někdy je analytik schopen problémy opravit sám. Je např. na první pohled již zřejmé, že hodnoty jsou o řád vyšší, o řád nižší, jsou v jiné měně nebo místo ročních hodnot jsou vyplněny měsíční ocenění. Po potvrzení korektnosti u všech firem, od kterých jsme schopni data sebrat, jsou data připravena k nahrání do databáze.

2.2.2 Statistická analýza

V této kapitole si předvedeme, jak použít teoretický základ na naše data a ukážeme si, jak aplikovat statistické metody v Excelu. Výpočty byly pro účely práce automatizovány v nástroji *udelej_meze.xlsm*. Funkce tohoto softwarového

prostředí, vzorce a buňky jsou níže popsány modře (vše zabarvené modře se vztahuje k souboru *udelej_meze.xlsm*). I když název může zavádět, tento nástroj neslouží pouze k vypočtení tolerančních mezí. Po vložení jakýchkoliv dat do B2 na List1 a stisknutím tlačítka *start* se spočtou všechny námi požadované statistiky (tlačítko *clear* pak slouží k vymazání starých výpočtů). Vyčíslí se základní charakteristiky polohy, charakteristiky variability, sestrojí se intervaly spolehlivosti pro kvantily a vypočítají se Wilksovy toleranční meze. Nyní si vymežíme prostor k popsání jednotlivých funkcí Excelu, kterých jsme v souboru *udelej_meze.xlsm* využili.

Poznámka: Program je vytvořen pro data rozsahu maximálně 100 000. Vzorce jsou upravitelné, rozsahy lze libovolně zvětšit. Zde uvažujeme, že jedna společnost nečítá více než 100 000 zaměstnanců.

Základní charakteristiky souboru dat:

- Počet dat n : $K2=POČET(B2:B100001)$.
- Aritmetický průměr \bar{x} : $K3=PRŮMĚR(B2:B100001)$.
- Minimum $x_{(1)}$: $K4=QURTIL(B2:100001;0)$.
- Dolní kvartil $x_{0,25}$: $K5=QURTIL(B2:100001;1)$.
- Medián $x_{0,5}$: $K6=QURTIL(B2:100001;2)$.
- Horní kvartil $x_{0,75}$: $K7=QURTIL(B2:100001;3)$.
- Maximum $x_{(n)}$: $K8=QURTIL(B2:100001;4)$.
- Výběrový rozptyl s^2 : $K9=VAR.VÝBĚR(B2:100001)$.
- Výběrová směrodatná odchylka s : $K10=SMODCH.VÝBĚR(B2:100001;4)$.
- Variační rozpětí R : $K11=QURTIL(B2:100001;4) - QURTIL(B2:100001;0)$.
- Kvartilové rozpětí $R_{(0,75;0,25)}$:
 $K12=QURTIL(B2:100001;3) - QURTIL(B2:100001;1)$.

Vše je v tabulce Tab. 1.

Intervaly spolehlivosti pro kvantily:

Snažíme se vytvořit model trhu pro klienta. Požadujeme, abychom s nejméně 95% pravděpodobností mohli tvrdit, že námi požadovaný kvantil je uvnitř zvolených mezí $x_{(i)}, x_{(j)}$, kde $j = n - i + 1$. Tedy oboustranný interval $x_{(i)} < x_p < x_{(j)}$, kde pravděpodobnost $1 - \alpha = P(x_{(i)} < x_p < x_{(j)}) > 95 \%$. Již v kapitole 1.6 jsme si ukázali, jak se teoreticky hodnota koeficientu spolehlivosti pomocí beta funkce $1 - I_{1-p}(n - i + 1, i) - I_p(n - i + 1, i)$ vypočítá. Nejdříve si stanovíme hodnotu p , tedy pro jak velký kvantil chceme interval spolehlivosti sestrojít. V nástroji *udelej_meze.xlsm* se tyto hodnoty nachází v tabulce Tab. 2:

- $p = 0,99$ K18
- $p = 0,95$ K19
- $p = 0,90$ K20
- $p = 0,75$ (kvartilové rozpětí) K21
- $p = 0,50$ (medián) K22
- Výpočetní hodnoty i (L25:L29)
- $1 - \alpha = P(x_{(i)} < x_p < x_{(j)})$:

$$K25 = 1 - \text{BETAINV}(1-K18;L25;K15-L25+1) - \text{BETAINV}(K18;L25;K15-L25+1)$$

Poznámka: V Excelu se musí u funkce BETAINV obrátit parametry p, q , jinak hodnoty vychází záporné.

Poslední vzorec nám vyčíslí koeficient spolehlivosti $1 - \alpha$ pro $p = 0,99$. Analogicky je tento vzorec použit pro všechna ostatní p (K26:K29). Nyní chceme stanovit hodnoty i, j takové, které nám splní podmínku $P(x_{(i)} < x_p < x_{(j)})$ námi požadované výše koeficientu spolehlivosti:

- $1 - \alpha = 99\%$ M23
- $1 - \alpha = 97,5\%$ P23
- $1 - \alpha = 95\%$ S23

Spuštěním programu se do tabulky Tab. 6 zapíše hodnoty $x_{(i)}, x_{(j)}$ splňující naše podmínky pro p a $1 - \alpha$. V následující tabulce je předvedena ukázka spočtených hodnot pro poziční úroveň 5 agregátu Total Cash.

Tab. 6									
$1 - \alpha$	99 %			97,5 %			95 %		
p	i	$x(i)$	$x(j)$	i	$x(i)$	$x(j)$	i	$x(i)$	$x(j)$
99 %	6	11 850	35 864	17	12 012	30 964	36	12 378	30 622
95 %	6	11 850	35 864	18	12 014	30 922	37	12 386	30 594
90 %	7	11 854	34 913	18	12 014	30 922	37	12 386	30 594
75 %	7	11 854	34 913	19	12 022	30 864	37	12 386	30 594
50 %	7	11 854	34 913	19	12 022	30 864	37	12 386	30 594

Tabulka 3: Ukázka vytvoření intervalu spolehlivosti.

Pro určení modelu trhu ocenění zaměstnanců pomocí intervalů spolehlivosti pracujeme s parametry $p = 90\%$ a $1 - \alpha = 97,5\%$. Klientům v našich analýzách tedy tvrdíme, že pravděpodobnost toho, že 90 % hodnot ocenění zaměstnanců na dané poziční úrovni je mezi $x_{(i)}$ a $x_{(j)}$, je alespoň 97,5 %. Pro poziční úroveň 5 agregátu Total Cash, kde máme k dispozici reprezentativní výběr rozsahu 1506, jsou tedy hodnoty mezí intervalu spolehlivosti rovny: $x_{(i)} = x_{(18)} = 12\,014$ CZK a $x_{(j)} = x_{(1489)} = 30\,922$ CZK.

Tento proces opakujeme pro všechny poziční úrovně ve všech třech agregátech. Dohromady tedy šedesátkrát. Tím získáme námi požadované modely trhu pro všechny tři agregáty. V tabulce 4 Tab. 3 jsou experimentálně napočítány nejnížší možné počty n dat, potřebné k docílení spočtení intervalů spolehlivosti pro hodnoty p od 50 % do 99 % a hodnoty $1 - \alpha$ od 95 % do 99 %. V tabulce 5 Tab. 4 potom nejnížší počty potřebné k tomu, aby tento interval neobsahoval všechny hodnoty výběru. Požadujeme, aby mimo interval spolehlivosti ležely alespoň minimum a maximum výběru.

Tab. 3			
$p / 1 - \alpha$	99 %	97,5 %	95 %
99 %	460	183	90
95 %	304	121	60
90 %	240	96	48
75 %	167	67	33
50 %	139	56	28

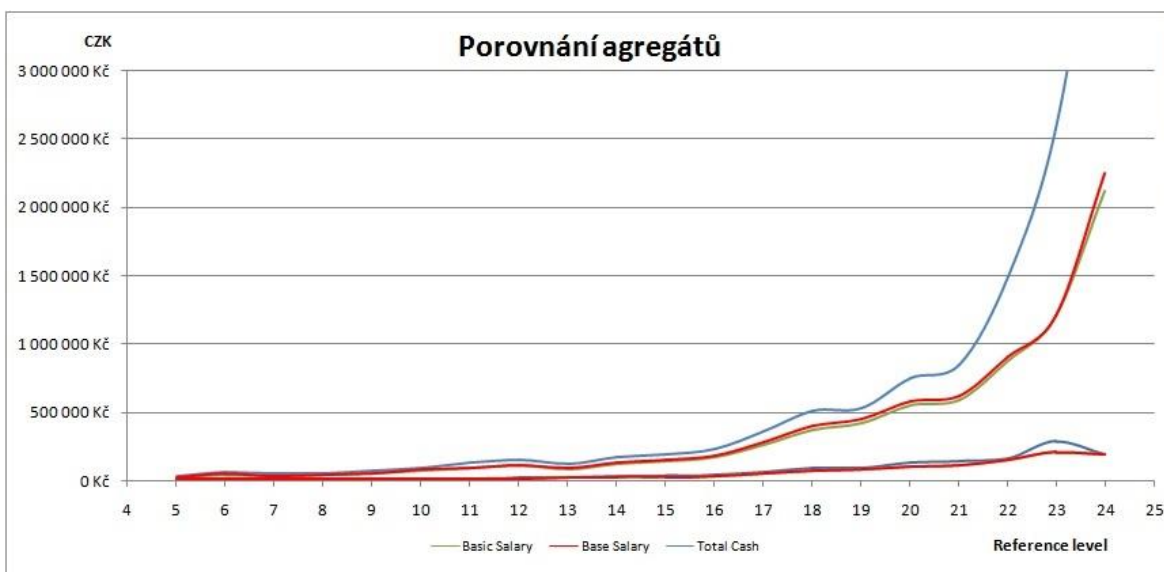
Tabulka 4: Nejnížší možné počty dat potřebné k výpočtu intervalů spolehlivosti.

Tab. 4			
$p / 1 - \alpha$	99 %	97,5 %	95 %
99 %	675	269	133
95 %	509	203	101
90 %	441	176	88
75 %	365	146	73
50 %	336	134	67

Tabulka 5: Nejnížší možné počty dat potřebné k výpočtu intervalů spolehlivosti, s podmínkou odstranění alespoň minima a maxima.

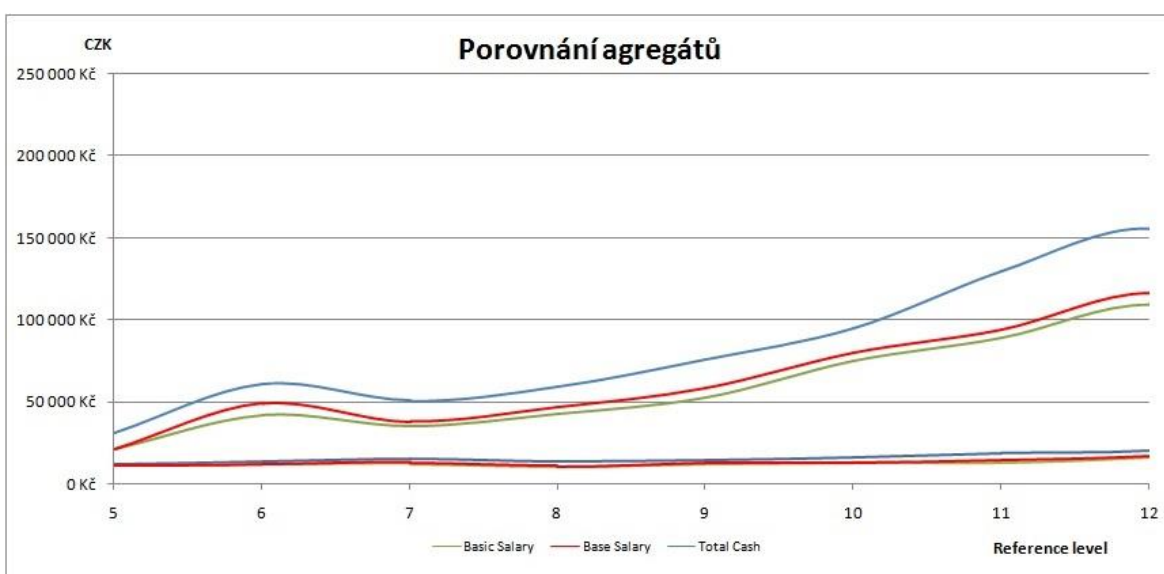
V této studii sestrojujeme model trhu na základě intervalů spolehlivosti pro 90% kvantil. Pravděpodobnost toho, že 90 % populace bude v tomto intervalu, je alespoň 97,5 %. Musíme mít tedy minimálně 96 hodnot ocenění zaměstnanců. Kdybychom požadovali, aby hodnota $i > 1$ a $j < n$, potom je nutné mít k dispozici alespoň 176 hodnot. Čím méně dat máme, tím se snižuje pravděpodobnost toho, že námi požadovaný kvantil bude mezi hodnotami $x_{(i)}$ a $x_{(j)}$.

Mějme tedy spočítány všechny intervaly pro 90% kvantil se spolehlivostí 97,5 %. Nejnížší počet dat, která vlastníme, je na úrovni 24 roven 114, což splňuje podmínky výpočtu. Na obrázku 16 je vidět sestrojený model trhu a porovnání mezi jednotlivými agregáty.

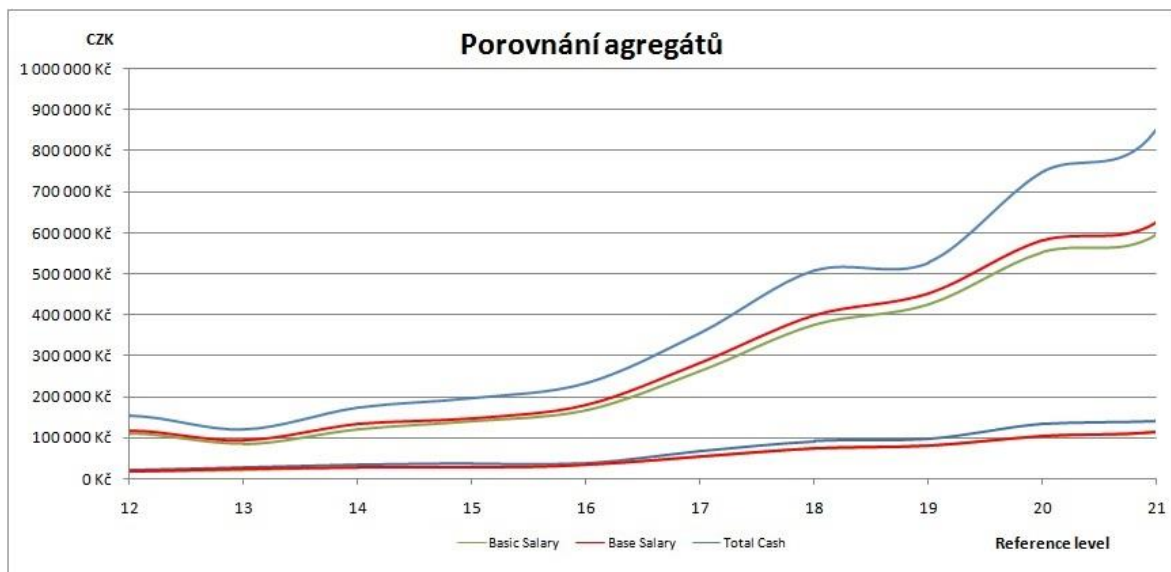


Obrázek 17: Porovnání agregátů.

Takové zobrazení nám dává informaci pouze o přibližném tvaru modelu. Zde vidíme celý model trhu, ale nemáme moc dobré informace o spodních hodnotách. Na dalších grafech je vidět spodní a střední část modelu.

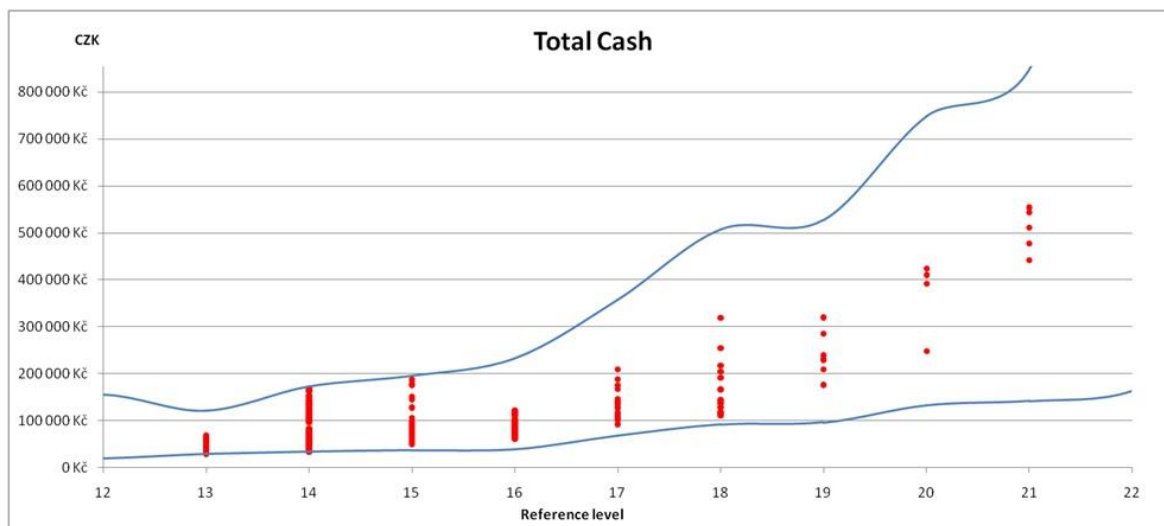


Obrázek 18: Poziční úrovně 5 až 12.



Obrázek 19: Poziční úrovně 12 až 21.

Intervaly spolehlivosti pro 90% kvantil trhu ocenění zaměstnanců slouží k vyhotovení zpětné vazby klientovi. Tomu poradenská firma za poskytnutí dat přislíbila analýzu všeobecného trhu. Použijeme k tomu právě tyto intervaly. Uvažujeme, že máme všech 341 dotazníků zkontrolovaných a data z nich jsou nahrána v databázi. Rozdělíme je do pozičních úrovní a následně je vložíme postupně do souboru *udelej_meze.xlsm*. Spočítané statistiky $x_{(i)}$ a $x_{(j)}$ slouží jako dolní a horní meze modelu trhu, který poskytujeme klientovi. Abychom vytvořili základní náhled ocenění zaměstnanců na trhu pro jednotlivé společnosti, musíme do složky „XLSTART“ zkopírovat soubor *Report.xlsm*. Meze intervalu spolehlivosti, které jsme spočítali pomocí nástroje *udelej_meze.xlsm*, vložíme do listu „GRAF“ souboru *Report.xlsm*. Spustíme poslední krok makra *Cleaning.xlsm Charts for report* (Obrázek 12). Tím se nám data poskytnutá od dané společnosti zobrazí v modelu našeho trhu. Hodnoty z obrázku 15, které jsou použity pro příklad zpětné vazby klientovi, vypadají následovně.



Obrázek 20: Intervaly spolehlivosti pro 90% kvantil populace ocenění zaměstnanců.

Poznámka: Data odpovídají firmě č. 9, která bude uvedena později v 2.3.2.

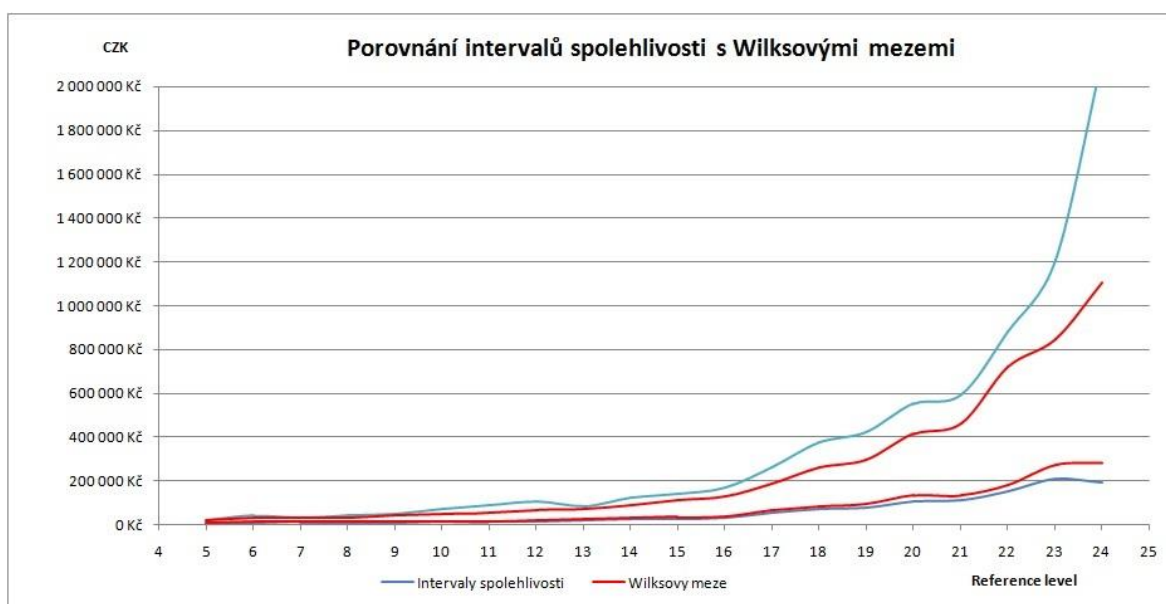
Nástroj *udelej_meze.xlsx* dále obsahuje výpočet pro určení Wilksových tolerančních mezí. V kapitole 1.7 jsme odvodili, jak mají vypadat parametry distribuční funkce beta rozdělení, která je nezbytná k docílení výpočtu statistik $L(x_1, x_2, \dots, x_n)$ a $U(x_1, x_2, \dots, x_n)$. Parametr $p = n - 2k + 1$ a $q = 2k$. Abychom vypočítali hodnotu statistik L, U , musíme nejdříve splnit předpoklady této metody.

Wilksovy toleranční meze vyžadují, aby pořádkové statistiky byly ostře rostoucí a aby pocházely ze spojitě náhodné veličiny. Ačkoliv ocenění zaměstnanců mají diskrétní charakter, uvažujeme, že pochází ze spojitě populace. Zaměstnanci tedy mohou reálně dostat nezaokrouhlené částky jakékoliv hodnoty (v nějakém rozumném smyslu). Jelikož se některé výplaty shodují s ostatními, musíme je podrobit umělému zašumění z rovnoměrného rozdělení $R(0,25; 0,25)$. Uvažujeme tedy, že hodnoty, které máme k dispozici, jsou zaokrouhlené.

V nástroji *udelej_meze* se nejdříve ve sloupci **A** vyplní hodnoty k (od 1 po n). Poté se náhodný výběr srovná do pořádkových statistik $x_{(i)}$ ve sloupci **C**. Jestliže se hodnota $x_{(i-1)} = x_{(i)}$ tak se hodnota $x_{(i)}$ uměle zašumí (sloupec **D**). Ve sloupci **E** jsou pak vyčísleny hodnoty empirické distribuční funkce. Představuje transformovaný náhodný výběr popsáný v kapitole 1.3. Hodnoty jsou již v souladu s předpoklady Wilksových mezí (pochází z rovnoměrného rozdělení na intervalu $(0,1)$ a jsou díky zašumění ostře rostoucí). Jsme tedy schopni vyčíslit hodnoty distribuční funkce beta rozdělení. Toho v Excelu docílíme funkcí **BETADIST**, konkrétně ve sloupci **F**: $F2=1-BETADIST(E2;2*A2;K2-2*A2+1)$, postupně ve všech řádcích až do poloviny rozsahu výběru. Zde záleží na tom, jakou požadujeme velikost hodnoty γ . Rozumně zvolenou hodnotou je 95 % a vyšší. Bývá pravidlem,

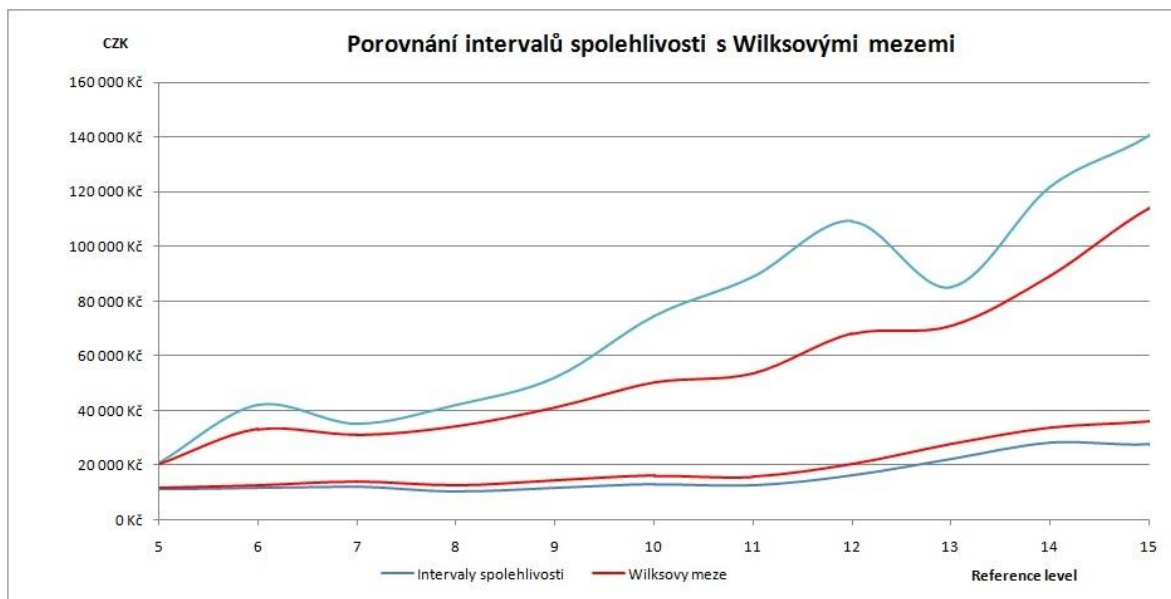
že při vhodném počtu dat (opět zhruba nad 150 hodnot), je pravděpodobnost správnosti výroku $\gamma > 99\%$ pro $k = 7$ a vyšší.

Wilksovy meze mají jiný charakter než intervaly spolehlivosti pro kvantily. U Wilksových tolerančních mezí vyslovujeme výrok o tom, že budoucí hodnota sběru dat v příštím roce bude mezi statistikami L, U s $100\beta\%$ pravděpodobností. Pravděpodobnost toho, že tento výrok platí, požadujeme alespoň $\gamma > 99\%$. Při zvyšujícím se počtu dat se nám zvyšuje také možnost si zvolit „užší“ interval mezi L a U . V analýze uvažujeme velikost pravděpodobnosti, že hodnota padne mezi naše toleranční meze, $\beta = 80\%$ U13. K docílení toho stačí mít data rozsahu alespoň již zmíněných 150 hodnot. Pro úroveň 24, kde máme pouze 114 pozorování, je $\gamma < 99\%$, přesto ale vyšší než 95%. Nově spočítané hodnoty tolerančních mezí vložíme do souboru *Cleaning.xlsm* na list „GRAF“. Tím máme čistící makro připraveno na sběr dat v příštím roce. Na následujícím obrázku je vidět porovnání intervalů spolehlivosti pro 90% kvantil, které slouží jako model trhu pro klienta a Wilksových tolerančních mezí pro sběr dat v příštím roce.



Obrázek 21: Porovnání modelu trhu s Wilkovými mezemi.

Vidíme, že Wilksovy meze jsou obecně „užší“ (díky volbě $\beta = 80\%$) než intervaly spolehlivosti. To je cílené. Požadujeme kontrolu korektnosti u více hodnot sběru příštího období, než kdybychom použili model trhu z minulého roku. Dlouhodobě tím docílíme toho, že rozptyly našich náhodných výběrů budou menší a model trhu v příštím roce bude kvalitnější. Pro nižší poziční úrovně vypadá porovnání následovně.



Obrázek 22: Porovnání intervalů spolehlivosti s Wilkovými mezemi pro úrovně 5 až 15.

Vidíme, že Wilkovy meze mají potenciál náš model trhu v budoucnosti „vyhlazovat“. Například horní mez $x_{(j)}$ intervalu spolehlivosti pro poziční úroveň 13 vyšla na základě zkoumaných dat nižší než horní mez $x_{(j)}$ intervalu spolehlivosti pro poziční úroveň 12. Tomu chceme v budoucnosti zabránit. Model trhu by měl obecně pro úrovně růst.

Tímto ukončujeme statistickou část. Metody k výpočtu máme automatizované. Nyní musíme vytvořit zpětnou vazbu pro klienty, kteří nám poskytli data.

2.3 Základní analýza pro klienta

V této chvíli jsou všechna data již zpracovaná, nahraná v databázovém systému, máme spočítané intervaly spolehlivosti pro kvantily, které považujeme za model trhu ocenění a máme i vypočítané Wilkovy toleranční meze ke sběru dat v příštím roce. Nyní přichází čas na to, abychom společně poskytli základní informace, jak si na trhu jejich zaměstnanci stojí. Uvede se, kolik zaměstnanců z kolika společností se analýzy zúčastnilo a poté firmě ukážeme, jak si její zaměstnanci stojí na trhu ocenění.

Jako příklad použijeme dvě společnosti (pro jednoduchost jsou všechny společnosti pojmenovány čísly od 1 do 341), data první z nich (firma č. 9) jsme používali již v předchozích kapitolách jako názornou ukázkou. Hodnoty ocenění této společnosti jsou nechybové, mezi mezemi trhu, a vhodně zařazené do příslušných pozičních úrovní. Druhé společnosti (firma č. 117) naznačíme kromě základních informací i doporučení, jak by měla své zaměstnance oceňovat, protože její zaměstnanci přesahují meze modelu trhu. V následující kapitole předvedeme, jak

by zpětná vazba měla vypadat. Celou ukázkou reportu ukážeme na druhé příkladové společnosti, která potvrdila korektnost všech informací. Pro první společnost bude ukázána jen zkrácená verze. Ukázka toho, jak si firma č. 9 stojí na trhu agregátu Total Cash, byla již ukázána na obrázku 20.

2.3.1 Ukázka protokolu pro firmu č. 117

Kromě úvodu, který je v rukou poradenské firmy uvedeme, že se průzkumu zúčastnilo 194 510 zaměstnanců z 341 společností. Klientova firma „117“ přispěla informacemi o 60 zaměstnancích, kteří spadají do 10 pozičních úrovní.

Poznámka: Na tomto místě by v reportu byl stejný graf jako na obrázku 11.

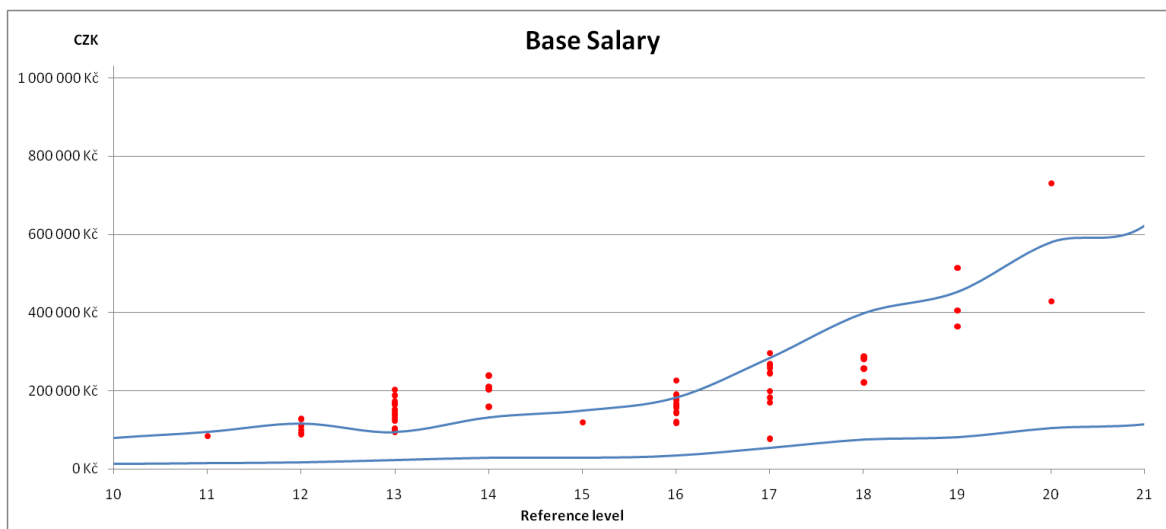
Data jsou zpracována do tří agregátů, *Basic Salary*, *Base Salary* a *Total Cash*. *Basic Salary* vyjadřuje model trhu základního ocenění zaměstnanců. Jsou zde zaneseny pouze hodnoty základu výplaty zaměstnance. Agregát *Base Salary* obsahuje ocenění zaměstnanců včetně příspěvků na jídlo, transport do zaměstnání, pracovní oblečení a všechny ostatní fixní přídatky na zaměstnance. V posledním modelu trhu jsou již zanesena celková ocenění včetně bonusové variabilní části. Na následujících obrázcích je vidět, jak si zaměstnanci firmy stojí na jednotlivých trzích. Meze trhu představují interval pro 90% kvantil se spolehlivostí 97,5 %. Pro vysvětlení, mezi horní a spodní mezí trhu leží 90 % všech ocenění (nejen těch, které pochází od účastníků šetření, ale celé populace ocenění zaměstnanců) s pravděpodobností 97,5 %.



Obrázek 23: Basic Salary společnosti č. 117.

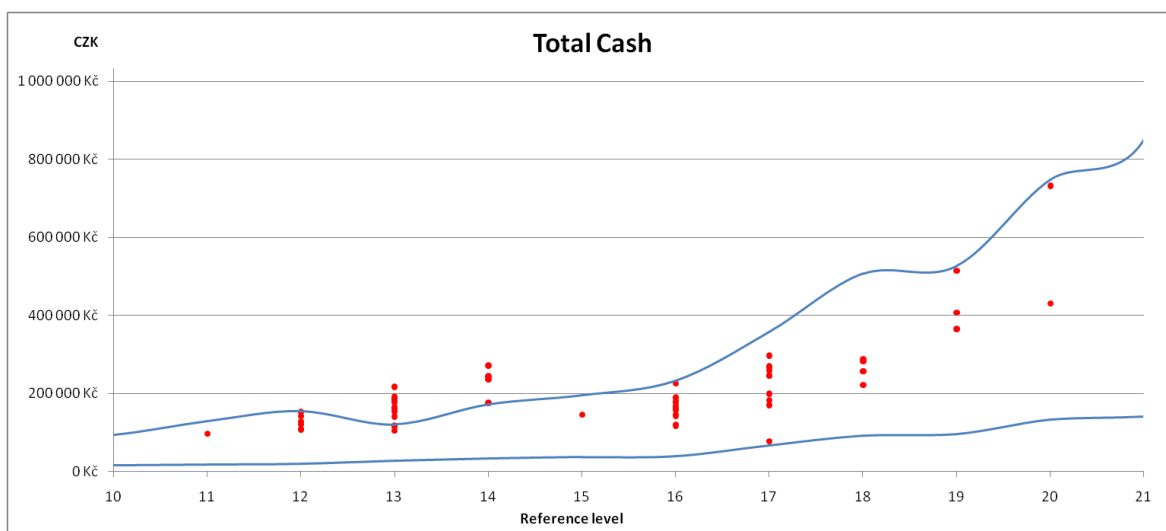
Na trhu základní výplaty pro zaměstnance si firma stojí velmi dobře. Ani jedna z hodnot není pod trhem a navíc je několik hodnot nad horní mezí trhu. Tito zaměstnanci jsou tedy oceněni lépe než minimálně 90 % všech zaměstnanců. Jediné, co bychom mohli vytknout, je velmi nízké ocenění zaměstnance na úrovni

17 a poté naopak až příliš vysoké ocenění zaměstnance na úrovni 19 ve srovnání se zaměstnancem na úrovni 20. Ovšem tento agregát nám dává pouze náhled na základ mzdy. Pevné příspěvky a variabilní bonusy mohou vše změnit. Na obrázcích 24 a 25 je vidět, jak si Vaši zaměstnanci stojí na trhu *Base Salary* a *Total Cash*.



Obrázek 24: Base Salary společnosti č. 117.

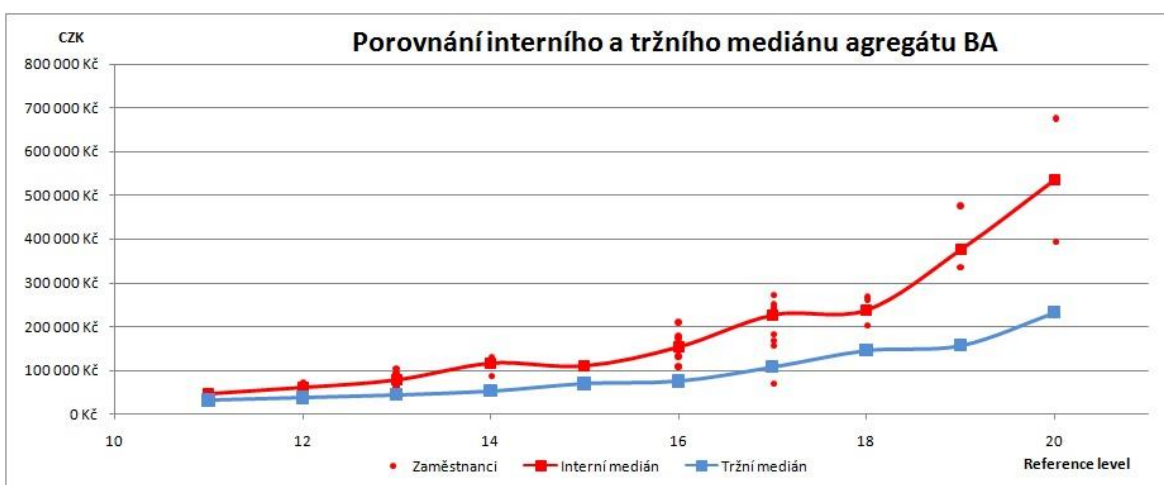
Fixní příspěvky zvedly o hodně hodnoty ocenění zaměstnanců na nižších úrovních. Pokud bychom brali v potaz pouze tento trh, tak bychom doporučili lépe přerozdělit fixní přídatky. Nejlépe placení zaměstnanci na pozici 13 jsou téměř tak dobře ocenění jako nejhůře ocenění zaměstnanci na pozici 18. Fixní část mzdy by se měla se zvyšující se poziciční úrovní úměrně zvětšovat. Nyní se podíváme, jak vypadá celkové ocenění zaměstnanců včetně variabilní složky.



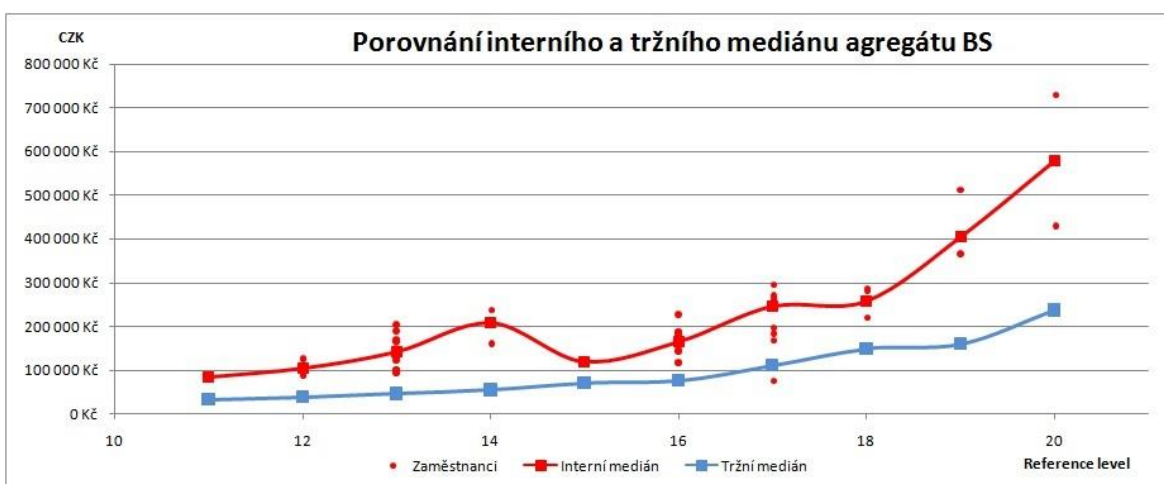
Obrázek 25: Total Cash společnosti č. 117.

Toto je finální verze trhu výplat zaměstnanců. Nelze přehlédnout, že nejhůře placený zaměstnanec na pozici 17 je vlastně nejhůře oceněný zaměstnanec ve

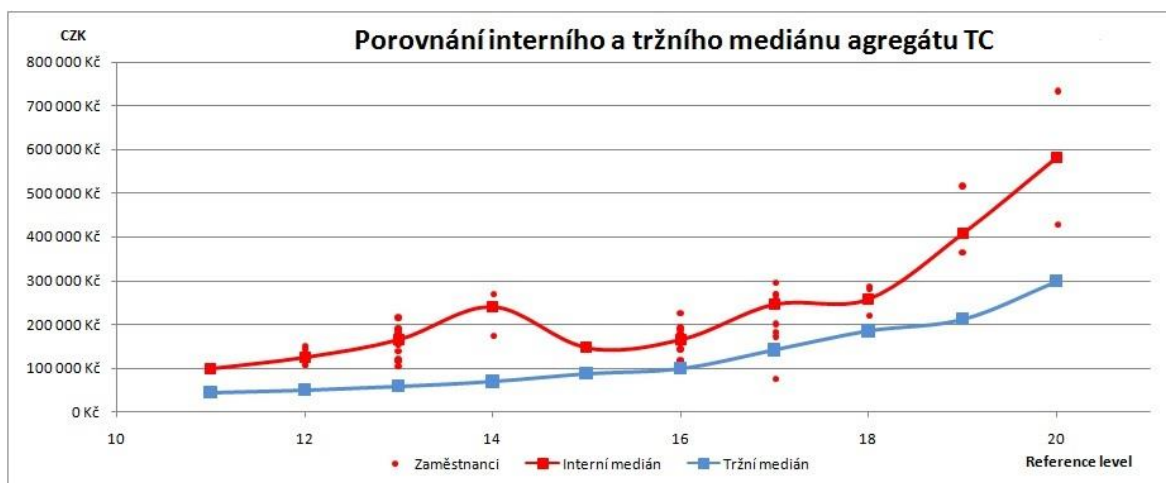
společnosti. Možná se jedná o nového zaměstnance, který dostává nejnižší možný plat a jen malé příspěvky a bonusy. Tento zaměstnanec by měl být ale v tomto případě na nižší poziční úrovni. Jestliže se opravdu jedná o pracovníka středního managementu, tak by měl buď být na pozici 15, nebo by se jeho výplata měla zvýšit o alespoň 100 000 Kč ročně. Dále je také nevhodné, že lidé na pozici 14 (nejvyšší úroveň nižšího managementu) mají stejné peníze jako pracující na úrovni 18. Nyní prozkoumáme interní medián v porovnání s tržním. Tržní výběrový medián je vypočítán na základě sebraných dat. Není to tedy obraz celé populace ocenění zaměstnanců, ale prostřední hodnota ocenění zaměstnanců z firem, které se zúčastnily průzkumu.



Obrázek 26: Porovnání mediánů pro BA.



Obrázek 27: Porovnání mediánů pro BS.



Obrázek 28: Porovnání mediánů pro TC.

Interní medián firmy je ve všech agregátech vyšší než medián tržní. To je příznakem toho, že zaměstnance oceňujete lépe než minimálně polovina ostatních firem. Musíme ale podotknout, že již zmíněný nejhůře oceněný člověk na pozici 17 je jako jediný pod hranicí tržního mediánu. Tento zaměstnanec je vůči ostatním ve firmě placen poměrně špatně. Měl by být nejspíše zařazen do jiné úrovně nebo by se jeho plat měl rapidně zvýšit. Všimáme si také, že povaha interního mediánu se v přechodu mezi BA a BS značně změnila. Fixní přídatky a variabilní část ocenění mají velký vliv na nárůst výplaty pro nižší úrovně, kdežto povaha mediánu vyšších pozic se od úrovně 15 nějak ztelně nezměnila. Tyto hodnoty prozkoumáme ještě individuálně.

Zaměstnanci		Fixní příspěvky		Variabilní bonusy	
úroveň	počet	minimum	maximum	minimum	maximum
11	1	37 150	37 150	13 078	13 078
12	6	36 610	55 600	15 117	24 171
13	15	42 000	143 980	10 281	30 651
14	4	71 360	107 590	17 087	40 467
15	1	8 925	8 925	27 795	27 795
16	13	8 818	17 182	243	746
17	10	5 812	22 107	114	1 010
18	5	16 987	21 634	701	1 000
19	3	27 838	39 286	928	1 431
20	2	32 834	56 070	1 605	1 934

Tabulka 6: Fixní přídatky a variabilní bonusy společnosti 117.

Již při nejjednodušším pohledu (variační rozpětí) na příspěvky a bonusy musíme podotknout, že nižší poziční úrovně dostávají oproti vyšším úrovním o hodně vyšší bonusy. To ovšem ihned neznamena, že jsou hodnoty špatně. Politikou firmy může být, že zaměstnanci dostávají bonusy ve formě provize z prodeje a prodejci jsou pouze do úrovně 15. Chybou je rozhodně rozdělení fixních příspěvků. Ty by se měly rozdělovat rovnoměrně nebo postupně růst s poziční úrovní.

Variabilní bonusy mohou být obecně různé. Při celkovém pohledu na společnost, doporučujeme přehodnotit rozdělování rozpočtu peněz určeného na fixní část nad rámec základní mzdy. Tedy příspěvky na jídlo, transport a jiné fixní přídatky. Ty by měly být buď pro všechny zaměstnance stejné, nebo by postupně se zvyšující se poziční úrovní měly růst. Hlavní část by měla být přidělena vyšším pozicím. Momentálně tomu tak ve Vaší firmě není. Úrovně 15 až 20 dostávají pouze necelých 25 % rozpočtu na fixní a variabilní složku mimo základ mzdy. Oproti tomu nižším úrovním 11 až 14 je přiděleno více než 75 % rozpočtu. Ve firmě máte zaměstnance na 10 různých pozičních úrovních. Doporučujeme tedy každé úrovni přidělit zhruba 10 % tohoto rozpočtu nebo nějakou váženou část s ohledem na počet zaměstnanců.

Takto vypadá základní analýza pro klienta, který nám poskytl informace o svých zaměstnancích. V následující kapitole si ukážeme pouze zkrácenou verzi analýzy trhu pro společnost č. 9.

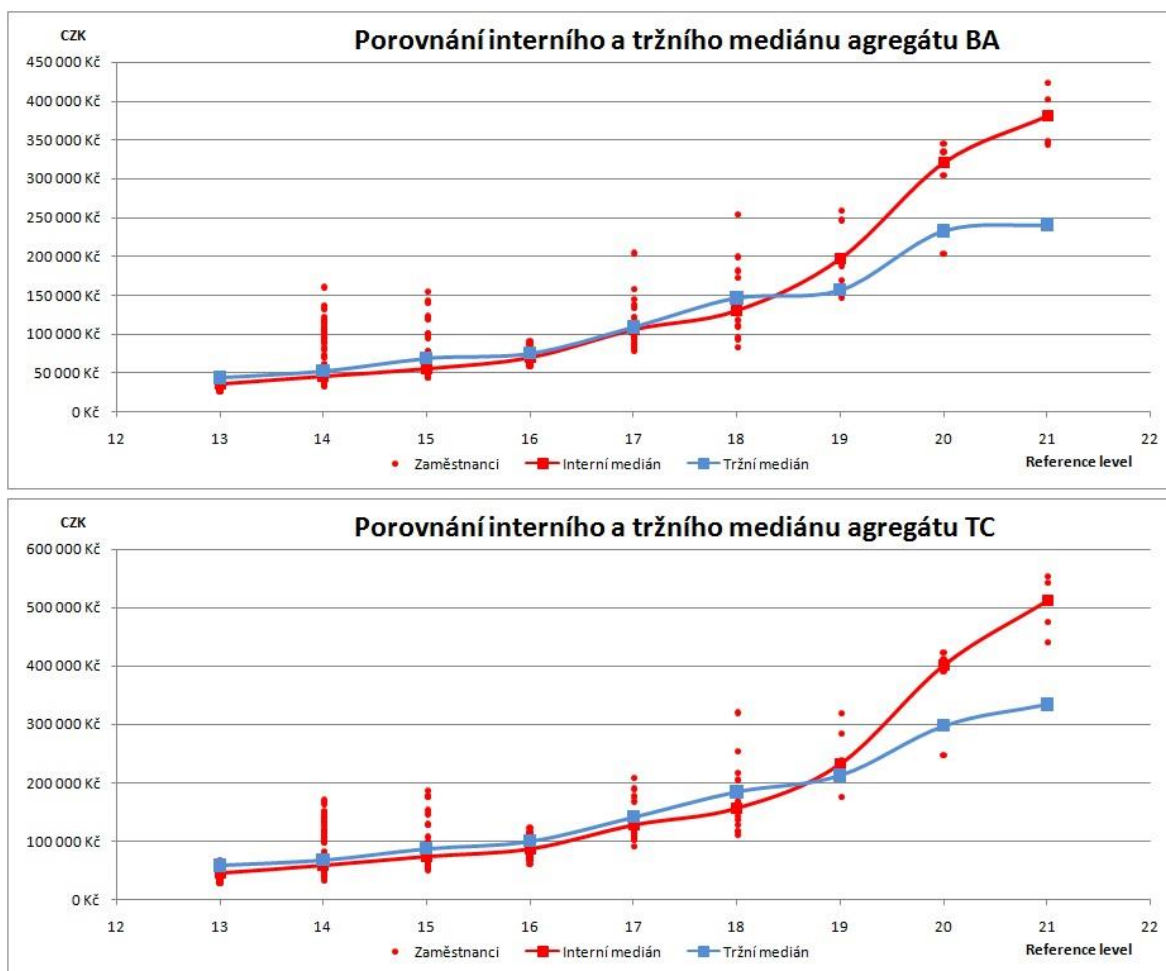
2.3.2 Ukázka protokolu pro firmu č. 9

Celý základní report jsme předvedli v předchozí kapitole. Nyní se ve zkrácené verzi podíváme na to, jak vypadá analýza pro druhou ukázkovou firmu č. 9. Ta neměla oproti firmě č. 117 téměř žádné chyby v čištění, resp. chyby byly po domluvě s klientem upraveny (viz. Kapitoly 2.2.1, 2.2.2). Firma přispěla do databáze 648 zaměstnanci. Známe u ní i průběh předzpracování dat.



Obrázek 29: Basic Salary pro firmu č. 9.

Takto si stojí společnost na trhu základní mzdy. Data vypadají v modelu trhu lépe i díky většímu počtu zaměstnanců. Porovnání interního mediánu s tržním vypadá následovně.



Obrázek 30: Porovnání mediánů pro BA a TC.

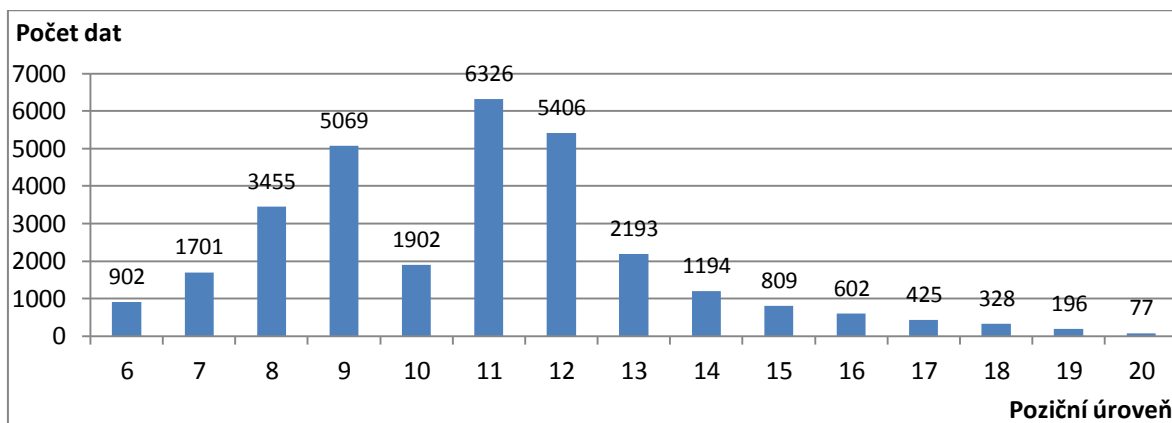
Pokud chce firma dosáhnout toho, že bude platit své zaměstnance okolo tržního mediánu, tak bude analýzou potěšena. Společnosti takového druhu platí své zaměstnance dobře a se svými finančními rozpočty manipulují rozumně. Hodnoty ocenění nižších pozic jsou těsně pod mediánem. Pouze top management se nachází nad ním.

2.4 Rozšířené analýzy

Nabízené analýzy lze rozšířit ve velkém rozsahu. V podstatě to znamená, že na požádání klienta je možné poskytnout další informace, týkající se našeho výběru dat. Data společnosti se porovnávají nejen s pravděpodobnostním modelem trhu a výběrovým mediánem, ale i s dalšími charakteristikami, kterými jsou např. aritmetický průměr, kvartilové rozpětí nebo jakékoliv jiné kvartilové rozpětí. Firma může také zažádat o analýzu trhu založeného pouze na výběru určité skupiny (selektivní výběr). Na následujících obrázcích je ukázka takto rozšířené analýzy.

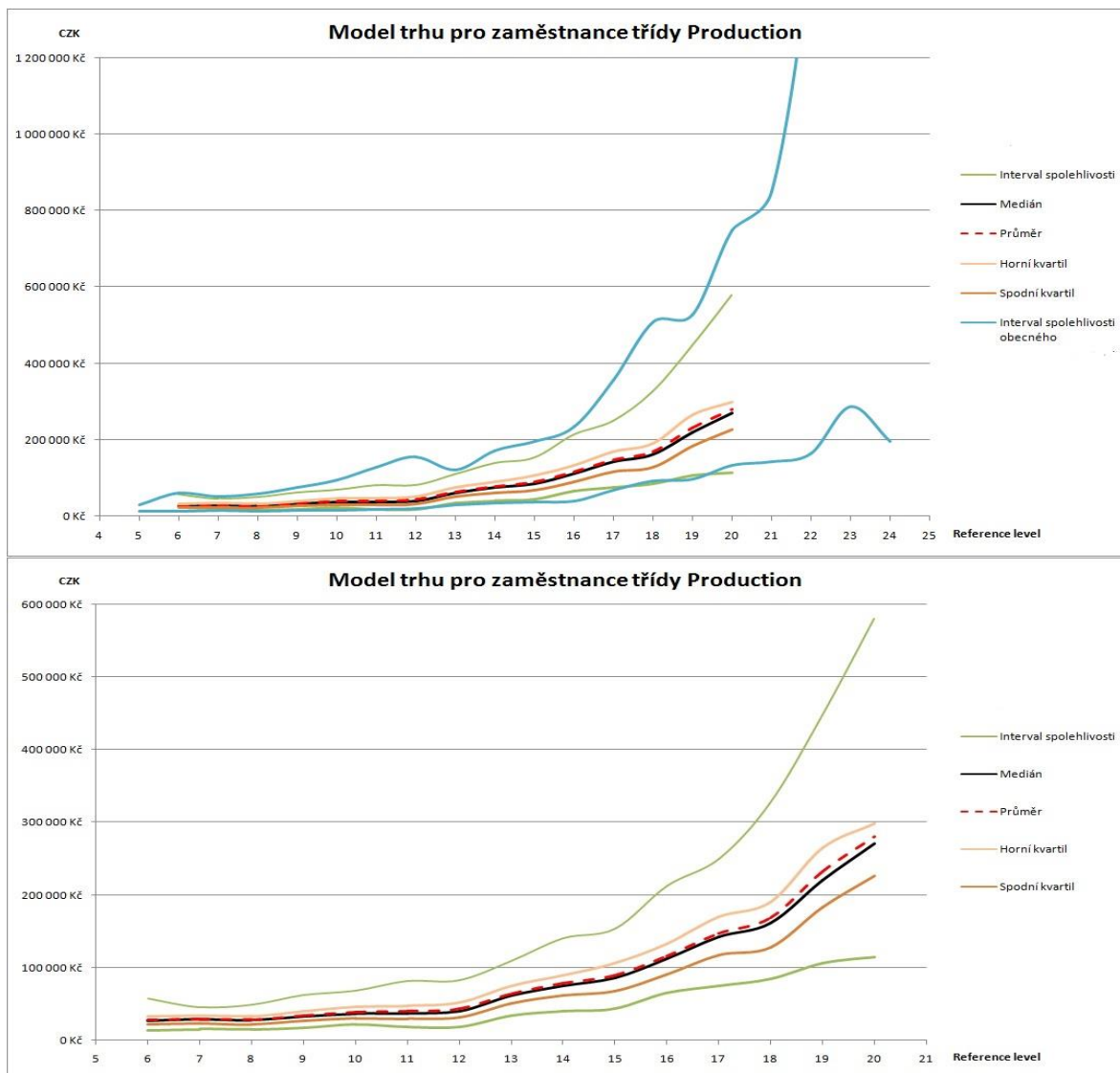
Klient opět dostane informace o počtu zúčastněných. Dokonce na základě smlouvy s klienty je někdy i možné uvádět jména společností. Pro názornou ukázkou

uvažujeme, že si jeden klient objednal analýzu trhu pracovní pozice *PR - Production* úrovní 7 až 12. Na trhu *PR - Production* je celkem 30 629 zaměstnanců z 203 společností. Pokud se omezíme ještě na úrovně 7 až 12, potom máme vzorek dat velikosti 23 859 zaměstnanců ze 171 firem. Klient přispěl informacemi o 68 zaměstnancích.



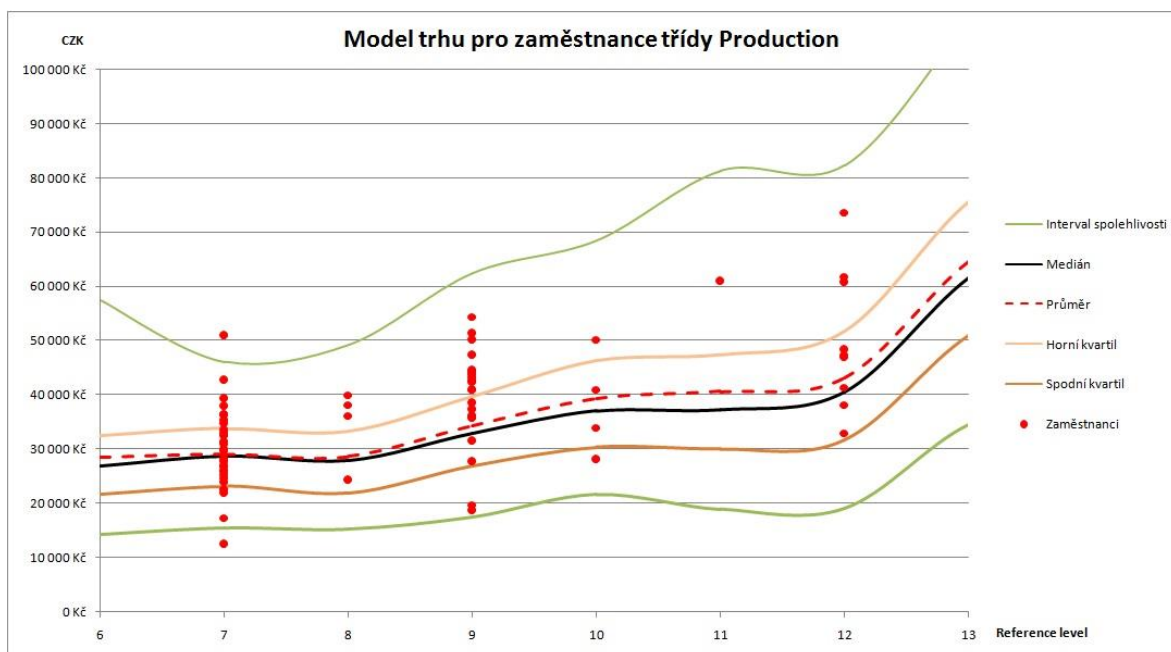
Obrázek 31: Počet dat trhu *Production*.

Na základě těchto dat provádíme analýzu požadovaného trhu. V následujících grafech je vidět jak model trhu *Production* vypadá. Pro ostatní poziční úrovně není možné model trhu sestavit, protože nemáme k dispozici dostatečně velký datový vzorek.



Obrázek 32: Model ocenění zaměstnanců trhu Production.

Klient má nyní představu o tom, jak jeho požadovaný model vypadá oproti obecnému. Samozřejmě jsou v analýze sepsány i charakteristiky polohy a variability, porovnání mediánů a průměrů, ale hlavní částí je zobrazení zaměstnanců firmy na trhu ocenění. Společnost má na požadovaném trhu 68 pracujících a chtěla by je oceňovat na základě průměru z tohoto selektivního výběru. Dále je tabelováno porovnání jednotlivých zaměstnanců vůči internímu mediánu, tržnímu mediánu a průměru, aby klient měl přesné informace.



Obrázek 33: Rozšířená analýza pro firmu.

Označení zaměstnance	Úroveň	Ocenění zaměstnance	Porovnání s interním mediánem	Porovnání s tržním mediánem	Porovnání s tržním průměrem
1	12	73626	155,24%	181,26%	170,46%
2	12	60704	127,99%	149,45%	140,54%
3	12	61818	130,34%	152,19%	143,12%
4	12	48347	101,94%	119,03%	111,93%
5	12	41392	87,27%	101,91%	95,83%
6	12	46953	99,00%	115,60%	108,71%
7	12	33005	69,59%	81,26%	76,41%
8	12	47428	100,00%	116,77%	109,81%
9	12	38164	80,47%	93,96%	88,36%
10	11	60876	100,00%	163,00%	149,77%
11	10	50233	133,92%	135,44%	127,57%
12	10	41004	109,31%	110,56%	104,13%
13	10	34016	90,69%	91,72%	86,39%
14	10	28026	74,72%	75,57%	71,18%
15	9	42502	102,12%	129,17%	123,55%
16	9	50089	120,35%	152,23%	145,61%
17	9	38499	92,50%	117,00%	111,92%
18	9	51281	123,22%	155,85%	149,07%
19	9	47260	113,55%	143,63%	137,38%
20	9	43104	103,57%	131,00%	125,30%
21	9	44450	106,80%	135,09%	129,22%
22	9	44248	106,32%	134,48%	128,63%
23	9	54402	130,71%	165,34%	158,15%
24	9	35994	86,48%	109,39%	104,63%
25	9	36093	86,72%	109,69%	104,92%
26	9	37252	89,51%	113,21%	108,29%

27	9	40977	98,46%	124,54%	119,12%
28	9	43837	105,33%	133,23%	127,43%
29	9	40896	98,26%	124,29%	118,88%
30	9	31552	75,81%	95,89%	91,72%
31	9	43690	104,98%	132,78%	127,01%
32	9	42261	101,54%	128,44%	122,85%
33	9	27744	66,66%	84,32%	80,65%
34	9	19785	47,54%	60,13%	57,51%
35	9	18628	44,76%	56,61%	54,15%
36	9	35706	85,79%	108,52%	103,80%
37	8	40026	107,56%	143,24%	139,77%
38	8	38217	102,70%	136,76%	133,45%
39	8	36210	97,30%	129,58%	126,45%
40	8	24339	65,40%	87,10%	84,99%
41	7	50967	171,05%	176,90%	175,10%
42	7	37946	127,35%	131,70%	130,37%
43	7	36235	121,61%	125,76%	124,49%
44	7	42747	143,47%	148,37%	146,86%
45	7	35339	118,60%	122,65%	121,41%
46	7	32918	110,48%	114,25%	113,09%
47	7	26868	90,17%	93,25%	92,31%
48	7	39266	131,78%	136,28%	134,90%
49	7	29676	99,60%	103,00%	101,96%
50	7	33532	112,54%	116,38%	115,20%
51	7	31305	105,06%	108,65%	107,55%
52	7	21879	73,43%	75,94%	75,17%
53	7	36337	121,95%	126,12%	124,84%
54	7	25360	85,11%	88,02%	87,13%
55	7	32602	109,42%	113,15%	112,01%
56	7	34648	116,28%	120,26%	119,04%
57	7	17361	58,27%	60,26%	59,65%
58	7	25907	86,95%	89,92%	89,01%
59	7	29916	100,40%	103,83%	102,78%
60	7	30932	103,81%	107,36%	106,27%
61	7	22542	75,65%	78,24%	77,45%
62	7	24758	83,09%	85,93%	85,06%
63	7	12494	41,93%	43,36%	42,92%
64	7	28394	95,29%	98,55%	97,55%
65	7	27659	92,83%	96,00%	95,03%
66	7	26622	89,35%	92,40%	91,46%
67	7	23944	80,36%	83,10%	82,26%
68	7	25942	87,07%	90,04%	89,13%

Tabulka 7: Porovnání zaměstnanců individuálně.

Majitel společnosti může nyní prozkoumat, jaké zaměstnance oceňuje lépe nebo naopak hůře oproti internímu mediánu, tržnímu mediánu a tržnímu průměru. Pro vizualizaci porovnání hodnot ocenění vůči výše zmíněným veličinám byl zvolen interval 90 až 110 %. Nižší hodnoty než 90 % z dané statistiky jsou zbarveny

červeně, vyšší než 110 % pak zeleně. Dále můžeme doporučit, aby se management podniku pokusil snížit rozptyly ocenění v jednotlivých úrovních. Může se tedy pokusit výplaty upravit tak, že všechny hodnoty budou ležet např. ve spočteném kvartilovém rozpětí.

Analýz tohoto druhu je velmi velké množství. Selektivní výběry mohou být založeny na pozicích, pozičních úrovních nebo jejich kombinaci, jak bylo ukázáno výše. Dále pak např. délka působení zaměstnanců ve firmách, specifický sektor firem nebo jejich roční obrát. Díky selekci dat z databáze jsme schopni vytvořit vysoký počet možných trhů. Pokud by se jednalo o společnost, která se našich šetření účastní již delší dobu (více let), lze potom pro klienta vyhotovit porovnání ocenění jeho zaměstnanců v čase.

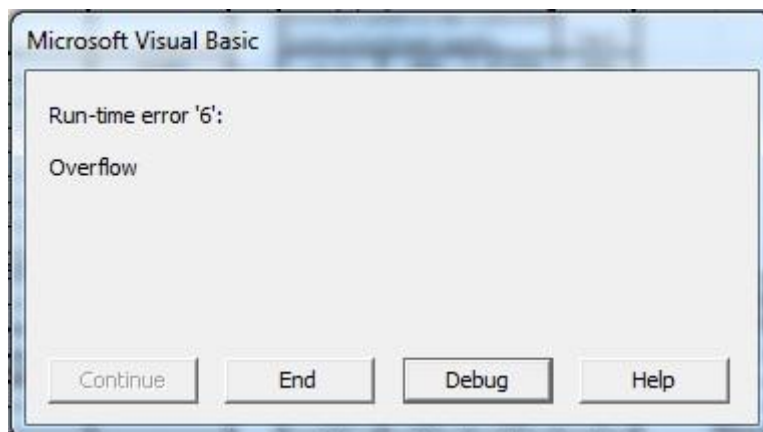
Tímto je celý cyklus naší práce ukončen. Celý od začátku se opět opakuje nadcházející rok. Pomocí Wilksových tolerančních mezí určených na základě letošních dat, máme stanoveny, jakých hodnot by data nového sběru v příštím roce měla nabývat. Všechna data podrobíme předčištění a kontrole. Vypočítáme hodnoty intervalů spolehlivosti a sestrojíme nový model trhu. Zvolenými statistickými metodami bychom postupem času měli dostávat neustále lepší obrazy trhů. Rozptyly našich náhodných výběrů by se měly snižovat a modely by měly být o to přesnější.

2.5 Diskuze nad výsledky a vzniklé problémy

Zjišťujeme, že studie přinesla množství problémů, které mají omezená řešení nebo je není možné rozumně nebo efektivně odstranit. V této kapitole jsou nastíněny nejzákladnější problémy a jejich krátký popis.

- **Začátek celého procesu.** Pro nově založenou poradenskou firmu je začátek celého procesu kritický. Nemá k dispozici žádná starší data, na jejichž základě je možné vypočítat toleranční meze a nemá tedy absolutní představu o tom, jak by hodnoty nového sběru měly vypadat. Obecně budou analýzy v prvních letech působení firmy na trhu nekvalitní. Bude také dlouho trvat, než si zajistíme dostačující velikost naší klientely.
- **Velikost statistického vzorku.** V tabulkách 3 a 4 jsou experimentálně napočítány nejmenší možné počty dat, které jsou potřebné k funkčnosti výpočtu modelu trhu. Pokud nejsme schopni reálně získat data těchto rozměrů, musíme buď snížit spolehlivost našich výroků o tržních modelech, nebo nejsme vůbec schopni požadovaný model trhu sestrojít. Malý počet dat ovšem není jediným problémem. Velký počet dat může také způsobit

komplikace. Např. metody používané v nástroji *udelej_meze.xlsm* nejsou schopny pracovat s větším rozsahem dat než je 40 000.



Obrázek 34: Overflow.

Excel také neumí do jednoho grafu vykreslit více než 32 000 bodových hodnot. Pokud bychom tedy potřebovali vyhotovit analýzu pro klienta, který má více zaměstnanců, museli bychom data rozdělit do více grafů nebo použít výkonnější výpočetní software.

- **Časová náročnost.** Už prvotní sběr dat a konzultace s klienty zabere velké množství času. Pokud bychom uvažovali např. data z našeho příkladu (kapitola 2.2.1), tedy 341 společností, a odhad doby strávené konzultacemi s klienty bychom stanovili na 3 hodiny, tak při těchto počtech naši poradci strávili u klientů dohromady přes 1 000 hodin v roce. Máme-li k dispozici 260 pracovních dní v roce a vezmeme-li v potaz i čas strávený na cestě ke klientovi, můžeme říci, že jeden poradce je schopen poskytnout v jednom dni pouze jednu konzultaci. Vidíme tedy, že již pro rozsah našich dat bylo nutné zaplatit nejméně 2 poradce, kteří naši firmu společně předvedli a sebrali od nich data o zaměstnancích.

Vyčištění dat a statistická analýza nezabere díky automatizacím mnoho času. Pokud bychom ovšem v datech nacházeli velké množství chyb, komunikace s klientem kvůli potvrzení korektnosti se opět protáhne. Čas potřebný k sepsání standardní analýzy by se mohl značně zkrátit také díky automatizacím. Základní report pro klienty je totiž až na doporučení vždy stejný. Rozšířené a selektivní analýzy se ovšem musí zpracovávat individuálně. To je také časově docela náročné.

Celkově je tedy celý proces velmi časově náročný. Jeden konzultant je za rok schopen komunikovat se zhruba sto klienty ročně a jeden analytik je schopen zpracovat data zhruba 150 společností. Pokud by firma působila např. v 50 zemích světa a v každé z nich bychom měli k dispozici podobný rozsah výběru, jaký máme v naší studii (nezohledňujeme tedy fakt, že

ekonomiky jsou navzájem jiné a každý výběr může obsahovat jiná data, ať už počtem nebo strukturou), potom bychom museli v poradenské firmě mít alespoň 170 poradců a alespoň 100 analytiků, kteří data zpracovávají v rozsahu, který byl uveden ve studii. Nehledě na management poradenské firmy.

- **Zařazení zaměstnance** do pozic a pozičních úrovní. Tato problematika je velmi obsáhlá a není jednoduché zaměstnance správně kategoricky zařadit. Modely trhu ocenění zaměstnanců by obecně měly v čase růst. Pracovníci by tedy měli být korektně zařazeni do pozičních úrovní tak, aby byl model ostře rostoucí. Na základě našich dat tomu ale tak není. Jak je vidět na obrázcích 17, 18 a 19, námi sestavené modely trhu nejsou ostře rostoucí. Tento problém by měl být co nejrychleji odstraněn. Mělo by se přehodnotit zařazení zaměstnanců do pozičních úrovní nebo najít jiný systém rozdělení. Řešením by mohlo být zavést větší počet pozičních úrovní, to ovšem vede k menšímu statistickému vzorku.
- **Výběr dat** pro vypočtení hodnot modelů trhu pro různé agregáty. V naší práci vystupují tři agregáty *Basic Salary*, *Base Salary* a *Total Cash*, kde hodnota $BS = BA +$ veškeré fixní příspěvky a $TC = BS +$ variabilní bonusy. Problém nastane při existenci zaměstnanců, kteří nepobírají od zaměstnavatele ani příspěvky ani bonusy. Tito zaměstnanci „tahají“ celý model trhu dolů. Myslíme tím, že spodní hodnota intervalu spolehlivosti $x_{(i)}$ pro námi zvolený kvantil agregátů BS a TC se o moc neliší v porovnání s agregátem BA , jak můžeme opět vidět na obrázcích 17, 18 a 19. Řešením je např. do výběru těchto agregátů zahrnout pouze zaměstnance s hodnotami příspěvků a bonusů větší než nula. Porovnání modelů trhu jednotlivých agregátů by potom mělo vypadat zhruba následovně.



Obrázek 35: Ukázka opraveného modelu trhu.

Omezením výběru ze základního souboru bychom mohli dosáhnout lepších výsledků. To vše je ale opět na úkor velikosti statistického vzorku. Musíme tedy uvádět, jak jsou agregáty sestaveny. Máme tři možnosti. První z těchto variant je ukázána ve studii, všechny agregáty jsou tedy počítány ze stejného počtu dat bez ohledu na to, jestli zaměstnanci dostávají fixní přídatky a variabilní bonusy nebo ne. Druhá varianta je zahrnout do vzorku dat pouze ty zaměstnance, u kterých jsou příspěvky a bonusy větší než 0. Poslední variantou je zahrnout zaměstnance, kteří dostávají fixní příspěvky větší než 0, ale do agregátu TC jsou zařazeni všichni, kteří mají na bonusy nárok i přes to, že na kladný bonus nedosáhli. Obecně se modely trhu mohou sestavit různými způsoby, mohou se zakládat pouze na náhodném výběru (model reprezentuje např. decilové rozpětí našeho výběru), pak ovšem nejsou brány za modely pravděpodobnostního typu. Tím klientovi ale sdělujeme informace pouze o trhu firem, které se průzkumu zúčastnily a ne o celé populaci ocenění zaměstnanců.

Toto jsou základní problémy, se kterými jsme se ve studii setkali. Ne všechny je reálně možné odstranit. Zařazování zaměstnanců, modely trhů a časová náročnost jsou věci, které jsme schopni sami odstranit nebo zlepšit. Oproti tomu velikost našich výběrů nemůžeme ovlivnit. Snažíme se o to, abychom sebrali data od co možná největšího počtu společností. Snaha nám ale vždy nezajistí požadovaný úspěch. Pro firmy, které na trhu finančního poradenství působí již déle a mají svoji stálou klientelu, se tyto problémy v průběhu času vyřešily. Poradenská firma tedy musí sehnat co možná v nejkratším čase velké množství klientů a překonat první roky působení na trhu.

Závěr

Cílem této práce bylo sestavení věrohodného modelu trhu ocenění zaměstnanců a vytvoření nástrojů potřebných k jeho uskutečnění. Celkovým tématem byla aplikace statistických metod na reálná data a poskytnutí rozumných doporučení a závěrů klientům, kteří se průzkumu zúčastnili.

Pomocí intervalů spolehlivosti pro kvantily spojitých veličin jsme vyčíslili hodnoty spodních a horních mezí modelů. Na základě takto sestrojených mezí jsme pak klientovi byli schopni poskytnout model trhu, který reprezentoval 90% kvantil populace ocenění zaměstnanců s pravděpodobností 97,5 %. Díky Wilksovým tolerančním mezím jsme pak připravili makra pro zachycení chyb a označení extrémních hodnot ve sběru dat následujícího roku.

V průběhu studie jsme sice narazili na několik problémů, ať už těch, které jsme schopni v budoucnosti opravit nebo těch, na které nemáme vliv. I přes to jsme ale na základě našich dat vytvořili věrohodný model trhu, který je možný prezentovat společností na oplátku za poskytnuté informace. Až na výjimky, byly sestrojené intervaly spolehlivosti pro poziční úroveň rostoucí a to svědčí o celkem dobrém zařazení zaměstnanců do příslušných pozic. Společnosti poté mohou zažádat o rozšířené analýzy nejružnějších selektivních výběrů.

Věříme, že tyto analýzy mohou pomoci hlavně novým společnostem, které si nejsou jisty, jak dobře své zaměstnance oceňují vůči konkurenci. Analýzy samozřejmě nepomáhají jen nově začínajícím společnostem. Klientům se naskýtá velké množství situací, kdy jsou pro ně tyto analýzy potřebné. Např. pokud povýší zaměstnance na novou pozici a chce ho ocenit nejprve 90 % tržního mediánu a až postupem času mu výplatu průběžně zvětšovat. Další situací může být otevření nové pobočky v jiné zemi, než v které klient dosud působil. Nemá představu o tom, jak by měl zaměstnance v dané zemi oceňovat. Zde a v mnoho dalších situacích se analýza trhu ocenění zaměstnanců stává cennou informací, která pomůže k úspěšnému rozvoji a řízení firem. Celou studii považujeme za úspěšnou.

Literatura a zdroje

- [1] HÁTLE, Jaroslav a Jiří LIKEŠ. *Základy počtu pravděpodobnosti a matematické statistiky: vysokoškolská učebnice*. 2., nezměn. vyd. Praha: Státní nakladatelství technické literatury, 1974.
- [2] REIF, Jiří. *Metody matematické statistiky*. 2. upr. vyd. V Plzni: Západočeská univerzita, Fakulta aplikovaných věd, 2004. ISBN 80-7043-302-7.
- [3] JÍLEK, Miloš. *Statistické toleranční meze*. Praha: SNTL - Nakladatelství technické literatury, 1988. Teoretická knihovna inženýra.
- [4] WILKS, S. S. *Order statistic*. Bull. Amer. Math. Soc. Volume 54, Number 1, Part 1, 1948, 6-50.
- [5] WILKS, S. S. *Determination of sample size for setting tolerance limits*, AMS, 12, 1941.
- [6] RÉNYI, A. *Teorie Pravděpodobnosti, Academia*, Praha, 1972.
- [7] SOMERVILLE, Paul N. *Tables for Obtaining Non-Parametric Tolerance Limits*, AMS, Vol. 29, Number 2, pages 599-601 (June 1958).
- [8] VÁVRA, František. *Studijní materiály z předmětů KMA/SA1, KMA/VSM*. 2016.
- [9] ŤOUPAL, Tomáš. *Studijní materiály z předmětu KMA/STAV*. 2012.
- [10] *Integrity Consulting* [online]. Odměňování jako strategický systém dlouhodobé motivace zaměstnanců. 2016. Dostupné z: http://www.integrity-consulting.cz/publ_odmenovani.php
- [11] *Proces EIA - online učebnice* [online]. Rozhodovací analýzy a variantní řešení. 2016. Dostupné z: <http://ucebnice-eia.zf.mendelu.cz/rozhodovaci-analyzy>
- [12] CLARDY, Alan B. *Managing human resources: Exercises, Experiments, and Applications Workbook*. Lawrence Erlbaum Associates, 1996.

- [13] RUMSEY, Michael G., Clinton B. WALKER a James H. HARRIS. *Personnel Selection and Classification*. Lawrence Erlbaum Associates, 1994.
- [14] ARTHUR, Diane. *Fundamentals of Human Resources Management*. American Management Association, 2004.

Zdrojová data a soubory přiložené na CD

Makra

- *Cleaning.xlsm*: Makro určené k předzpracování dat.
- *Dotazník.xlsm*: Určen ke sběru dat.
- *Report.xlsm*: Slouží k vytvoření obrazu modelu trhu.
- *Udelej_meze.xlsm*: Výpočet základních statistických charakteristik, intervalů spolehlivosti a Wilksových mezí.

Trhy

- *Base.xlsx*: Model trhu agregátu BS.
- *Basic.xlsx*: Model trhu agregátu BA.
- *DATA.xlsx*: Zdrojová data.
- *Porovnání agregátů.xlsx*: Porovnání agregátů a Wilksových mezí.
- *Total cash.xlsx*: Model trhu agregátu TC.

Ukázka firmy č. 9

- *Dotazník.xlsm*: Dotazník naplněný daty této firmy.
- *Firma_09.xlsx*: Statistiky a porovnání pro firmu č. 9.
- *Sešit1.xlsx*: Zkontrolovaná data firmy, připravena pro nahrání do databáze.

Ukázka firmy č. 117

- *Dotazník.xlsm*: Dotazník naplněný daty této firmy.
- *Firma_117.xlsx*: Statistiky a porovnání pro firmu č. 117.
- *Sešit1.xlsx*: Zkontrolovaná data firmy, připravena pro nahrání do databáze.

Ukázka selektivní analýzy pro třídu PR-Production

- *Select PR.xlsx*: Statistiky a porovnání pro ukázkovou firmu, model trhu PR.

Ukázka SQL

- *Tabulky.sql*: Vytvoření databázových tabulek.
- *Select.sql*: Ukázka jednoduchého „selectu“ z databáze pro vytvoření trhu PR.