Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR

Metody pro získávání dat, informací a poznatků v rámci veřejné strategie

Výstup v rámci aktivity KA1
Implementace metod a nástrojů pro strategické řízení a plánování

Aktivita 1.1 Metodická podpora pro strategické řízení a plánování

Nástroj č. 9/12: Metody pro získávání dat, informací a poznatků v rámci veřejné strategie

Název projektu: Podpora strategického řízení a plánování ve veřejné správě ČR
s důrazem na rozšiřování metod, nástrojů, inovací, znalostí a zkušeností v praxi
Registrační číslo CZ.03.4.74/0.0/0.0/15_025/0016924

červen 2023

www.verejne-strategie.cz

MINISTERSTVO
PRO MÍSTNÍ
ROZVOJ ČR

Název: Metody pro získávání dat, informací a poznatků v rámci veřejné strategie.

Autoři: Milan Křápek, Milan Jan Půček

Recenzenti: Radim Misiaček

Počet stran: 60

Vzor citace: KŘÁPEK, Milan, PŮČEK Milan Jan. *Metody pro získávání dat, informací a poznatků v rámci veřejné strategie.* Praha: Ministerstvo pro místní rozvoj, 2023

Text vznikl v rámci projektu „Podpora strategického řízení a plánování ve veřejné správě ČR s důrazem na rozšiřování metod, nástrojů, inovací, znalostí a zkušeností v praxi“, registrační číslo CZ.03. 4. 74/0.0/0.0/15_025/0016924, který byl podpořen finančními prostředky Evropského sociálního fondu, které byly na realizaci projektu poskytnuty z Operačního programu Zaměstnanost.

V rámci projektu „Podpora strategického řízení a plánování ve veřejné správě ČR“ realizovaného Ministerstvem pro místní rozvoj byl vytvořen soubor metodických nástrojů zaměřujících se na specifická témata strategického řízení a plánování. Vedle samotných textů metodických doporučení jsou k dispozici instruktážní videa a interaktivní šablony. Výstupy všech metodických nástrojů jsou zveřejněny na „Portále strategické práce v ČR“ (www.verejne-strategie.cz) v sekci Nástroje pro strategickou práci.

Série metodických nástrojů strategické práce se zaměřuje na následující témata:

1 Analýza, řízení a registr rizik
2 Využívání příležitostí ve vazbě na strategickou práci
3 Rozhodování a strategický přístup
4 SWOT Analýza v rámci veřejné strategie
5 Komunikace, participace a marketing strategií a projektů
6 Plánování a řízení projektů pomocí trojimperativu projektu
7 Plánování a řízení času v rámci strategické práce
8 Zdravé finanční řízení ve vazbě na strategické řízení
9 Metody pro získávání dat, informací a poznatků v rámci veřejné strategie
10 Využití procesního přístupu (procesní řízení) v rámci strategické práce
11 Posuzování vlivů koncepcí na životní prostředí (SEA) pro zpracovatele strategických dokumentů
12 Implementace veřejných strategií

Tato publikace neprošla redakční jazykovou úpravou.

Praha: Ministerstvo pro místní rozvoj, 2023

ISBN (pdf): 978-80-7538-502-4

Obsah

Souhrn **4**
Summary **5**
1. Úvod k problematice získávání dat, informací a poznatků **6**
2. Přehled metod, postup a přístup k získání dat **9**
2.1 Přehled metod pro získávání dat, informací nebo poznatků 9
2.2 Vizualizace dat, informací nebo poznatků 10
3. Odkazy na vybrané databáze a zdroje dat **11**
4. Výzkum od stolu (Desk research), rešerše **12**
4.1 Jak se provádí 12
4.2 Příklad – hledání informací o strategickém plánování 13
4.3 Citování použitých zdrojů 14
5. Dolování dat (Data mining), vyhledávání dat **16**
5.1. Jak se provádí 16
5.2. Monitor 16
5.2.1. Základní monitor 16
5.2.2. Analytická část 18
5.2.3. Datový katalog 21
5.2.4. Informační systém o veřejných zakázkách (ISVZ) 22
5.2.5. Věstník veřejných zakázek 23
5.3. Český statistický úřad 25
5.3.1. Statistiky 25
5.3.2. Vše o území 27
5.3.3. Vlastní výběr 28
5.3.4. Ukazatele 31
6. Využití matematických a statistických metod **33**
6.1 Popisná statistika 33
6.1.1. Průměry 33
6.1.2. Kvantily 35
6.1.3. Charakteristiky variability 36
6.2 Hledání podobností 37
6.2.1. Určení podobnosti teorie 37
6.2.2. Určení podobnosti praxe 38
6.3 T-Test 39
6.3.1. T-test – Kdy a proč používat 40
6.3.2. T – test – Teorie 40
6.3.3. T-Test – Praxe 43
6.4 F-Test 44
6.4.1. F-Test – Teorie 44
6.4.2. F-test – Praxe 44
6.5 Chi-square test 45
6.5.1. $\chi 2$-test dobré shody – teorie 45
6.5.2. $\chi 2$-test nezávislosti – teorie 46
6.5.3. $\chi 2$-test– praxe 46

6.6 Výpočet regresního modelu 47
6.6.1. Regresní analýza – volba modelu 48
6.6.2. Regresní analýza – provedení 48
6.6.3. Regresní analýza – interpretace výsledků 50
6.7 Výpočty používaných statistik 51
7. **Příklady konkrétních postupů** 52
7.1 Analýza dlouhodobého trendu počtu obyvatel obce 52
7.2 Porovnávání finanční ukazatelů obcí 53
7.3 Veřejné zakázky na strategický plán 54
8. Popis šablony v Excelu pro získávání dat 56
9. Závěr 57
Přílohy 58
Použitá literatura 59

Souhrn

Text popisuje problematiku získávání a následné zpracování dat. Z hlediska strategické práce jsou správná data a jejich pochopení nutnou podmínkou pro úspěšné vytvoření funkční strategie. Data využijeme jak ve fázi přípravy strategie, tak při jejím schvalování. Během realizace potřebujeme data aktualizovat a na základě jejich analýzy hodnotit, zda připravená strategie funguje, případně některé parametry upravit, aby se zvýšila efektivita. Také při plánování a řízení projektů je úloha dat a jejich pochopení nezastupitelná. Dnešním problémem není nedostatek informací, ale spíš přehlcení zapojených osob nepodstatnými daty a snížení významnosti těch důležitých. Správně provedený datamining nám pomůže se v datech lépe orientovat a vytáhne ty podstatné znalosti, které můžeme následně využít.

Text se člení na 9 kapitol. Kapitola 1 je úvodem k dolování a zpracování dat, jsou vysvětleny vybrané pojmy. V kapitole 2 jsou popsány nejčastěji používané metody a vizualizační techniky. Kapitola 3 obsahuje přehled nejčastěji využívaných zdrojů dat. Kapitola 4 popisuje výzkum od stolu, jedná se hlavně o možnosti získávat data z výzkumů publikovaných ve vědeckých časopisech. V kapitole je popsáno, jak můžeme tyto publikace vyhledávat a také jsou zde popsány zásady správného citovaní. Kapitola 5 podrobněji popisuje některé z nejčastěji používaných a zároveň nejrozsáhlejší zdroje dat. Jsou zde ukázány způsoby, jak s těmito zdroji pracovat a jak z nich data získávat. Kapitola 6 obsahuje přehled některých statistických metod, které mohou být využity pro analýzu získaných dat. V kapitole 7 jsou příklady možných problémů, které můžete chtít řešit s popisem možného řešení s odkazy na jednotlivé kapitoly, které danou problematiku řeší. Obsahuje tak spojení jednotlivých metod pro lepší představu, jak mohou být popsané metody efektivně propojeny. V kapitole 8 jsou informace k šabloně. Kapitola 9 uzavírá celý text. K textu je též zpracovaná video přednáška, která shrnuje nejdůležitější poznatky, šablona v Excelu pro zpracování dat a návod k šabloně.

Summary

The text describes the issue of data acquisition and subsequent processing. From the point of view of strategic work, correct data and their understanding are necessary conditions for the successful creation of a functional strategy. We will use the data both during the strategy preparation phase and during its approval. During implementation, we need to update the data and, based on their analysis, evaluate whether the prepared strategy works or adjust some parameters to increase efficiency. Also, in the planning and management of projects, the role of data and its understanding is irreplaceable. Today's problem is not a lack of information, but rather the drowning of the people involved with irrelevant data and the reduction of the significance of the important ones. Properly performed data mining will help us better navigate the data and extract essential knowledge that we can then use.

The text is divided into nine chapters. Chapter 1 is an introduction to data mining and processing; selected concepts are explained. Chapter 2 describes the most frequently used methods and visualization techniques. Chapter 3 contains an overview of the most frequently used data sources. Chapter 4 describes research from the table; it is mainly about the possibility of obtaining data from research published in scientific journals. The chapter describes how we can search for these publications and also describes the principles of correct citation. Chapter 5 describes in more detail some of the most frequently used and, at the same time, the most extensive data sources. Ways to work with these sources and how to get data from them are shown here. Chapter 6 contains an overview of some statistical methods that can be used to analyze the obtained data. In Chapter 7, there are examples of possible problems that you may want to solve, along with a description of a possible solution and links to the individual chapters that solve the problem in question. It thus contains the connections between individual methods for a better idea of how the described methods can be effectively connected. See Chapter 8 for template information. Chapter 9 concludes the entire text. The text is accompanied by a video lecture that summarizes the most important findings, an Excel template for data processing, and instructions for the template.

1. Úvod k problematice získávání dat, informací a poznatků

Data, informace, poznatky v rámci strategické práce potřebujeme:

- Při tvorbě veřejné strategie, a to jak její analytické části, variantních řešení nebo scénářů budoucího vývoje, tak návrhu cílů a měřítek strategie, projektů nebo akčních plánů. Jsou též potřebné k pochopení potřebných souvislostí (kontextu) a k porozumění místním nebo specifickým podmínkám.
- Při hodnocení dosažení / nedosažení měřítek (ukazatelů) cílů, projektů, akčních plánů nebo jiných prováděcích dokumentů.
- Při evaluaci či změnách veřejné strategie.
- Při běžném rozhodování v rámci strategické práce.

Význam dat, informací a poznatků pro správné rozhodování

Při tvorbě veřejné strategie, její implementaci a obecně při strategické práci je potřeba provádět potřebné analýzy a přijímat správná rozhodnutí. Analýzy a rozhodování musí být založeny na **relevantních datech, informacích, poznatcích**, ale také na zkušenostech a znalostech. To umožní **rozhodovat na základě faktů** (na základě logiky), nezbytné zkušenosti (intuice), místních podmínek a souvislostí. Každé (často i drobné) rozhodnutí při strategickém řízení ovlivňuje splnění cílů nebo plnění úkolů týkajících se strategické práce, a to buď kladně (je-li rozhodnutí částečně či zcela správné) nebo záporně (je-li částečně či zcela chybné). Najít (identifikovat, přijmout) správné rozhodnutí je v dnešní komplexní a složité době nesnadné – obvykle vybíráme na základě dostupných dat, informací a poznatků z více variant, které se zdají srovnatelné (stejně špatné či stejně dobré). Za každé rozhodnutí a jeho následky nese rozhodovatel odpovědnost. **Správné rozhodnutí má následující znaky**, které platí současně:

1) **Je morální** – opírá se o hodnotový (morální) rámec rozhodovatele nebo o etický kodex dané instituce. Jako základ morálky v činnosti instituce lze vyjít z těchto tří oblastí: (a) při činnosti bychom neměli škodit jiným lidem, společnosti či přírodě, (b) je třeba dodržovat platná pravidla (legislativu, interní směrnice, běžné postupy) a současně (c) neseme plnou odpovědnost za každé rozhodnutí, a to včetně negativních následků z něj plynoucích. Nesoulad mezi rozhodnutím a morálkou signalizuje, že dochází k manažerskému selhání a morálnímu hazardu. Respektování morálky v praxi znamená, že vyloučíme všechny varianty možného rozhodnutí, které bychom považovali za nemorální.
2) **Je zákonné** – je v souladu s předpisy, tedy se zákony a vnitřní předpisy; zákony mají přednost před vnitřními předpisy, interními postupy (např. směrnicemi). Nesoulad mezi rozhodnutím a zákonností signalizuje, že dochází k manažerskému selhání. Respektování zákonnosti v praxi znamená, že vyloučíme všechny varianty možného rozhodnutí, které bychom považovali za nezákonné.
3) **Vede účinně k dosažení cílů** nebo ke splnění zadaného úkolu (správné rozhodnutí nás k cíli přibližuje, chybné vzdaluje).

Relevantnost (ověřitelnost, kvalita) dat, informací, poznatků

V textu pracujme s pojmem „relevantní" informace, mělo by jít o pravdivá fakta, ne o „fámy" (fámy = neověřené, zkreslené informace). Musíme brát v úvahu, že lidé nereagují na skutečnost, ale na svoji představu (přesvědčení) o skutečnosti. Tato představa je často vytvořena na základě neověřených či zkreslených informací. Proto je ověření pravdivosti dat, informací nebo poznatků při strategické práci nesmírně důležité. Navíc každý prožívá či chápe situaci, ve které se nachází (včetně s ní souvisejících informací), jinak. Má z toho jiný dojem. Problémem je, že lidé vyčtou nebo i slyší jen to, co sami chtějí – lidé se „zavrtávají" do vlastních myšlenek (uzavírají do své „informační bubliny") a nechtějí nic jiného slyšet než to, co odpovídá jejich názoru. Tuto skutečnost je při získávání a využití dat nezbytné brát v úvahu. Je tedy nezbytná data, informace a poznatky vhodně argumentovat, vizualizovat, případně použít potenciál kritického myšlení.

Vztah data – informace – znalosti

Mít správná data, informace a znalosti ve správném okamžiku na správném místě může výrazně přispět k dosahování cílů. Je nezbytné pochopit rozdíly a souvislosti mezi daty, informacemi a znalostmi (viz tabulka).

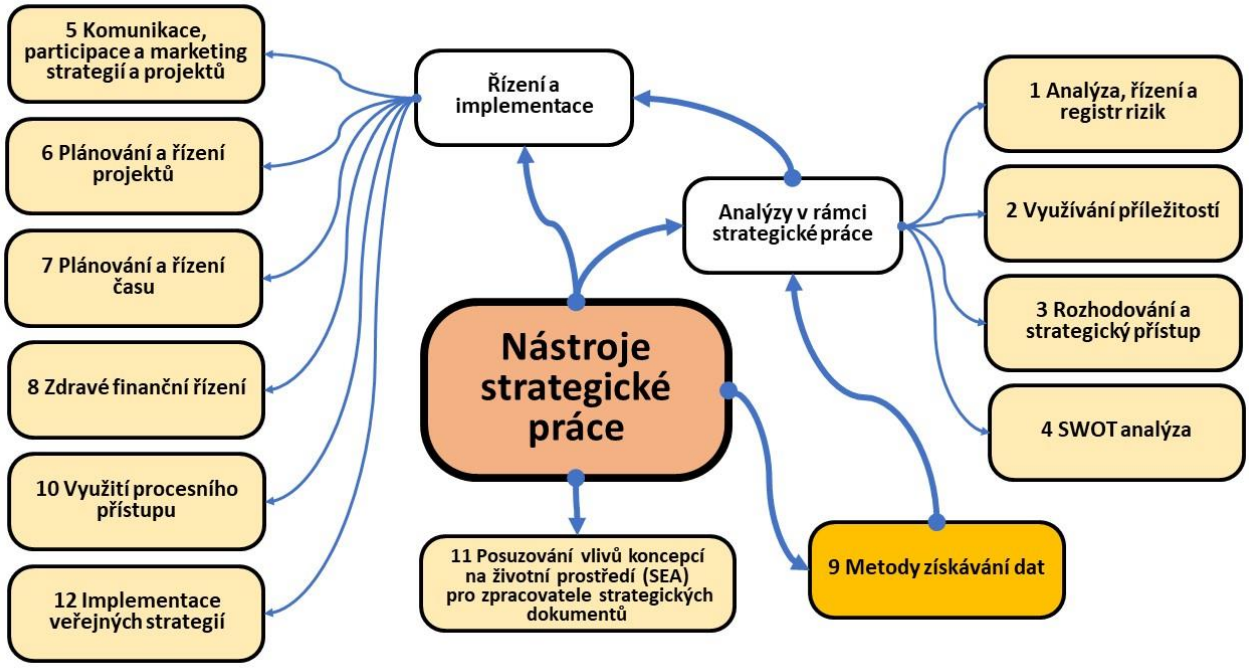
Tab. 1 Data – informace – znalosti – moudrost

Č.	Oblast	Popis
4.	Moudrost (vědění) ↑	**Moudrost (vědění)** jsou znalosti spojené s určitým postojem a vyjadřuje komplexní hodnocení (poznání) okolního světa jednotlivcem. Je to **schopnost nalézat správná rozhodnutí**. Moudrost vychází z minulosti i současnosti, ale týká se také budoucnosti (**intuice**). Moudrost **je neoddělitelně spojena s pokorou**, neboli kdo nezískal pokoru, nezískal moudrost. Moudrost je též **spojena s jednoduchostí** (moudrý člověk se dokáže vyjadřovat jednoduše).
3.	Znalosti ↑	**Znalost je tedy schopnost řešit problémy**, „vědět jak". Znalost je schopnost využít své vzdělání, zkušenosti, hodnoty a odbornost jako rámec pro řešení problémů, pro vyhodnocení dat, informací a jiných zkušeností k výběru odpovědi na danou situaci. Znalosti poskytují odpověď na otázky „Proč?" a „Jak?". Znalosti se získávají procesem aktivního učení. Znalost má pro svého uživatele kvalitativní stránku a jistou hodnotu.
2.	Informace ↑	**Informace jsou účelově zpracovaná (utříděná) data**, kterým jejich uživatel v procesu interpretace přisuzuje určitý význam. Pojem informace se používá v mnoha oborech lidské činnosti i v mnoha vědeckých disciplínách, a proto existuje mnoho definic i výkladů tohoto pojmu. Jsou výsledkem interpretace dat na základě individuálních schopností hodnot a znalostí. Informace poskytují odpovědi na otázky kdo?, co?, kde?, kdy?
1.	Data (údaje)	**Data jsou vyjádřena symboly** (čísla, písmena, text, zvuk, obraz), ale může jít i o **smyslové vjemy** (čich, hmat).

Zdroj: autor dle Ochrana, Půček 2012

Na následujícím obrázku jsou zachyceny **vazby mezi nástroji strategické práce**, které jsou zpracovávány v rámci projektu. Nástroj k plánování a řízení času má s ohledem na skutečnost, že plánovat a efektivně využívat čas je nezbytné při každé činnosti, vazbu na všechny ostatní nástroje. Velmi silná vazba je k rozhodování a strategickému přístupu – faktor času je při každém rozhodování nezbytné v přiměřeném rozsahu zvažovat. Samozřejmě také v rámci řízení času je nezbytné provádět před rozhodnutím týkající se času potřebné analýzy.

Obr. 1 Vazby na ostatní nástroje zpracovávané v rámci projektu



Zdroj: autoři

2. Přehled metod, postup a přístup k získání dat

2.1 Přehled metod pro získávání dat, informací nebo poznatků

Pro efektivní strategické řízení je nezbytné mít schopnost zajistit a ověřit potřebná data, informace a poznatky. V následující tabulce jsou uvedeny vhodné metody pro získávání dat, informací či poznatků.

Tab. 2 Metody pro získávání dat, informací či poznatků při řízení rizik

Název	Popis metody nebo komentář
Výzkum od stolu (Desk research), **rešerše**	**Výzkum od stolu** a rešerše jsou podobné metody. Lze zjednodušeně říci, že výzkum od stolu je širší – zaměřuje se na strategické či podobné dokumenty či texty, zabývá se otázkami: *Řešil to už někdo? Jak postupoval? S jakými riziky se setkal? Jak je hodnotil? Jaká opatření přijímal?* Hlavním účelem výzkumu od stolu je zorientovat se v dané problematice a navrhnout další postup. **Rešerše** je zaměřena více na lieraturu a co z ní převzít. Odpovídá na otázku: *Kdo a jak se tím zabýval?* Účelem rešerše je sestavit přehled literatury a vybrané citace z ní.
Dolování dat (Data mining), **vyhledávání dat**	Zaměřuje se na: (1) Databáze či jiná veřejně dostupná data, knihovny a podobně (2) Internet, respektive na vyhledávání dostupných dat či informací či poznatků na internetu, (3) Vnitřní data, informace či poznatky šetřené instituce / obce / kraje. Při získávání dat je třeba dodržovat etiku odborné práce – zejména čerpat jen z povolených dat a dodržovat citační pravidla.
Pozorování	**Pozorování** je **empirickou deskriptivní metodou**, která popisuje zjevné (jevové) problémy, kdy potřebná informace je získávána **různými formami pozorování**. **Předností** metody je, že se **zaměřuje na evidenci (zřejmosti, patrnosti).** **Problémem** je, že teprve za jevem je skryta podstata problému. Na ni se zaměřuje vědecké zkoumání. Z pozorování tedy musíme umět konstruovat vědeckou výpověď vypovídající o podstatě.
Měření	Je **kvantitativní empirickou** metodou využitelnou zejména v přírodních či technických vědách (např. měříme fyzikální veličiny). Metoda měření je využitelná i ve společenských vědách. Zde má měření své **limity**. Obvykle **stanovujeme referenční jednotku**, na jejímž základě pak provádíme normativní hodnocení (měřítkem je norma stanovená vzhledem k určité skupině) zkoumaného předmětu P, který srovnáváme s referenční jednotkou R („*riziko nedostatku prostředků na provoz veřejné služby P je dvakrát vyšší než veřejné služby R*" nebo „*riziko události P je poloviční než R*") nebo pomocí nastavené škály (např. *dopad rizika velmi malý 1; malý 2; střední 3; vysoký 4; velmi vysoký 5*). Fakticky jde o kvalitativní komparaci.
Šetření	Strukturované či částečně strukturované rozhovory, dotazníky, připomínkování, metoda Delphi, brainstorming, panel expertů, ….

Název	Popis metody nebo komentář
Expertní odhad	Podstatou expertního odhadu je předpověď výsledku (výsledného stavu) ze strany jednotlivce (experta na danou věc). V tomto smyslu jde o prozatímní vyhodnocení řešeného jevu, respektive předběžný úsudek postavený na metodě pozorování a hlubokých znalostech dané věci ze strany experta.
Vizualizační techniky	Kontrolní seznamy (Check lists), vývojový diagram, diagram příčin a následku (Ishikavův diagram), myšlenková mapa, Ganttův diagram (pro vizualizaci harmonogramu) a podobně
Další možnosti	Dalších specifických metod může být celá řada (například metoda experimentu).

Zdroj: autoři s využitím Ochrana, Půček a kol. (2019, s. 85)

2.2 Vizualizace dat, informací nebo poznatků

Získaná data, informace nebo poznatky je nezbytné prezentovat vhodnou formou – nejen vysvětlujícím textem, ale také pomocí vhodných vizualizačních technik jako jsou:

- tabulky,
- grafy,
- obrázky,
- modely,
- diagramy,
- myšlenkové mapy a další vizualizační techniky.

Vizualizační techniky výrazně přispívají k porozumění informacím. U každé tabulky, obrázku atd. je třeba mít uvedeno číslo, název, zdroj. Taktéž je třeba v textu před tabulkou či obrázkem tyto tabulky či obrázky avizovat (minimálně například větou: *„Na následujícím obrázku je zobrazen …."* nebo *„Tabulka č. 5 popisuje data týkající se …."*. Před nebo za obrázkem či tabulkou je třeba mít také její vysvětlení či vhodný komentář k ní.

3. Odkazy na vybrané databáze a zdroje dat

Zdroje dat můžeme rozčlenit do těchto oblastí, vždy je uveden odkaz:

1. Ministerstvo financí

1) Databáze Ministerstva financí: MONITOR (http://monitor.statnipokladna.cz/); zde jsou informace o hospodaření státu, jeho organizací a všech samospráv od roku 2013 do současnosti. **Databáze MONITOR** je specializovaný informační portál Ministerstva financí, který umožňuje veřejnosti volný přístup k rozpočtovým a účetním informacím ze všech úrovní státní správy a samosprávy.
2) Data Ministerstva financí: Zadluženost územních rozpočtů, monitoring obcí http://www.mfcr.cz/cs/verejny-sektor/monitoring/zadluzenost-uzemnich-rozpoctu
3) Fiskální výhled České republiky: zahrnuje predikci na běžný a také na následující rok, u některých ukazatelů pak i výhled na 2 roky. Dokument dostupný na https://www.mfcr.cz/cs/verejny-sektor/makroekonomika/fiskalni-vyhled. Důležité pro stanovování očekávaného vývoje veřejných financí (například odhad budoucích daňových příjmů pro obec).
4) Makroekonomická predikce (je zveřejňována se čtvrtletní periodicitou). Dostupná na: https://www.mfcr.cz/cs/verejny-sektor/makroekonomika/makroekonomicka-predikce. Důležité pro stanovování očekávaného vývoje veřejných financí.

2. Další zdroje státních institucí, například

1) Data Českého statistického úřadu https://www.czso.cz/
2) Data k veřejným zakázkám http://www.isvz.cz/ISVZ/Podpora/ISVZ.aspx
3) Data o projektech čerpající prostředky z fondů EU https://dotaceeu.cz/cs/statistiky-a-analyzy nebo: https://www.risy.cz/cs/vyhledavace/projekty-eu
4) Data ke smlouvám lze získat v **Registru smluv** (Ministerstvo vnitra) https://smlouvy.gov.cz/

3. Data EU, například:

1) Portál veřejně přístupných dat EU. Dostupné z https://data.europa.eu/euodp/cs/data/
2) Hospodaření státu EU a další data https://ec.europa.eu/eurostat/data/database

4. Další zdroje – vytvořené neziskovým sektorem, například:

1) Ekonomická data v databázi „Hlídač státu" https://www.hlidacstatu.cz/
2) Databáze veřejných strategií http://databaze-strategie.cz/

4. Výzkum od stolu (Desk research), rešerše

Při provádění rešerší jde vlastně o vyhledání dříve publikovaných výstupů. Ideálně samozřejmě z ověřených zdrojů, případně vědeckých výsledků publikovaných ve kvalitních časopisech.

4.1 Jak se provádí

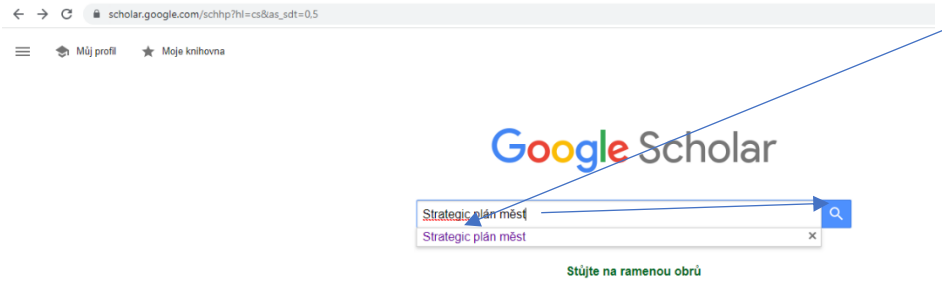
K provedení správné rešerše obvykle využíváme různé vědecké databáze, z nichž nejznámější jsou Web of Science, Scopus, Erih plus, … Přístup k těmto zdrojů je ovšem zpoplatněn, a i když určité části těchto databází jsou dostupné zdarma, velmi to omezuje jejich využití, pokud si danou databázi neplatíte, to v případě měst a obcí není pravděpodobné.

Z tohoto důvodu se zde zaměříme na jinou databázi, která je dostupná zdarma a pomůže nám vyhledávat odborné texty dle zadaných klíčových slov. Jedná se o službu google scholar, která je dostupná z adresy https://scholar.google.com. Vyhledávání zde funguje stejně jako při „klasickém" vyhledávání v google. Tedy napíšete do zobrazeného pole, co hledáte a získáte přehled odpovídajících publikací.

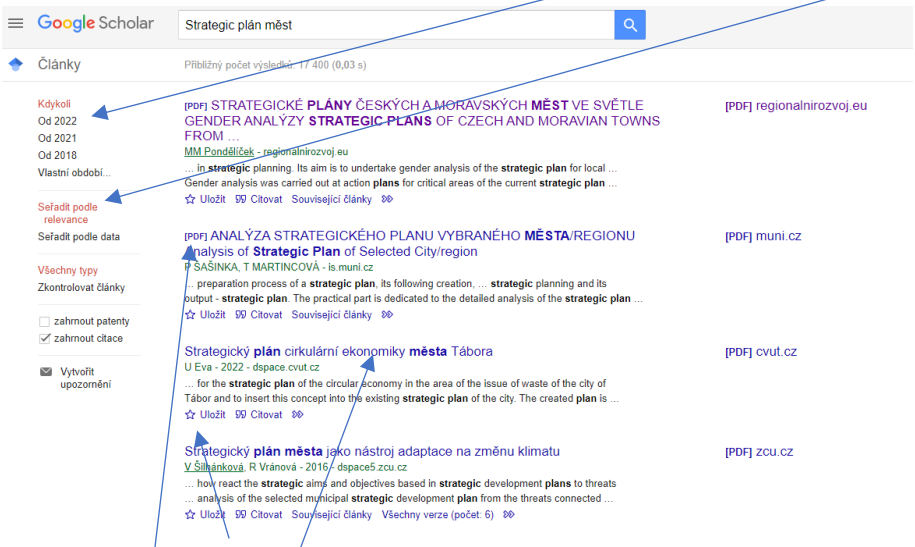
Tento nástroj také umožňuje ukládání vybraných publikací do knihovny pro jejich pozdější využití. Samozřejmě je třeba počítat s tím, že ne všechny nalezené dokumenty jsou veřejně dostupné. Různí vydavatelé mají také různé podmínky pro přístup, a někdy je tedy nutné za stažení publikace zaplatit. Velká část publikací je ovšem v módu „Open access" tedy dostupné veřejně bez nutnosti přihlašování nebo placení.

4.2 Příklad – hledání informací o strategickém plánování

V tomto příkladu si zkusíme vyhledat dokumenty týkající se strategického plánování měst a obcí. Nejdříve si zobrazíme web scholar.google.com. Poté do vyhledávače zadáme klíčová slova. Klíčová slova by měla být co nejpřesnější, aby nebylo výsledků mnoho, může jít i o tisíce výsledků, a zároveň nesmí být moc přesná, aby během vyhledávání nacházel různé tvary slov. Proto zde můžeme použít například klíčová slova „strategické plánování měst“, v tomto případě totiž google vyhledá různá slova odpovídající hledaným základům. Například „měst“ vyhledá všechny publikace, které obsahují jak jednotné číslo „město“, tak množné „města“. Google již často umí vyhledávat i různé tvary českých slov, ale ne vždy je to dokonalé, proto může být vhodné použít i vyhledávání **„strategic plán měst“**, a opět budou vyhledány požadované publikace



Výsledky se poté zobrazí v seznamu, kde vlevo můžeme filtrovat **dle data vydání**, abychom například seznam omezili jen na nejnovější výsledky. Případně nastavit řazení **„dle relevance“** a **„dle data“**.



Po kliknutí na **hvězdičku** je zvolená položka uložena do Vaší knihovny, samozřejmě v případě, že máte vytvořen účet na google. K těmto položkám pak máte přístup přímo bez nutnosti znovu vyhledávat.

Kliknutím na **název publikace**, budete přesměrováni na příslušný web, případně pokud je u názvu příznak **[pdf]** je rovnou zobrazen text práce.

4.3 Citování použitých zdrojů

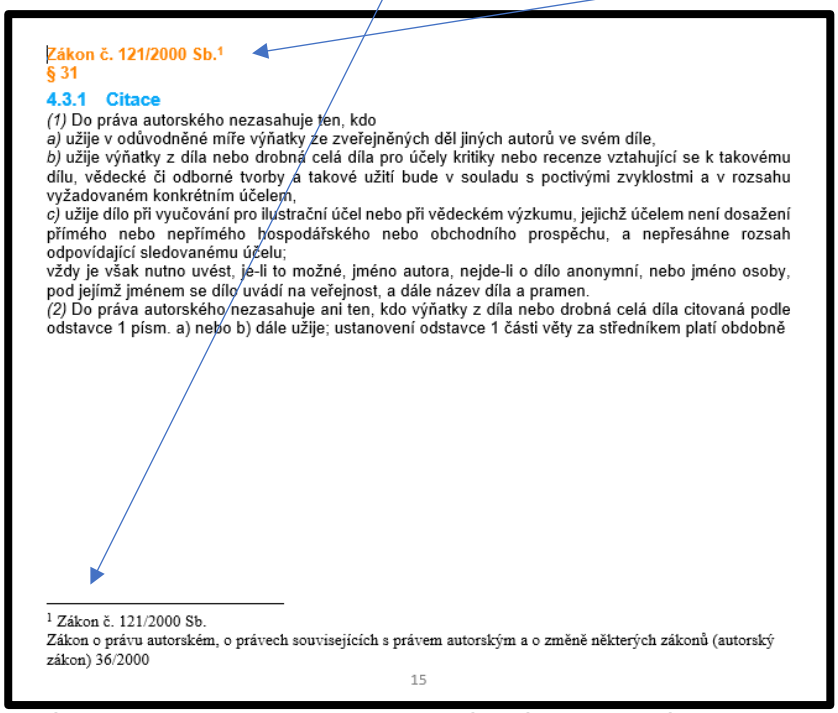
Při využívání takto vyhledaných publikací si ovšem musíme dát pozor, abychom neporušili autorský zákon. Pokud bychom části těchto textů použili ve vlastním dokumentu, je nutné je správně citovat, jinak se dopouštíme porušení zákona. V případě správné citace je použití části díla v odůvodněné míře v pořádku.

Zákon č. 121/2000 Sb.
§ 31

1) Do práva autorského nezasahuje ten, kdo
   a) užije v odůvodněné míře výňatky ze zveřejněných děl jiných autorů ve svém díle,
   b) užije výňatky z díla nebo drobná celá díla pro účely kritiky nebo recenze vztahující se k takovému dílu, vědecké či odborné tvorby a takové užití bude v souladu s poctivými zvyklostmi a v rozsahu vyžadovaném konkrétním účelem,
   c) užije dílo při vyučování pro ilustrační účel nebo při vědeckém výzkumu, jejichž účelem není dosažení přímého nebo nepřímého hospodářského nebo obchodního prospěchu, a nepřesáhne rozsah odpovídající sledovanému účelu;

   vždy je však nutno uvést, je-li to možné, jméno autora, nejde-li o dílo anonymní, nebo jméno osoby, pod jejímž jménem se dílo uvádí na veřejnost, a dále název díla a pramen.
2) Do práva autorského nezasahuje ani ten, kdo výňatky z díla nebo drobná celá díla citovaná podle odstavce 1 písm. a) nebo b) dále užije; ustanovení odstavce 1 části věty za středníkem platí obdobně.

Citovat můžeme mnoha způsoby. V praxi se setkáváme s dvěma přístupy. První přístup využívá citace v poznámkách pod čarou. Pokud v textu použijeme nějaké dílo, přidáme k němu **příznak** a **poznámku pod čarou s tímto příznakem** na stejné stránce dokumentu. Tento postup je vhodný v případě, že náš dokument je kratší a obsahuje malý počet citací.


Zákon č. 121/2000 Sb.[1]
§ 31

4.3.1 Citace

(1) Do práva autorského nezasahuje ten, kdo
a) užije v odůvodněné míře výňatky ze zveřejněných děl jiných autorů ve svém díle,
b) užije výňatky z díla nebo drobná celá díla pro účely kritiky nebo recenze vztahující se k takovému dílu, vědecké či odborné tvorby a takové užití bude v souladu s poctivými zvyklostmi a v rozsahu vyžadovaném konkrétním účelem,
c) užije dílo při vyučování pro ilustrační účel nebo při vědeckém výzkumu, jejichž účelem není dosažení přímého nebo nepřímého hospodářského nebo obchodního prospěchu, a nepřesáhne rozsah odpovídající sledovanému účelu;
vždy je však nutno uvést, je-li to možné, jméno autora, nejde-li o dílo anonymní, nebo jméno osoby, pod jejímž jménem se dílo uvádí na veřejnost, a dále název díla a pramen.
(2) Do práva autorského nezasahuje ani ten, kdo výňatky z díla nebo drobná celá díla citovaná podle odstavce 1 písm. a) nebo b) dále užije; ustanovení odstavce 1 části věty za středníkem platí obdobně

[1] Zákon č. 121/2000 Sb.
Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) 36/2000

15

Druhý postup spočívá v tom, že za každou citaci vložíme textový odkaz na literaturu, a na konec dokumentu pak přehled literatury, ve kterém dle odkazu můžeme vždy vyhledat tu, která byla

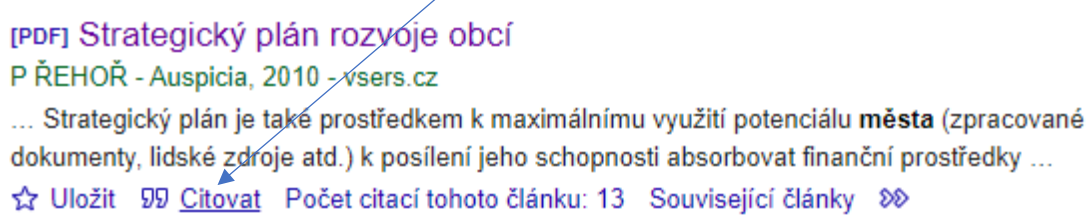
použitá v daném místě dokumentu. Jak mají vypadat odkazy i správný způsob citace pak určují různé normy. V České republice je nejpoužívanější normou ISO-690, která popisuje, jak mají citace vypadat. Ale i tato norma umožňuje dva různé přístupy. První spočívá v tom, že za text vložíme číslo do závorky (1) a toto číslo pak označuje danou publikaci, v celém dokumentu se poté používá stejné číslo. V seznamu literatury jsou pak publikace seřazeny dle čísla, tedy dle pořadí, v jakém se ve vytvořeném dokumentu objevují.

Druhou možností je že se do závorky vloží jméno autora a rok (Křápek 2022). Seznam publikací je pak seřazen dle abecedy.

Seznam literatury u každé publikace obsahuje jména autorů, název publikace, rok vydání, název vydavatele a další informace které umožňují jednoznačnou identifikaci díla.

S vytvářením těchto odkazů nám google scholar opět pomůže.

Pokud v seznamu klikneme na odkaz „**citovat**“, zobrazí se různé možnosti správných citací dle různých norem. Tu pak stačí zkopírovat do dokumentu.



5. Dolování dat (Data mining), vyhledávání dat

Pojmem dolování dat rozumíme postupy pro získávání, ukládání, třídění, filtrování a další zpracování velkého objemu dat, jehož výsledkem jsou užitečné informace využitelné v praxi.

5.1. Jak se provádí

Základem pro úspěšné dolování dat je vytvoření vhodné databáze. Data musíme tedy nejdříve získat a uložit abychom je následně mohli zpracovávat. Získávání dat může probíhat jak přímo, tedy vlastním měřením, hodnocením, tak nepřímo, kdy využíváme různé dostupné databáze.

Po získání potřebných dat můžeme přistoupit k další části a tou je zpracování těchto dat. K tomu je možné použít speciální programy, které využívají statistické metody, neuronové sítě, rozhodovací stromy a mnoho dalších metod.

5.2. Monitor

„MONITOR je specializovaný informační portál Ministerstva financí, který umožňuje veřejnosti volný přístup k rozpočtovým a účetním informacím ze všech úrovní státní správy a samosprávy. Prezentované informace pocházejí ze systému Státní pokladny (IISSP – Integrovaný informační systém státní pokladny) a Centrálního systému účetních informací (CSÚIS) a jsou čtvrtletně aktualizovány.

MONITOR dále zajišťuje zveřejňování účetních závěrek organizačních složek státu podle § 21a zákona č. 563/1991 Sb., o účetnictví, ve znění pozdějších předpisů.“

Webová stránka je rozdělena na části, z nichž každá zobrazuje výsledky jiným způsobem a umožňuje jiný způsob práce s nimi.

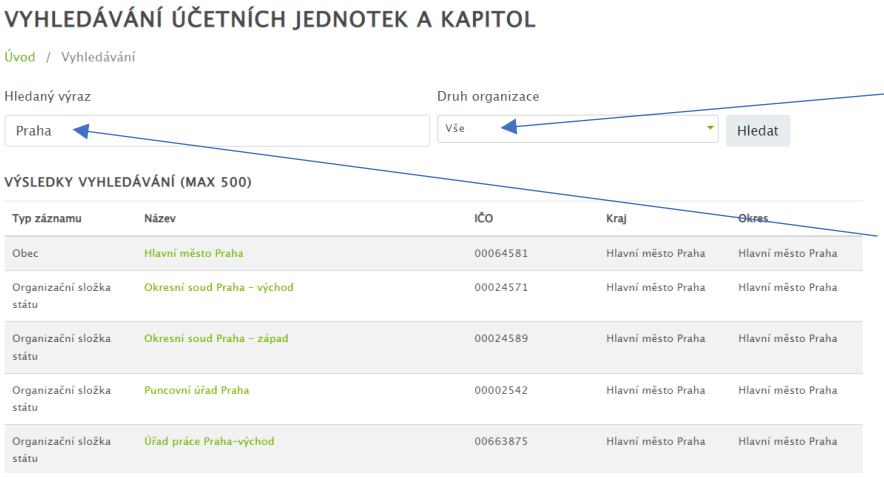
1. Základní monitor
2. Analytická část
3. Datový katalog

https://monitor.statnipokladna.cz/

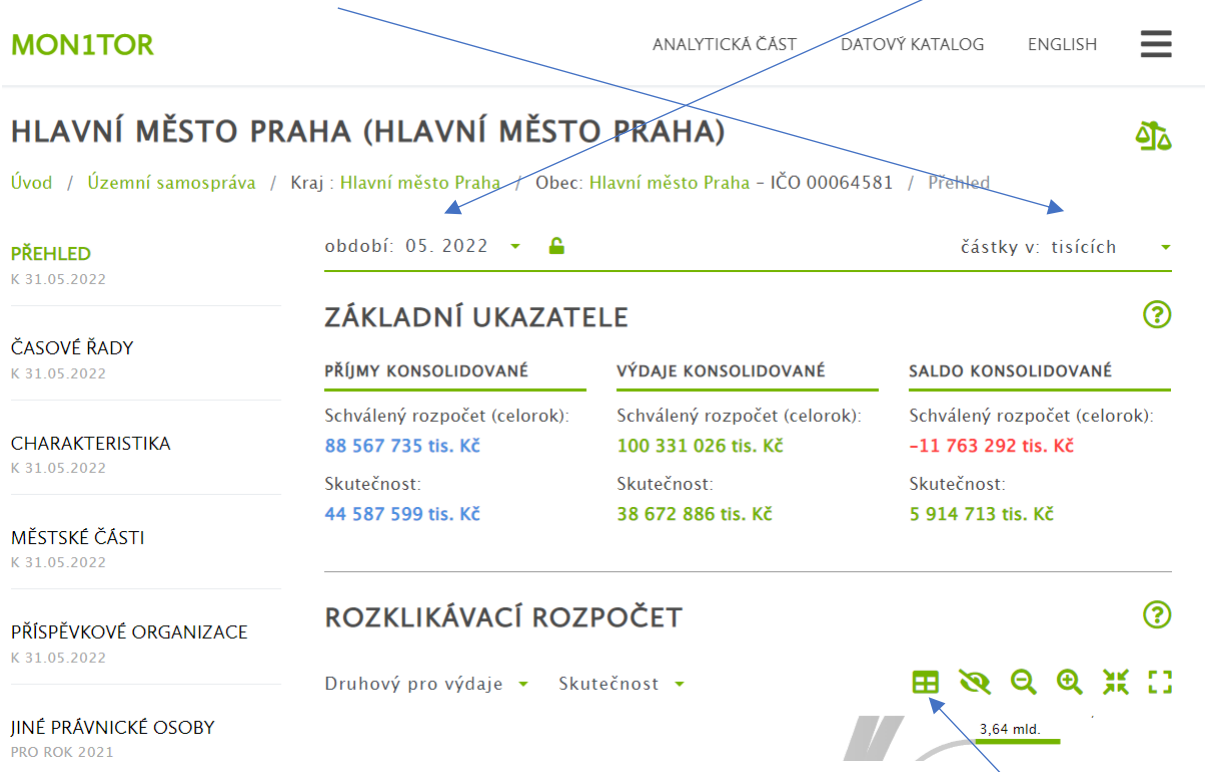
5.2.1. Základní monitor

Základní monitor se Vám zobrazí po otevření tohoto webu. Jeho použití je snadné a je pouze třeba vyhledat požadovaný subjekt. Vyhledání provedené zadáním Jména obce, názvu organizace nebo IČO do **vyhledávacího pole** a potvrzení tlačítkem **Hledat**.

Následně se zobrazí vyhledávací okno, kde jsou zobrazeny nejrelevantnější organizace. Pokud je množství vyhledaných záznamů velké je možné volbou **druh organizace** počet položek snížit. Další možností, jak snížit počet nalezených položek je **přesněji specifikovat** hledání.



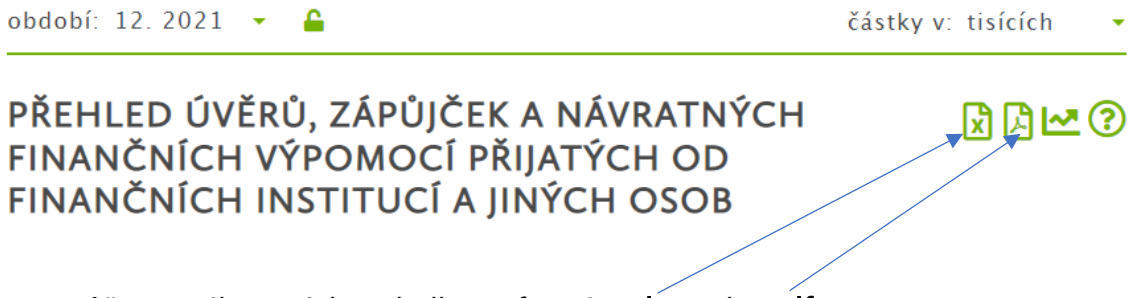
Po zvolení jedné organizace se zobrazí přehledová stránka organizace. Ta obsahuje finanční přehled a základní identifikaci. Pozor vždy je nutné zkontrolovat, za který rok jsou údaje zobrazovány a případně **rok** změnit. Také je třeba hlídat použité **jednotky**.



Zde také vidíme rozklikávací rozpočet pomocí stromového diagramu, kdy můžeme u jednotlivých položek provést rozklad. Jestliže preferujete tabulkové zobrazení, můžete jej **zobrazit tlačítkem** nad diagramem.

V levé části si můžeme zobrazit podrobnější údaje, jako je vývoj finančních ukazatelů za poslední roky, podrobnou charakteristiku, přehledy městských částí, příspěvkových organizací, jejichž zřizovatelem obec je a také přehled jiných organizací, ve kterých má obec vlastnický podíl. Samozřejmostí jsou pak finanční data (účetní závěrka, účetní ukazatele, přehled úvěrů,…).

U většiny přehledů najdeme skupinu tlačítek, které nám umožní data stáhnout a dále pracovávat jakýmikoliv vlastními způsoby.

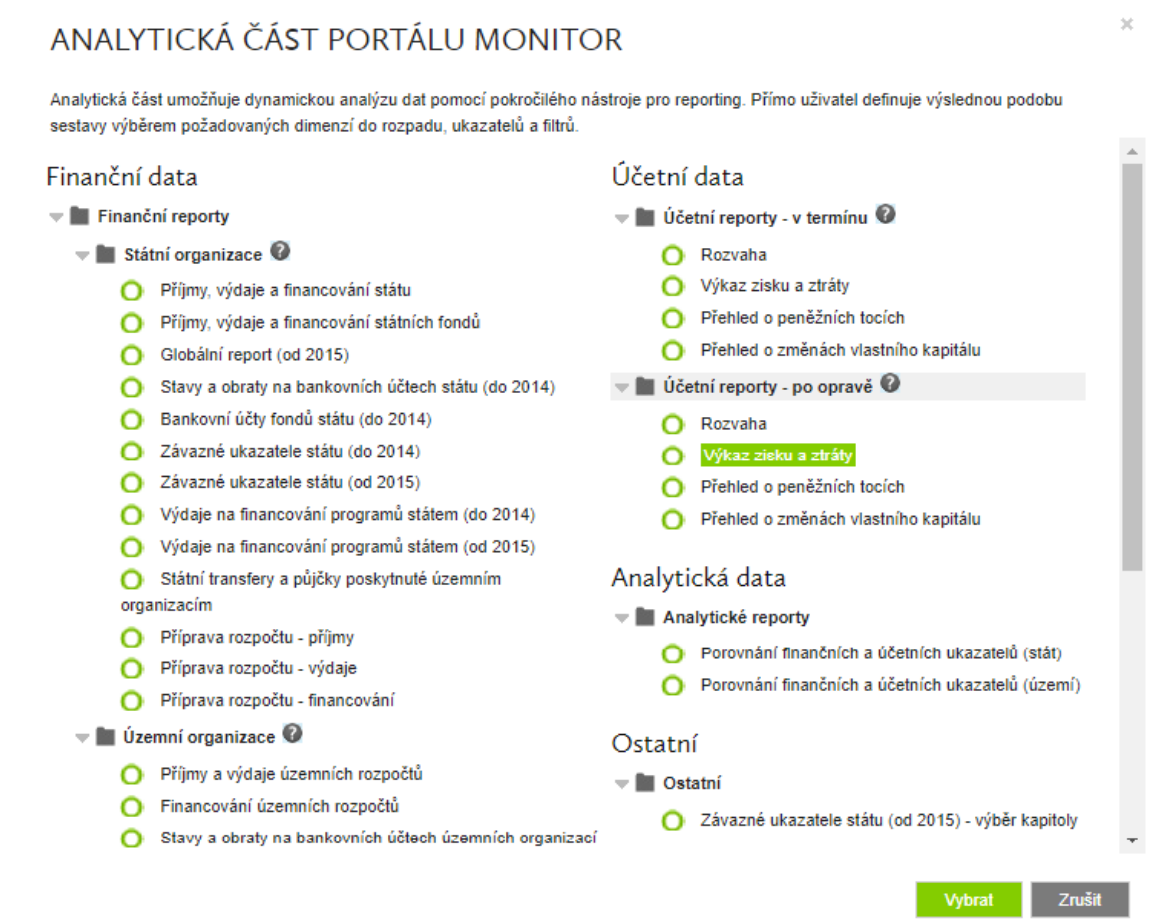


Data můžeme stáhnout jako tabulku ve formátu **xlsx**, nebo **pdf**.

5.2.2. Analytická část

Pomocí analytické části si můžeme sami určit, jaké údaje mají být zobrazeny. Tento nástroj je velmi komplexní, a tak často dojde k jeho zahlcení a vyřizování požadavku může být časové náročné.

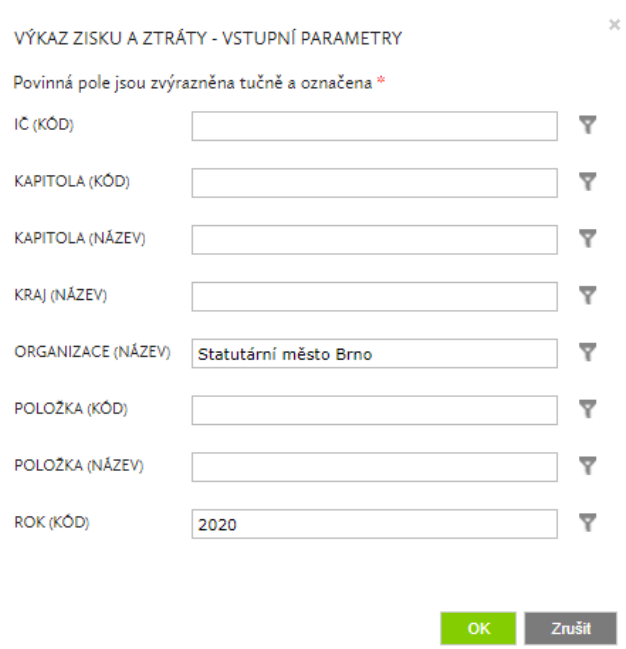
Po spuštění je zobrazen přehled agend, které lze zobrazit. Z tohoto seznamu je třeba jednu vybrat a výběr potvrdit tlačítkem *Vybrat*.



Následně bude zobrazeno úvodní okno zvoleného reportu. Každý report má unikátní možnosti, které jsou přednastaveny.

Zde můžeme vybrat položky, které chceme zobrazit. Nastavíme tedy filtry tak, aby obsahovaly požadované údaje. Není vhodné nechat pole prázdná, protože by systém zobrazoval všechny údaje a načítání by bylo velmi dlouhé (z pohledu uživatele to vypadá, že se načítání zcela zaseklo).

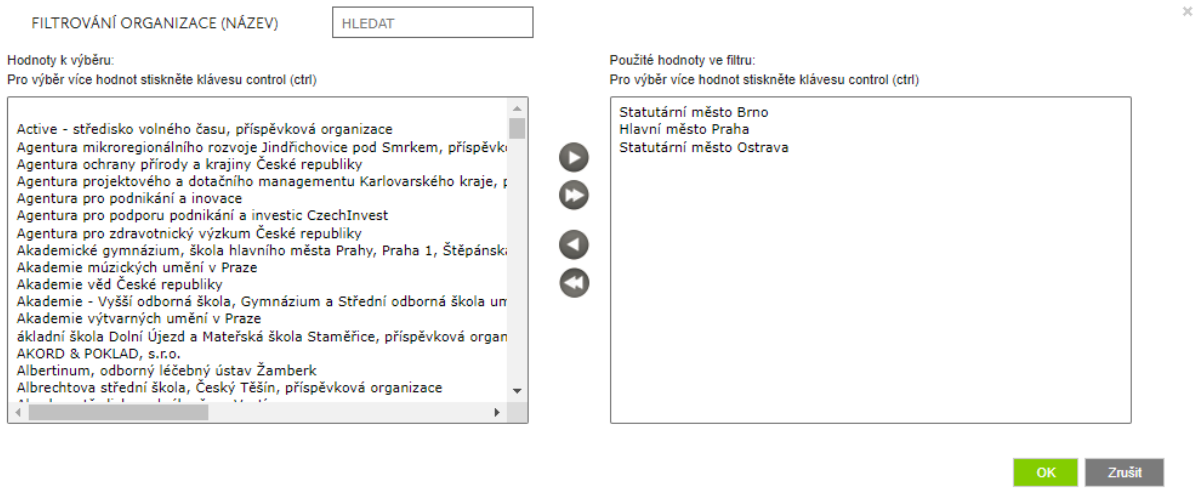
Místo zápisu jednotlivých položek, je lepší vždy kliknout na symbol filtru u položky.



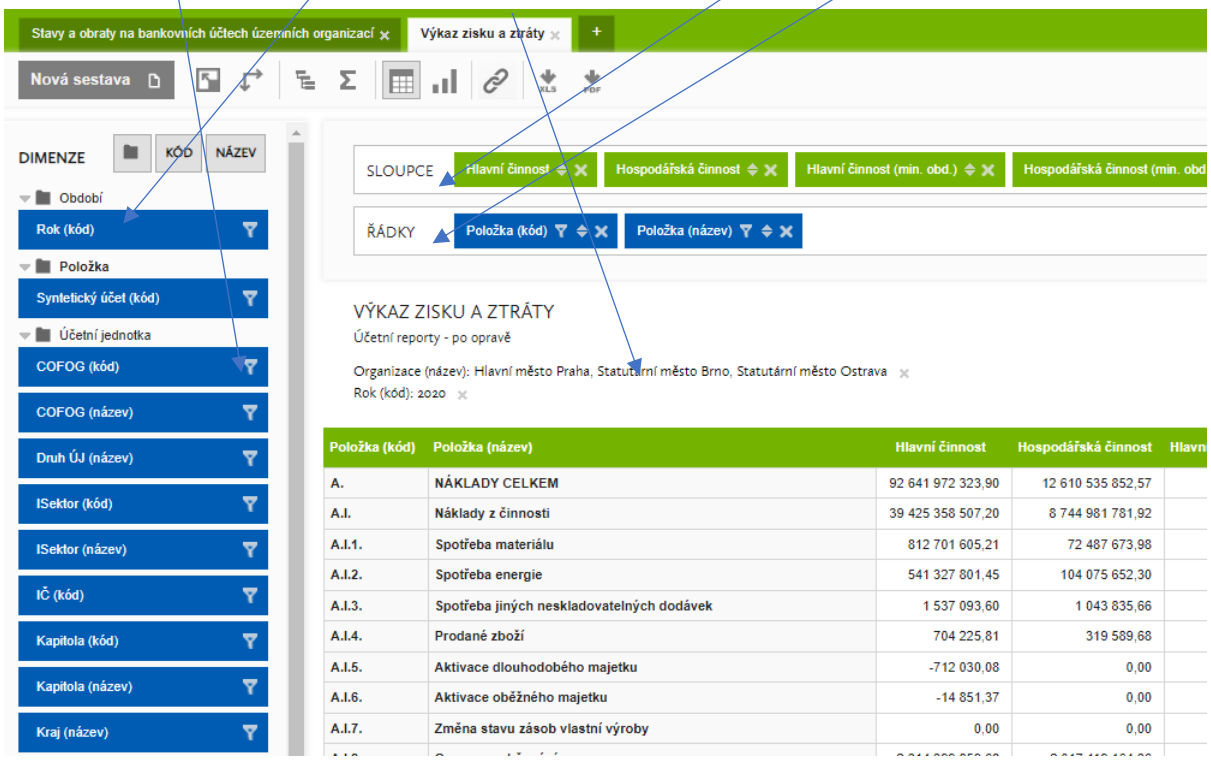
Ve filtru pomocí vyhledávání najdeme požadované hodnoty. Je možné jich přidávat více najednou. Výběr potvrdíme tlačítkem OK.

Jestliže použijeme více filtrů, jsou zobrazeny jen ty položky, které splňují všechny podmínky najednou.

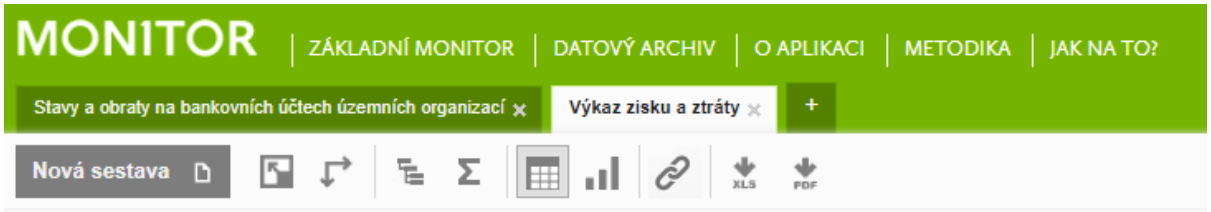
Po ukončení nastavování filtrů je zobrazena výsledná sestava (někdy je třeba obrnit se trpělivostí).



Vytvořená sestava obsahuje výsledky za všechny zvolené organizace, a to v součtu. Toto se ovšem dá dále upravit. Nejdůležitější jsou dva řádky v horní části tabulky nadepsané **SLOUPCE a ŘÁDKY**. Ty ukazují, jak bude výsledná tabulka vypadat. Sloupce ukazují zobrazované hodnoty a řádky pak jejich rozklad, rozklad je vždy **řazen** dle nastavení, záleží tedy na pořadí položek. Z levého podokna můžeme další položky rozkladu přidat (přetažením myši mezi vybrané řádky), případně kliknutím na značku **filtr** změnit původní nastavení filtrů. Vymazat nějakou vybranou položku pak můžeme kliknutím na mazanou položku v **přehledu nastavených filtrů** nad tabulkou.



V horní části okna pak máme několik dalších ovládacích prvků.



Tlačítko **Nová sestava** – vytvoří novou záložku a spustí úvodní okno analytického nástroje. Můžete si tak připravit víc přehledů a jen pomocí záložek mezi nimi přepínat.

Skrýt/zobrazit postranní panel – tato možnost ovládá viditelnost levého panelu s přehledem možných filtrů a možností rozkladů.

Prohodit osy – tato možnost by měla způsobit výměnu řádků za sloupce, tedy transpozici tabulky, bohužel u velké části reportů způsobí nefunkčnost analytické stránky.

Hierarchie – vytvoří report, který obsahuje hodnoty za hlavní skupiny a podskupiny si můžeme kliknutím na znaménko + rozbalit. Tlačítko funguje jako přepínač, dalším kliknutím na něj tedy hierarchii opět zrušíme.

Součtové řádky – kliknutí na tuto možnost způsobí přidání řádků pro jednotlivé skupiny, ve kterých jsou součty položek všech podskupin.

Zobrazit tabulku – jedná se o výchozí nastavení, tedy vidíme tabulku s daty, tímto tlačítkem ji zobrazíme znovu v případě, že si přepneme zobrazení na graf. Tato možnost po najetí nabídne další varianty, které přidají minigrafy ke každému řádku.

Zobrazit graf – možnost po kliknutí nabídne různé varianty grafů a po zvolení jsou data zobrazena místo tabulky v grafu.

Export do xls/pdf/png – jedná se o tři různá tlačítka, všechny způsobí vytvoření a stažení souboru zvoleného typu. Tabulky je možné exportovat do xls (ve skutečnosti je to novější formát xlsx) a do pdf. Grafy pak do png a pdf.

5.2.3. Datový katalog

Datový katalog umožňuje stáhnout data v podobě souboru a zpracovat je ve vlastní režii. Pro nás jsou nyní podstatné první dvě položky, tedy **Číselníky** a **Transakční data**.

Číselníky obsahují přehledy variant pro jednotlivá datová pole použitá při v datech. Jedná se tedy o seznamy pro všechny sloupce, které mají výčtový charakter. Například zde najdeme **přehled paragrafů**, seznam všech územních jednotek, rozpočtové položky a mnoho jiných.

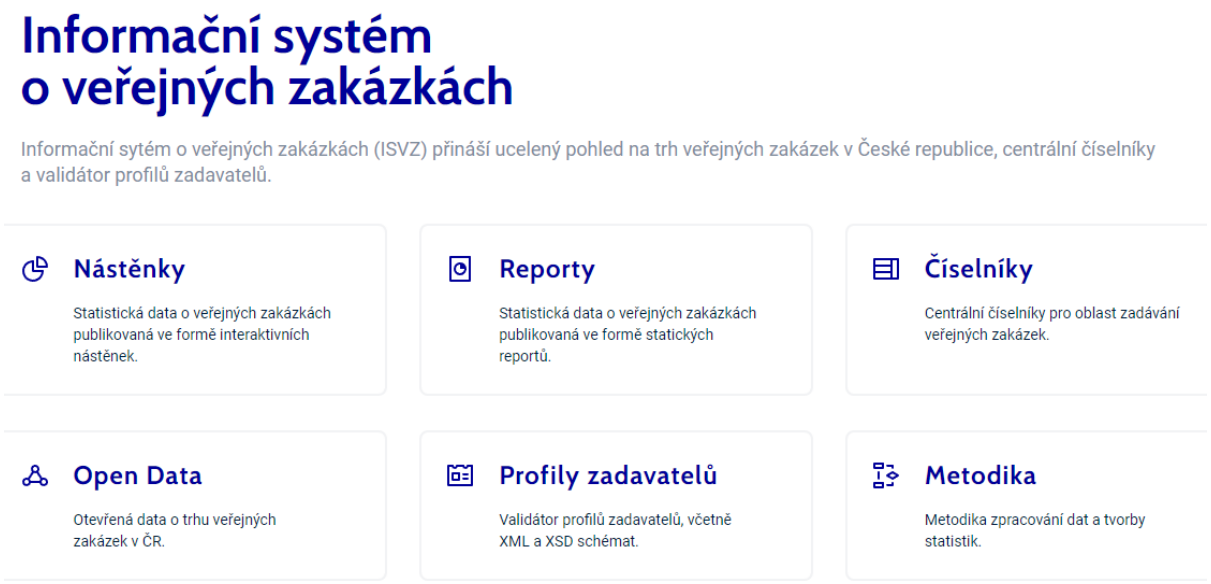
Číselníky si můžeme přímo prohlížet, nebo stáhnout ve formátu csv nebo xml.

Transakční data umožňují stažení datových sad, a to ve formátu csv. Pro stažení datové sady stačí rozbalit příslušnou položku a kliknout na časové období, které nás zajímá.

Formát csv je textový formát, který ukládá tabulková data, je tedy zpracovatelný textovými editory i tabulkovými kalkulátory a dalšími programy. Problémem těchto dat může být jejich rozsah. Některé tabulky obsahují i více než 10 miliónů řádků, a to běžný tabulkový kalkulátor nezvládne otevřít. Je tedy potřeba počítat i s možností stažená data předzpracovat ve specializovaném programu a poté je teprve použít v obvyklém kancelářském softwaru.

5.2.4. Informační systém o veřejných zakázkách (ISVZ)

Informační systém je dostupný na adrese https://isvz.nipez.cz/uvod a obsahuje všechny důležité informace o veřejných zakázkách. Data do tohoto systému vycházejí z věstníku veřejných zakázek, který spravuje ministerstvo pro místní rozvoj.



Jak vidíme je databáze rozdělena na 6 částí, v nichž najdeme souhrnné přehledy o veřejných zakázkách. Na tomto webu nenajdeme jednotlivé veřejné zakázky, ale pouze souhrny a statistiky.

Data o jednotlivých veřejných zakázkách pak můžeme vyhledávat ve výše zmíněném věstníku veřejných zakázek na adrese https://vestnikverejnychzakazek.cz/

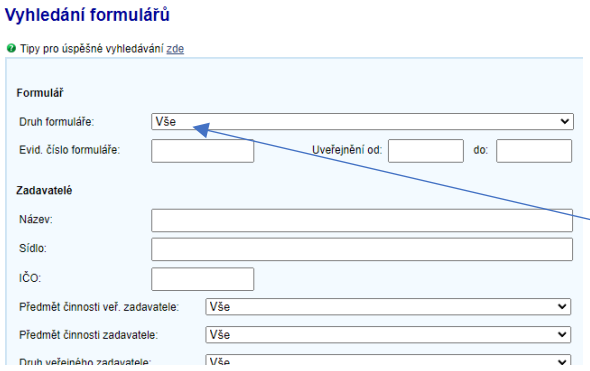
Najdeme zde jak veřejné zakázky, které již proběhly, nebo právě probíhají, tak veřejné zakázky, které byly vyhlášeny, a výběrové řízení zatím probíhá. Databázi tak lze použít jak pro kontrolu hospodaření organizace, nebo určení obvyklých cen pro vlastní veřejnou zakázku ale pro dodavatele primárně slouží k získání dokumentace pro účely podání vlastní nabídky.

5.2.5. Věstník veřejných zakázek

Vyhledávání ve věstníku může být z pohledu municipalit důležité z mnoha důvodů. Jedním může být průzkum před vypsáním výběrového řízení. Z věstníku můžeme zjistit, které společnosti již obdobné zboží, či služby poskytuje jiným veřejným organizacím, můžeme určit obvyklou cenu na trhu, obvyklý druh výběrového řízení, a další podobné informace.

Na úvodní stránce zvolíme možnost **Vyhledat formulář**

Formulář pro vyhledávání obsahuje velké množství možností.



Při vyplňování je třeba být dostatečně přesný, aby výsledků nebylo mnoho, z nichž většina nebude odpovídat Vaším požadavkům. Na druhou stranu, pokud vyhledávání omezíme moc je možné, že nedojde k žádné shodě.

Jestliže hledáme již uzavřené zakázky, abychom se mohli inspirovat již úspěšně ukončeným výběrovým řízením při vypisování vlastní zakázky, zvolíme formulář oznamující výsledky veřejné **zakázky**. Můžeme také filtrovat podle druhu zadavatele a dalších informací o zadavateli.

Jestliže naopak chceme znát historii společnosti, u které uvažujeme o zadání veřejné zakázky, je možné si zde vyhledat všechny veřejné zakázky, které tato společnost vyhrála.

Jestliže plánujete vypsat výběrové řízení na konkrétní věc, službu, či výstavbu, je vhodné nejdříve zjistit CPV kód a dle něj vyhledávání zadat. Vyhledávání pak najde odpovídající položky a my můžeme snáze zjistit obvyklý průběh tohoto výběrového řízení.

Přehled CPV je možné v číselnících Informačního systému o veřejných zakázkách:

https://isvz.nipez.cz/ciselniky

Zde najdete položku *Klasifikace kódů Common Procurement Vocabulary* a stáhnete soubor v některém z formátů, nejlépe xlsx. V tabulce, kterou poté otevřete, vyhledáte potřebný údaj a použijete CPV kód.

Například pokud chceme vybavit počítači učebnu školy najdeme řádek ***Počítače***.

30	Název kor	Elektromotory, generátory a transformátory	25939	05.02.2021	29	28	31100000-7
31	Název kor	Elektromotory	25941	05.02.2021	30	29	31110000-0
32	Název kor	Adaptéry	25943	05.02.2021	31	30	31111000-7
33	Název kor	Počítače	25945	05.02.2021	32	22	30200000-1
34	Název kor	Zařízení související s počítači	25947	05.02.2021	33	32	30230000-0
35	Název kor	Součásti, příslušenství a doplňky pro počítače	25949	05.02.2021	34	33	30237000-9

Odpovídající CPV kód je ve sloupci vpravo, s tím že kód, který zadáváme do formuláře je pouze první část, bez pomlčky a čísla za ní (8 číslic).

Vyhledávání tedy mohu zadat následovně

Druh formuláře: F03 – Oznámení o výsledku zadávacího řízení

Druh zakázky: Dodávky

CPV kód: 30200000

Ostatní položky necháme prázdné.

Evidenční číslo formuláře	Evidenční číslo zakázky	Název zadavatele	Název zakázky	Druh	Datum uveřejnění
F2022-027177	Z2022-016820	Město Hořovice	Modernizace jazykové a ICT učebny a rekonstrukce konektivity v budovách 2. ZŠ Hořovice	F03	18.07.2022
F2022-003344	Z2021-031309	Centrum uznávání a celoživotního učení Pardubického kraje o.p.s.	Jazykové učebny pro projekt I KAP II Pardubického kraje	F03	24.01.2022
F2021-035659	Z2021-035659	Univerzita Jana Evangelisty Purkyně v Ústí nad Labem	Vybaveni uceben IT technikou PF v DNS	F03	04.10.2021
F2021-035467	Z2021-018000	Univerzita Karlova, Lékařská fakulta v Plzni	U2 vybavení - dodávka datové infrastruktury	F03	04.10.2021
F2021-030055	Z2021-003818	Statutární město Zlín	Vybavení odborných učeben pro 10 základních škol ve Zlíně	F03	19.08.2021
F2021-028797	Z2021-019373	Městské gymnázium a Základní škola Jirkov	MG a ZŠ Jirkov, vybavení odborných učeben a stavební úpravy objektu	F03	09.08.2021
F2021-026595	Z2021-013436	STATUTÁRNÍ MĚSTO KARVINÁ	„REKONSTRUKCE ODBORNÝCH UČEBEN ZŠ A MŠ BOROVSKÉHO A ZŠ A MŠ CIHELNÍ V KARVINÉ - DODÁVKY“	F03	26.07.2021

Je nalezeno množství výsledků, které můžeme ještě redukovat například pomocí roku, klíčového slova, zde pomůže zadat slovo učebny, případně vyhledávání udělat znovu i se slovem učebna.

Nalezneme následující výběrová řízení, která můžeme rozkliknout a uvidíme všechny potřebné informace. Je samozřejmě vhodné udělat si větší přehled a nepoužít první položku, kterou nalezneme. Tyto informace by měly sloužit k odhadu ceny například pomocí průměru cen několika výsledků.

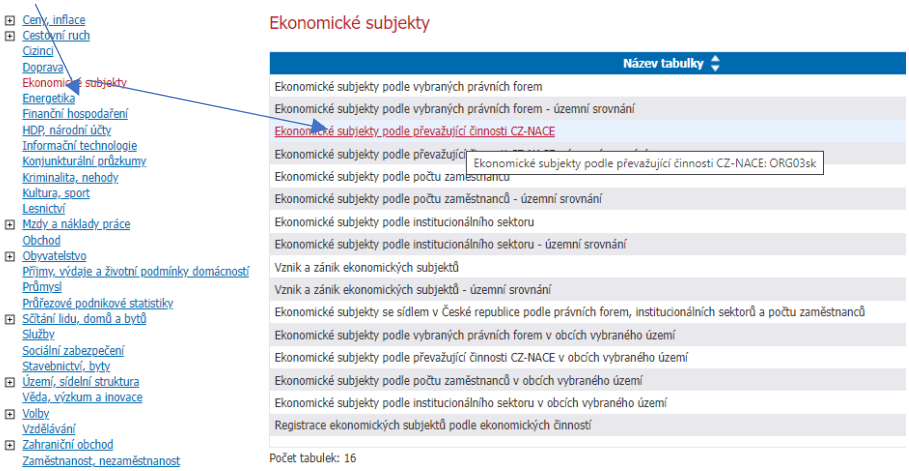
5.3. Český statistický úřad

Český statistický úřad spravuje a poskytuje nejrozsáhlejší databázi využitelných informací. Pravidelně vydává různé přehledy, statistiky, studie, či reporty. I ty je možné využít, ale pro vlastní zpracování pro strategickou práci, je často nutné využít základní data. Ty najdeme ve **veřejné databázi**.

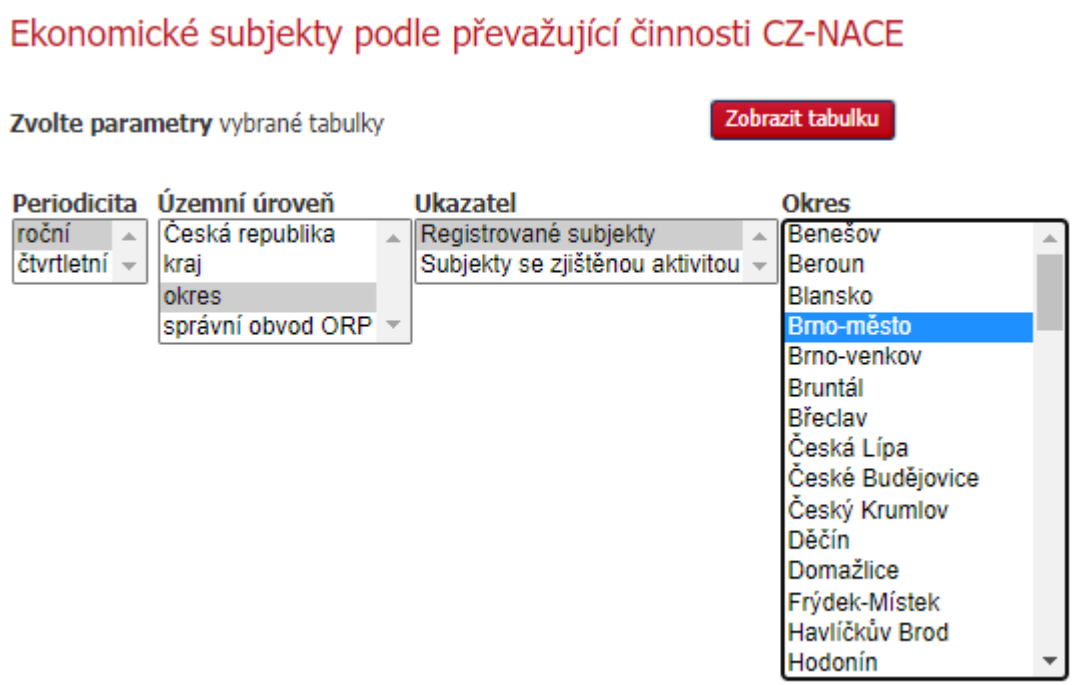
Veřejná databáze obsahuje 4 možnosti jak získávat data.

5.3.1. Statistiky

Statistiky obsahují předem definované přehledy, nám tedy stačí vybrat z mnoha typů **ukazatelů**. V zobrazené tabulce reportů poté vybereme požadovaný

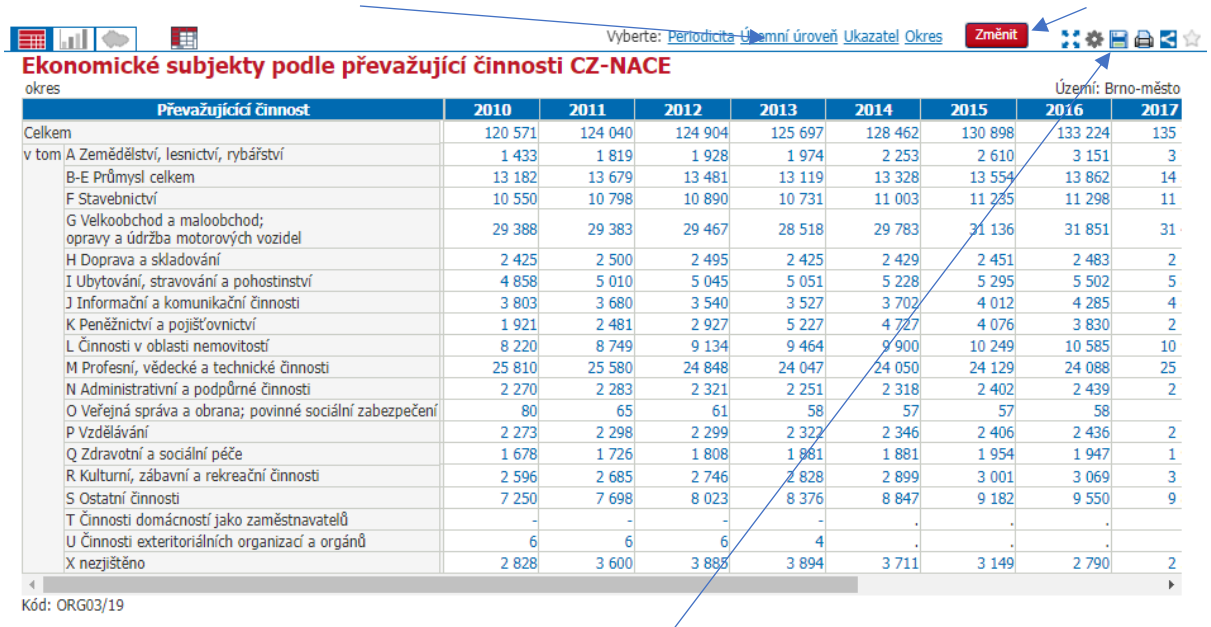


Následně se zobrazí stránka s nastavením podrobnějších parametrů. Vždy záleží na konkrétním ukazateli a předpřipraveném typu reportu, jaké parametry jsou nastavovány. Mohou to být, periodicita, požadovaná oblast (ČR, kraj, okres, město), či zpřesnění ukazatele.



Po nastavení požadovaných parametrů zvolíme tlačítko Zobrazit tabulku.

Výsledná tabulka obsahuje požadované údaje v dané oblasti. Předem nastavené parametry zde můžeme měnit a to **jednotlivě**, nebo se vrátit k předchozí stránce a nastavit **vše najednou**.



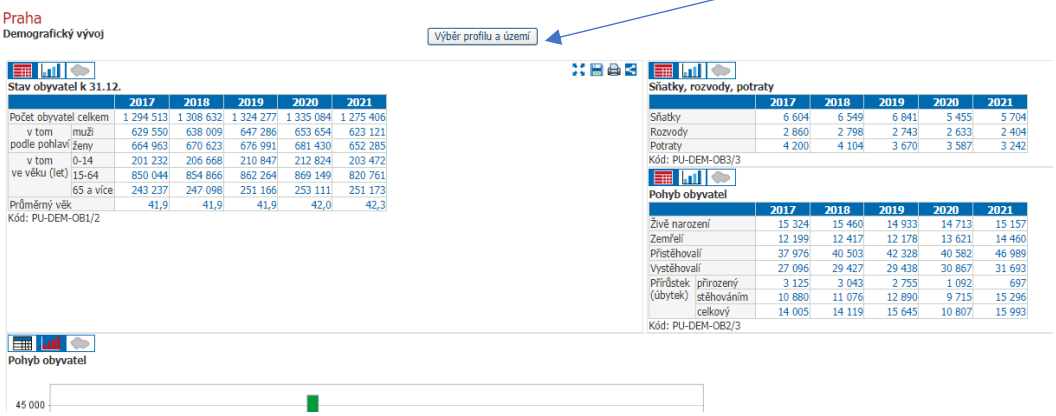
Převažující činnost	2010	2011	2012	2013	2014	2015	2016	2017
Celkem	120 571	124 040	124 904	125 697	128 462	130 898	133 224	135
v tom A Zemědělství, lesnictví, rybářství	1 433	1 819	1 928	1 974	2 253	2 610	3 151	3
B-E Průmysl celkem	13 182	13 679	13 481	13 119	13 328	13 554	13 862	14
F Stavebnictví	10 550	10 798	10 890	10 731	11 003	11 235	11 298	11
G Velkoobchod a maloobchod; opravy a údržba motorových vozidel	29 388	29 383	29 467	28 518	29 783	31 136	31 851	31
H Doprava a skladování	2 425	2 500	2 495	2 425	2 429	2 451	2 483	2
I Ubytování, stravování a pohostinství	4 858	5 010	5 045	5 051	5 228	5 295	5 502	5
J Informační a komunikační činnosti	3 803	3 680	3 540	3 527	3 702	4 012	4 285	4
K Peněžnictví a pojišťovnictví	1 921	2 481	2 927	5 227	4 727	4 076	3 830	2
L Činnosti v oblasti nemovitostí	8 220	8 749	9 134	9 464	9 900	10 249	10 585	10
M Profesní, vědecké a technické činnosti	25 810	25 580	24 848	24 047	24 050	24 129	24 088	25
N Administrativní a podpůrné činnosti	2 270	2 283	2 321	2 251	2 318	2 402	2 439	2
O Veřejná správa a obrana; povinné sociální zabezpečení	80	65	61	58	57	57	58	
P Vzdělávání	2 273	2 298	2 299	2 322	2 346	2 406	2 436	2
Q Zdravotní a sociální péče	1 678	1 726	1 808	1 881	1 881	1 954	1 947	1
R Kulturní, zábavní a rekreační činnosti	2 596	2 685	2 746	2 828	2 899	3 001	3 069	3
S Ostatní činnosti	7 250	7 698	8 023	8 376	8 847	9 182	9 550	9
T Činnosti domácností jako zaměstnavatelů	-	-	-	-	.	.	.	
U Činnosti exteritoriálních organizací a orgánů	6	6	6	4	.	.	.	
X nezjištěno	2 828	3 600	3 885	3 894	3 711	3 149	2 790	2

Výslednou tabulku můžeme snadno **exportovat do souboru**, kdy zvolíme z formátů xlsx, xml a pdf.

5.3.2. Vše o území

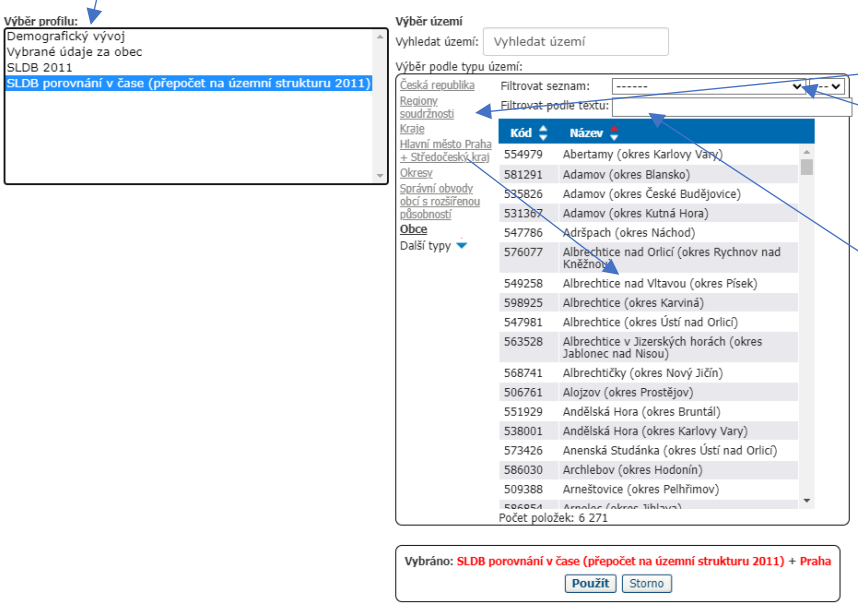
Tento nástroj umožňuje snadno zobrazit údaje o konkrétním území a opět je možné data stáhnout ve zvoleném formátu.

Územím v tomto případě rozumíme celou Českou republiku, jednotlivé kraje, regiony soudržnosti, správní obvody, okresy, obce, nebo městské části. Které z uvedených možností jsou možné, závisí na požadovaných datech. Některé se sbírají jen za kraje jiné, jsou **detailnější** a mohou být až do jednotlivých městských částí. Tuto volbu provedeme po kliknutí na tlačítko, Výběr profilu a území.



Poté **vybereme profil**. Profilem rozumíme sadu ukazatelů, které chceme zobrazit. Demografická data obsahují základní přehled o obyvatelstvu ke konci předchozího roku.

Vybrané údaje za obec obsahují mimo základního přehledu o obyvatelích i informace o pozemcích, podnikatelských subjektech, a dalších ukazatelích. SLDB obsahuje data ze sčítání lidu, domů a bytů, s tím že můžeme volit zobrazení údajů z posledního sčítání, nebo přehled za všechna sčítání od roku 1970.

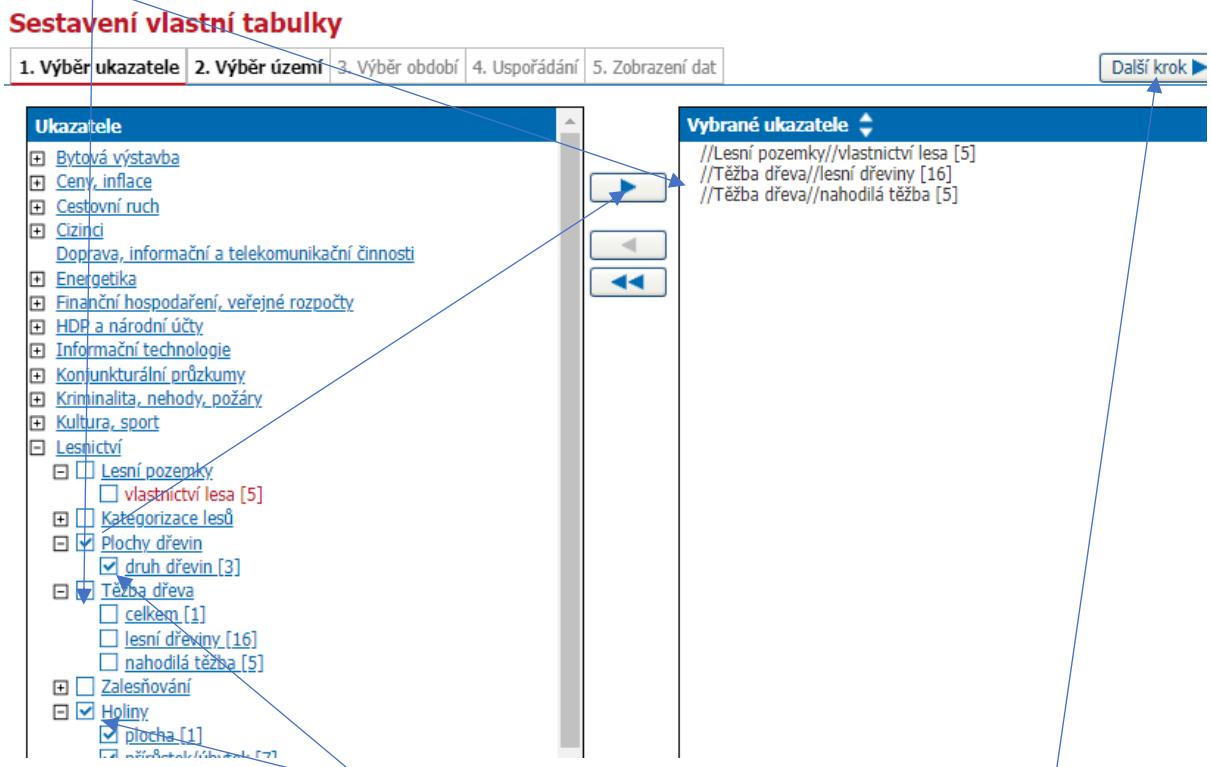


V pravé části pak vybereme **typ území** a poté konkrétní územní jednotku.

Pro snazší vyhledávání můžeme použít **filtry**, které nám umožní zobrazit například obce jednoho okresu, nebo vyhledávat pomocí **zadaného textu**.

5.3.3. Vlastní výběr

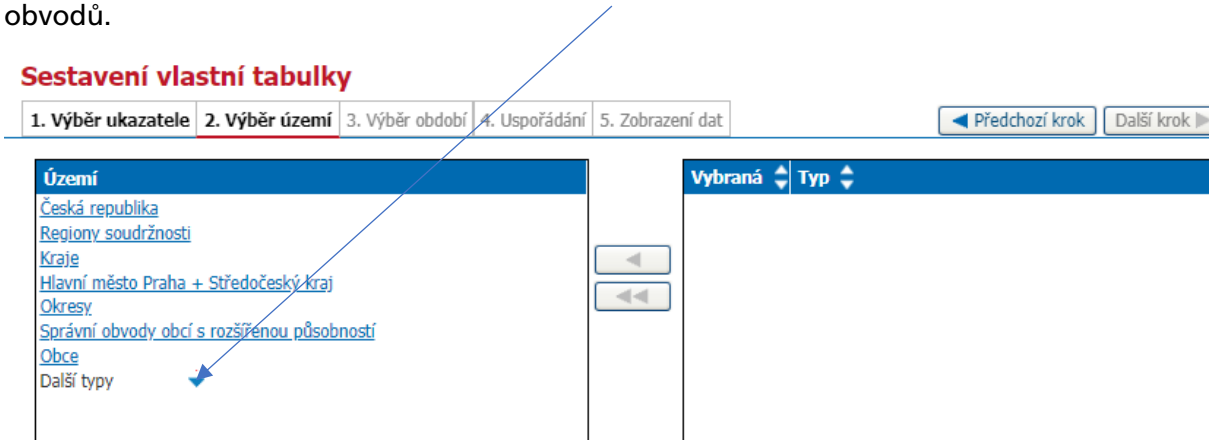
Vlastní výběr je způsob, jak dostat přesně ta data, která požadujeme, ale za cenu nutnosti sami nastavit vzhled tabulky. V prvním kroku je třeba vybrat požadované ukazatele. Ty jsou členěny dle typu do různých skupin, které lze rozkliknout a vybrat požadované ukazatele. Ukazatel je třeba **vždy zvolit** a pomocí ikony šipky vpravo přidat do seznamu. Zvolené ukazatele poté vidíme v pravé **straně**

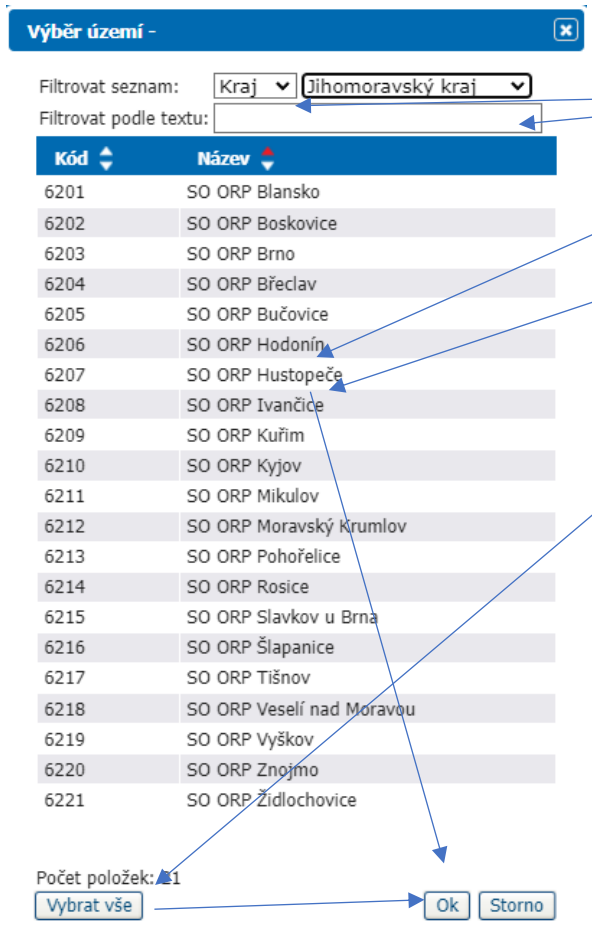


Ukazatelů můžeme zvolit i **více najednou**, nebo každý z nich přidávat jednotlivě.

Pokud máme vybrány všechny požadované ukazatele, klikneme na tlačítko **Další krok**.

Druhý krok spočívá ve výběru území zájmu. Můžeme volit dle různých rozsahů od celé ČR až po jednotlivé obce. Kliknutím na zvolený rozsah se zobrazí okno ve kterém zvolíme konkrétní území. Například pokud zvolíme ***Správní obvody obcí s rozšířenou působností***, zobrazí se přehled všech obvodů.





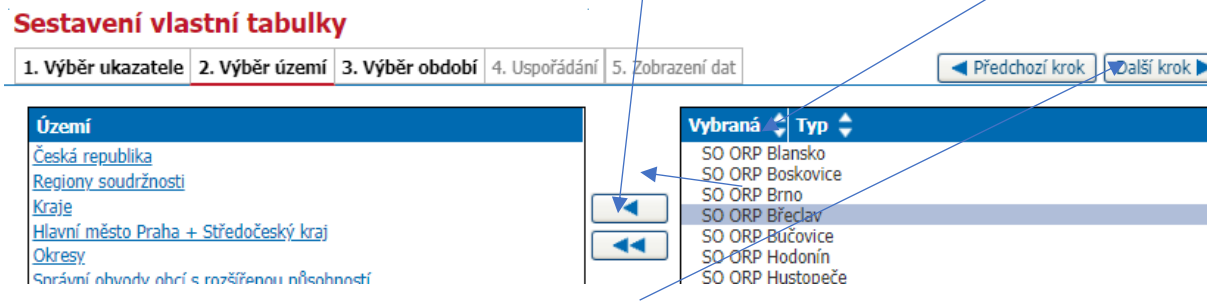
Z nich můžeme vybrat **procházením seznamu**, ale také můžeme vyhledávat dle zadaného textu, anebo můžeme **zvolit filtr** v tomto případě dle kraje a zobrazí se nám všechny obvody ve zvoleném kraji.

Z vybraného seznamu můžeme klikáním označit **jeden nebo více najednou** a kliknutím na tlačítko **Ok**, je můžeme přidat do seznamu. Také můžeme použít tlačítko **Vybrat vše**, kdy dojde k výběru všech položek, které splňují zadaný filtr, a v případě kdy filtr není zadán, jsou vybrány úplně všechny položky, výběr opět potvrdíme tlačítkem **Ok**.

U výběru všech území je třeba dát si pozor při výběru velkého množství položek najednou, například všech obcí ČR, systém má omezení počtu zobrazených položek a takto zadaný dotaz poté nezpracuje. Je tedy třeba položky zadávat postupně například po okresech.

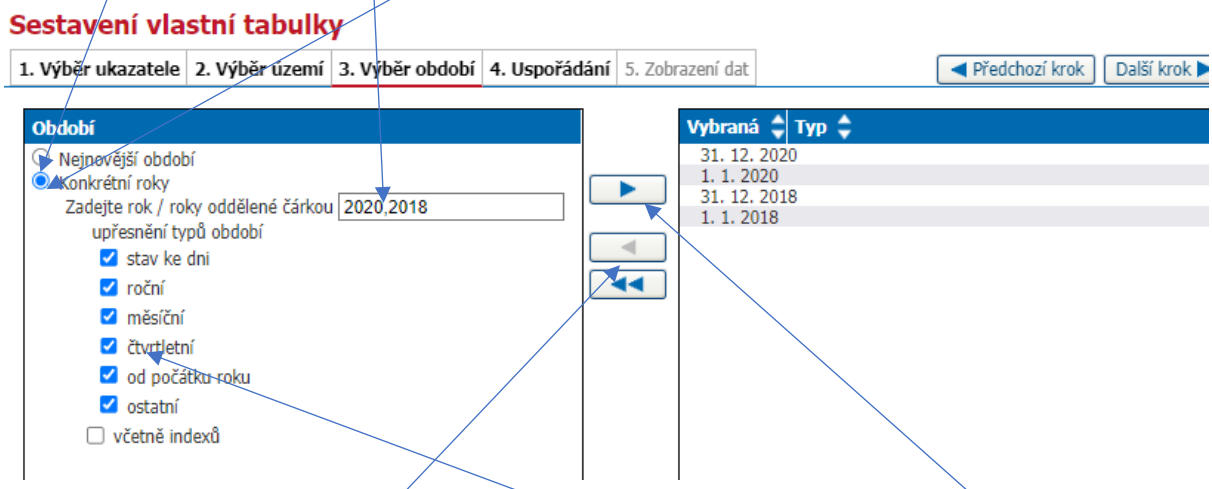
Vybraná území, jsou pak zobrazena v pravé části okna. Jestliže se rozhodneme, že některá z nich použít nechceme, můžeme jej kliknutím **označit** a pomocí **šipky** ze seznamu odstranit.

Dvě šipky odstraní všechna vybraná území.



Výběr ukončíme kliknutím na tlačítko **Další krok**.

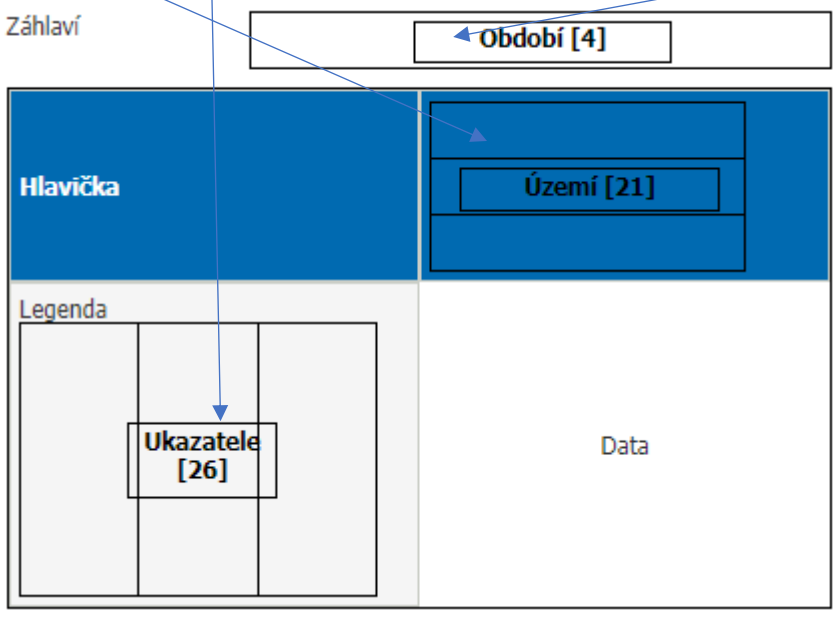
Po výběru území zvolíme časové období. Zde vybíráme ze dvou základních možností. První je **nejnovější období**, a druhá pak **konkrétní roky**. U volby konkrétní roky je pak třeba určit rok, **zapsáním do připraveného pole**, let můžeme zadat i více, s tím, že mezi nimi musíme napsat čárku.

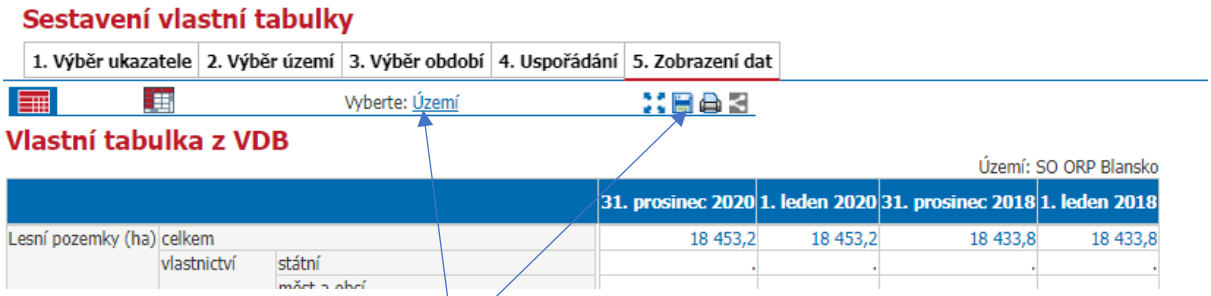


Dále zvolíme jaké typy časových řad v daném roce nás zajímají. Vzhledem k tomu, že často předem nevíme v jakých intervalech jsou údaje sledovány (každý ukazatel může mít jinou frekvenci), bývá při prvním zobrazení ukazetele vhodné **zvolit všechny možnosti**. Kliknutím na **šipku vpravo** pak vložíme všechna období, které odpovídají výběru.

Pomocí **šipek vlevo** můžeme opět ta období, která byla vložena navíc a jejichž výsledky zobrazit nepotřebujeme z výběru odstranit, před odstraněním je opět musíme označit kliknutím. Výběr období potvrdíme tlačítkem **Další krok**.

Krok 4 pak určuje výsledný vzhled tabulky, vybíraná data jsou podle tři kritérií: ukazatele, území a období. Tato kritéria jsou nyní zobrazena ve schématu výsledné tabulky. Ta obsahuje **Záhlaví, Hlavičku a Legendu**. My můžeme jednotlivá kritéria tažením myši přesout v těchto třech polích. S tím, že co vložíme do záhlaví, bude mít funkci filtru, tedy bude ve výsledné sestavě možné přepínat jednotlivé tabulky dle tohoto kritéria. Hlavička pak značí, co bude v popiscích sloupců a Legenda pak určuje, co bude v řádcích. Lze také vložit více kritérií do jedné části, kromě záhlaví, které může obsahovat nejvýše jedno kritérium. Pořadí pak určuje způsob členění. Tlačítkem **Zobrazení dat** pak již zobrazíme výslednou tabulku.



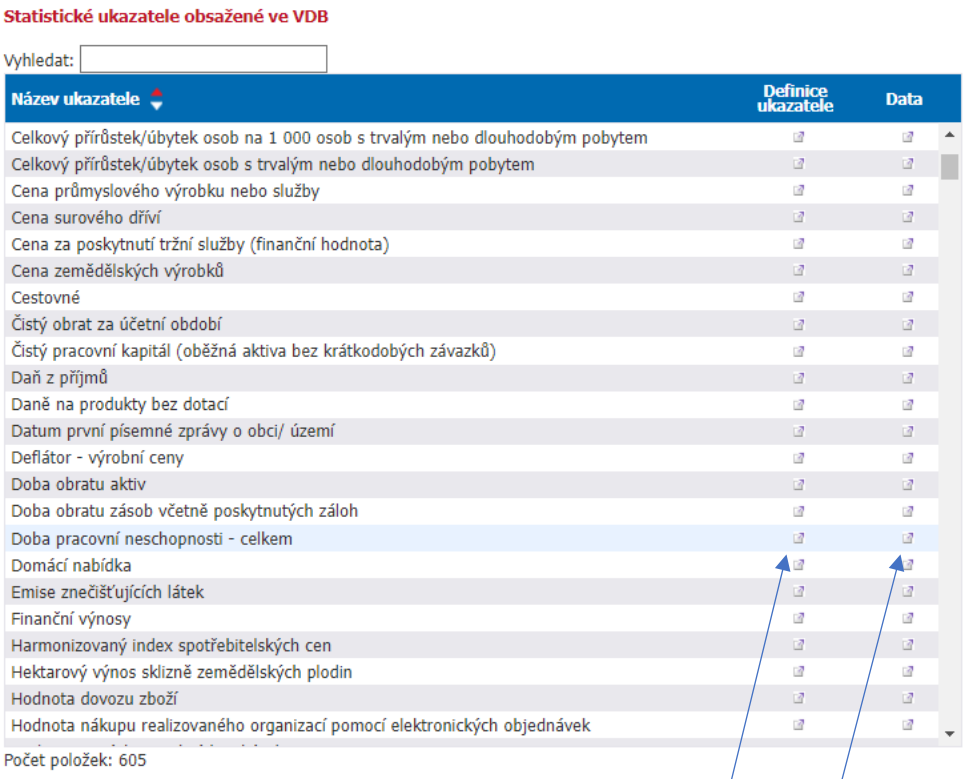


Výsledná sestava může obsahovat **filtr**, pomocí kterého zobrazujeme jednotlivé tabulky. Zobrazené výsledky také můžeme **exportovat** do souboru, který již zpracujeme ve vlastním programu.

5.3.4. Ukazatele

Položka Ukazatele slouží hlavně k získání informací o jednotlivých ukazatelích, které jsme v předchozích kapitolách získávali. Z názvu ukazatele nemusí být jednoznačné, co přesně udává. Tento přehled pro každý ukazatel obsahuje popis dat a další informace o času, případně podrobnosti územního dělení.

První stránka, kterou vidíme, obsahuje přehled všech ukazatelů, které jsou obsaženy v databázi CZSO. Vidíme, že je jich aktuálně 605. Nad tabulkou je vyhledávací pole, které již při zadávání textu filtruje ukazatele a zobrazuje ty, které zadaný text obsahují.



U každého ukazatele máme ikony pro zobrazení jeho **definice a dat**.

Definice poskytuje základní informace o ukazateli. Velmi důležitá je platnost ukazatele, která udává, v jakém období byly data evidována. Dále obsahuje definici, typ a seznam souvisejících ukazatelů.

Doba pracovní neschopnosti - celkem

Detaily:

Plný název ukazatele:	Doba pracovní neschopnosti - celkem
Kód:	3605
Téma:	Zdraví
Platnost:	01.01.2009 - 09.09.9999
Definice:	1. Počet kalendářních dnů pracovní neschopnosti zahrnuje neschopnosti se zjistí na základě hlášení o počátku a ul
Synonyma:	1. Počet kalendářních dnů pracovní neschopnosti celkem

Atributy:

Časová charakteristika:	toková
Datový typ:	numerický
Skupina:	NEM - Pracovní neschopnost, nemoci a úrazy

Seznam souvisejících ukazatelů:

Doba pracovní neschopnosti - celkem
- Doba pracovní neschopnosti - ostatní úrazy
- Doba pracovní neschopnosti - ostatní úrazy - ženy
- Doba pracovní neschopnosti - pracovní úrazy
- Doba pracovní neschopnosti - pracovní úrazy (15 -17 let)
- Doba pracovní neschopnosti - pracovní úrazy - ženy
- Doba pracovní neschopnosti - pro nemoc
- Doba pracovní neschopnosti - pro nemoc z povolání
- Doba pracovní neschopnosti - pro nemoc z povolání - ženy
- Doba pracovní neschopnosti - ženy

Při zobrazení dat pak vidíme přehled všech tabulek, které tento ukazatel obsahují. U každé tabulky vidíme období, tedy, za které roky jsou data zobrazena, území, to nám udává informaci o tom, jak podrobná data obsahuje. Dále vidíme periodicitu (roční čtvrtletní, měsíční,…), informaci zda tabulka obsahuje základní data, ty které je neobsahují, jsou přepočítány pomocí dalších ukazatelů. ČR pak značí, zda je v tabulce i sledované časová řada, a tedy je možné z ní vyčíst vývoj ukazatele v čase.

Tabulky pro Doba pracovní neschopnosti - celkem

Název tabulky	Období	Území	P	Z	ČR
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ OSVČ v krajích ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ v ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ v ČR dle sekcí, oddílů CZ-NACE	2017-2021	Stát	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ v krajích a okresech ČR	2017-2021	Stát, KR, OKR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ v nefinančních podnicích v ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz na 1 nově hlášený případ ve zpracovatelském průmyslu v ČR	2018-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz OSVČ v krajích ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz v ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz v ČR dle sekce, oddílů CZ-NACE	2017-2021	Stát	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz v krajích a okresech ČR	2017-2021	Stát, KR, OKR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz v nefinančních podnicích v ČR	2017-2021	Stát, KR	R		
Počet kalendářních dnů pracovní neschopnosti pro nemoc a úraz ve zpracovatelském průmyslu v ČR	2018-2021	Stát, KR	R		
Pracovní neschopnost pro nemoc a úraz	2006-2018	Stát, KR, OKR	R		
Pracovní neschopnost pro nemoc a úraz - výběr období	2006-2020	Stát, KR, OKR	R	✓	
Pracovní neschopnost pro nemoc a úraz - výběr území	2006-2020	Stát, KR, OKR	R	✓	✓

Kliknutím na některou z tabulek zobrazíme požadovaná data. U některých tabulek je před zobrazením dat ještě potřeba zvolit podrobnosti, obvykle území, které nás zajímá.

6. Využití matematických a statistických metod

Použití statistických metod je nutné, abychom dokázali odlišit skutečné informace od zdánlivých. Velmi často se můžeme setkat s tím, že jsou nesprávně určovány vlastnosti zkoumaných dat, které jsou ovšem ve skutečnosti nevýznamné a založené jen na náhodných výkyvech. Hlavní chybou tak bývá zobecňování určité vlastnosti i v případě, kdy toto zobecnění není možné. Můžeme si tento problém přiblížit na jednoduchém příkladu.

Pokud chceme vědět, zda je určitá vlastnost statisticky významná, použijeme některou ze statistických metod. Těch existuje celá řada a zde si představíme ty nejčastěji používané.

6.1 Popisná statistika

Popisná statistika, jak již naznačuje název má za cíl popsat chování sledované proměnné ve stručném formátu, který ovšem sdělí hlavní charakteristiky proměnné. Charakteristiky popisují určitou vlastnost proměnné, obvykle pomocí jednoho čísla, které poté můžeme porovnávat s jinými charakteristikami, a to jak stejné proměnné získané z jiného vzorku, nebo jiné proměnné.

Základní a nejčastěji používané charakteristiky jsou charakteristiky polohy. Velmi zjednodušeně můžeme říct, že tyto charakteristicky popisují hodnotu na číselné ose, kde se vyskytují naměřené hodnoty. Například zjistíme, že hodnoty jsou spíše v desítkách a ne stovkách. Do této skupiny patří velké množství charakteristik, kdy každá je vhodná pro jiná data. Nejčastěji se používají průměry a kvantily.

Další již méně často používanými charakteristikami, ale pořád velmi důležitými, jsou charakteristiky variability. Ty nám udávají, jaká je rozdílnost naměřených dat. Jejich hodnoty jsou obvykle od 0 a platí, že pokud je variabilita nula, znamená to, že naměřené hodnoty se vůbec neliší, všechny jsou zcela stejné. Naopak čím vyšší variabilita, tím rozdílnější výsledky jsme získaly. Můžeme si to snadno uvést na příkladu platů. Pokud mzdy určité skupiny zaměstnanců vycházejí z daných tabulek a všichni patří do stejné skupiny, kdy nedochází k velkým změnám udělováním odměn, bude variabilita zjištěných platů malá, pokud by nebyly odměny vůbec, bude variabilita nulová. Naopak pokud platy v organizace neupravuje žádný předpis, ale je to dáno jen dohodovou mezi zaměstnancem a zaměstnavatelem můžeme variabilitu platů očekávat větší.

Existují i další druhy charakteristik, jako jsou charakteristiky tvaru, které nám udávají například šikmost a špičatost proměnné, ale ty již nejsou tak často používány, a proto je zde vynecháme.

6.1.1. Průměry

Pokud uslyšíme slovo průměr, obvykle se tím myslí aritmetický průměr, bohužel stále velká část populace ani netuší, že průměrů je ve skutečnosti mnohem víc a každý je vhodný na jiný typ dat. Zde se podíváme na tři nejčastěji používané.

Aritmetický průměr

Jak již bylo popsáno, tento průměr je nejčastěji používaný. Je vhodný pro většinu dat, a tak platí pravidlo, že pokud nevidíme důvod použít jiný průměr, sáhneme po aritmetickém průměru.

Aritmetický průměr se určí tak, že sečteme všechny naměřené hodnoty a výsledek vydělíme počtem těchto hodnot. Existuji dvě varianty výpočtu tohoto průměru. První je tzv. Prostý průměr a druhý

Vážený průměr. Je třeba si uvědomit, že tyto průměry se odkazují na použití vzorce, ale vždy platí, že pokud použijeme jakýkoliv postup správně, dostaneme vždy sejný průměr. Aritmetický průměr je tedy jen jeden a rozdílnost je dána jen použitým vzorcem.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{n} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n} = \sum_{i=1}^{k} x_i \cdot p_i$$

Zde x_i značí i-tou hodnotu proměnné (konkrétní zjištěné číslo), případně konkrétní obměna. Hodnoty n_i jsou pak absolutní četnosti odpovídající obměny, tedy konkrétní počet kolikrát se dané číslo v souboru vyskytuje, p_i jsou pak relativní četnosti, které dávají poměr počtu určité obměny k celkovému počtu hodnot. Relativní četnost se dá snadněji chápat jako číslo, které po vynásobení 100 udává informaci o počtu procent hodnot, které odpovídají dané obměně. Například pokud $x_1 = 1$ a $p_1 = 0{,}3$, tak to znamená, že 30 procent ze všech naměřených hodnot byly jedničky (jinak řečeno, tři desetiny všech čísel jsou jedničky).

Pokud pro výpočet použijeme Excel, můžeme využít funkci PRŮMĚR.

Geometrický průměr

Geometrický průměr je vhodný v případě, kdy data jsou navzájem ovlivňována a funguje u nich myšlenka složeného úrokování. Jedná se obvykle o indexy v časové řadě, například inflace v jednotlivých letech, nárůst mezd sledovaný třeba měsíčně, a podobné.

Od průměru očekáváme, že pokud by hodnoty byly v celém období stejné (průměrné) tak by ve výsledku na konci vše odpovídalo skutečnosti. Tedy například pokud je průměrné zvýšení ceny konkrétního výrobku za 10 let 5 procent, tak by to mělo značit, že pokud pro každý rok zvýšíme původní cenu o 5 procent, dostaneme cenu, která je po 10 letech skutečná. Bohužel se ukazuje, že aritmetický průměr tuto vlastnost nemá, a proto není vhodné jej v takovém případě použít.

Využijeme tedy geometrický průměr. Zde je nutné, aby hodnoty, které používáme, nebyly v procentech ale jako koeficient růstu nebo poklesu, což je hodnota, kterou pokud vynásobíme původní číslo, dostaneme nové. Například při zvýšení ceny o 5 procent je nutné původní cenu násobit hodnotou 1,05. Všechny procenta tedy musíme na tento tvar převést. Poté již můžeme použít následující vzorce.

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[(n_1+n_2+\ldots+n_k)]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}} = \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}}$$

Zde x_i značí i-tou hodnotu proměnné (je nutné používat výhradně koeficienty růstu/poklesu), případně konkrétní obměna. Hodnoty n_i jsou pak absolutní četnosti odpovídající obměny, tedy konkrétní počet, kolikrát se dané číslo v souboru vyskytuje.

Pokud pro výpočet použijeme Excel, můžeme využít funkci GEOMEAN.

Harmonický průměr

Harmonický průměr se používá v případě, kdy máme data udávající časy výroby u opakující se produkce. Také jej lze použít pří výpočtu průměrných rychlostí, pokud známe ujeté vzdálenosti.

Použít jej můžeme také tehdy, když jsou data nerovnoměrně rozložené kolem aritmetického průměru, nebo obsahují extrémně malé případně extrémně velké hodnoty.

$$\bar{x}_H = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n_1+n_2+\ldots+n_k}{\frac{n_1}{x_1}+\frac{n_2}{x_2}+\ldots+\frac{n_k}{x_k}} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k}\frac{n_i}{x_i}}$$

Zde x_i značí i-tou hodnotu proměnné (je nutné používat výhradně koeficienty růstu/poklesu), případně konkrétní obměna. Hodnoty n_i jsou pak absolutní četnosti odpovídající obměny, tedy konkrétní počet, kolikrát se dané číslo v souboru vyskytuje.

Pokud pro výpočet použijeme Excel, můžeme využít funkci HARMEAN.

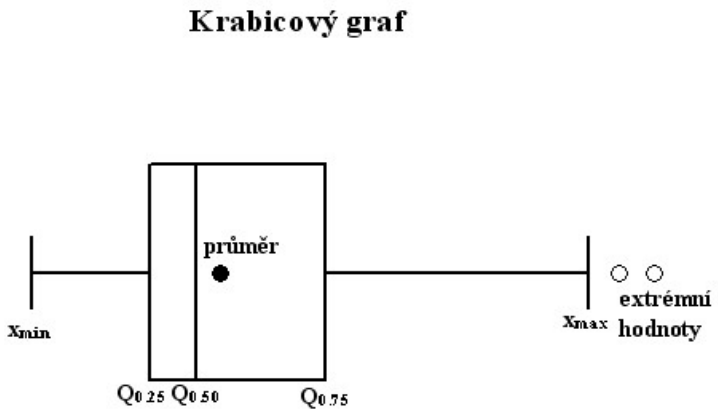
6.1.2. Kvantily

Jestliže máme hodnoty, které jsou nerovnoměrně rozdělené, mohou být pro jejich popis vhodné kvantily. Kvantil je hodnota proměnné, u které můžeme říct, že určité procento všech hodnot je menší než tento kvantil a zbytek hodnot je větší. Použité procento je variabilní a můžeme si sami určit, jaké potřebujeme. Kvantil značíme Q_k, kde k po vynásobení stem odpovídá procentu menších hodnot.

Nejčastěji používaným kvantilem je medián značíme jej $Q_{0.5}$ případně $\tilde{x}$. Jak vidíme i z označení, medián je kvantil, u kterého platí, že padesát procent všech hodnot je menších a padesát procent je větších. Proto se také nazývá prostřední hodnota. Například u platů bývá vhodnější než často uváděný průměr.

Často se počítají také kvartily, což jsou celkem tři kvantily, které rozdělí data na části o 25 procentech. Jedná se tedy o kvantily $Q_{0.25}$, $Q_{0.5}$, $Q_{0.75}$. Vidíme, že medián mezi ně patří. Pokud se budeme bavit například o cenách například benzínu v určitý den na různých pumpách, a určíme kvantily, tak získáme přehled, do jaké ceny můžeme benzín koupit u 25 procent nejlevnějších pump. Jaké je rozpětí cen u 25 procent dražších ale stále patřících do levnější poloviny, rozpětí cen 25 procent, které nejsou nejdražší, ale již patří mezi polovinu dražších, až po poslední skupinu pump, které benzin prodávají dráž, než je horní kvantil, a tedy patří mezi 25 procent nejdražších.

Kvartily můžeme využít k vytvoření krabicového grafu, který nám toto graficky zobrazí rozdělení dat na 25% části.



Kvantily často používáme pro určení a následné vynechání extrémních hodnot, které by nám výsledky zkreslovali.

Kvantily můžeme určit, pokud si data seřadíme a určíme si, jaká hodnota odpovídá hledaným procentům. Případně je možné použít funkce QUARTIL nebo PERCENTIL v Excelu, případně naši šablonu.

6.1.3. Charakteristiky variability

Tyto charakteristiky popisují rozdílnost hodnocených dat. Zde se zaměříme na nejčastěji používané charakteristiky.

Variační rozpětí

Variační rozpětí je nejjednodušší charakteristikou variability, jedná se o rozdíl největší hodnoty a nejmenší hodnoty. Dá nám tak první představu o rozdílnosti dat, i když v tomto případě se jedná o rozdílnost extrémních hodnot a tedy nemusí být moc směrodatná, i když pro první představu o datech stačí.

Rozptyl

Rozptyl udává rozdílnost dat tak, že počítáme rozdíl každé hodnoty od aritmetického průměru. Hodnoty umocníme, abychom měli všechny kladné, a zároveň jsme tím dali větší váhu těm odlehlejším pozorováním.

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i}{n}$$

Aritmetický průměr $\bar{x}$ musíme spočítat jako první a poté určíme zbytek vzorce.

Nevýhodou rozptylu je, že variabilitu udává v druhých mocninách jednotky.

Existuje druhá varianta rozptylu, která se nazývá výběrový rozptyl a liší se jen tím, že místo n dělíme hodnotou n-1. Tento rozptyl se využívá v případech, kdy nemáme kompletní data ale jen nějaký výběr, ale chceme výsledky použít pro zobecnění na větší soubor.

$$s_x'^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i}{n-1}$$

V Excelu použijeme funkci VAR.P pro výpočet rozptylu a VAR.S pro výpočet výběrového rozptylu.

Směrodatná odchylka

Směrodatná odchylka řeší výše uvedený problém, který má rozptyl, tedy to že udává výsledky v druhé mocnině původní jednotky, což se velmi špatně interpretuje. Směrodatnou odchylku spočítáme jako odmocninu rozptylu, to stejné platí u výběrové směrodatné odchylky, kterou získáme jako odmocninu z výběrového rozptylu.

$s_x = \sqrt{s_x^2}$ $s_x' = \sqrt{s'^2_x}$

V Excelu můžeme využít funkci SMODCH.P pro získání směrodatné odchylky a SMODCH.VÝBĚR.S pro získání výběrové směrodatné odchylky.

6.2 Hledání podobností

Při srovnávání statistických jednotek, například obcí, není vhodné náhodně vybrat ze všech ostatních obcí a provést srovnání s nimi. Například pokud budeme zkoumat vývoj počtu obyvatel a provedeme srovnání s mnohem větší, nebo naopak menší obcí, nebude důležité to, co můžeme ovlivnit, ale rozdíly budou spíš záviset na celkovém vývoji společnosti. Samozřejmě toto srovnání nám může vyhovovat, pokud bychom se snažili prokázat, že například ve městech počet obyvatel stagnuje nebo klesá, a naopak v menších obcích počet obyvatel roste. Takový vývoj je pak evidentně celospolečenský a pro srovnání v tom smyslu, jestli naše rozhodování děláme správně nebo ne to velký význam mít nebude. Tady potřebujeme najít obce v podobné situaci jako je naše a poté porovnat, zda se od ostatních podobných lišíme, ať už v dobrém, nebo naopak špatném chování. Cílem této kapitoly je tedy ukázat možnosti, jak takové obce vybrat.

Možností, jak eliminujeme ty jednotky, které nejsou podobné, je pomocí podmínek.

Podmínky mohou být striktní, tedy při nesplnění z výběru jednotka vypadne, nebo volné, a v takovém případě netrváme na tom, aby byla splněna, ale při nesplnění bude podobnost jednotky s referenční jednotkou hodnocena hůř.

Například si dáme podmínku, že chceme taková města, která mají pokryté alespoň střední školství, a k tomu chceme, aby vybraná města měla podobný počet obyvatel, a podobnou rozlohu. Ze všech obcí tedy vyřadíme ty, které nesplňují podmínku týkající se školství. Poté u ostatních zjistíme zbývající hodnoty a z nich pak určíme, jak jsou podobné.

6.2.1. Určení podobnosti teorie

Podobnost můžeme určit matematicky, a to pomocí zvolené metriky. Důležité je, abychom dokázali rozdílnost jednotek převést na číslo udávající míru rozdílnosti, která čím je menší, tím jsou si jednotky podobnější. Nejčastější a zároveň nejsnadnější způsob spočívá v ohodnocení rozdílnosti jednotlivých parametrů zvlášť (počet obyvatel, výměra obce) a výsledek poté sečíst. Sčítané rozdílnosti ovšem nemohou být v absolutních hodnotách, protože by se tak mohlo stát, že některá proměnná by měla mnohem větší vliv než jiná. Například rozdíl v počtu obyvatel bude nabývat mnohem větších čísel než rozdíl v počtu škol. Proto je vhodné rozlišovat výše uvedenou rozdílnost a matematickou operaci odčítání (rozdíl). Místo rozdílu je tedy vhodnější spočítat podíl hodnot, který nám bude udávat o kolik procent má jednotka větší hodnotu x_j než referenční jednotka x_r. Protože je pro nás rozdílné jak to, že je hodnota menší, tak to že je větší musíme obě strany zohlednit, a to nejlépe tak, že od tohoto podílu odečteme 1 (1 vychází, pokud jsou hodnoty stejné), a z výsledku určíme absolutní hodnotu.

Jestliže si označíme $x_1^r, x_2^r, \cdots, x_k^r$ jako hodnoty jednotlivých proměnných u referenční jednotky a $x_1^j, x_2^j, \cdots, x_k^j$ jako hodnoty proměnných u porovnávané jednotky, můžeme výše uvedený postup zapsat následujícím vzorcem.

$$M\left(x_i^r, x_i^j\right) = \left|\frac{x_i^j}{x_i^r} - 1\right|$$

Jestliže ovšem zkoumané proměnné nejsou čísla, ale nějaké slovní hodnocení, je situace složitější. V takovém případě pokud lze hodnocení seřadit od nejmenší po největší, musíme sami určit vhodné očíslování hodnot (např, 1,2,3,….) a určit jejich podíl, nebo rozdíl, který poté dělíme maximálním rozdílem. Pokud si označíme *a* nejmenší ohodnocení a *b* největší ohodnocení můžeme míru rozdílnosti určit například takto:

$$M(x_i^r, x_i^j) = \left| \frac{x_i^r - x_i^j}{b - a} \right|$$

Případně pokud hodnoty není možné rozumně seřadit a jde o to, zda jsou stejné nebo nejsou stejné, můžeme použít binární proměnnou, tedy jako rozdílnost u této proměnné použijeme číslo 0 v případě, že jsou hodnoty stejné a 1 pokud jsou rozdílné.

$$M(x_i^r, x_i^j) = \begin{cases} 0 & pokud\ x_i^r = x_i^j \\ 1 & pokud\ x_i^r \neq x_i^j \end{cases}$$

Celkovou míru rozdílnosti pak získáme součtem rozdílností jednotlivých proměnných.

$$M(x^r, x^j) = \sum_{i=1}^{k} M(x_i^r, x_i^j)$$

Tuto míru učíme pro všechny jednotky a výsledky seřadíme. Ty obce, u kterých je míra rozdílnosti nejmenší, jsou ty nejpodobnější.

6.2.2. Určení podobnosti praxe

Nyní si ukážeme konkrétní příklad, jak vybrat podobné obce. Byla náhodně vybrána obec Soběslav. Podobnost budeme hledat pomocí počtu obyvatel, výměry obce, příjmů v rozpočtu a informací, zda se jedná o obec s pověřeným obecním úřadem, případně s rozšířenou působností.

Pro srovnání obcí tedy byla získána data ze statistického úřadu a monitoru veřejné správy.

Pro naprostou jistotu správného postupu je potřebné mít uvedená data pro všechny obce v ČR, což je obtížné. Proto základní sadu dat najdete jako součást šablony.

Na obrázku vidíme velmi malou část dat, ve skutečnosti celá databáze sestává z 1470 měst, protože byla omezena na města nad 1000 obyvatel.

	A	B	C	D	E	J
1	nazev obce	pověře	rozšíře	počet c	výměra	příjmy
2	Adamov (okres Blansko)	1	0	4517	378,1066	84987881,2
3	Albrechtice (okres Karvi	0	0	3848	1268,785	64830628
4	Albrechtice nad Orlicí (o	0	0	1003	522,7533	16350917,6
5	Aš (okres Cheb)	1	1	13210	5585,612	306258890
6	Babice (okres Praha-výcl	0	0	1194	556,7154	22134675
7	Babice (okres Uherské H	0	0	1832	661,3402	33781730,9
8	Babice nad Svitavou (okr	0	0	1284	1741,973	21651457,4
9	Bakov nad Jizerou (okres	0	0	5156	2700,831	101048524
10	Bánov (okres Uherské Hr	0	0	2119	1617,839	46685305,8
11	Bartošovice (okres Nový	0	0	1703	2414,618	44039372,3
12	Baška (okres Frýdek-Mís	0	0	3892	1283,372	74881843,9
13	Bašť (okres Praha-výcho	0	0	2696	780,8398	36981485
14	Batelov (okres Jihlava)	0	0	2361	4269,69	49943068,2
15	Bavorov (okres Strakonic	0	0	1572	3540,491	36407925,4
16	Bečov (okres Most)	0	0	1445	2822,868	51990006,2
17	Bečváry (okres Kolín)	0	0	1031	1627,775	21968876,2
18	Bedihošť (okres Prostějc	0	0	1077	646,5616	22616906,1

Pro tyto sloupce byly určeny průměry všech hodnot, nejmenší a největší hodnota. Pomocí těchto hodnot byla určena rozdílnost v daném parametru s použitím vzorce

$$M(x_i^r, x_i^j) = \left| \frac{x_i^r - x_i^j}{b - a} \right|$$

Dostáváme pak tabulku rozdílů.

	A	K	L	M	N	O	T
1	nazev obce	odchylka	pověřený o	rozšířená	počet ob	výměra	příjmy
2	Soběslav (okres Tábor)	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3	Český Brod (okres Kolín)	0,0009	0,0000	0,0000	0,0001	0,0006	0,0003
4	Holice (okres Pardubice)	0,0016	0,0000	0,0000	0,0003	0,0007	0,0007
5	Kravaře (okres Opava)	0,0020	0,0000	0,0000	0,0002	0,0013	0,0006
6	Stod (okres Plzeň-jih)	0,0037	0,0000	0,0000	0,0026	0,0001	0,0011
7	Lanškroun (okres Ústí na	0,0039	0,0000	0,0000	0,0023	0,0013	0,0002
8	Hořice (okres Jičín)	0,0044	0,0000	0,0000	0,0013	0,0029	0,0002
9	Nový Bor (okres Česká Lí	0,0049	0,0000	0,0000	0,0036	0,0011	0,0001
10	Náměšť nad Oslavou (ok	0,0052	0,0000	0,0000	0,0016	0,0028	0,0008
11	Broumov (okres Náchod	0,0056	0,0000	0,0000	0,0004	0,0046	0,0006
12	Nýřany (okres Plzeň-sev	0,0061	0,0000	0,0000	0,0000	0,0056	0,0004
13	Železný Brod (okres Jabl	0,0063	0,0000	0,0000	0,0007	0,0051	0,0005
14	Frýdlant nad Ostravicí (o	0,0066	0,0000	0,0000	0,0023	0,0039	0,0004

Ve sloupci K je pak součet odchylek jednotlivých proměnných a tabulka je seřazena dle tohoto sloupce. Obce, které jsou na prvních místech, jsou ty, které jsou Soběslavi nejpodobnější.

6.3 T-Test

Tento test se používá pro odhalení rozdílností středních hodnot (průměrů), a to buď dvou výběrových souborů, nebo jednoho souboru dat s konstantou.

6.3.1. T-test – Kdy a proč používat

Z laického pohledu se může zdát, že přece není třeba při porovnávání dvou průměrů cokoliv testovat. Pokud spočítáme průměry dvou vzorků, například průměrné výše udělených pokut za rok ve dvou různých obcích, vidíme, že jeden průměr je větší než druhý. Problém nastává v určování toho, jaký rozdíl je skutečně dostatečný pro to abychom mohli tvrdit, že jedno číslo je větší než jiné a co je naopak takto podobné, že to můžeme považovat za téměř stejné a žádný praktický rozdíl tam není. Pokud bychom to určovali dle svých pocitů, může to být pokaždé hodnoceno jinak a výsledky by byly sporné. Pro někoho je velký rozdíl, když je průměr alespoň v tisících pro někoho to můžou být i koruny. Zároveň velmi záleží na druhu proměnné. Jestliže uvažujeme výše zmíněné pokuty, půjde nejspíše řádově o tisíce korun, naopak pokud budeme porovnávat výši investice, budeme spíše u statisíců či milionů.

Navíc můžeme zkoumat ne všechny hodnoty, ale můžeme udělat jen nějaký výběr, třeba z důvodu, že ostatní data nejsou dostupná. Pak můžeme otisky porovnat průměry těchto výběrů, ale v žádném případě nemůžeme zjištěné poznatky zobecnit a tvrdit že v první obci dávají vyšší pokuty než v druhé, protože rozdíl je pouze na nějakém vzorku a my si nemůžeme být jistí, zda by při jiném výběru vyšly hodnoty obdobně či ne.

Tento, a i jiné testy nám v takovém hodnocení mohou pomoci, protože nám místo tvrzení, které nemusí odpovídat skutečnosti, určí pravděpodobnost, s jakou je tvrzení správné. Výsledkem pak může být například tvrzení, že s pravděpodobností 99% jsou průměrné pokuty udělené zkoumanými obcemi rozdílné.

Obecně se dá konstatovat, že pokud porovnáváte průměrné hodnoty, nemůžete udělat chybu, když tento test použijete a tím svá tvrzení podpoříte statistikou a pravděpodobností.

6.3.2. T – test – Teorie

Tento test má velké množství variant a je tedy třeba správně pochopit rozdíly, abychom použili správný test.

1) **Jednovýběrový t-test** – testuje rozdílnost střední hodnoty souboru dat a dané konstanty.
2) **Párový t-test** – testuje rozdíly středních hodnot dvou skupin dat, které jsou navzájem spárovány.
3) **Dvouvýběrový t-test** – testuje rozdílnost středních hodnot dvou rozdílných skupin.

Všechny testy se provádějí v následujících krocích, ve kterých jen dosazujeme do různých vzorců dle volby testu a alternativní hypotézy.

1. Volba testu, hypotéz a hladiny významnosti α. – Dle porovnávaných dat a zkoumaného výsledku zvolíme příslušný test a hypotézy. Hladina významnosti pak značí, jak velkou pravděpodobnost chybného výsledku jsme ochotní akceptovat. Obvyklá hodnota α=0,05, což značí 5%. Pokud požadujeme silnější výsledek, volíme α=0,01, naopak pokud nám dostačuje slabší, volíme α=0,1.
2. Výpočet testového kritéria – dosadíme hodnoty do odpovídajícího vzorce. Hodnoty spočítáme ze získaných dat. Pokud některé z použitých statistik neznáte, nebo nevíte, jak je spočítat podívejte se na část 7.9.

3. Určení kritického oboru – kritický obor je tvořen intervaly, či intervalem. Jejich okraje jsou určeny kvantily studentova rozdělení a druhé strany nekonečnem. Kvantily určíme pomocí statistických tabulek nebo pomocí vhodného softwaru.
4. Závěr – Hypotézu H_1 potvrzujeme na hladině významnosti α, pokud testové kritérium leží v kritickém oboru. Naopak pokud neleží v kritickém oboru, tak hypotézu H_1 na hladině významnosti α nemůžeme potvrdit.

Jednovýběrový t-test

Tento test použijeme v případě, kdy máme určité hodnoty a porovnáváme je s předem daným číslem. Například máme hodnotu průměru za celou ČR, ale nemáme přístup k jednotlivým datům, zatímco data za naši obec máme dostupné. Tímto testem pak určíme, zda jsou naše hodnoty silně podprůměrné, přibližně průměrné, či silně nadprůměrné.

Potřebujeme soubor *n* čísel, které označujeme x_i. Z těchto hodnot spočteme potřebné statistiky.

Následující tabulka uvádí přehled všech variant tohoto testu.

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Střední hodnota výběru (μ) je různá od hodnoty μ_0	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$T = \frac{\bar{x} - \mu_0}{s'_x} \cdot \sqrt{n}$	$(-\infty, t_{n-1}(\alpha/2)) \cup (t_{n-1}(1-\alpha/2), \infty)$
Střední hodnota výběru (μ) je větší než hodnota μ_0	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$T = \frac{\bar{x} - \mu_0}{s'_x} \cdot \sqrt{n}$	$(t_{n-1}(1-\alpha), \infty)$
Střední hodnota výběru (μ) je menší než hodnota μ_0	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$T = \frac{\bar{x} - \mu_0}{s'_x} \cdot \sqrt{n}$	$(-\infty, t_{n-1}(\alpha))$

Párový t-test

Párový t-test použijeme tehdy, když porovnáváme dvojice výsledů vždy na stejném subjektu. Například hodnocení stejných respondentů před a po určité události. Například průzkum, kde mají obyvatelé hodnotit vzhled určité městské části před její úpravou a ti stejní pak hodnotí i vzhled po úpravě. Důležité je, aby byly odpovědi spárovány a tedy vždy docházelo k porovnání výsledků od stejného respondenta. Z rozdílů hodnocení všech respondentů uděláme průměr a testujeme jej proti zvolenému číslu, obvykle hodnotě 0, která udává průměrně nezměněný stav, naopak výrazně kladné nebo záporné hodnoty udávají určitý směr změny.

V tomto případě tedy máme dva soubory dat, které značíme x_i a y_i. Vždy musí platit, že pro konkrétní hodnotu *i* pocházejí x_i a y_i od stejného respondenta. Tyto hodnoty od sebe odečteme a získáme novou proměnnou m_i, kde $m_i = x_i - y_i$.

Poté postupujeme dle následující tabulky, dle toho co požadujeme potvrdit.

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Odpovědi se změnily jedním směrem. Většinou tedy dochází ke zlepšení, nebo naopak ke zhoršení.	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$T = \frac{\bar{m} - 0}{s'_x} \cdot \sqrt{n}$	$(-\infty, t_{n-1}(\alpha/2)) \cup (t_{n-1}(1 - \alpha/2), \infty)$
Průměrná změna odpovědí ve v kladném směru.	$H_0: \mu = 0$ $H_1: \mu > 0$	$T = \frac{\bar{m} - 0}{s'_x} \cdot \sqrt{n}$	$(t_{n-1}(1 - \alpha), \infty)$
Průměrná změna odpovědí je v záporném směru	$H_0: \mu = 0$ $H_1: \mu < 0$	$T = \frac{\bar{m} - 0}{s'_x} \cdot \sqrt{n}$	$(-\infty, t_{n-1}(\alpha))$

Dvouvýběrový t-test

Dvouvýběrový t-test je nejčastěji používanou verzí t-testu. Využijeme jej v případě, kdy máme data ze dvou zdrojů. Mohou to být dvě různé skupiny respondentů, data dvou různých obcí, …. Tento test pak určí, zda jsou střední hodnoty dat z těchto zdrojů rozdílné, případně zda je jedno větší než druhé, či naopak.

Potřebujeme tedy dva soubory dat. První budeme označovat x_i a druhý y_i. Není nutné, aby měly tyto soubory stejné množství údajů. Počty údajů z prvního souboru označme n_x a z druhého pak n_y.

Opět dle dokazované vlastnosti určíme vhodnou hypotézu a z toho vyplývající další vlastnosti dle následující tabulky.

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Střední hodnoty zkoumaných skupin jsou různé.	$H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$	$T = \frac{\bar{x} - \bar{y}}{SE(x_i - y_i)}$	$(-\infty, t_v(\alpha/2)) \cup (t_v(1 - \alpha/2), \infty)$
Střední hodnota první skupiny je větší než střední hodnota druhé skupiny.	$H_0: \mu_x = \mu_y$ $H_1: \mu_x > \mu_y$	$T = \frac{\bar{x} - \bar{y}}{SE(x_i - y_i)}$	$(t_v(1 - \alpha), \infty)$
Střední hodnota první skupiny je menší než střední hodnota druhé skupiny.	$H_0: \mu_x = \mu_y$ $H_1: \mu_x < \mu_y$	$T = \frac{\bar{x} - \bar{y}}{SE(x_i - y_i)}$	$(-\infty, t_v(\alpha))$

Konkrétní výpočty se pak liší dle toho, zda můžeme předpokládat rovnost rozptylů s_x^2 a s_y^2 u těchto dvou skupin, nebo naopak předpokládáme jejich nerovnost. Ověření této skutečnosti může provést pomocí F-testu popsaného v části 7.3.

Hodnoty $SE(x_i - y_i)$ a stupeň volnosti v, jsou pak určeny následovně:

Rozptyly jsou shodné	$SE(x_i - y_i) = \sqrt{\frac{(n_x - 1)s_x^{2\prime} + (n_y - 1)s_y^{2\prime}}{n_x + n_y - 2}} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$ $v = n_x + n_y - 2$
Rozptyly nejsou shodné	$SE(x_i - y_i) = \sqrt{\frac{s_x^{2\prime}}{n_x} + \frac{s_y^{2\prime}}{n_y}}$ $v = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{s_x^{2\prime 2}}{n_x^2 \cdot (n_x - 1)} + \frac{s_y^{2\prime 2}}{n_y^2 \cdot (n_y - 1)}}$

6.3.3. T-Test – Praxe

Použití výše uvedené teorie může být trochu náročnější, je třeba provést řadu zdlouhavých výpočtů, a proto je v praxi příjemnější využít vhodný software. Ten může být přímo určen pro statistické výpočty, ale také je možné pro výpočty využít tabulkové kalkulátory například Excel.

Všechny zde uvedené výpočty naleznete i v přiložené šabloně, do které můžete zadat data, a hodnoty Vám budou spočítány.

Jednovýběrový t-test – provedení v Excelu

V případě jednorozměrného t-testu musíme nejdříve znát počet prvků *n*, aritmetický průměr $\bar{x}$ a směrodatnou odchylku s_x. Výpočty všech těchto statistik naleznete v části 6.7.

Samotné testové kritérium pak spočítáme pomocí vzorce $T = \frac{\bar{x} - \mu_0}{s_x'} \cdot \sqrt{n}$. Tento vzorec do Excelu zadáme takto =*(I7-I13)/I14*ODMOCNINA(I12)*.

Párový a Dvouvýběrový t-test – provedení v Excelu

Dvouvýběrový i párový t-test je v excelu velmi snadný provést. Pro jeho provedení potřebujeme mít data rozdělená do dvou oblastí, nejčastěji ve dvou sloupcích.

Poté můžeme použít funkci =*T.TEST(matice1;matice2;chvosty;typ)*. Za parametry matice1 a matice2 nastavíme souřadnice, kde máme uloženy data prvního, resp. druhého souboru, či měření. Třetí parametr chvosty, obsahuje buď hodnotu 1 nebo 2 a uvádí, zda určujeme jednostrannou hypotézu, či oboustrannou hypotézu.

Poslední parametr obsahuje jednu z hodnot 1 2 3, a to dle následujícího popisu:
1- použijeme při párovém testu.
2 – jestliže předpokládáme rovnost rozptylů, což ověříme F-testem (část 6.4.2).
3 – jestliže předpokládáme nerovnost rozptylů, což opět ověříme F-testem (část 6.4.2).

Výsledkem funkce je pak p-hodnota, což je vlastně pravděpodobnost, že nebudeme mít pravdu, pokud budeme tvrdit, že střední hodnoty výběrů či měření jsou odlišné. Obvykle rozdílnost potvrzujeme, pokud je p-hodnota menší než 0,05.

6.4 F-Test

Tento test se používá v případě, kdy chceme určit rozdíly rozptylů dvou statistických souborů. V praxi k tomu nejčastěji dochází v případech, kdy máme dva soubory dat a porovnáváme jejich střední hodnoty. Jak víme z části 0 je u dvou výběrového t-testu nutné zvolit jeden ze dvou postupů. A to podle toho, zda máji srovnávané dva soubory stejný rozptyl, nebo jej mají rozdílný. Před tímto rozhodnutím provedeme právě F-test, který nám odpoví na otázku rozdílnosti rozptylů.

6.4.1. F-Test – Teorie

Test má pouze jednu variantu a ta spočítá v tom, že srovnáváme rozptyl dvou souborů dat. Technicky jsou možné tři alternativní hypotézy podobně jako v případě jiných testů ale v praxi se většinou používá a také doporučuje pouze hypotéza o nerovnosti.

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Rozptyly jsou různé	$H_0: \sigma_x = \sigma_y$ $H_1: \sigma_x \neq \sigma_y$	$F = \frac{s_x^{2\prime}}{s_y^{2\prime}}$	$\left(-\infty, F_{(n_x-1,n_y-1)}(\alpha/2)\right) \cup$ $\left(F_{(n_x-1,n_y-1)}(1-\alpha/2), \infty\right)$

Postupy pro výpočet hodnot použitých v těchto vzorcích najedete v části 6.7

Výsledek určíme tak, že pokud je hodnota testového kritéria v některém z intervalů, které tvoří kritický obor, pak můžeme tvrdit, že na hladině významnosti α potvrzujeme rozdílnost rozptylů, tedy šance že jsou rozptyly totožné je menší než α. V opačném případě je závěrem testu, že rozdílnost nepotvrzujeme, což pro nás v praxi znamená, že nemáme dostatečný důkaz o rozdílnosti a předpokládáme tedy, že jsou rozptyly totožné.

6.4.2. F-test – Praxe

Pokud chceme tento test provést, je třeba pomocí vhodného postupu spočítat výběrové rozptyly obou skupin $s_x^{2\prime}, s_y^{2\prime}$ a poté buď pomocí vhodného programu, nebo ve statistických tabulkách najít kvantily Fisherova rozdělení F pro výpočet kritického oboru.

Jestliže používáme excel je provedení tohoto testu velmi snadné, stačí nám použít vhodnou funkci, tedy do některé buňky zadat =F.TEST(matice1;matice2), kde parametry označené jako matice1 a matice2 obsahují adresy dat získaných z první a druhé skupiny. Při zadávání stačí ten po napsání jména funkce označit první data, zadat středník a označit druhá dat.

Výsledkem této funkce je číslo, které označujeme jako p-hodnota (p-value), které udává pravděpodobnost, že je alternativní hypotéza nepravdivá. Čím menší hodnota je, tím větší šanci na to, že platí alternativní hypotéza, máme. Jestliže jsme si tedy předem určili hodnotu α, musíme s ní získanou p-hodnotu porovnat. Pokud je α menší než p-hodnota předpokládáme platnost alternativní hypotézy, tedy, že jsou rozptyly rozdílné, pokud je tomu naopak, předpokládáme, že rozptyly jsou totožné.

6.5 Chi-square test

Tento test se používá pro určení, zda hodnoty určité proměnné odpovídají předem zvolenému rozdělení. Tento test je dá prakticky rozdělit na dvě varianty.

χ^2-test dobré shody a χ^2-test nezávislosti. Každý z nich použijeme v jiném případě, ale hlavní myšlenku mají stejnou. Vždy předpokládáme, že získané údaje rozdělíme do skupin a vždy určíme kolik je jich v každé skupině. Poté si určíme předpokládané (teoretické) počty, které by v těchto skupinách měly být, pokud by platilo to, co chceme prokázat. χ^2-test pak určí, zda se skutečné hodnoty statisticky významně odlišují od těch teoretických, nebo nikoliv.

6.5.1. χ^2-test dobré shody – teorie

Test dobré shody se používá v případě, že zjišťujeme určitou hodnotu a ověřujeme, zda má předpokládané chování. Například si provedeme průzkum mezi voliči týkající se nadcházejících voleb a tímto testem můžeme ověřit, zda se výsledky významně odlišují od předchozích voleb.

Je tedy třeba předem určit rozdělení náhodné veličiny, se kterým budeme aktuální data dále porovnávat. Toto rozdělení obvykle vychází z nějakých výsledků, které jsou již známé, tedy v minulosti, nebo z teoretického popisu odpovídající určité vědecké teorii.

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Empirické rozdělení neodpovídá teoretickému rozdělení	H_0: *empirické rozdělení odpovídá teoretickému rozdělení* H_1: *empirické rozdělení neodpovídá teroetickému rozdělení*	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}$	$(\chi_v^2(1-\alpha), \infty)$

Zpracovávaná data rozdělíme do skupin a vytvoříme tabulku s počty hodnot v jednotlivých skupinách.

V uvedených vzorcích používáme následující značení:

n_i je empirická absolutní četnost v i-té skupině, tedy udává počet hodnot, které náleží do jednotlivých skupin. n_i' je teoretická absolutní četnost náležící do dané skupiny, kterou si určíme na základě dokazované hypotézy.

Vidíme, že vzorec určitým způsobem hodnotí součet rozdílů počtů v jednotlivých skupinách.

Pomocí kvantilu rozdělení χ^2 o v stupních volnosti pak určíme kritický obor, kde určíme jako počet skupin (*k*) od kterého odečteme *h+1*, kde *h* odpovídá počtu odhadovaných parametrů zvoleného rozdělení, u uváděného příkladu s volbami by *h* bylo 0.
Tedy máme $v = k - h - 1$.

Pokud je hodnota spočítaného testového kritéria větší než kvantil v kritickém oboru, můžeme tvrdit, že na hladině významnosti α potvrzujeme rozdílnost teoretického a empirického rozdělení. Naopak pokud je testové kritérium menší než kritická hodnota, na hladině významnosti α nepotvrzujeme rozdílnost empirického a teoretického rozdělení a v praxi tedy můžeme předpokládat, že jsou totožné.

6.5.2. χ^2-test nezávislosti – teorie

Test závislosti na rozdíl od předchozího testu určuje, zda dvě proměnné na sobě nějakým způsobem závisí. Například počet škol ve městě jistě závisí na celkovém počtu obyvatel města, naopak počet škol asi nezávisí na počtu stromu v parcích města. Je tedy třeba při sběru dat vždy od každého respondenta získat a udržovat dva údaje. Poté si připravíme dvourozměrnou tabulku, kdy v řádcích budou hodnoty x_i jedné proměnné, a ve sloupcích hodnoty y_i druhé proměnné. V případě že je hodnot některé proměnné velké množství můžeme vytvořit intervaly a tím snížit počet řádků (k) a sloupců (l). Hodnoty v tabulce (n_{ij}) pak odpovídají počtu dvojic, které patří do každé kombinace skupin. Přepočtem hodnot této tabulky získáme novou tabulku, kterou nazýváme tabulka teoretických četností, které značíme n'_{ij}. Tuto tabulku získáme přepočtem dle vzorce $n'_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$, kde $n_{i\cdot}$ značí součet hodnot i-tého řádku a $n_{\cdot j}$ značí součet i-toho sloupce. Hodnota n pak udává celkový počet zkoumaných objektů (respondentů).

Co dokazujeme	Hypotézy	Testové kritérium	Kritický obor
Proměnné na sobě nějakým způsobem závisí	H_0: *proměnné nejsou statisticky závislé* H_1: *proměnné jsou statisticky závislé*	$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$	$(\chi^2_v(1-\alpha), \infty)$

Vidíme, že vzorec určitým způsobem hodnotí odlišnosti všech buněk v těchto dvou tabulkách. Pomocí kvantilu rozdělení χ^2 o v stupních volnosti pak určíme kritický obor, kde $v = (k-1)\cdot(l-1)$.

Pokud je hodnota spočítaného testového kritéria větší než kvantil v kritickém oboru, můžeme tvrdit, že na hladině významnosti α potvrzujeme závislost uvedených proměnných, tedy jedna proměnná určitým způsobem ovlivňuje druhou. Naopak pokud je testové kritérium menší než kritická hodnota, tak na hladině významnosti α nepotvrzujeme závislost těchto proměnných a v praxi tedy můžeme předpokládat, že jsou nezávislé.

6.5.3. χ^2-test– praxe

V praktickém použití je vždy nejdůležitější vytvořit si dvě tabulky s totožným záhlavím, kde jedna obsahuje skutečně získaná data a druhá pak teoretické hodnoty. Důležité je, že součet všech hodnot v obou tabulkách musí být totožný.

Tabulka empirických (skutečných) četností je obvykle ta, kterou řešíme jako první. Tabulku teoretických četností pak spočítáme tak aby odpovídala našemu předpokladu. Důležité je, omezení tohoto testu, které říká, že alespoň 80 % buněk kontingenční tabulky má mít teoretickou četnost

větší než 5 a všechny buňky tabulky musí mít očekávanou četnost větší než 2. Jestliže tabulka teoretických četností toto nesplňuje, je třeba tabulky určitým způsobem upravit, než test použijeme. Prakticky máme dva postupy, jak problém vyřešit. Prvním je navýšení počtu respondentů, tím vzroste celkový počet hodnot v tabulce empirických četností a díky tomu se po přepočítání zvýší i hodnoty v tabulce teoretických četností. Druhou možností pak je snížení počtu řádků a sloupců tabulky. Díky tomu že součet hodnot v tabulce se nezmění, musí po přepočítání tabulky teoretických četností dojít ke zvětšení čísel uvnitř tabulky.

Nyní můžeme použít testová kritéria uvedená v předchozích kapitolách a pomocí kvantilu určit výsledek testu. Tyto výpočty jsou sice náročnější ale stále prakticky zvládnutelné. Snazším postupem opět může být použití funkce Excelu. V tomto případě se jedná o funkci CHISQ.TEST. Tato funkce lze použít v případě testu nezávislosti a také v případě testu dobré shody, kdy známe přesné parametry rozdělení a není nutné je dopočítávat.

Funkce CHISQ.TEST má dva parametry, jako první použijeme tabulku empirických četností a jako druhý tabulku teoretických četností. Vždy vybíráme jen vnitřní hodnoty bez záhlaví. Výsledkem tohoto testu je p-hodnota, která udává pravděpodobnost, že platí hypotéza H_0.

P-hodnotu porovnáme se zvolenou hladinou významnosti α. Pokud je p-hodnota menší než α, platí alternativní hypotéza H_1, tedy můžeme na hladině významnosti tvrdit, že proměnné jsou závislé (v případě že provádíme test nezávislosti), nebo proměnná nemá testované rozdělení (v případě že provádíme test dobré shody).

6.6 Výpočet regresního modelu

Regresní model slouží k matematickému popisu skutečnosti, který získáme na základě již známých údajů. Vždy potřebujeme alespoň dvě proměnné, ale může jich být víc. Jedna proměnná je vysvětlovaná, protože naším cílem je vysvětlit chování této proměnné na základě ostatních proměnných, kterým říkáme vysvětlující. Vysvětlujících proměnných může být víc než jedna.

Po zvolení použitých proměnných musíme určit vhodný typ modelu. Prakticky můžeme využít jakoukoliv matematickou funkci $y = f(x_1, x_2, x_3, \dots, x_k)$, kde v místě proměnné y je zvolená vysvětlovaná proměnná a jednotlivé x pak určují vysvětlující proměnné.

Například si můžeme za y zvolit počet škol ve městě, vysvětlit tuto proměnnou můžeme pomocí počtu obyvatel města x_1 a počet obyvatel spádové oblasti x_2. Regresní analýza nám pak dá rovnici například ve tvaru $y = 0{,}002 \cdot x_1 + 0{,}005 \cdot x_2$. Rovnici pak lze použít k určení, zda se některé město vymyká obvyklým hodnotám, které jsou založené na městech ostatních, ale také ji můžeme použít k predikcím. Například pokud máme odhady, že se počet obyvatel za 5 let zvýší o dva tisíce, bylo by, pokud by tato rovnice byla skutečná vhodné začít uvažovat o výstavbě nové školy, protože za 5 let nejspíš bude potřeba, pokud je aktuální počet nyní optimální. Výsledky tedy mohou sloužit k predikcím a tím i k možnému odhalení rizik dříve, než k nim skutečně dojde.

V případě regresní analýzy jde o velmi složitý proces, který zahrnuje celou řadu testů a náročných výpočtů, proto se zde zaměříme pouze na základy. Popíšeme si možnosti modelů, jejich výpočet pomocí softwaru a poté vyhodnocení výsledků.

6.6.1. Regresní analýza – volba modelu

Modely mohou mít velké množství různých tvarů, ale nejčastěji používaný je lineární regresní model. Tento model spočívá v tom, že použitá funkce je součtem funkcí jednotlivých vysvětlujících proměnných.

$$y = b_1 \cdot f_1(x_1) + b_2 \cdot f_2(x_2) + \cdots + b_k \cdot f_k(x_k)$$

Za jednotlivé funkce pak volíme prakticky libovolné matematické funkce, volba záleží na našem předpokladu o druhu vlivu proměnné na vysvětlovanou proměnnou, případně můžeme volit funkci dle určité teorie, kterou v daném případě máme.

Nejčastěji ovšem používáme lineární funkce, tedy rovnice má tvar.

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + \cdots + b_k \cdot x_k$$

Volba modelu je velmi zásadní pro získané výsledky a zároveň je volba nejednoznačná. Nedá se říct jaká křivka je správná. Pokud nevíme, jakou křivku zvolit někdy se používá možnost spočítat jich víc a dle výsledků pak určit tu, která se jeví jako nejlepší.

6.6.2. Regresní analýza – provedení

V praxi určitě nebudete provádět klasickým počítáním, vždy k tomu využijete nějaký software, ať už je to přímo statistický software, nebo tabulkový kalkulátor. Vždy je ovšem potřeba připravit si data do správného tvaru.

Pro provedení regresní analýzy potřebujeme identifikovat proměnnou, která bude vysvětlovaná a proměnné které budou vysvětlující. Bývá vhodné si potřebné sloupce zkopírovat zvlášť tak, aby vysvětlovaná proměnná byla mimo vysvětlující a ty tvořili souvislou tabulku. Vždy je ovšem nutné dodržet pořadí řádků, data se tedy nesmí přeházet a stále platí, že musí být data od stejné statistické jednotky v jednom řádku. V regresním modelu je možné použít několikrát stejnou proměnnou v různých funkcích a je pak třeba odpovídající sloupce upravit.

Například při použití tohoto modelu

$$T = \beta_0 + \beta_1 V + \beta_2 Z + \beta_3 M + \beta_4 M^2$$

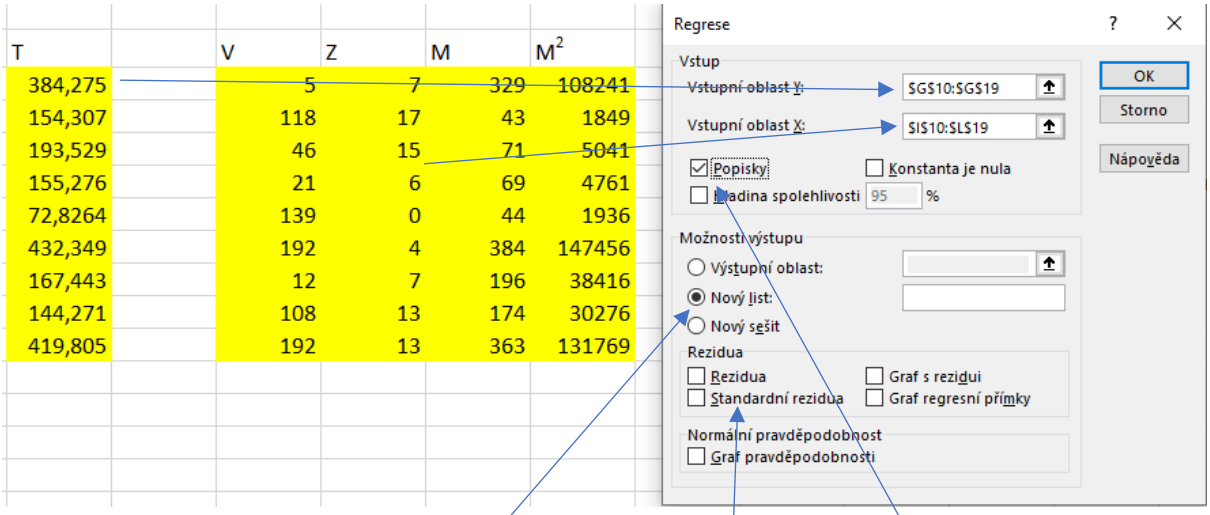
si připravíme následující data.

T		V	Z	M	M^2
384,275		5	7	329	108241
154,307		118	17	43	1849
193,529		46	15	71	5041
155,276		21	6	69	4761
72,8264		139	0	44	1936
432,349		192	4	384	147456
167,443		12	7	196	38416
144,271		108	13	174	30276
419,805		192	13	363	131769

Kde hodnoty ve sloupci M^2 získáme pomocí druhé mocniny hodnot ve sloupci M.

Na takto upravená data již můžeme použít metodu nejmenších čtverců, a to v různých programech. My se opět zaměříme na možnosti Excelu.

Excel obsahuje doplněk Analytické nástroje, které je možno spustit v nastavení Excelu v záložce Doplňky. Analytické nástroje pak najdeme v záložce data pod názvem Analýza dat. Zde zvolíme možnost Regrese a klikneme na tlačítko OK.



V následujícím okně nastavíme do buňky pro vstupní oblast Y sloupec s vysvětlovanou proměnnou a U vstupní oblast X pak vybereme celou tabulku vysvětlovaných proměnných.

Jestliže označujeme tabulky i s popisky nezapomeneme **zaškrtnout příslušné políčko**. Poté zvolíme, kam chceme **uložit výsledky**. Můžeme nastavit **další parametry**, kterými se v tuto chvíli zabývat nebudeme a klikneme na OK.

Na zvoleném místě se zobrazí výsledky regresní analýzy.

6.6.3. Regresní analýza – interpretace výsledků

Po provedení regresní analýzy vidíme velké množství výsledků a může být obtížné se v nich zorientovat. Zde se zaměříme na ty nejdůležitější výsledky.

Nejdříve se podíváme na hodnotu označenou jako Významnost F. Tato hodnota udává, zda je celý model statisticky významný, tedy dává smysl. Hodnota nám udává pravděpodobnost nevýznamnosti modelu. V tomto případě je to hodnota 0,0000812678, pozor na zobrazování malých hodnot v Excelu, symbol E na konci říká, že je potřeba posunout desetinou čárku o počet míst za tímto znakem, tedy -5, což znamená posun o pět míst doleva. Vidíme, že nevýznamnost modelu je nepravděpodobná. Obvykle model považujeme za významný, pokud je tato hodnota menší než 0,05.

Druhou zajímavou statistikou je **Hodnota spolehlivosti R**, která nám ukazuje, jak přesně model popisuje vstupní data, čím blíž k 1 tím větší procento shody model s daty má. Tato hodnota se obvykle považuje za směrodatnou v případě použití rovnice pro budoucí předpovědi. Pokud je tato hodnota blízká 1, můžeme ji použít pro odhady chování s tím, že chyba odhadu by neměla být velká proti skutečnosti. Naopak malá hodnota ukazuje nevhodnost modelu pro počítání odhadů.

VÝSLEDEK

Regresní statistika	
Násobné R	0,99738969
Hodnota spol	0,9947862
Nastavená ho	0,98957239
Chyba stř. hoc	13,9387172
Pozorování	9

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	4	148279,3361	37069,83	190,798533	8,12678E-05
Rezidua	4	777,1513435	194,2878		
Celkem	8	149056,4875			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	191,80267	20,43575235	9,385643	0,000718	135,063925	248,541414	135,063925	248,5414142
V	-0,4945808	0,089983965	-5,49632	0,00534101	-0,744416372	-0,2447453	-0,74441637	-0,24474529
Z	5,00282244	0,903685835	5,53602	0,00520401	2,493788326	7,51185655	2,493788326	7,511856554
M	-1,3308507	0,233006708	-5,71164	0,00464716	-1,977781019	-0,6839204	-1,97778102	-0,68392035
M2	0,00559053	0,000575553	9,713317	0,00062893	0,00399254	0,00718852	0,00399254	0,007188524

Další skupina hodnot, která nás zajímá, jsou **hodnoty P**. Tyto hodnoty pro každou složku modelu říkají zda je významná, či nevýznamná, tedy zda by tato složka mohla mít nulový vliv a pak by se v modelu neměla vyskytovat. Opět tyto hodnoty porovnáváme s číslem 0,05. Jestliže je některá hodnota větší než 0,05 je vhodné vrátit se zpět před výpočet modelu a danou složku vypustit (smazat z připravené tabulky) a regresní model spočítat znovu, již bez tohoto řádku.

Pokud jsou výše zmiňované hodnoty v pořádku, můžeme se zaměřit na Koeficienty. Ty nám popisují naši hledanou regresní funkci (po zaokrouhlení). Do použitého modelu tato čísla vložíme a dostaneme tuto rovnici.

$$T = 191{,}8 - 0{,}49 \cdot V + 5 \cdot Z - 1{,}33 \cdot M + 0{,}006 \cdot M^2$$

Jednotlivé koeficienty pak určují, o jakou hodnotu se změní vysvětlovaná proměnná, jestliže dojde ke zvýšení vysvětlující proměnné o 1. Například zde pokud bychom měli statistickou jednotku, která

bude stejná jako jiná, ale u proměnné Z bude mít hodnotu o 1 větší, předpokládáme, že hodnota T by u ní měla být o 5 vyšší.

6.7 Výpočty používaných statistik

Aritmetický průměr $\bar{x}$, můžeme spočítat jako prostý nebo vážený. U prostého máme pouze seznam hodnot a u váženého máme přehled obměn a jejich četnosti.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} (n_i \cdot x_i)}{n}$$

Hodnoty x_i značí postupně všechny naměřené hodnoty proměnné, případně jednotlivé obměny, n_i značí četnosti obměn, *n* pak značí celkový počet hodnot, v případě že známe četnosti je *n* jejich součet.

Výběrová směrodatná odchylka s'_x, opět máme dvě možnosti výpočtu. První vychází ze znalosti hodnot a druhá z obměn a jejich četností. Obě varianty využívají aritmetický průměr $\bar{x}$, ten je třeba spočítat předem.

$$s'_x = \sqrt{\frac{\sum_{i-1}^{n} (\bar{x} - x_i)^2}{n-1}} = \sqrt{\frac{\sum_{i-1}^{n} (n_i \cdot (\bar{x} - x_i)^2)}{n-1}}$$

Hodnoty x_i značí postupně všechny naměřené hodnoty proměnné, případně jednotlivé obměny, n_i značí četnosti obměn, *n* pak značí celkový počet hodnot, v případě že známe četnosti je *n* jejich součet.

Kvantily Studentova rozdělení $t_v(\alpha)$ – kvantily určíme pomocí statistických tabulek, nebo vhodným softwarem. Z praktického hlediska je vhodnější software. Určení kvantilů umí prakticky každý statistický software, my opět můžeme použít Excel. Studentovo rozdělení má dva parametry, stupeň volnosti v a pravděpodobnost α. Po napsání vzorce $= T.INV(\alpha, v)$ do buňky excelu získáme hledaný kvantil. Místo parametrů samozřejmě musíme vložit konkrétní hodnoty, nebo odkazy na buňky obsahující čísla.

Kvantily Fisherova rozdělení $F_{(v_1,v_2)}(\alpha)$ - Opět je můžeme určit pomocí tabulek nebo softwarově. Tento kvantil má tři parametry, dva stupně volnosti v_1, v_2 a pravděpodobnost α. V excelu můžeme použít vzorec $= F.INV(\alpha, v_1, v_2)$. Opět parametry nahradíme konkrétními čísly.

Kvantily rozdělení Chi-kvadrát $\chi^2_v(\alpha)$ – Stejně jako u ostatních kvantilů i zde můžeme určit hodnoty pomocí tabulek, nebo programů. Rozdělení má dva parametry stupeň volnosti v a pravděpodobnost α. Hodnotu rozdělení určíme pomocí funkce $= CHISQ.INV(\alpha, v)$, hodnoty parametrů opět musíme nahradit konkrétními čísly nebo odkazy na buňky.

7. Příklady konkrétních postupů

V této kapitole popíšeme několik aplikací, které se často objevují a navrhneme jednu z možných cest jak problém řešit. Většinou je však třeba počítat s vyšší časovou zátěží, pokud byste analýzu prováděli bez pomocí vhodného softwaru. S výpočty Vám také může pomoci naše šablona.

7.1 Analýza dlouhodobého trendu počtu obyvatel obce

Tento příklad předpokládá, že například z důvodu vytváření registru rizik, případně swot analýzy potřebujeme určit, jak probíhá vývoj počtu obyvatel obce. Samozřejmě hodnoty za poslední roky nejsou problém zjistit, ale my potřebujeme víc. Ideální je, pokud určíme chování obdobných obcí, a to porovnáme s naší. S tím, že podobnými obcemi budeme chápat takové, které mají podobnou strukturu obyvatel a počet obyvatel.

Postup by mohl být následující

1. Stáhneme potřebné údaje – tato data můžeme získat z databáze Českého statistického úřadu – ukazatel, který bychom použili, by byl Počet obyvatel (trvalý pohyb) věk. Postup najdete v části 5.3.3. Samozřejmě potřebujeme údaje ideálně za všechny obce, což není rychlé ani snadné, databáze má omezení počtu řádků.
2. Určíme podobné obce – postup najdete v části 6.2 – počet obcí které zvolíme je na nás, ale je vhodné, aby jich nebylo méně než 5. U těchto obcí si stáhneme informace o počtu obyvatel za více let.
3. Určíme trend vývoje pro každou obec zvlášť – například regresní analýzou (část 6.6), kdy jako vysvětlující proměnnou dáme počet obyvatel a jako vysvětlující proměnnou dosadíme rok.
4. Můžeme také určit rozdíly za jednotlivé roky (1 diference), což jsou vlastně změny počtu obyvatel v každém roce. Tyto rozdíly pak můžeme testovat jednovýběrovým t-testem proti hodnotě nula a tím určit, jestli můžeme dlouhodobě tvrdit, že počet obyvatel roste nebo klesá. Další možností je pak pomocí dvouvýběrového t-testu provést porovnání naší obce postupně s ostatními a zjistit, jestli se v průměrných změnách velmi lišíme od ostatních, nebo všechny obce mají podobnou změnu. (část 6.3)
5. Vyhodnocení výsledků – porovnáme výsledky, tedy zda je regresní koeficient proměnné rok větší než nula, tedy dlouhodobě počet obyvatel roste, nebo je menší než nula a v tu chvíli klesá. Jestliže je nevýznamný značí to, že počet spíše stagnuje a není možné určit trend. Z toho, zda se naše obec chová podobně jako jiné nebo naopak odlišně určíme, zda je to naše příležitost, hrozba, riziko,… samozřejmě zde je třeba brát v potaz další informace o obci a také rychlost růstu.
6. Grafické znázornění – počty obyvatel i diference můžeme zobrazit pomocí spojnicového grafu, kdy každá obec bude zobrazena jako jedna datová řada. Také můžeme vytvořit diagram rozdílů zjištěných t-testem.

Taková analýza je často potřeba při určování vlivů, které mohou být různé pro obce různých velikostí, například (Plaček 2019).

7.2 Porovnávání finanční ukazatelů obcí

Finanční analýzu si s vlastními ukazateli a výsledky, které vezmete z účetnictví, můžete udělat velmi snadno. Občas však může být vhodné porovnat svou finanční situaci s okolím, a zjistit, v čem bychom se mohli zlepšit, případně co patří mezi naše rizika, nebo naopak zjistit v čem je naše silná stránka. Většina výsledků hospodaření za posledních 10 let je již dostupná a vhodným zpracováním tak můžeme provést srovnání.

1. Určení porovnávaných obcí – dle různých kritérií určíme množinu obcí, s kterými se chceme srovnávat. Můžeme použít například postup popsaný v části 6.2
2. Určení porovnávaných dat – porovnávat můžeme například celkové příjmy, nebo příjmy rozdělené do konkrétních skupin, totéž u výdajů. Porovnávat můžeme také zadlužení a další ukazatele. V tomto příkladu budeme srovnávat náklady na školství.
3. Získání potřebných údajů – vlastní údaje můžeme zjistit přímo v našich účetních/finančních záznamech. Informace o ostatních obcích stáhneme z Monitoru (část 5.2). Pokud chceme porovnat jen poslední období, stačí stáhnout informace za jeden rok, pokud bychom chtěli dlouhodobější analýzu, stáhneme víc let. Důležité je abychom u všech obcí měli pokryté stejné roky. Z důvodu makroekonomických změn by srovnávání různých období nemuselo dávat smysl. Výdaje na školství získáme, pokud přepneme na „odvětvový pro výdaje“, zde rozklikneme skupinu „služby pro fyzické osoby“ a poté již vidíme výdaje na školství.
4. Pro porovnání jednoho roku, můžeme využít jednovýběrový t-test (část 6.3). Zde bychom postupovali tak, že naším cílem bude určit, zda průměrná hodnota srovnávaných obcí je statisticky odlišná od výdajů naší obce. Tedy porovnáváme data ostatních obcí s naším číslem. Jestliže test vyjde jako významný, víme, že se naše výdaje významně odlišují od průměru výdajů ostatních. To, zda jsou vyšší nebo menší můžeme určit pomocí jednostranné alternativní hypotézy.
5. Pokud máme stažené údaje za více let, můžeme porovnávat jak průměrné dlouhodobé náklady, mezi dvojicemi obcí. V praxi obvykle mezi naší obcí a postupně různými obcemi, které jsme zvolili ke srovnání. Zde bychom tedy použili dvouvýběrový t-test (část 6.3). Také bychom mohli zkoumat trend vývoje, tedy provést regresní analýzu (část 6.6) těchto dat s tím že výdaje jsou vysvětlovaná proměnná a rok je vysvětlující. To můžeme udělat pro každou obec zvlášť a porovnávat výsledky. Také můžeme udělat regresi pro naši obec a pak druhou pro data všech ostatních obcí a tím získat průměrný trend v podobných obcích, s kterým naše výsledky porovnáme.
6. Grafické znázornění – náklady v jednotlivých letech můžeme zobrazit pomocí sloupcového, nebo spojnicového grafu. Pokud bychom určili trend jednotlivých obcí, je možné získané regresní křivky zobrazit pomocí spojnicového grafu.

Postup byl použit při výzkumech určujících existencí úspor z rozsahu u obcí, například (Matějová 2017)

7.3 Veřejné zakázky na strategický plán

Při zadávání veřejných zakázek je třeba předem prozkoumat aktuální situaci na trhu. Pokud bychom vycházeli ze zastaralých údajů, nemuseli bychom postihnout aktuální trendy, mít chybný přehled o nejčastějších dodavatelích a hlavně chybně odhadnout předpokládanou cenu. To může být ve svém důsledku velký problém, pokud ji nasadíme příliš vysoko, může to vest k tomu, že nabídky budou určitým způsobem nadsazené, aby se této ceně přizpůsobily. Naopak při nastavení nízké ceny, může být nabídek malé množství, případně žádná, protože společnosti si spočítají, že za uvedenou cenu se jim nevyplatí do soutěže vstupovat a připravovat nabídku. Z tohoto důvodu je třeba předem určit jak situaci na trhu (e-shopy a další dostupné tržní ceny, dodavatelé, …) ale hlavně vyhledat obdobné veřejné zakázky probíhající v nedávné době.

1. Získání potřebných údajů – Údaje získáme hlavně z věstníku veřejných zakázek. Zde je vhodné nastavit správný filtr, tedy například zakázky za poslední rok, nebo i méně. Vhodný popis veřejné zakázky, případně pokud známe tak CPV kód. Viz část 5.2.5.
2. Určení potřebných proměnných – Určíme si, jaké údaje jsou pro nás podstatné, a tedy které budeme z vyhledaných položek získávat. Může to být předpokládaná cena, kterou je třeba určitým způsobem sjednotit, případně k ní uvádět i složitost zadání. Druh zadávacího řízení nám může pomoci určit, jak nejčastěji postupovaly ti, jejichž veřejná zakázka byla úspěšně uzavřena a dokončena. Přehled obvyklých dodavatelů, kteří vypsané zakázky získali, a zároveň si udělat přehled o těch kde nastávají častěji potíže, či více práce. Tento přehled můžeme využít pro jejich přímé kontaktování, případně nastavení podmínek tak abychom se vyhnuli budoucím problémům.
3. Vytvoření přehledu proměnných pro jednotlivé zakázky vyhledané v kroku 1. Ideálně pomocí datové tabulky, kdy každý řádek je jedna zakázka, a sloupce jsou zjišťované proměnné. Měli bychom mít snahu, aby byly vyplněné všechny údaje, ale ne vždy je to možné, takže určité procento prázdných polí není překážkou dalšího zpracování dat.
4. Získaná data můžeme doplnit spotřebitelským průzkumem trhu, to ovšem využijeme hlavně při nákupu obvyklého zboží či služeb. V případě zakázky na tvorbu strategií, nemá smysl se o toto dále pokoušet.
5. Data z věstníku zpracujeme základními statistickými postupy. Vytvoříme si přehled možností jednotlivých proměnných, a určíme základní statistiky (viz část 6.1).
6. Dále můžeme data rozdělit dle některého z parametrů a obdobně určit charakteristiky i rozdělení vzniklých skupin. Například určit, jak vypadaly zakázky, které proběhly bez problémů a čím se lišily od těch, které byly s problémy (nebyly vůbec uzavřeny smlouvy, nebylo dílo odevzdáno včas a bez chyb, …) V tomto kroku můžeme využít porovnání t-testem (část 0), případně zkoumat rozdělení pomocí chi-testu (část 6.5).
7. Poslední možností, kterou můžeme provést, je regresní analýza, kdy můžeme určit, které proměnné ovlivňují úspěšnost veřejné zakázky a tuto znalost využít pro nastavení vlastní zakázky tak aby se zvýšila pravděpodobnost, že také proběhne v pořádku. Vysvětlovaná proměnná by pak byla binární a obsahovala 1 v případě, že zakázka byla v pořádku a 0 pokud u ní byly problémy. Vysvětlující proměnné nemusí být jen ty, které jsme získali, ale mohou být i přepočtené. Například můžeme vytvořit proměnnou, která nabývá hodnoty 1, pokud byla předpokládaná cena vyšší než skutečná, 0 pokud byl tento odhad odpovídající, a -1 pokud byla předpokládaná cena nižší než skutečně vysoutěžená. (část 6.6)

8. Pokud by se nám podařilo najít statisticky významný model, či rozdíly, můžeme nastavit jednotlivé parametry tak aby zvyšovaly pravděpodobnost úspěchu.

Podobný postup byl využit v mnoha výzkumech a při přípravě mnoha veřejných zakázek, např. pří přípravě publikace (Ochrana 2015).

8. Popis šablony v Excelu pro získávání dat

Součástí tohoto nástroje pro strategickou práci je šablona v Excelu, která každému zájemci umožní lépe pracovat s potřebnými daty.

K využití šablony potřebujete mít k dispozici Microsoft Excel.

Šablona je přílohou číslo 3 tohoto textu. K šabloně je zpracován

- video návod, který je přílohou číslo 4 tohoto textu,
- textový návod, který je přílohou číslo 5 tohoto textu.

9. Závěr

Tento nástroj se zabývá získáváním a následným zpracováním dat. Pro strategickou práci jsou správná data a jejich adekvátní vizualizace například pomocí grafů nebo tabulek zásadní nejen pro přípravu strategií, ale také pro jejich realizaci a evaluaci. V dnešní době je velké množství dat veřejně dostupných (například databáze MONITOR, data Českého statistického úřadu, databáze k veřejným zakázkám, databáze veřejných strategií a podobně). Problémem tak často není nedostatek dat, ale spíše jejich nevhodná struktura nebo zahlcení nepodstatnými údaji. Správně provedené dolování dat (data mining) nám může pomoci se v datech lépe orientovat a lépe je využít. Jde o to, že osoby, které se na strategické práci podílejí, musí mít k dispozici potřebná a relevantní data, informace a poznatky. Taktéž musí znát nezbytné souvislosti (kontext). Jsou lidé, kteří mají schopnost (talent) přirozeně řešit problémy a předcházet jim, dokáží úspěšně uplatňovat strategický přístup[1], nicméně bez relevantních informací to však obvykle není možné.

Nástroj (tento text a video přednáška) je dostupný na stránkách projektu „Podpora strategického řízení a plánování ve veřejné správě ČR s důrazem na rozšiřování metod, nástrojů, inovací, znalostí a zkušeností v praxi“[2]. Vedle samotných textů metodických doporučení jsou k dispozici instruktážní videa a interaktivní šablony. Výstupy všech metodických nástrojů jsou zveřejněny na „Portále strategické práce v ČR“ (www.verejne-strategie.cz) v sekci Nástroje pro strategickou práci.

[1] Abychom mohli cokoli efektivně (úspěšně) řídit, musíme naplnit čtyři podmínky (1) danou činnost dobře znát včetně všech potřebných souvislostí a místních podmínek. Je třeba znát u dané činnosti silné i slabé stránky a taktéž příležitosti a rizika. (2) Musíme být schopni přijímat a prosadit „správná“ rozhodnutí a nést za ně odpovědnost. Za hodnotící kritérium toho, co je a co není správné lze považovat (a) soulad s cíli, (b) soulad s platnými předpisy a současně (c) soulad s morálními (etickými) hodnotami. Z hlediska prosazení to znamená mít podporu a důvěru. (3) Musíme vynakládat zdroje potřebné k naplňování cílů hospodárně, efektivně a účelně (ekonomicky racionální alokace zdrojů). K tomu je také nezbytné zvládat rizika a využívat příležitosti. (4) Musíme dosahovat stanovené cíle (a naplňovat potřeby společnosti) v souladu s právními předpisy a morálními hodnotami. Tyto čtyři podmínky lze také nazvat podmínkami strategického přístupu.

[2] Odkaz na projekt: https://mmr.cz/cs/microsites/portal-strategicke-prace-v-ceske-republice/nastroje-a-metodicka-podpora/nastroje-pro-strategickou-praci

Přílohy

Příloha č. 1: Prezentace: Metody získávání dat

Příloha č. 2: Video přednáška: Metody získávání dat

Příloha č. 3: Šablona v Excelu pro metody získávání dat

Příloha č. 4: Video přednáška: Jak použít šablonu v Excelu

Příloha č. 5: Návod, jak použít šablonu v Excelu

Přílohy metodického nástroje jsou dostupné na „Portále strategické práce v ČR“ (www.verejne-strategie.cz) v sekci Nástroje pro strategickou práci/Metody pro získávání dat.

Použitá literatura

MATĚJOVÁ, L, J NĚMEC, Milan KŘÁPEK a D KLIMOVSKÝ. Economies of Scale on the Municipal Level: Fact or Fiction in the Czech Republic? NISPAcee Journal of Public Administration and Policy. 2017, roč. 10, č. 1, s. 39–59. ISSN 1337-9038. doi:10.1515/nispa-2017-0002.

OCHRANA, F, Milan Jan PŮČEK, Michal PLAČEK, L MATĚJOVÁ a Milan KŘÁPEK. Předpokládaná hodnota veřejné zakázky, návrh metod a postupů ke stanovení předpokládané hodnoty veřejné zakázky. Veřejné zakázky a PPP projekty: Brno: Institut pro veřejné zakázky a PPPs, 2015, roč. 6. ročník, s. 84-95. ISSN 1803-9553.

PLAČEK, Michal, David ŠPAČEK, František OCHRANA, Milan KŘÁPEK a Petra DVOŘÁKOVÁ. Does excellence matter? National quality awards and performance of Czech municipalities. JEEMS Journal of East European Management Studies. 2019, roč. 24, č. 4, s. 589-613. ISSN 0949-6181. doi:10.5771/0949-6181-2019-4-589.

PŮČEK, J. M. (2021) *Jak zvládnout bakalářskou či diplomovou práci (metody a postupy odborné činnosti a smluvního výzkumu)*. Praha: Ambis Vysoká škola a.s.

PŮČEK, M., J. (2015) *Udržitelné finanční řízení*. Praha: NSZM. Dostupné na: https://www.dataplan.info/img_upload/f96fc5d7def29509aeffc6784e61f65b/publikace_pucek_udrz.fin.rizeni.pdf

PŮČEK, M., J., KOPPITZ, D. (2012) *Strategické řízení a plánování pro obce, města a regiony*. Praha: NSZM. Dostupné na: https://www.dataplan.info/img_upload/f96fc5d7def29509aeffc6784e61f65b/publikace-nszm-el_2.pdf

PŮČEK, Milan Jan. *Techniky efektivního řízení měst a obcí – část „SWOT analýza" (metodika).* Praha: NSZM, 2020. Dostupné z https://dobrapraxe.cz/cz/video/nszm-videokurz-analyza-rizik

Zákon č. 121/2000 Sb. Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) 26/2000.

Ministerstvo pro místní rozvoj ČR
Staroměstské náměstí 6
110 15 Praha 1
www.verejne-strategie.cz
www.mmr.cz