Studijní opory předmětu MT 003 část STATISTIKA v kombinovaném studiu Vysoké školy hotelové v Praze, bakalářský studijní program všech oborů

Obsahová náplň předmětu MT003 část statistika

1. Statistika - zajímavosti a historie statistiky
2. Statistika - vědní disciplína, základní statistické pojmy
3. Rozdělení statistických charakteristik, míry polohy
4. Míry variability
5. Typy proměnných, rozdělení četností
6. Statistické třídění
7. Metody zkoumání závislosti – kontingence, asociace
8. Metody zkoumání závislosti – regrese
9. Metody zkoumání závislosti – korelace
10. Absolutní přírůstek a index
11. Indexy úrovně a množství, indexní řady
12. Souhrnné indexy

Studijní literatura

<u>Základní:</u>
Novák, I.: Statistika. 2001,VŠH, ISBN 80-86578-56-9
Malec, M.: Elementární matematika. VŠH, 2008, Vysoká škola hotelová, ISBN 978-80-86578-62-0
<u>Doporučená:</u>
Jirásek, F.; Benda J.: Matematika pro bakalářské studium. Ekopress Praha, 2006, Tiskárny Havlíčkův Brod, ISBN 80-86929-02-7
Pecáková, I., Novák, I., Herzmann, J.:Pořizování a vyhodnocování dat. VŠE Praha, 2004,Oeconomica ISBN 80-245-0753-6
Hindls, R., Hronová, S., Novák, I.: Analýza dat v manažérském rozhodování. VŠE Praha, 1999 ,Grada, ISBN 80-7169-255-7
Kaňka, M., Henzler, J.: Matematika pro ekonomy. Ekopress, Praha 1997.
Hindls, R. a kol.: Statistika pro ekonomy. Professional Publishing, Praha 2007.

Průvodce studiem jednotlivých MODULŮ

A) ČÁST STATISTIKA

1. Modul

Modul tvoří tři tématické okruhy. Každý je probírán samostatně, jako kapitola v učebním materiálu.

<u>Tématické okruhy:</u>

1.1. Pojem statistika; Historie statistiky; Český statistický úřad
1.2 Statistické charakteristiky
1.3. Statistické třídění

Studijní cíle

V této kapitole se studenti seznámí se základními statistickými pojmy a historickými základy vědní disciplíny „statistika“. Bude objasněna práce a význam Českého statistického úřadu včetně uvedení kontaktních údajů na tuto státní organizaci. Dále budou studenti seznámeni se základními statistickými charakteristikami, postupy výpočtu a metodami statistického třídění.

Klíčová slova: pojem statistika, historie statistiky, ČSÚ, statistický soubor, normované normální rozdělení, typy proměnných, míry polohy, míry variability, četnostní tabulky

1.1. Pojem statistika, Historie statistiky, Český statistický úřad

Slovo **statistika** vzniklo z latinského slova
„**status**“ = stav
Pod pojmem statistika lze rozlišit následující významy:
•Číselné údaje
•Praktická činnost
•Vědní disciplína
V povědomí lidí se běžně vyskytují výrazy: statistický úřad, statistický vzorec, statistický výpočet.
Statistika se zabývá zkoumáním pravidelností a zákonitostí, projevujících se v tzv. **hromadných jevech** a vyjadřuje je **číselně**. Řada symbolů ve statistice čerpá z řecké abecedy.

<u>Statistika jako vědní disciplína:</u>
•Pracuje s hromadnými jevy
•Hledá zákonitosti hromadných jevů, ke kterým využívá hromadná pozorování
•Používá kvantitativní metody hodnocení s využitím matematiky
• Je aplikovatelná ve většině oborů kvantitativního výzkumu
<u>Statistika zahrnuje:</u>
•Sběr dat
–Průzkum
•Prezentování dat
–Grafy
–tabulky
•Popis dat
–Rozptyl
–Modus…

1.2. Historie statistiky

•Úřední zjišťování

- Vojenské a finanční účely panovníků, sčítání obyvatel
(Egypt 3000 let př.n. l., Čína)

•Univerzitní statistika

- 18.století, pouze slovní výrazy;německý prof. *Gottfried Achenwall* (1719-1772) – rozšíření slova „statistika“ (státní zvláštnosti)

•Politická aritmetika

- Nestačí jevy pouze popisovat, je nutné hledat zákonitosti jejich fungování
Adolphe Jacques Quételet (1796-1874) – zavedl pojem průměrný (ideální) typ člověka, koncept normálního rozdělení, střední hodnoty a rozptylu

•Teorie pravděpodobnosti + Matematická statistika

Matematici 17-19. století
- Ve 20. století dochází k výběrovému zjišťování s využitím teorie pravděpodobnosti
- Karl Pearson (1857-1936)

Statistika 19. a počátku 20. století Þ vytváření rozsáhlých souboru dat, sběr mnoha informací od co nejširšího okruhu respondentů, se zjevným cílem: obsáhnout ve svém šetření celou populaci a tím získat maximálně přesný obraz stavu společnosti

Úvaha z časových a finančních důvodů: je opravdu třeba zkoumat celou populaci, nebo postačí vybrat pouze její reprezentativní vzorek???

Na základe této myšlenky se počátkem 20. století zrodila **matematická statistika**, disciplína, jejímž charakteristickým rysem je hledání metod, jež by umožnily vytvoření závěru o celku na základě výběru

Česká republika má z historického hlediska ve statistice velmi silné kořeny. Za vůbec nejstarší dochovaný soupis je považován soupis majetku litoměřického kostela z roku **1058**, který je součástí zakládací listiny knížete Spytihněva II.

Významné osobnosti ve statistice lze nalézt na následujících www stránkách:http://www.economics.soton.ac.uk/staff/aldrich/Figures.htm

Český statistický úřad (ČSÚ)

K 1. 1. 1993 se vznikem ČR převzal ČSÚ všechny kompetence národního statistického úřadu (zákon č. 89/1995 Sb., O státní statistické službě, novelizace k 1. 1. 2001, ve znění pozdějších předpisů) ČSÚ nabízí přístup k mnoha významným statistickým informacím, klientům je k dispozici knihovna, studovna. ČSÚ poskytuje zásadní literární prameny se statistickými výstupy ve formě ročenek i necyklických publikací.

Adresa a kontaktní údaje ČSÚ:
Na padesátém 81
100 82 Praha 10
Tel: 274 051 111 (ústředna)
http://www.czso.cz/

1.2. Statistické charakteristiky

Základní pojmy

•**Hromadný jev –** předmět statistiky:

Hromadné jevy = jakékoliv přírodní nebo společenské **jevy** (skutečnosti), týkající se **souboru prvků** určitým způsobem definovaných – neboli jevy, které se vyskytují u velkého počtu jednotek, přičemž jejich konkrétní forma na individuální jednotce je výsledkem působení určitého seskupení činitelů

•Soubory = **statistické soubory –** množina statistických jednotek (mající společné vlastnosti)
•Prvky statistických souborů = **statistické jednotky –** základní objekt pozorování, na kterém je možné zkoumat konkrétní projevy sledovaného hromadného jevu (osoba, hotel, domácnost, událost…)
Rozsah souboru – počet jednotek, tvořících statistický soubor

Statistický soubor

•Vymezení:
–Věcné (druhové) (Příklad: průměrná měsíční mzda žen)
–Prostorové (Příklad: zaměstnankyně hotelu „Sen“)
–Časové (Příklad: srpen 200n)
•Rozsah:
–Základní soubor (populace), rozsah **N**
–Výběrový soubor (vzorek), rozsah **n**•Obsah
–Je určen znaky statistických jednotek

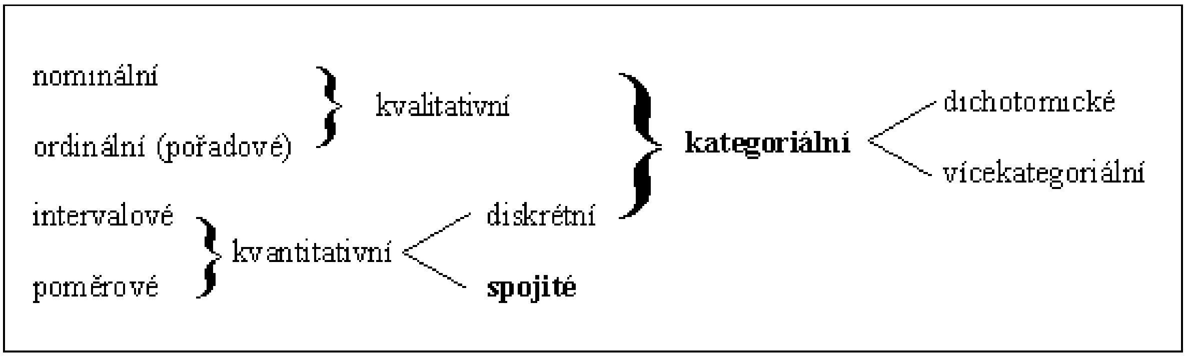
Statistický znak = proměnná – vnější, pozorovatelný, měřitelný projev vlastností statistické jednotky. **Variabilní statistický znak** – vlastnosti, v nichž se jednotky souboru mohou lišit

Statistické třídění

Třídění jednotek podle jednoho znaku v rámci statistického souboru umožňuje popis jeho charakteristických skupin
Třídění jednoduché = jednostupňové
(Příklad: rozdělení populace na muže a ženy)
vícestupňové (vícenásobné) (Příklad: muži do 50 let…)
Třídící znaky (kriteria) - znaky umožňující roztřídění souboru do skupin (věk…)
Statistické charakteristiky – charakteristika vlastnosti množiny hodnot = charakteristika vlastnosti souboru daných jednotek, například aritmetický průměr
Techniky pořizování dat

•Dotazování
- ústní – přesné odpovědi, minimalizace odmítání
- písemné – levné, nízká návratnost dotazníků
- telefonické – mnoho dotazovaných odmítá odpovědět
- elektronické…
•Měření (laboratorní výsledky…)

Proměnné ve statistice

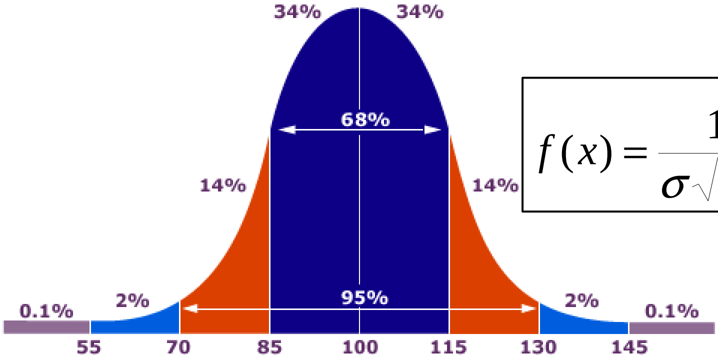


Nejčastější typy proměnných ve statistických výpočtech

- **Spojité proměnné** – mohou nabývat všech hodnot z konečného nebo nekonečného intervalu (tělesná teplota, cena zboží)
- **Diskrétní (nespojité) proměnné** - nabývají konečně nebo spočetně mnoha od sebe vzájemně oddělených hodnot (počet srdečních stahů za minutu, počet míst v restauraci)

Nejčastěji využívaný typ rozdělení ve statistice je **Gaussovo** rozdělení, které po transformaci převedeme na N(0; 1) normované normální rozdělení s následujícími vlastnostmi:

– zásadní význam ve statistické teorii i aplikacích
– je nejdůležitějším a nejfrekventovanějším rozdělením spojitých náhodných veličin
– lze jím nahradit i rozdělení diskrétní
– určujeme pravděpodobnost, že náhodná veličina X z normálního rozdělení bude nabývat hodnot z nějakého intervalu (a, b)



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

(*Převzato z* http://someonecz.blogspot.com/2009/10/iq-aneb-proc-je-chytry-kluk-sam.html)

Grafické znázornění normálního rozdělení je dáno touto symetrickou jednovrcholovou hustotou, která je zvonovitého tvaru a nikde neprotíná vodorovnou osu. Normované normální rozdělení je tabelizované vis část matematika kapitola Normální rozdělení. Tabulky bývají součástí základní statistické literatury nebo PC programů. Průměr μ - parametr ležící pod vrcholem hustoty. Parametr σ - směrodatná odchylka a jeho druhá mocnina σ^2 je rozptyl veličiny X. Plocha pod křivkou hustoty normálního rozdělení je rovna jedné.
Pravděpodobnost, že náhodná veličina nabude hodnot z určitého intervalu, je rovna ploše pod hustotou nad tímto intervalem.

Příklad:
Pro interval s hranicemi μ-1,96σ a μ+1,96σ má tato plocha velikost 0,95. Náhodná veličina *X* nabývá tedy hodnot z tohoto intervalu s 95% pravděpodobností a pouze s 5% pravděpodobností leží její hodnoty mimo uvedený interval

Statistické charakteristiky

Základní statistické charakteristiky studujeme u dvou typů statistických souborů:
•Základní statistický soubor (symbolika je vyjádřena řeckou abecedou) – nekonečné (hypotetické) nebo velmi rozsáhlé konečné soubory (statisíce jednotek)
•Výběrový statistický soubor(symbolika je vyjádřena latinskou abecedou) – malé (desítky jednotek) a velké výběry (stovky až tisíce jednotek)
Na základě údajů o výběrovém souboru, **na základě výběrových dat**, formulujeme **závěry o základním souboru**!

Míry polohy

Rozsah statistického souboru **- n**

$$\bar{x} = \frac{\sum x_i}{n}$$

•**Aritmetický průměr** $\bar{x}$ – střední hodnota kvantitativního statistického znaku (součet hodnot, dělený jejich počtem)

•**Medián** $\tilde{x}$ – je-li **n** (rozsah souboru) liché číslo, medián je prostřední hodnota; je-li n sudé číslo je medián aritmetickým průměrem dvou prostředních hodnot

•**Modus** $\hat{x}$ – hodnota nejčastěji se v souboru vyskytující

•**Kvantily**

Kvantily jsou míry polohy rozdělení pravděpodobnosti náhodné veličiny. Popisují body, ve kterých distribuční funkce náhodné proměnné prochází danou hodnotou.

V statistice kvantily rozdělují seřazený soubor na několik (zhruba) stejně velkých částí. Kvantily pro některé význačné hodnoty jsou označovány zvláštními jmény a pro nejdůležitější rozdělení jsou hodnoty základních kvantilů uváděny v tabulkách.

Percentil - dělí statistický soubor na setiny. 1% kvantil je 1. percentil.

Decil - dělí statistický soubor na desetiny. 10% kvantil je 1. decil.

Kvartily - oddělují ze statistického souboru čtvrtiny. Rozlišuje se dolní kvartil a horní kvartil. 25% kvantil je 1. kvartil (dolní kvartil) a 75% kvantil je 3 kvartil (horní kvartil).

Medián - kvantil rozdělující statistický soubor na dvě stejně početné množiny. Medián je totéž co 50% kvantil, 2. kvartil, 5. decil nebo 50. percentil.

Dobrý popis rozdělení pravděpodobnosti dostaneme stanovením dostatečného počtu kvantilů.

Příklad: Kvantily lze používat např. pro vyhodnocování přijímacích testů: bodové výsledky všech zájemců tvoří statistický soubor, zatímco příslušné kvantily označují, jaká část zájemců dosáhla daného výsledku. Pokud například kvantil 90 % má hodnotu 150 bodů a některý student v testu získal právě 150 bodů, ví, že má lepší hodnocení než 90 % všech studentů (je tedy mezi 10 % nejlepších a pokud má být přijato např. 15 % zájemců, měl by se kvalifikovat).

•**Dolní kvartil –** horní mez jedné čtvrtiny nejmenších hodnot v uspořádaném výběru
Výpočet n/4
- je-li výsledek celé číslo, kvartil je aritmetický průměr hodnoty n/4-té a n/4+1-ní
-není-li výsledek celé číslo, hledáme nejmenší celé číslo větší než n/4 a kvartilem je hodnota s tímto pořadovým číslem v uspořádaném výběru

•**Horní kvartil –** dolní mez jedné čtvrtiny největších hodnot v uspořádaném výběru
Výpočet 3n/4, postup stejný jako pro dolní kvartil

Poznámka: *Při výpočtu charakteristik míry polohy, vyjma aritmetického průměru, je nezbytné data seřadit do tzv. uspořádaného výběru, kdy data řadíme od minimální po maximální hodnotu. Rozsah souboru n musí být po seřazení zachován.*

Míry variability nedefinující proměnlivost uvnitř souboru dat

- ✓ **Variační rozpětí** R - je rozdílem mezi maximální a minimální hodnotou znaku: R = $x_{max} - x_{min}$

Používá se jako základní informace pro návrh hranic intervalů při statistickém třídění.

- ✓ **Mezikvartilové rozpětí** – rozdíl mezi horníma dolním kvartilem.

Udává délku intervalu, ve kterém leží zhruba polovina pozorovaných hodnot.

Míry variability definující proměnlivost uvnitř souboru dat

•Rozptyl – nejpoužívanější míra variability. Rozptyl je průměrná hodnota ze součtu čtverců odchylek jednotlivých hodnot souboru od aritmetického průměru (µ respektive $\overline{X}$); charakterizuje střední stupeň kolísání hodnot v souboru kolem aritmetického průměru. Je vyjádřen ve druhých mocninách jednotek sledovaného znaku.

Pro základní statistický soubor se označuje σ^2, σ_n^2 s^2,

pro výběrový statistický soubor σ_{n-1}^2, S^2.

x – numerická proměnná, x_i – hodnoty numerické proměnné u vybraných jednotek, i = 1,2…n

$$\sigma^2 = \frac{\sum x_i^2}{n} - \overline{x}^2$$

$$S^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}$$

•Směrodatná odchylka σ (základní soubor) respektive S (výběrový soubor) – je kladná hodnota druhé odmocniny rozptylu. Vyjadřuje střední kolísání hodnot znaku v souboru okolo aritmetického průměru ve stejných jednotkách v jakých je vyjádřen aritmetický průměr.

$$\sigma = \sqrt{\sigma^2}$$

$$S = \sqrt{S^2}$$

•Variační koeficient V_k – je relativní charakteristikou variability. Vyjadřuje variabilitu ve srovnatelném měřítku. Využívá se pro porovnání variabilit většího počtu u znaků, které často nabývají nejen rozdílné úrovně hodnot, ale jsou i v rozdílných jednotkách.

$$V_k = \frac{\sigma}{\bar{x}}.100\%$$

$$V_k = \frac{S}{\bar{x}}.100\%$$

***Poznámka:** pro statistické výpočty je nezbytné využívat vědecký kalkulátor s alespoň jednorozměrnou statistickou funkcí. Pro ovládání kalkulátoru využijte návod přikládaný výrobcem.*

1.3. Statistické třídění – rozdělení četností

Pod pojmem rozdělení četností chápeme uspořádání dat (hodnot) do skupin za účelem vyniknutí charakteristické vlastnosti sledovaných jevů. Nejčastěji se při statistickém třídění spojitých a diskrétních proměnných využívají tzv. četnostní tabulky.
Četnostní tabulky:

Četnostní tabulka diskrétní číselné proměnné zahrnuje:
Absolutní četnosti – hodnoty proměnné se řadí do tabulky od nejmenší k největší, každé hodnotě se připíše počet statistických jednotek ve výběru s danou hodnotou
Modus – hodnota proměnné s největší četností
Relativní četnosti – poměr četnosti a rozsahu výběru
Relativní četnosti nezávisí na rozsahu výběru, lze porovnávat dva výběry různého rozsahu.
Kumulativní četnosti a **kumulativní relativní četnosti** – kolik statistických jednotek nebo jaká část souboru má hodnoty nanejvýše rovné hodnotě, k níž jsou přiřazeny
Četnostní tabulka spojité číselné je principiálně podobná tabulce pro diskrétní proměnné s následujícími odlišnostmi:

•Hodnoty spojité proměnné se nemusí opakovat, proto se četnosti jednotlivých hodnot se nahradí četnostmi, patřících do jednotlivých intervalů – **intervalové četnosti**
•Intervaly volíme stejného rozsahu
•Nízký počet intervalů zkresluje výsledky, příliš vysoký počet intervalů výsledky znepřehledňuje.
Při statistickém třídění s využitím četnostních tabulek s výhodou počítáme vážený aritmetický průměr a vážený rozptyl.

Vážený aritmetický průměr: $= \dfrac{n_i x_i}{\Sigma n_i}$

Vážený rozptyl $= \dfrac{\Sigma n_i x_i^2}{\Sigma n_i} - \left(\dfrac{\Sigma n_i x_i}{\Sigma n_i}\right)^2$

Poznámka: U diskrétní proměnné jsou váhou hodnot proměnné četnosti, u spojité proměnné jsou váhou hodnot proměnné četnosti n_i, vlastní hodnoty x_i jsou nahrazeny hodnotou „střed intervalu“.

Grafické výstupy z četnostních tabulek

Pro znázornění četností u spojitých proměnných se používají **histogramy (četnosti se přiřazují intervalům)**. Pro znázornění četností u diskrétních proměnných se používají **polygony (četnosti se přiřazují jednotlivým hodnotám)**.

Shrnutí kapitoly

V kapitole byly vysvětleny základní statistické pojmy a nejzajímavější historické mezníky. Z práce ČSÚ vyplývá, že statistika nás provází dnes a denně doslova na každém kroku. Podmínkou úspěšného řešení statistických výpočtů je získávání kvalitních výběrových dat v dostatečném množství. Na základě zásadních statistických výpočtů lze popsat data získaná statistickým šetřením základními statistickými charakteristikami a je možné tato data dále hlouběji analyzovat. Rovněž statistické třídění, sestrojování četnostních tabulek, je nedílnou součástí před hlubší analýzou dat nadstavbovými statistickými metodami.

Pojmy k zapamatování:

Hromadný jev, statistický soubor, statistické šetření a třídění, proměnná, normované normální rozdělení, rozsah souboru, míry polohy (aritmetický průměr, kvantil, modus, medián), míry variability (variační a mezikvartilové rozpětí, rozptyl, směrodatná odchylka, variační koeficient), četnostní tabulky spojité a diskrétní proměnné, histogram a polygon.

Úkoly k zopakování a procvičení

Příklad 1.1.

Historie statistiky na území České republiky se datuje od:

a) středověku
b) novověku
c) 20 století

Řešení: a

Hlavní sídlo Českého statistického úřadu se nachází:

a) ve Zlíně
b) v Ostravě
c) v Praze

Řešení: c

Modus je:

a) hodnota nejčastěji se v souboru vyskytující
b) střední hodnota kvantitativního statistického znaku
c) prostřední hodnota kvantitativního statistického znaku

Řešení: a

Příklad 1.2.:

Rozhodněte, zda uvedené náhodné veličiny jsou diskrétní nebo spojité:

1. Počet servírek na území obce Plzeň
2. Obsah vitamínu B v pivu
3. Počet 5* hotelů na území Evropské unie
4. Procentické zastoupení bílkovin v hovězím mase

Řešení:

1. diskrétní proměnná, 2. spojitá proměnná, 3. diskrétní proměnná, 4. spojitá proměnná

Příklad 1.3.:
V obchodních řetězcích byla zjišťována cena vepřové plece v Kč. Byly získány následujíce data:
98,00; 94, 50; 89,50; 101,00; 92,00; 89,50; 90,50.
Určete typ souboru a proměnné, minimální a maximální hodnotu a vypočítejte aritmetický průměr, modus, medián, dolní kvartil, horní kvartil, variační rozpětí, mezikvartilové rozpětí, směrodatnou odchylku, rozptyl, variační koeficient u daného souboru dat.
Řešení:
Jedná se o výběrový statistický soubor.
Typ proměnné: spojitá proměnná
Uspořádaný výběr:
89,50; 89,50; 90,50;92,00; 94, 50; 98,00; 101,00
Minimum = 89,50
Maximum = 101,00
Dolní kvartil: n/4 = 7/4 = 1,75 → 2 hodnota = 89,50
Horní kvartil: 3n/4 = 21/4 = 5,25 → 6 hodnota = 98,00Variační rozpětí: max. – min. = 101,00 – 89,50 = 11,50
Mezikvartilové rozpětí: horní kvartil – dolní kvartil = 6 – 2 = 4
Aritmetický průměr: 93,57,-Kč
Modus: 89,50,-Kč
Medián: 92,00,-Kč
Směrodatná odchylka S: 4,49,-Kč
Rozptyl S^2 : 20,12,-Kč^2
Variační koeficient V_k: 0,048 = 4,8%

Příklad 1.4.:
Sestavte četnostní tabulku z následujících dat diskrétní proměnné -počty míst k sezení: 2, 4, 4, 8, 2, 2, 4, 4, 4, 3, 6, 2, 4, 4, 3, 4, 4, 8, 6, 4, 3, 8, 1, 2, 1.
Vypočítejte vážený aritmetický průměr a vážený rozptyl.
Řešení:
Uspořádaný výběr: 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 6, 6, 8, 8, 8
Četnostní tabulka diskrétní proměnné:

Hodnota x_i	Četnost n_i	Relativní četnost n_i/n	Kumul. četnost	Kumu. relativní četnost	n_ix_i	$n_ix_i^2$
1	2	2/25=0,08	2	0,08	1.2=2	1.2=2
2	5	5/25=0,20	2+5=7	0,28	2.5=10	4.5=20
3	3	3/25=0,12	7+3=10	0,40	3.3=9	9.3=27
4	10	10/25=0,40	10+10=20	0,80	4.10=40	16.10=160
6	2	2/25=0,08	20+2=22	0,88	6.2=12	36.2=72
8	3	3/25=0,12	22+3=25	1	8.3=24	64.3=192
Σ	n=25	1			97	473

Poslední dva sloupce tabulky slouží jako pomocné výpočty pro zjištění váženého aritmetického průměru a váženého rozptylu. Vážený aritmetický průměr = 3,88 míst, vážený rozptyl = 3,87míst^2·

Sestavte četnostní tabulku z následujících dat spojité proměnné – cena zboží v Kč: 15,50; 21,0; 18,50; 16,0; 28,50; 14,50; 24,0; 19,50; 25,50; 16,0; 17,50; 17,0; 29,0; 21,50; 22,0; 11,0;13,0; 24,50;23,50;25,0.

Vypočítejte vážený aritmetický průměr a vážený rozptyl.

Řešení:

Uspořádaný výběr: 11,0; 13,0; 14,50; 15,50; 16,0; 16,0; 17,0; 17,50; 18,50; 19,50; 21,0; 21,50; 22,0; 23,50; 24,0; 24,50; 25,0; 25,50; 28,50; 29,0

Četnostní tabulka spojité proměnné:

Dolní mez <	**Horní mez)**	**Střed intervalu x_i**	**Četnost n_i**	**$n_i x_i$**	**$n_i x_i^2$**
10,0	15,0	12,5	3	3.12,5=37,5	468,75
15,0	20,0	17,5	7	7.17,5=122,5	2143,75
20,0	25,0	22,5	6	6.22,5=135,0	3037,50
25,0	30,0	27,5	4	4.27,5=110,0	3025,00
-	-	-	Σ20	Σ405,0	Σ8675,00

Poslední dva sloupce tabulky slouží jako pomocné výpočty pro zjištění váženého aritmetického průměru a váženého rozptylu. Vážený aritmetický průměr = 20,25,-Kč, vážený rozptyl =· 23,69,-Kč^2·

Hodnocení

Každá správná odpověď nebo výsledek výpočtu je hodnoceno jedním bodem. Sebehodnocením je žádoucí dosáhnout alespoň 70% úspěšnost správných odpovědí, výsledků výpočtů. Jestliže jste nedosáhli požadované úspěšnosti, pokuste se zlepšit svůj studijní výsledek pozornějším studiem kapitoly, popřípadě se spojit s tutorem předmětu.

Další studijní zdroje

http://new.euromise.org/czech/tajne/ucebnice/html/html/node9.html
http://www.economics.soton.ac.uk/staff/aldrich/Figures.htm
http://www.czso.cz/

<u>Korespondenční úkol</u>

Následující korespondenční úkoly přinese student na následující soustředění (soustředění číslo 2). Korespondenční úkoly jsou rovněž umístěny v odpovědníku IS VŠH. S hodnocením se student seznámí během tutoriálu. Je požadována 70% úspěšnost v řešených úkolech.

Příklad:
Byly získány následující data výběrového statistického souboru – jedná se o ceny zboží v Kč:

15,70; 16,30; 16,90; 16,90; 16,90; 17,00; 17,00; 18,00; 18,30; 19,50; 19,60; 19,90; 20,50; 20,50; 21,40; 21,80; 22,10.

Určete typ souboru a proměnné, minimální a maximální hodnotu a vypočítejte aritmetický průměr, modus, medián, dolní kvartil, horní kvartil, variační rozpětí, mezikvartilové rozpětí, směrodatnou odchylku, rozptyl, variační koeficient u daného souboru dat. Jednotlivé statistické charakteristiky slovně okomentujte.

Sestrojte četnostní tabulku a **vypočíte**jte vážený aritmetický průměr a vážený rozptyl.
Sestavte vhodný grafický výstup z četnostní tabulky.

2. Modul

Modul tvoří tři tématické okruhy. Každý je probírán samostatně, jako kapitola v učebním materiálu.

<u>Tématické okruhy:</u>
2.1. Metody zkoumání závislosti
2.2. Kontingenční koeficienty, koeficient asociace
2.3. Regresní a korelační analýza

Studijní cíle

V této kapitole se studenti seznámí se zásadními metodami zkoumání závislosti. Upřesní si pojmy závisle a nezávisle proměnné. Kromě orientačních metodických výpočtů koeficientů kontingence a asociace se studenti seznámí se základy stěžejní statistické metody a to regresní a korelační analýzy. Uvedená metoda je vysvětlena na klasickém lineárním modelu. Rovněž jsou uvedeny a vysvětleny metody hodnocení těsnosti závislosti a další doplňkové definice vztahující se k této, v praxi často využívané, problematice.
Klíčová slova: závisle a nezávisle proměnná, dvojrozměrné tabulky, kontingenční koeficient, koeficient asociace, regrese a korelace, regresní přímka, korelační koeficient, koeficient determinace, index determinace

2.1. Metody zkoumání závislosti

Při zkoumání závislosti mezi **proměnnými** je nejdříve nutné posoudit, zda závislost **existuje**, tedy lze-li vysvětlovat změny hodnot jedné proměnné –vysvětlované = **závisle proměnné**, změnami hodnot proměnné druhé – vysvětlující = **nezávisle proměnné**.
U systému dvou proměnných obecně platí následující symbolika:
x – nezávisle proměnná (vysvětlující proměnná)
y – závisle proměnná (vysvětlovaná proměnná)
<u>Typy závislostí dvou proměnných:</u>
➢ **Jednostranná závislost** – závisle proměnnou může být pouze jedna z řešených proměnných (závislost velikosti mzdy na počtu odpracovaných hodin).**Vzájemná závislost** – obě proměnné lze volit za závisle nebo nezávisle proměnnou (výdaje domácnosti na cestování a na vzdělání).

Síla závislosti

Mezi proměnnými se zkoumá **síla – těsnost závislosti**. Závislost lze považovat za **silnou** – velmi těsnou, jestliže změny hodnot jedné proměnné jsou plně **vysvětlitelné** změnami druhé proměnné. Síla závislosti se popisuje různými **koeficienty**.Při nezávislosti proměnných jsou hodnoty koeficientů rovné nule, s růstem závislosti rostou jejich absolutní hodnoty (maximální hodnotou je jednička).

Dvojrozměrné tabulky

Naměřené (zjištěné) údaje – hodnoty závisle a nezávisle proměnné se uspořádávají do **dvojrozměrné tabulky.** V záhlaví tabulek se uvedou hodnoty proměnných, buňky tabulek obsahují četnosti kombinací obou proměnných (sdružené četnosti). Tabulky se doplňují součty za řádky a za sloupce (okrajové četnosti).

Můžeme rozlišit dva typy dvojrozměrných tabulek:

- **Kontingenční tabulky** – dvojrozměrná tabulka slovních proměnných.
- **Korelační tabulky** – dvojrozměrné tabulky číselných proměnných.

Příklad kontingenční tabulky s rozsahem souboru n = 130:

Spokojenost s novým vozidlem Věková kategorie	**Ano**	**Ne**	**Celkem Σ**
18 - 40 let	35	12	47
41- 60 let	42	11	53
nad 60 let	16	14	30
Celkem Σ	93	37	**130**

2.2. Kontingenční koeficienty, koeficient asociace

Kontingenční koeficienty se používají k měření závislosti. Některé kontingenční koeficienty jsou založeny na výpočtu hodnoty χ^2 (čti chí-kvadrát):

Pearsonův kontingenční koeficient P; P ∈ <0; 1)

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Cramérův kontingenční koeficient V; V ∈ <0; 1>

$$V = \sqrt{\frac{\chi^2}{n.h}}$$

; kde h = minimum (počet řádků-1, počet sloupců-1)

Výpočet hodnoty χ^2:

$$\boxed{\chi^2 = \sum \frac{(s-e)^2}{e}}$$

s – sdružené četnosti
e - teoreticky očekávané četnosti, za předpokladu nezávislosti proměnných.
e = (součet četností řádku . součet četností sloupce) / rozsah výběru n

Koeficient asociace
Asociační závislost stanovujeme mezi kvalitativními znaky .
Typicky se analýza asociace provádí pro *dichotomické znaky*, což jsou znaky, které principiálně nabývají pouze dvou hodnot, (ano, ne) a které se navzájem vylučují.
Asociační výpočet má za úkol:

- ze známých variant jednoho znaku odhadnout varianty znaku druhého.
- změřit intenzitu (stupeň těsnosti) vlastní asociace.

Hodnoceni koeficientu asociace: $R_A \in$ <-1;1>; přímá závislost: $R_A > 0$, nepřímá závislost: $R_A < 0$, nezávislost $R_A = 0$.
Čím více se hodnota R_A blíží k 1, tím je asociace silnější. Formálně se jedná o koeficient korelace pro 0,1 hodnoty proměnných.Asociační tabulka obsahuje četnosti výskytu jednotlivých kombinací uspořádané do čtyřpolní tabulky. Je to tedy speciální případ kontingenční tabulky typu 2x2 .
Příklad asociační tabulky:

b a	**ano**	**ne**	Celkem
ano	(ab)	(aβ)	**(a)**
ne	(αb)	(αβ)	**(α)**
Celkem	**(b)**	**(β)**	n

(ab); (aβ); (αb); (αβ) – sdružené četnosti
(a); (α); (b); (β) – okrajové četnosti
n – rozsah souboru

Výpočet koeficientu asociace R_A:

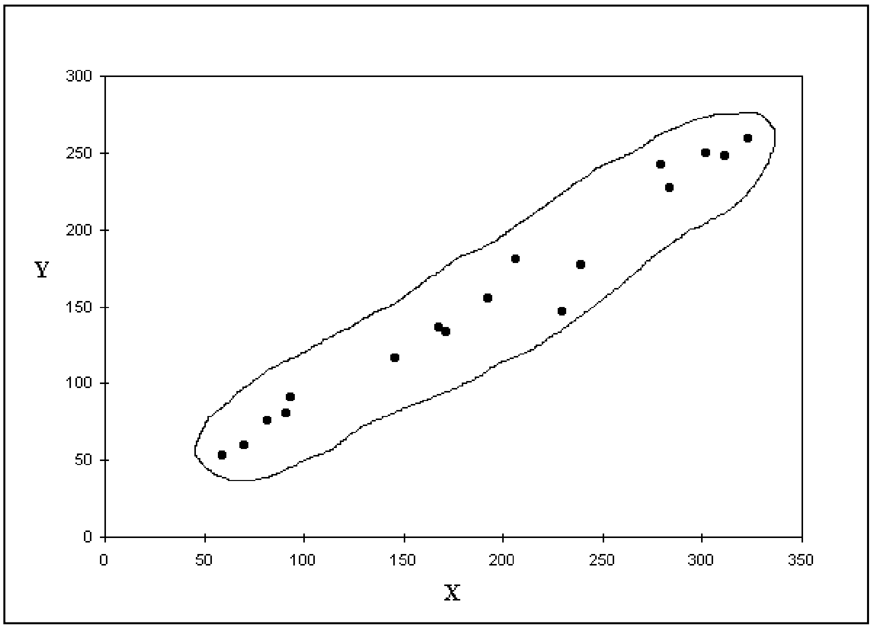
$$\boxed{R_{ab} = \frac{n(ab)-(a)(b)}{\sqrt{(a)(b)(\alpha)(\beta)}} =}$$

2.3. Regresní a korelační analýza

Regresní analýzou zkoumáme **průběh** a **korelační analýzou** zkoumáme **těsnost závislosti** mezi kvantitativními znaky - vztahy závisle proměnných (y) na nezávisle proměnných (x).
Před každým výpočtem, je nezbytné se ujistit, že mezi proměnnými závislost **existuje**!
Regrese je vyjádřena **matematickou funkcí**, která udává vlastní **průměrný průběh** sledované **závislosti** mezi proměnnými x a y. Parametry regresní funkce jsou počítány.

Před vlastním výpočtem regresní funkce je nutno zvolit vhodný typ funkce pro vyjádření průměrného průběhu závislosti.
Základním modelem je lineární regrese, kde matematickou funkcí je přímky s obecnou rovnicí y = a + bx.
Data, získaná statistickým šetřením, vytvářejí uspořádané dvojice [x; y]. Jednotlivé body se vynášejí do pravoúhlého osového systému a vzniká tak tzv. korelační pole.

Příklad korelačního pole:



U lineárních regresních funkcí, kam řadíme rovněž přímku, se číselné hodnoty parametrů počítají metodou nejmenších čtverců. (Detailní postup této metody nespadá do rámce předmětu statistika pro bakalářské studium. Zájemci se o metodě mohou dozvědět více v doporučené literatuře).

<u>Výpočet parametrů regresní přímky</u>
Níže uvedené pracovní vzorce jsou výstupem metody nejmenších čtverců a umožňují výpočet parametrů a, b regresní přímky s obecnou rovnicí y = a + bx.
<u>b- regresní koeficient</u> - udává, o kolik se v průměru změní hodnota závisle proměnné $\mathbf{y_i}$ v rovnici, jestliže hodnotu nezávisle proměnné $\mathbf{x_i}$ zvýšíme o jednotku. **Znaménko** před hodnotou regresního koeficientu určuje **průběh funkce**. Je-li **b kladné** číslo, funkce je lineárně **rostoucí**, je-li **b záporné** číslo, funkce je lineárně **klesající**. Regresní koeficient je **směrnicí přímky**.

$$\boxed{b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}}$$

, kde n - počet uspořádaných dvojic $[x_i;y_i]$.

<u>a - absolutní člen</u> - udává formální počátek regresní funkce, tj. hodnotu závisle proměnné y_i v rovnici při nulové hodnotě nezávisle proměnné x_i.

Absolutní člen je bodem, kde přímka protíná osu y.

$$\boxed{a = \frac{\sum y_i - b\sum x_i}{n}}$$

, kde n - počet uspořádaných dvojic $[x_i;y_i]$.

Korelační analýza

Pro hodnocení těsnosti lineární závislosti mezi proměnnými x, y, vyjádřené regresní funkcí například rovnicí přímky y = a + bx, je používán **<u>koeficient korelace - r.</u>** **Koeficient korelace** může nabývat hodnot z uzavřeného intervalu od (-1) do (+1), tj. $r \in \langle -1;+1 \rangle$. **Znaménko** před hodnotou korelačního koeficientu (stejně jako před hodnotou regresního koeficientu) **určuje směr závislosti**. Je-li **r kladné číslo**, je regresní funkce lineárně **rostoucí**, mezi proměnnými x, y je **přímo úměrná závislost**. Je-li **r záporné číslo**, je regresní funkce lineárně **klesající**, mezi proměnnými x, y je **nepřímo úměrná závislost**.

Poznámka: Znaménko před hodnotou regresního a korelačního koeficientu se vždy shoduje.
Absolutní hodnota korelačního koeficientu |r| udává **těsnost** hodnocené **závislosti**. Čím je absolutní hodnota koeficientu korelace **blíže** k **1**, tím je závislost **silnější**. Závislost lze hodnotit podle 3-5 bodové stupnice:

r = 0 – závislost neexistuje
|r| ∈ (0; 0,3) - slabá závislost
|r| ∈ <0,3; 0,6) - střední závislost
|r| ∈ <0,6; 0,8) – silná (těsná závislost)
|r| ∈ <0,8; 1) – velmi silná (velmi těsná závislost)
|r| = 1 – perfektní závislost

Pracovní tvar vzorce pro výpočet korelačního koeficientu:
x_i ; y_i - proměnné
n - počet uspořádaných dvojic $[x_i;y_i]$

$$\boxed{r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right].\left[n\sum y_i^2 - (\sum y_i)^2\right]}}}$$

Pro výpočet korelačního koeficientu lze využít i následující vzorec:

$$r = \frac{\text{kovariance } xy}{\sqrt{\text{rozptyl } x \,.\, \text{rozptyl } y}}$$

$$\text{, kde kovariance } xy = \frac{\Sigma x_i y_i}{n} - \bar{x}.\bar{y}$$

Výsledky výpočtu hodnoty r, za použití obou výše uvedených rovnocenných vzorců, jsou vždy shodné.

Koeficient determinace - r^2 Je druhou mocninou hodnoty koeficientu korelace. Koeficient determinace vyjádřený v procentech (r^2.100%) udává, z kolika procent jsou změny hodnot závisle proměnné **y** v regresní rovnici vysvětlovány hodnotami nezávisle proměnné **x**.

Index determinace

I^2 hodnotí kvalitu regresního modelu. Udává kolik procent rozptylu závisle (vysvětlované) proměnné **y** je vysvětleno modelem a kolik zůstalo nevysvětleno. Nabývá hodnot **od nuly do jedné** (teoreticky i včetně těchto krajních mezí), přičemž hodnoty **blízké nule** značí **špatnou kvalitu regresního modelu**; hodnoty **blízké jedné** značí **dobrou kvalitu regresního modelu**. Index determinace se udává se většinou v procentech.

Shrnutí kapitoly

V této kapitole byli studenti detailně seznámeni se základními statistickými metodami zkoumání závislostí a přípravou dat pro tyto statistické analýzy jako je například třídění dat do dvourozměrných tabulek. Byly nastíněny základní matematicko statistické operace v oblasti kontingence, asociace a zejména lineární regrese a korelace. Kapitola rovněž ukazuje práci s náročnějším početními postupy.

Pojmy k zapamatování

Dvojrozměrné statistické tabulky, kontingence, asociace, korelační pole, regrese, rovnice přímky, lineární regrese, metoda nejmenších čtverců, korelace, koeficient korelace, hodnocení těsnosti závislosti, koeficient determinace, index determinace.

Úkoly k zopakování a procvičení

Příklad 2.1.

Uveďte příklady koeficientů, které se ve statistické praxi užívají pro hodnocení těsnosti závislosti.

Řešení: Cramerův kontingenční koeficient, Pearsonův kontingenční koeficient, koeficient asociace, korelační koeficient

Výpočet Cramerova kontingenčního koeficientu je založen na hodnotě:

a) aritmetického průměru

b) χ^2 (chí-kvadrát)

c) směrodatné odchylky

Řešení: b

Příklad 2.2.
Byla sledována závislost spokojenosti hostů se službami bazénového baru hotelu na jejich věku. Na základě výpočtu **Pearsonova** a **Cramerova** kontingenčního koeficientu určete těsnost závislosti. Zdrojová data jsou uvedena v následující tabulce:

Spokojenost Věk	**Ano**	**Ne**	**Celkem** **Σ**
Do 45 let	25	7	
Nad 45 let	12	21	
Celkem **Σ**			

Řešení:
Nejdříve je nutné vypočítat **okrajové četnosti**.

Spokojenost Věk	**Ano**	**Ne**	**Celkem** **Σ**
Do 45 let	25	7	**32**
Nad 45 let	12	21	**33**
Celkem **Σ**	**37**	**28**	**65**

Dále je nutné vypočítat teoreticky očekávané četnosti ***e***, kdy spokojenost hosta nebude záviset na jejich věku (četnosti odpovědí za předpokladu nezávislosti obou proměnných).
***e* = (Součet četností řádku.součet četností sloupce) / rozsah výběru n**
s - skutečně zjištěné četnosti
e - teoretické sdružené četnosti – spokojenost hosta nezávisí na jeho věku *e*= (Součet četností řádku.součet četností sloupce) / rozsah výběru n

Spokojenost Věk	**Ano - s** **(*e*)**	**Ne - s** **(*e*)**	**Celkem** **Σ**
Do 45 let	**25** **(18,215)**	**7** **(13,785)**	32
Nad 45 let	**12** **(18,785)**	**21** **(14,215)**	33
Celkem **Σ**	37	28	**65**

n=65

χ^2 = 2,527+3,340+2,451+3,239 = **11,557**

$$P = \sqrt{\frac{11,557}{11,557 + 65}} = 0,389$$

$$V = \sqrt{\frac{11,557}{65.1}} = 0,422$$

Vypočítané hodnoty **obou kontingenčních koeficientů** svědčí o **střední závislosti** studovaných proměnných. Spokojenost hostů s bazénovým barem středně **závisí** na jejich věku.

Příklad 2.3.:

Byla studována závislost mezi spokojeností hostů se stravováním a s čistotou v hotelu. Celkem bylo osloveno 20 hostů. Vypočítejte hodnotu koeficientu asociace a okomentujte těsnost závislosti Data (odpovědi) jsou shrnuty v kontingenční tabulce:

Čistota Stravování	Ano	Ne	Celkem
Ano	11	1	**12**
Ne	5	3	**8**
Celkem	**16**	**4**	**20**

Řešení:

$$R_A = \frac{20.11 - 12.16}{\sqrt{12.16.8.4}} = \frac{28}{\sqrt{6144}} = \frac{28}{78,38} = 0,36$$

Závislost mezi spokojeností hostů se stravováním a čistotou je slabá – nepříliš těsná asociační závislost.

Příklad 2.4.:

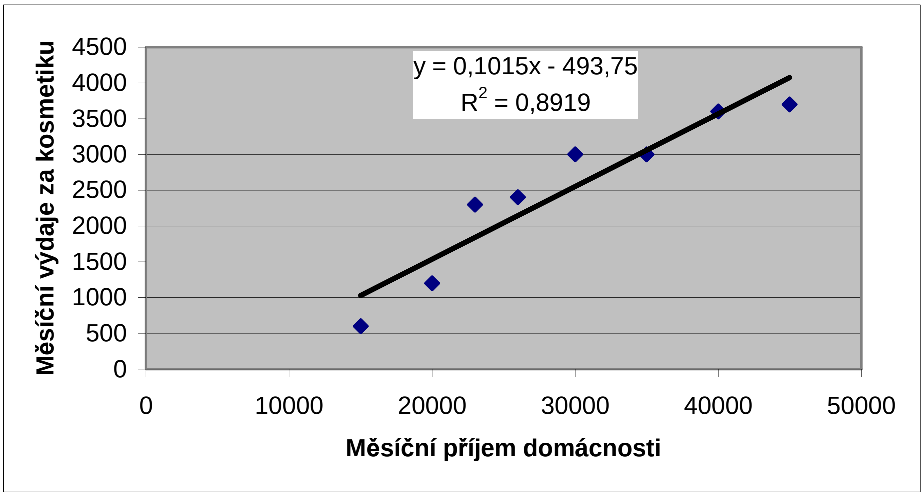
Byla studována závislost mezi měsíčním příjmem domácnosti a měsíčními výdaji za kosmetické výrobky. Předpokládáme, že regresní funkcí je přímka.

Získaná data jsou uvedena v následující tabulce:

Měsíční příjem domácnosti v tisících Kč	15	20	23	26	30	35	40	45
Měsíční výdaje za kosmetiku v tisících Kč	0,6	1,2	2,3	2,4	3,0	3,0	3,6	3,7

➢Určete závisle a nezávisle proměnnou➢Zakreslete korelační pole
➢Vypočítejte parametry rovnice funkce
➢Sestavte rovnici přímky
➢Vypočítejte odhad, kolik korun měsíčně vydá domácnost s příjmem 29000,-Kč
➢Vypočítejte hodnotu korelačního koeficientu
➢Určete těsnost závislosti
➢Určete typ úměry závislosti
➢Vypočítejte a okomentujte koeficient determinace

Řešení:
Nezávisle proměnná x – měsíční příjem v domácnosti v Kč
Závisle proměnná y – měsíční výdaje za kosmetiku v Kč



Domácnost s příjmem 29000,-Kč v průměru vydá **2450,-Kč** za kosmetiku.
Hodnota korelačního koeficientu **r = 0,944** svědčí o velmi silné závislosti průměrných výdajů za kosmetiku na průměrných příjmech domácnost.
Mezi proměnnými platí **přímá úměra**.
Změny hodnot závisle proměnné – průměrné měsíční výdaje jsou z **89,19%** vysvětlovány hodnotami nezávisle proměnné – průměrné měsíční příjem.

Hodnocení
Každá správná odpověď nebo výsledek výpočtu je hodnoceno jedním bodem.
Sebehodnocením je žádoucí dosáhnout alespoň 70% úspěšnost správných odpovědí, výsledků výpočtů. Jestliže jste nedosáhli požadované úspěšnosti, pokuste se zlepšit svůj studijní výsledek pozornějším studiem kapitoly, popřípadě se spojit s tutorem předmětu.

Další studijní zdroje:
http://iastat.vse.cz/

Korespondenční úkol
Následující korespondenční úkoly přinese student na následující soustředění (soustředění číslo 3). Korespondenční úkoly jsou rovněž umístěny v odpovědníku IS VŠH. S hodnocením se student seznámí během tutoriálu. Je požadována 70% úspěšnost v řešených úkolech.

Příklad:
Byla sledována závislost spokojenosti hostů hotelu na jejich vzdělání. Na základě výpočtu Cramerova a Pearsonova kontingenčního koeficientu určete těsnost závislosti. Data jsou uvedena v následující tabulce.

Spokojenost / Vzdělání	Ano	Ne	Celkem
Středoškolské	10	5	
Vysokoškolské	15	8	
Celkem			

Příklad
Byla studována závislost velikosti tržby v milionech Kč na počtu hotelových hostů. Data jsou uvedena v následující tabulce:

Tržba	5,2	5,8	6,4	6,5	6,8	7,0	7,3	7,5	7,8	8,1
Počet hostů	250	300	330	336	350	359	369	375	383	390

a) Určete závisle a nezávisle proměnnou
b) Zakreslete korelační pole
c) Vypočítejte parametry rovnice funkce
d) Sestavte rovnici přímky
e) Vypočítejte, jaká je tržba, jestliže byl počet hotelových hostů 370
f) Vypočítejte hodnotu korelačního koeficientu
g) Určete těsnost závislosti
h) Určete typ úměry závislosti
i) Vypočítejte a okomentujte koeficient determinace

3. Modul

Modul tvoří tři tématické okruhy. Každý je probírán samostatně, jako kapitola v učebním materiálu.

Tématické okruhy:

3.1. Absolutní přírůstek a index
3.2 Indexní řady
3.3. Souhrnné indexy

Studijní cíle

V závěrečné kapitole se studenti seznámí se zásadní terminologií a statistickými výpočty z problematiky absolutních přírůstků a indexů, kde nedílnou součástí kapitoly je rovněž klasifikace indexů. Pracovní tvary vzorců jsou, z důvodů snadnější orientaci studenta v problematice, detailně rozepsány. Kapitola řeší rovněž problematiku indexních řad a postupových možností práce s pouze bazickými nebo pouze řetězovými indexy. Ze složitějších metod jsou uvedeny zásady v klasifikaci a výpočtových postupech v oblasti souhrnných indexů s důrazem na agregátní cenové indexy.

Klíčová slova: absolutní přírůstek, index, klasifikace indexů, indexy množství, indexy úrovně, extenzitní ukazatel, intenzitní ukazatel, individuální indexy, indexní řady, souhrnné indexy (Laspeyresův, Paascheho, Fisherův souhrnný index), vážený průměr, cenový index

3.1. Absolutní přírůstek a index

Starší období = **základní období** (základ pro porovnávání)
Porovnávané období = **běžné období**
Absolutní přírůstek Δ = rozdíl mezi dvěma časovými ukazateli – o kolik jednotek se změní (+zvětší, -zmenší) hodnota v běžném období oproti základnímu období
Index I = podíl mezi dvěma ukazateli – kolik procent hodnoty základního období činí hodnota běžného období
Indexy a absolutní přírůstky se vzájemně doplňují a měly by být uváděny společně.

Klasifikace indexů
Indexy množství – porovnávají hodnoty extenzitních ukazatelů **- *q*,** tj. ukazatelů vyjadřujících množství, velikost, objem (počet hostů, tržba, prodané množství zboží určitého druhu)
Indexy úrovně – porovnávají hodnoty intenzitních ukazatelů **- *p*,** tj ukazatelů vyjadřujících úroveň, hladinu, intenzitu (cena, tržba na pracovníka)
Každý intenzitní ukazatel je poměrem ukazatelů extenzitních:
Tržba na pracovníka = tržba / počet pracovníků
Jmenovatel je extenzivní ukazatel - nositel intenzity
Indexy lze rovněž rozdělit na:

- Indexy individuální
- Indexy souhrnné

Individuální indexy – indexy stejnorodých ukazatelů
a) dílčí hodnoty lze druhově a prostorově shrnout **součtem** – extenzitní ukazatele (počet hostů restaurací řetězce „Eurest“)
b) dílčí ukazatele lze druhově a prostorově shrnout **průměrem** – intenzitní ukazatele (cena 0,5l piva v hotelech „Holiday Inn“)
Souhrnné indexy – popisují změny množství či úrovně v celku, složeném z nestejnorodých částí (změna ceny mléčných výrobků v síti „Tesco“)

Jednoduchý index množství:
Symbolika:
Základní období q_0Běžné období q_1
$I(q_i) = q_{1i}/q_{0i}$
Odpovídající absolutní přírůstek: $\Delta(q_i) = q_{1i} - q_{0i}$

Složený index množství:- je váženým průměrem jednoduchých indexů, váhou jsou dílčí hodnoty ze základního období q_{0i} $I(\Sigma q_i) = \Sigma q_{1i}/\Sigma q_{0i}$Odpovídající absolutní přírůstek:
$\Delta(\Sigma q_i) = \Sigma q_{1i} - \Sigma q_{0i}$
- je součtem jednoduchých absolutních přírůstků

Individuální jednoduché indexy úrovně
Každý intenzitní ukazatel je poměrem ukazatelů extenzitních:
p = Q/q
p - intenzitní ukazatel
Q - extenzitní ukazatel
q - extenzitní ukazatel nositel intenzity
Q = pq
Intenzitní ukazatel je stejnorodý jsou-li stejnorodé oba extenzitní ukazatele Q, q (jejich dílčí hodnoty lze sčítat).
Dílčí hodnoty stejnorodého intenzitního ukazatele $p_i = Q_i/q_i$ lze shrnout poměrem součtů dílčích extenzitních ukazatelů $\Sigma SQ_i/Sq_i$

$Q_i=p_i.q_i$ $\Sigma Q_i/\Sigma q_i = \Sigma\, p_i q_i/\Sigma q_i = \boxed{\overline{p}}$

Základní období:

$$\boxed{\overline{p}_0 = \frac{\Sigma Q_0}{\Sigma q_0}}$$

Běžné období:

$$\boxed{\overline{p}_1 = \frac{\Sigma Q_1}{\Sigma q_1}}$$

Jednoduchý index úrovně a jemu odpovídající absolutní přírůstek
$I(p_i) = p_{1i}/p_{0i} \Delta\, (p_i) = p_{1i} - p_{0i}$
Individuální složené indexy úrovně

Složený index úrovně odráží změny dílčích hodnot ukazatele, ale i změny ve struktuře nositele intenzity – **je indexem proměnlivého složení**
Složený index úrovně a jemu odpovídající absolutní přírůstek:

$$\boxed{I(\overline{p}) = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\dfrac{\Sigma p_{1i} q_{1i}}{\Sigma q_{1i}}}{\dfrac{\Sigma p_{0i} q_{0i}}{\Sigma q_{0i}}}}$$

Odpovídající absolutní přírůstek:

$$\boxed{\Delta(\overline{p}) = \overline{p}_1 - \overline{p}_0 = \frac{\Sigma p_{1i} q_{1i}}{\Sigma q_{1i}} - \frac{\Sigma p_{0i} q_{0i}}{\Sigma q_{0i}}}$$

3.2. Indexní řady

Indexní řady jsou tvořeny **bazickými** a **řetězovými indexy.Bazické indexy** ukazují změnu průměrné hodnoty proměnné x (například počet zaměstnanců) ve srovnání s rokem výchozím – srovnávací období.
Řetězové indexy ukazují jak se změní hodnota proměnné x (například počet pracovníků pivovarů) ve srovnání s předcházejícím rokem.
Pro období 200n – 200n+i, kde i = 1, 2…, platí pro hodnoty x:
Bazické indexy: x200n/x200n; x200n+1/x200n; x200n+2/x200n….
Řetězové indexy: x200n+1/x200n; x200n+2/x200n+1; x200n+3/x200n+2…

Z řetězových indexů lze vypočítat indexy bazické a naopak:

- **Řetězové indexy** jsou podílem dvou za sebou jdoucích bazických indexů, čitatelem je bazický index pro vyšší (mladší) ročník.
- **Bazické indexy** lze vypočítat postupným násobením řetězových indexů.

3.3. Souhrnné indexy

Souhrnný index množství
Souhrnný index množství - popisuje změny množství či úrovně v celku, složeném z nestejnorodých částí. **Souhrnný index množství lze vypočítat jako:**
•**Vážený aritmetický průměr** individuálních indexů množství
•**Laspeyresův** souhrnný index množství
•**Paascheho** souhrnný index množství
•**Fisherův** souhrnný index množství

Souhrnný index množství vyjádřený jako vážený aritmetický průměr individuálních indexů množství$I_q = \Sigma I(q_i)v_i I(q_i) = q_{1i}/q_{0i} I(q_i)$ – individuální index množství
q_{1i} – prodané množství v běžném období
q_{0i} – prodané množství v základním období
v_i – „váha“ jednotlivých druhů zboží zvolená tak, aby její součet byl roven jednéLaspeyresův souhrnný index množství$I_{q,L} = \Sigma q_{1i}p_{0i}/\Sigma q_{0i}p_{0i} q_{1i}$ – prodané množství v běžném období
q_{0i} – prodané množství v základním období
p_{0i} – cena v základním období
Paascheho souhrnný index množství$I_{q,P} = \Sigma q_{1i}p_{1i}/\Sigma q_{0i}p_{1i} q_{1i}$ – prodané množství v běžném období
q_{0i} – prodané množství v základním období
p_{1i} – cena v běžném období

Fisherův souhrnný index množství
Fisherův souhrnný index množství je geometrickým průměrem indexů Laspeyresova a Paascheho:

$$\boxed{I_{q,F} = \sqrt{I_{q,L}.I}_{q,P}}$$

Souhrnné indexy úrovně

Nejčastěji užívané jsou souhrnné **indexy cenové**.

Souhrnné indexy cenové jako **vážené aritmetické průměry individuálních cenových indexů**

$\mathbf{I_p = \Sigma I(p_i)v_i}$

$I(p_i)$ – individuální cenové indexy

v_i - „váhy“, jejichž součet je roven jedné

Index **souhrnně** charakterizuje změnu cen zboží prodávaného například obchodním řetězcem – váhou je podíl tržby za jednotlivé druhy zboží na tržbě z prodeje, zjištěné za určité období.

Agregátní cenové indexy:

Prodaná množství za určité období se ocení nejdříve cenami základního a poté cenami běžného období a výsledky se pak porovnávají Zvolí-li se prodaná množství ze základního období je indexem **Laspeyresův souhrnný cenový index:**

$I_{q,L} = \Sigma p_{1i}q_{0i}/\Sigma p_{0i}q_{0i}$Zvolí-li se prodaná množství z běžného období je indexem **Paascheho cenový index:**

$I_{q,P} = \Sigma p_{1i}q_{1i}/\Sigma p_{0i}q_{1i}$Kombinací je **Fisherův souhrnný cenový index**

Shrnutí kapitoly

Poslední kapitola seznámila studenty se zásadami v klasifikaci a výpočtech absolutních přírůstků a indexů využívaných v ekonomické praxi. Detailní výklady vzorců umožňují studentovi realizovat výpočty ze zdrojových dat získaných statistickým šetřením a aplikovat dané výstupy v navazujících studijních úkolech. Student je podrobně seznámen s problematikou a praktickým využitím indexních řad. V závěru kapitoly je řešena náročná problematika souhrnných indexů s uvedením zásadních postupových výpočtů. Pozornost je věnována rovněž agregátním cenovým indexům.

Pojmy k zapamatování

Absolutní přírůstek, index, základní a běžné období, extenzitní a intenzitní ukazatel, nositel intenzity, individuální a souhrnný index, indexní řady, bazický a řetězový index, vážený aritmetický průměr, Laspeyresův, Paascheho, Fisherův souhrnný index, cenový index

Úkoly k zopakování a procvičení

Příklad: 3.1.

Indexy množství porovnávají hodnoty:

a) extenzitních ukazatelů

b) intenzitních ukazatelů

c) absolutních přírůstků

Řešení: a

Příklad 3.2.:
Byl studován počet hostů v pěti hotelech ve dvou po sobě jdoucích letech. Zjištěné hodnoty jsou uvedeny v tabulce:

Hotel	Počet hostů v roce 200n	Počet hostů v roce 200n+1	Absolutní přírůstky **Δ (q_i)**	**Indexy I (q_i)**
Aida	721	895		
Bílý Lev	452	521		
Libuše	529	498		
Platan	890	948		
U Staré Paní	356	326		
Všechny hotely				

Vypočítejte:-jednoduché absolutní přírůstky a indexy
-složený absolutní přírůstek a index
Dokažte, že složený index je váženým aritmetickým průměrem jednoduchých indexůŘešení:

Hotel	Počet hostů v roce 200n **q_{0i}**	Počet hostů v roce 200n+1 **q_{1i}**	Absolutní přírůstky **Δ q_i)**	**Indexy I (q_i)**
Aida	721	895	174	1,24
Bílý Lev	452	521	69	1,15
Libuše	529	498	-31	0,94
Platan	890	948	58	1,07
U Staré Paní	356	326	-30	0,92
Všechny hotely	2948	3188	**240**	1,08

Složený absolutní přírůstek = 240
Složený absolutní index = 1,08
Důkaz:
Složený absolutní přírůstek je součtem jednoduchých absolutních přírůstků: **240** = 3188 - 2948 = 174+69-31+58-30.

Příklad 3.3.
Byly zjištěny průměrné počty zaměstnanců pivovarů v České republice v posledních pěti letech. Z následujících hodnot uvedených v tabulce **sestavte bazické a řetězové indexy**. Pro bazické indexy zvolte základ ročník 200n. **Vysvětlete** význam obou typů indexů.

Ročník	Počet zaměstnanců -x	Indexy bazické	Indexy řetězové
200n	6932		
200n+1	6280		
200n+2	6364		
200n+3	7205		
200n+4	7186		

Řešení:

Ročník	Počet zaměstnanců -x	Indexy bazické	Indexy řetězové
200n	6932	1,00	-
200n+1	6280	0,91	0,91
200n+2	6364	0,92	1,01
200n+3	7205	1,04	1,13
200n+4	7186	1,04	1,00

Bazické indexy ukazují změnu průměrného počtu zaměstnanců ve srovnání s rokem 200n. Řetězové indexy ukazují jak se změnil počet pracovníků pivovarů ve srovnání s předcházejícím rokem.
Příklad 3.4:

V odborné literatuře byly publikovány bazické indexy pro období 200n-200n+4 pro počty pokojských v penzionech:
1,00; 1,15; 1,07; 0,96; 1,11. **Vypočítejte** hodnoty řetězových indexů.Řešení:
Řetězové indexy jsou podílem dvou za sebou jdoucích bazických indexů, čitatelem je bazický index pro vyšší ročník.
<u>Hledané řetězové indexy:</u>
200n: -; 200n+1: 1,15/1 = **1,15**; 200n+2:1,07/1,15 = **0,93**; 200n+3:0,96/1,07 = **0,90**; 200n+4:1,11/0,96 = **1,16**

Hodnocení
Každá správná odpověď nebo výsledek výpočtu je hodnoceno jedním bodem. Sebehodnocením je žádoucí dosáhnout alespoň 70% úspěšnost správných odpovědí, výsledků výpočtů. Jestliže jste nedosáhli požadované úspěšnosti, pokuste se zlepšit svůj studijní výsledek pozornějším studiem kapitoly, popřípadě se spojit s tutorem předmětu.

Další studijní zdroje:
Hindls, R., Hronová, S., Seger, J.: Statistika pro ekonomy. Professional Publishing, Praha 2002, druhé vydání, ISBN 80-86419-30-4

Korespondenční úkol

Následující korespondenční úkoly odešle student do odevzdávárny. Korespondenční úkoly jsou rovněž umístěny v odpovědníku IS VŠH. S hodnocením se student seznámí na základě elektronické komunikace případně konzultace. Je požadována 70% úspěšnost v řešených úkolech.

Příklad:

Průměrná délka pobytu hosta v penzionu ve dnech (p) je poměrem celkového počtu pobytových dnů (Q) a počtu hostů (q). Zjištěné hodnoty za dva roky jdoucí po sobě jsou uvedeny v následující tabulce:

Penzion	Pobytové dny		Počet hostů		Délka pobytu hosta		Přírůstek délky pobytu	Index délky pobytu
	Q_{0i}	Q_{1i}	q_{0i}	q_{1i}	p_{0i}	p_{1i}	$\Delta(p_i)$	$I(p_i)$
Babka	1950	2390	490	580				
Merlin	2980	2650	489	434				
Oba penziony								

Vypočítejte:

a) jednoduché absolutní přírůstky a indexy

b) složený absolutní přírůstek a indexPříklad

Ve sborníku byly publikovány řetězové indexy pro období 200n-200n+4 pro počet kuchařů v závodních podnikových jídelnách: -; 0,98; 1,05; 1,19; 0,91. **Vypočítejte** hodnoty bazických indexů.