

Literatura

učebnice z minulého semestru

Jarošová, Pecáková – sbírka příkladů pro statistiku B (2000 a novější)

1. Testování hypotéz

H_0 – testovaná (nulová) hypotéza

H_1 – alternativní hypotéza (dvoustranná, jednostranná, pravostranná)

Stojí proti sobě – nemají průnik

Ze základního souboru vybereme výběrový soubor, na základě jeho vlastností **odhadujeme** vlastnosti celku (bodové a intervalové odhady)

Testy hypotéz – testujeme odhad, hypotézu

Výsledek testu – přijaté jedné ze dvou hypotéz a zamítnutí té druhé.

Protože máme údaje jen z výběru, můžeme se dopustit chyby – dvě možné odpovědi
è dvě možné chyby:

Chyba 1. druhu – nesprávné zamítnutí H_0 , pravděpodobnost chyby - α – hladina významnosti

Chyba 2. druhu – nesprávné přijetí H_0 – pravděpodobnost chyby – β .

Hladinu významnosti (α) volíme rozumně nízkou (např. 5%, někdy 1%)

Pravděpodobnost druhého druhu nevolíme, může být poměrně velká.

è Zamítnutí nulové hypotézy a přijetí nulové hypotézy:

zamítnutí nulové a přijetí alternativní, lepší – riskujeme menší chybu (α) Nulová hypotéza téměř jistě neplatí.

přijetí nulové hypotézy a zamítnutí alternativní – říkáme „nezamítáme nulovou hypotézu“ – hrozí větší chyba (β), riziko neznáme, jen tušíme, že může být velká pravděpodobnost chyby. Říkáme např. „zjištěné údaje nejsou v rozporu s H_0 “ nebo „test neprokázal neplatnost hypotézy H_0 “ Při velkém rozsahu výběru má nezamítnutí větší význam.

Testové kritérium – charakteristika vypočtená z výběrových dat. Pro každý test určitý vzorec. Má nějaké známé rozdělení

obor hodnot testového kritéria – obor přijetí (V) a kritický obor (W)

Padne-li hodnota testového kritéria do kritického oboru W è zamítáme nulovou hypotézu.

1.1. Postup při testování hypotéz - příklad

Výrobce tvrdí, že dodávané žárovky mají průměrnou životnost 400 hodin

Při kontrole 25 náhodně vybraných žárovek, jsme zjistili průměrnou životnost 380 hodin. Rozptýl byl 6000. Předpokládáme pro životnost normální rozdělení.

Pokuste se vhodným testem prokázat, že střední životnost dodávaných žárovek je nižší než udává výrobce.

Zformulujeme hypotézy:

Nulová hypotéza: průměrná životnost je 400

Alternativní hypotéza: průměrná životnost je nižší

Zvolíme $\alpha = 5\%$

Nalezneme testové kritérium a jeho rozdělení – viz vzorce

Toto testové kritérium má v případě platnosti nulové hypotézy studentovo t-rozdělení, kde počet stupňů volnosti (n) se vypočítá jako $n-1$ (rozsah výběru - 1)

Kritický obor a obor přijetí

Kritický obor: průměr z výběru je o hodně menší než výrobce stanovená životnost. Dolních 5% ze studentova rozdělení – kritickou hodnotou bude 5% kvantil t-rozdělení 24 stupňů volnosti (-1,71)

Vyšší hodnoty – obor přijetí

Dosadíme do vzorce testového kritéria

$$t = 380 - 400 / (6000 / 25)^{1/2} = - 1,29$$

Formulace závěru

-1,29 je v oboru přijetí \Rightarrow nezamítáme nulovou hypotézu \Rightarrow tvrzení výrobce se na základě údajů z výběru nepodařilo vyvrátit.

1.2. Velký rozsah výběru

více než 30

můžeme použít normální rozdělení

1.3. Test hypotézy o rozptylu normálního rozdělení

viz učebnice

1.4. Test hypotézy o podílu (o pravděpodobnosti), velký výběr

např. testujeme podíl kvalitních výrobků v dodávce nebo testujeme hypotézu o pravděpodobnosti padnutí čísla na kostce

1.4.1. Příklad

Vedení určité strany tvrdí, že ji v daném regionu podporuje 15% voličů.

Pokuste se toto tvrzení vyvrátit, víte-li, že při výzkumu veřejného mínění se pro tuto stranu vyslovilo 105 z 880 náhodně vybraných osob (11,93181818%)

$$H_0: p = 0,15$$

$$H_1: p < 0,15$$

Hladina významnosti – zvolíme $\alpha = 1\%$, protože 880 je relativně hodně

Normované normální rozdělení (symetrické podle nuly)

P – čím menší podíl ve výběru, tím je pravděpodobnější alternativní hypotéza

Kritický obor – menší hodnoty než kritická hodnota: 1% kvantil normovaného normálního rozdělení (-2,33)

Vypočítáme hodnotu testového kritéria – dosadíme do vzorce

$$U = ((105 / 880) - 0,15) / (0,15 * (1 - 0,15) / 880) = -2,55$$

Hodnota padla do kritického oboru \Rightarrow na 1% hladině významnosti zamítáme nulovou hypotézu – s jednocentním rizikem chyby můžeme tvrdit, že podíl voličů této strany je menší, než udávaných 15%

p-hodnota ve statgraphicsu – minimální hladina významnosti – minimální riziko chyby při kterém lze ještě zamítnout nulovou hypotézu
nižší p-hodnota než naše hladina významnosti \Rightarrow zamítáme nulovou hypotézu

2. přednáška

test hypotézy o střední hodnotě normálního rozdělení
test hypotézy o střední hodnotě libovolného rozdělení při velkém výběru
test o rozptylu
test o pravděpodobnosti – velký výběr

1.5. Test o shodě dvou středních hodnot

1.5.1. Nezávislé výběry

Např. dva dodavatele a testujeme, že životnost výrobku jednoho je v průměru větší než druhého

Příklad: Nezávislý výběr velkého rozsahu

H0: $\mu_1 = \mu_2$

H1: $\mu_2 > \mu_1$

Testové kritérium U

Pro alternativní hypotézu svědčí vysoké hodnoty U (kritický obor)

1.5.2. Závislé výběry

Např. U stejných jednotek zkoumáme skutečnost ve dvou obdobích. Testujeme hypotézu, že majitelé automobilů jezdí po zdražení benzínu méně než před zdražením.

Např. Testujeme hypotézu, že spotřeba alkoholu u otce a syna spolu souvisí.

1.6. Chí kvadrát test dobré shody

Předpokládáme normální rozdělení.

Vybereme např. 50 jednotek \Rightarrow vznikne nám histogram

Porovnáme ho s křivkou normálního rozdělení

Nebo: máme hrací kostku, je kostka v pořádku? Padá každé číslo se stejnou pravděpodobností? Hodíme 60x, mohou být rozdíly \Rightarrow je kostka v pořádku nebo ne? Na to odpoví chí kvadrát test dobré shody.

Nebo: uděláme reprezentativní výběr, několik lidí nám odmítne odpovědět. Zjišťujeme jestli je zbytek výběru stále reprezentativní.

Příklad: Házení kostkou

číslo	kolikrát padlo (n_i)	P_{i0}	nP_{i0}	Ch^2 (výsledek vzorce)
1	5	1/6	10	2,5
2	11	1/6	10	0,1
3	15	1/6	10	2,5
4	8	1/6	10	0,4
5	12	1/6	10	0,4
6	9	1/6	10	0,1
celkem	60	1	60	6

Pokuste se na základě výsledků prokázat špatnost kostky.
 Srovnáváme skutečné četnosti (n_i) s hypotetické četnostmi (nP_{i0})
 Velké rozdíly \Rightarrow špatná kostka

Počet stupňů volnosti = $k - c - 1$
 k = počet kategorií = 6
 c = počet odhadovaných parametrů = 0

Zvolíme hranici 5%
 $Ch^2_{0,95}$ = kritická hodnota = 11,1
 Větší hodnoty = kritický obor

Nám vyšlo 6 ... $6 < 11,1 \Rightarrow$ nezamítáme nulovou hypotézu, neprokázali jsme neregulérnost kostky.

Podmínkou je, že jednotlivá políčka musí být dostatečně obsazená – musíme provést dostatek hodů kostkou. Hypotetické hodnoty musí být ≥ 5
 Pokud není podmínka splněna, můžeme slučovat skupiny (všechny výpočty provádíme až po slučování)

2. Regresní a korelační analýza

Zkoumání závislostí mezi proměnnými

2.1. Analýza závislostí – základní pojmy

Závislost

Příčinná (kauzální)
 Složitější (různě zprostředkované)

Jednostranná
 Vzájemná

Pevná – funkční: z hodnoty jedné proměnné můžeme vypočítat hodnotu druhé
 Volná – statistická: změna hodnoty jedné proměnné znamená tendenci ve vývoji druhé proměnné

Koncentrační tabulka

Pro každou jednotku (např. kraj) máme v jednotlivých sloupečcích jednotlivé hodnoty (počet obyvatel, nemocnic...)

Korelační tabulka

např. věk ženichů v legendě a v jednotlivých sloupcích věk nevěsty
 v tabulce jsou hodnoty počtů sňatků mezi ženichy a nevěstami daného věku

Kontingenční tabulka

jako korelační, ale ne číselné hodnoty
 např. v legendě – co čtou a v jednotlivých sloupcích na co se dívají v televizi
 v jednotlivých políčkách jsou počty takových lidí (čtou daný tisk a dívají se a danou televizi)
 Rozdíl: noviny, pořady se nedají seřadit jako čísla

Čtyřpolní tabulka – asociční

jen 4 políčka – jen alternativní políčka v legendě i ve sloupečkách
 např. muž/žena a byl jste na dovolené ano/ne

2.2. Závislost v kontingenční tabulce

závislost dvou kategoriálních proměnných
 testujeme, zda je závislost prokazatelná
 χ^2 test o nezávislosti v kontingenční tabulce

H_0 : proměnné jsou nezávislé
 H_1 : proměnné jsou závislé

$P(A \text{ a } B) = P(A) * P(B)$ pro NEZÁVISLÉ jevy – tím získáme hypotetické hodnoty pro jednotlivá políčka

Počet stupňů volnosti = (počet řádků – 1) * (počet sloupců – 1)

Podle vzorce spočítáme χ^2 hodnotu a porovnáme s kritickou hodnotou.

Všechna políčka musí být dostatečně obsazená – hypotetická hodnota musí být dostatečně velká – alespoň 5 (nebo někdy se uvádí: všechna políčka musí mít hypotetickou hodnotu větší než 1 a v 80% políček musí být více než 5)

3. přednáška

2.3. Podmíněné charakteristiky

X_i	Y_{ij}	n_i	X_i průměr	S_i^2
X_1	Y_{11}, Y_{12}, Y_{13}	n_1		
X_2	Y_{21}, Y_{22}, Y_{23}	n_2		

X_3	Y_{31}, Y_{32}, Y_{33}	n_3		
...

X_i – např. rok výroby

Y_{ij} – např. cena prodaného auta

n – celkový počet pozorování

Zjišťujeme, jestli proměnná Y závisí na faktoru X .

vnitro-skupinový a mezi-skupinový rozptyl

determinační poměr = mezi-skupinová variabilita / celková variabilita

vyjde např. $0,24 \hat{=} 24\%$ z celkové variability je vysvětlitelné změnami faktoru X (např. stáří automobilu prodávaného v bazaru) a více než $\frac{3}{4}$ je způsobeno jinými vlivy.

P = odmocnina z determinačního poměru = korelační poměr.

Vyšší hodnota znamená silnější závislost. Nabývá hodnot od 0 do 1.

Počty pozorování musí být dostatečné – aspoň 5.

Pokud je X číselná, měla by to být nespojitá proměnná (neměla by to být ani spojitá proměnná rozdělená do skupin)

2.4. Jedno-faktorová analýza rozptylu

Na základě výběrových dat máme přesně, jasně rozhodnout o závislosti proměnné Y na faktoru X .

Nulová hypotéza: střední hodnoty m_1, m_2, \dots, m_k se sobě rovnají
hypotéza o neúčinnosti faktoru. Průměrná cena automobilů je stejná, bez ohledu na roky.

Alternativní hypotéza: alespoň dvě střední hodnoty se od sebe liší

stupně volnosti = počet veličin – počet funkčních vztahů mezi nimi = $(n-1)$

Testové kritérium = F

3. Regresní analýza

Bodový diagram

např. na ose X je příjem domácnosti, na ose Y jsou výdaje za jídlo a pití.
jednotlivé domácnosti vyznačíme jako body v diagramu.

snažíme se určit průběh funkce, která vystihuje závislost (lineární růst, hyperbola, klesající...)

Zkoumání těchto závislostí říkáme regrese – zkoumání průběhu závislosti. Cílem je stanovit regresní funkci, která vystihuje průběh bodového diagramu.

Zkoumáme i těsnost-intenzitu závislosti (blízkost bodů kolem regresní funkce).
Nejtěsnější je funkční/pevná závislost – všechny body leží na regresní funkci.

3.1.1. Regresní analýza

vysvětlující proměnná – nezávislá proměnná - X
vysvětlovaná proměnná – závislá proměnná – Y

jednoduchá-párová závislost – závislost mezi 2 proměnnými
vícenásobná závislost – závislost Y na několika vysvětlujících proměnných

3.1.2. Historie

Gregory King – 1696

Francis Galton (1822-1911) antropolog, výška dospělých dětí v závislosti na výšce rodičů. Použil termín regrese. Dílo: Ústup k průměrnosti

Karl Pearson (1857-1936) – časopis Biometrika, měl syna Egona.

4. přednáška

3.2. Párová závislost

tabulka dvojic hodnot – X_i a Y_i

zkoumáme závislost Y na X – párová závislost.

X je jediný faktor ovlivňující Y, pokud bychom vyloučili náhodnost ležely by všechny body na křivce dané funkcí $y = f(x)$

$$Y_i = \eta(X_i) + \text{Epsilon } i$$

$\eta(X_i)$ – deterministická složka, dovedeme jí vypočítat

η – teoretická regresní funkce

Epsilon i – náhodná složka

Parametry v teoretické regresní funkci značíme $\beta_0 \dots \beta_{p-1}$

Používáme funkce, které jsou lineární kombinací parametrů.

$f_0 \dots f_n$ – regresory – jednoduché funkce vysvětlující proměnné X

Podmínky – týkají se náhodné složky Epsilon

Střední hodnota Epsilon = 0 (křivka funkce prochází středem roje teček)

Rozptyl odchylek je konstantní

Někdy je podmínkou i normální rozdělení (klasický lineární model)

3.3. Volba typu regresní funkce

přímka, parabola, hyperbola, exponenciála...

3.3.1. Apriorní volba

vycházíme z věcného rozboru závislosti

nezkoumáme konkrétní dvojice čísel

často vychází z ekonomické teorie

3.3.2. Empirická volba

vycházíme z daných hodnot
namalujeme si bodový diagram a hledáme funkci, která by se tam vešla

V praxi se používají oba postupy, apriorní volba se ověřuje se pomocí empirie.
Kritéria přiléhavosti – vypočítá Statgraphics z dvojic hodnot – která funkce se nejvíce hodí k dvojicím hodnot.

Výsledkem je rozhodnutí, jakou funkci vybereme: přímka, parabola...

3.4. Odhad parametrů regresní funkce

Teoretická regresní funkce: $Y = \beta_0 + \beta_1 * X$

Výběrová regresní funkce: $Y = b_0 + b_1 * X$

b_0, b_1 – odhadnuté parametry

odhadujeme metodou nejmenších čtverců

nejlepší funkci najdeme:

jednotlivé odchylky e_i^2 sečteme – $Q = \text{Suma}(y_i - Y_i)^2$

Zvolíme funkci jejíž součet čtvercových odchylek je nejmenší.

Matematik La Place používal absolutní hodnotu místo mocniny.

Za Y_i dosadíme podle $Y = b_0 + b_1 * X$ pro jednotlivé přímky.

Q je funkcí dvou nezávisle proměnných b_0 a b_1

Provedeme dvě parciální derivace podle b_0 a b_1 a hledáme minimum

Výsledkem je p rovnic o p neznámých – soustava normálních rovnic (z nich zjistíme parametry).

Pomocí Kramerova pravidla vyjádříme b_1

Vynásobíme čitatele i jmenovatele $1/n$ a ve jmenovateli získáme **rozptyl vysvětlující proměnné X**. V čitateli: **průměr(xy) – průměr(x)*průměr(y)** což je **kovariance** = s_{xy}

$$b_1 = \text{kovariance}(xy) / \text{rozptyl}(x)$$

Kovariance je charakteristika lineární závislosti dvou proměnných. Hodnotu většinou neinterpretujeme, ale používáme jí k dalším výpočtům. Nabývá libovolných hodnot.

Přímá závislost \Rightarrow kladná kovariance

Nepřímá závislost \Rightarrow záporná kovariance

Nezávislé proměnné \Rightarrow nulová kovariance

$$\text{kovariance}(xy) = \text{kovariance}(yx)$$

$$\text{rozptyl}(x+y) = \text{rozptyl}(x) + \text{rozptyl}(y) + 2 * \text{kovariance}(xy)$$

Ze dvou normálních rovnic vyjádříme: $b_0 = \text{průměr}(y) - b_1 * \text{průměr}(x)$

Výběrová regresní funkce je odhadem teoretické regresní funkce.

Beta₁ – regresní koeficient – směrnice regresní přímky – neznáme ale odhadneme
Beta₀ – hodnota Y pro X = 0

Někdy nevíme, co je faktor a co je vysvětlovaná proměnná – pak to zkoumáme oboustranně, jako dvě závislosti.

5. přednáška

3.5. Parabola

Parabola: $Y = b_0 + b_1 * X + b_2 * X^2$

$Q = \text{suma} (y_i - Y_i)^2$

Dosadíme Y, hledáme minimum Q

$Q = \text{suma} (y_i - b_0 - b_1 * X - b_2 * X^2)^2$

Suma Y = $n*b_0 + b_1*\text{suma}(X) - b_2*\text{suma}(X^2)$

Suma XY = $b_0*\text{suma}(x) + b_1*\text{suma}(X^2) - b_2*\text{suma}(X^3)$

3.6. Funkce nelineární v parametrech

Např. Törnquistovy křivky pro modelování poptávky

3.6.1. Metoda linearizující transformace

upravíme funkci tak, aby byla lineární
zlogaritmujeme obě strany rovnice.

Po nahrazení odhady získáme:

$$Y^* = b_0^* + b_1^* * X$$

řešíme dvě normální rovnice, odlogaritmujeme a získáme b_0 a b_1
příklad na straně 200

3.6.2. Metoda vybraných bodů

lepší metoda

vybereme několik typických bodů, tolik bodů, kolik má funkce neznámých parametrů
Tři body dosadíme do rovnice $\hat{=}$ n rovnic o n neznámých
pomocí počítače najdeme lepší řešení

3.7. Vícenásobná regrese

Výběr funkce je složitější

$$\text{eta} = \text{Beta}_0 + \text{Beta}_1 * x_1 + \text{Beta}_2 * x_2 + \dots$$

Regresní rovina nebo nadrovina

Např. závislost výdajů na jídlo a pití v závislosti na počtu členů a příjmu.

Problém je i volba vysvětlujících nezávislých proměnných – co na čem závisí. Doporučuje se zařadit spíše méně vysvětlujících proměnných, a to tak, aby mezi vysvětlujícími proměnnými byla co nejslabší závislost.

Multikolinearita – závislost mezi vysvětlujícími proměnnými. Má být co nejslabší.

Nebo: $\eta = \beta_0 + \beta_{yx1.x2} * x_1 + \beta_{yx2.x1} * x_2 + \dots$

Dílčí regresní koeficienty: $\beta_{yx1.x2}$ a $\beta_{yx2.x1}$

$\beta_{yx1.x2}$ = jak se v průměru zvýší y, pokud se x_1 zvýší o 1, je-li vyloučen vliv x_2 za tečkou. jsou proměnné, jejichž vliv je vyloučen.

Postupujeme zase metodou nejmenších čtverců, odhadneme hodnoty parametrů.

3.8. Měření těsnosti závislosti

strana 202

3.8.1. Korelační koeficient (Pearsonův)

$r_{xy} = r_{yx} = S_{xy} / S_x * S_y$ = kovariance xy / součin směrodatných odchylek
nabývá hodnoty od -1 do 1

Interpretace

záporná hodnota = nepřímá závislost

kladná hodnota = přímá závislost

čím vyšší absolutní hodnota tím silnější závislost

Je odvozený a má se používat **pro přímkovou závislost**. Nevhodný pro parabolu...a závislosti výrazně odlišné od přímkové.

3.8.2. Determinační koeficient

Druhá mocnina korelačního koeficientu

Statgraphics ho vyjadřuje v %.

Říká nám, z kolika procent se dá variabilita vysvětlit pomocí vysvětlující proměnné.

6. přednáška

dá se napsat jako součin regresních koeficientů: $r_{xy}^2 = b_{yx} * b_{xy}$

3.8.3. Test o nulové hodnotě korelačního koeficientu

H_0 : výběrový soubor pochází ze souboru, kde neexistuje závislost.

H_1 : korelační koeficient není nulový

Test provede Statgraphics.

testové kritérium má T-rozdělení s počtem stupňů volnosti $n-2$

Kritický obor: extrémně vysoké a nízké hodnoty testového kritéria.

Stejný test se používá i pro jiné korelační koeficienty.

3.9. Těsnost u vícenásobné závislosti

vícenásobný determinační koeficient:

$r^2_{y,x1x2}$ - těsnost závislosti Y na X1 a X2 zároveň
za tečkou jsou vysvětlující proměnné.

Hodnota 0,62 znamená, že variabilitu hodnot Y je možné z 62% vysvětlit naší regresní rovinou (středně těsná až těsnější závislost)

$(1 - r^2_{y,x1x2}) / (1 - r^2_{y,x1})$ = jak velká část variability nevysvětlené přímkou zůstala nevysvětlená i regresní rovinou – tím zjistíme, jestli mělo smysl přidat další vysvětlující proměnnou.

$r_{yx2.x1}$ = koeficient dílčí korelace – jak těsná je závislost mezi proměnnými před tečkou (y a x_2) jestliže je vyloučen vliv proměnných za tečkou (x_1)

závislost Y na X1 a X2: $r_{y,x1x2}$

závislost Y na X1 při vyloučení vlivu X2: $r_{yx1.x2}$ (dílčí korelace)

3.10. Výběr vysvětlujících proměnných do modelu

zařazených proměnných bývá menší počet: 1,2,3...

takové, které co nejvíce ovlivňují Y a zároveň, aby v modelu byla nízká multikolinearita (závislost mezi vysvětlujícími proměnnými)

3.10.1. Postup

vypočteme závislosti Y na jednotlivých vysvětlujících proměnných ($r_{yx1}, r_{yx2}...$)

vybereme největší

$$Y = b_0 + b_{yx1} X_1$$

Vypočteme korelační koeficient – jak velká část variability Y je vysvětlena působením faktoru X1

vypočítáme koeficienty dílčí korelace – zjišťujeme, kterou další proměnnou přidat:

$r_{yx2.x1}, r_{yx3.x1}$, jednu vybereme

určíme rovnici regresní roviny

pokračujeme dál – vypočítáme jaká část variability je vysvětlena 2 vysvětlujícími proměnnými. Pokud to pořád nestačí přidáme další proměnnou.

Sekvenční F-Test – ve Statgraphicsu (Regrese po krocích) – tím zjistíme, kdy má ještě smysl přidávat další proměnnou.

3.11. Spearmanův korelační koeficient (pořadové korelace)

Zvláštní případ Pearsonova korelačního koeficientu, pokud hodnotami Y a X jsou pořadová čísla i_x, i_y

nabývá hodnot od -1 do 1

-1 znamená, že pořadí jsou přesně opačná

Např. pro posouzení shody dvou pořadí v soutěži. Nebo jde použít pokud nechceme počítat s konkrétními hodnotami, zjistíme jen pořadí, a srovnáme tato pořadí. Výsledek není tak přesný a hodnoty se nesmí opakovat.

3.12. Determinační index

strana 203

odmocninou je korelační index.

nabývá hodnot od 0 do 1

můžeme ho vyjádřit v %

část variability Y, která je vysvětlitelná naší regresní funkcí.

vyšší závislost \Rightarrow vyšší hodnota

lepší funkce \Rightarrow vyšší hodnota

pro parabolu, hyperbolu...

7. přednáška

3.13. T-testy

Zjišťujeme, jestli získané parametry patří do rovnice nebo ne.

dílčí T-testy – tolik, kolik má funkce neznámých parametrů, pro přímku 2

H_0 : parametr je ve skutečnosti nulový

H_1 : parametr není nulový a má být obsažen v rovnici

má T-rozdělení

Kritický obor tvoří extrémně vysoké a nízké hodnoty.

3.14. Celkový F-test

Test o celém modelu

H_0 : celý model je nesmyslný, protože regresní funkce je přímka rovnoběžná s osou X (ať je Y jakékoli, x je průměrně konstantní)

H_1 : alespoň některé z parametrů jsou nenulové, model má smysl.

má F-rozdělení

p – počet parametrů

n – počet pozorování

Kritický obor tvoří nejvyšší hodnoty.

Jestliže zamítáme H_0 říkáme, že F-test je významný (stejně jako u T-testu)

Dobrý model

Celkový F-test by měl být významný a všechny nebo skoro všechny T-testy by měly být významné. Neznamená to ale, že tento model je nejlepší, pouze není špatný.

Pokud by byl celkový F-test nevýznamný, neměli bychom používat model.

Pokud by byl celkový F-test významný a řada T-testů by byla nevýznamných, měli bychom přehodnotit regresní funkci – je tam příliš mnoho regresorů, příliš složitá funkce.

3.15. Sekvenční F-testy

Testuje jestli přidání dalšího regresoru významně zlepší regresní funkci. Jestli je přidání regresoru efektivní.

Čím vyšší hodnota, tím spíše bychom měli zařadit nový regresor, protože lepší model.

4. Analýza časových řad

Časová řada – posloupnost věcně a prostorově srovnatelných pozorování (dat), která jsou uspořádána v čase. Analýza dosavadního vývoje umožňuje předpovídat budoucí vývoj.

4.1. Úvod do časových řad

4.1.1. Druhy časových řad

okamžikové – stav zásob, počet studujících studentů, k nějakému okamžiku
intervalové – tržba... za nějakou dobu

Podle periodicity:

dlouhodobé – roční

krátkodobé – měsíční, týdenní, denní...

4.1.2. Srovnatelnost ukazatelů v časové řadě

věcná srovnatelnost

prostorová srovnatelnost – údaje za stejné území, stejný podnik (bez fúze, rozdělení)

časová srovnatelnost – očišťování od důsledků kalendářních variací (v únoru může být menší tržba než v březnu) strana 247 (přepočteme na standardní měsíc)

otázka užívání běžných nebo stálých cen

4.1.3. Průměr v časových řadách

intervalové řady – prostý aritmetický průměr

okamžikové řady – chronologický průměr

chronologický průměr: $Y = (y_{1/2} + y_2 + y_3 + \dots + y_{k-1} + y_{k/2}) / (k - 1)$ strana 249

4.1.4. Elementární charakteristiky časových řad

Diference – přírůstky – rozdíl dvou sousedních hodnot

$$\Delta_t = Y_t - Y_{t-1}$$

Y_t – hodnota v okamžiku / intervalu

absolutní hodnota, nebo průměrný přírůstek.

Druhé difirence – přírůstky z přírůstků

$${}^2\Delta_t = \Delta_t - \Delta_{t-1}$$

Koeficienty růstu (tempa růstu) – hodnota v příslušném období / hodnota v předcházejícím období

řetězové indexy

Průměrný koeficient růstu počítáme jako **geometrický průměr**.

Stačí nám znát poslední a první hodnotu:
průměr = $n-1$ odmocnina(poslední hodnota / první hodnota)

4.1.5. *Přístupy k modelování časových řad*

Jednorozměrný model

Zkoumáme vývoj ukazatele jako funkci času.

Ukazatel je funkcí času a náhodné složky: $y_t = f(t, \text{epsilon } t)$

epsilon t – náhodná složka

$$y_t = T_t + S_t + C_t + \text{Epsilon}_t$$

T – trend, trendová složka, základní vývojová tendence, růst, pokles

S – sezónní složka, pravidelné kolísání během roku

C – cyklická složka, kolísání kolem trendu, kde délka vlny je delší než rok, demografické cykly, inovační cykly

Epsilon_t – čistě náhodná složka, nedá se modelovat

Teoretická složka – model: $Y_t = T_t + S_t + C_t$

často se používá multiplikační tvar modelu (můžeme převést zlogaritmováním)

$$y_t = T_t * S_t * C_t * \text{Epsilon}_t$$

Box-Jenkinsonova metodologie – za základ se považuje náhodná složka

strana 255

Spektrální analýza – strana 256

Vícerozměrné modely časových řad

Ukazatel nezávisí jen na čase, ale je ovlivněn i dalšími faktory (alespoň jedním)

Často má smysl zkoumat závislost na jednom nebo několika faktorech o několik období zpátky (opožděné působení)

8. přednáška

4.2. *Trendová složka*

Vyrovnaní časové řady – nahrazení lomené čáry grafu nějakou rovnější křivkou, matematickou funkcí.

4.2.1. *Analytické vyrovnaní*

Pokud existuje jedna tendence pro celou časovou řadu (bez zlomů, cyklů...)

Lineární trend

Přímka: $T_t = \text{Beta}_0 + \text{Beta}_1 * t$

T – teoretická hodnota

použijeme metodu nejmenších čtverců, jako v regresní analýze, odvodíme soustavu normálních rovnic – strana: 258

použijeme-li časovou proměnnou t':

Lichý počet let ...-3,-2,-1,0,1,2,3...

Sudý počet let ...-5,-3,-1,1,3,5...

Vypadne tak sumace let z normálních rovnic, dostaneme tak dvě rovnice, každou o jedné proměnné.

$$b_0 = \text{suma}(Y_t) / n$$

$$B_1 = \text{suma}(t^i) * Y_t / \text{suma}(t^{i^2})$$

n = počet let, za které máme údaje

Interpretace: $T_t = (\text{průměr minulých let}) + (\text{průměrný přírůstek tržeb za rok}) * t$.
V případě sudého počtu let to znamená průměrný přírůstek za půl roku.

Odhady-předpovědi – extrapolační, provádíme dosazením do rovnice.

Parabolický trend

$$\text{Parabola 2. stupně: } T_t = \text{Beta}_0 + \text{Beta}_1 * t + \text{Beta}_2 * t^2$$

Jako v regresní analýze odvodíme 3 normální rovnice, metoda nejmenších čtverců.
strana: 262

Exponenciální trend

Taky jako u regresní analýzy.

$$\text{Exponenciála: } T_t = \text{Beta}_0 * \text{Beta}_1^t$$

není lineární v parametrech, metodu nejmenších čtverců můžeme použít až po linearizující transformaci (zlogaritmování).

Interpretace: $T_t = (\text{geometrický průměr hodnot}) * (\text{průměrný koeficient růstu})^t$

Metoda linearizující transformace nedává ideální výsledky – strana:268, lepší metoda je v učebnici (používá váhy y_t^2) vážená metoda nebude u zkoušky.

Modifikovaný exponenciální trend

Posunutá exponenciála, funkce s nenulovou asymptotou

$$T_t = \text{Theta} + \text{Beta}_0 * \text{Beta}_1^t$$

Theta – parametr posunutí

Používají se metody:

částečných součtů

dílčích průměrů

vybraných bodů

Modifikovaný exponenciální trend nebude u zkoušky.

Logistický trend

S-křivka. Různé funkční předpisy.

Původně odvozena jako model růstu populace hmyzu za podmínek omezených zdrojů. Kladná horní asymptota. Křivka je symetrická, uprostřed obsahuje inflexní bod

Používá se pro modelování poptávky po předmětech dlouhodobé spotřeby.

Gompertzova křivka

S-Křivka, asymetrická – část za inflexním bodem je delší.

9. přednáška

Volba vhodného modelu trendu

Věcně ekonomická analýza – strana 286 – zamysleme se nad limitami, mezemi, růst, pokles, inflexní bod (s-křivka)

Analýza grafu – nakreslíme spojnicový graf, hledáme nejvhodnější funkci

Rozbor empirických kritérií – hledáme empirická kritéria přiléhavosti, zkoumáme naše data.

M.S.E. – střední kvadratická odchylka = průměr z $((Y_t - T_t)^2)$ = průměr rozdílů z empirických a teoretických (vyrovnaných hodnot) na druhou = reziduální součet čtverců děleno n. Čím nižší, tím lépe funkce přiléhá.

M.A.E. – místo na ² se použije absolutní hodnota.

F-test – čím vyšší, tím lepší funkce

Sekvenční F-test – jako u regresní analýzy. Složitější funkce mají lepší výsledky. Jde o to jestli zvýšení přiléhavosti (u složitější funkce) je dostatečné, aby to bylo výhodné.

Učebnice-strana: 290

I když najdeme funkci (polynom), která prochází všemi n body, nemusí to být správná funkce, vhodnější může být nějaká jednodušší funkce.

Analýza diferencí

Přímka – každý rok roste o stejnou konstantu. Spočítáme řadu přírůstků z dat, pokud zjistíme, že kolísají kolem nuly, pak se zdá, že vhodnou funkcí je přímka. Nebo také: druhé difference kolísají kolem nuly.

Parabola druhého stupně

Δt přírůstky rostou (lineárně)

$^2\Delta t$ druhé difference kolísají kolem konstanty

$^3\Delta t$ třetí difference kolísají kolem nuly

Exponenciála

řada koeficientů růstu je konstantní (kolísají kolem konstanty, ale nerostou systematicky ani neklesají)

Analýza růstových charakteristik – počítáme průměrné růstové charakteristiky - strana 292

Pomocí extrapolčních kritérií – časovou řadu zkrátíme na konci o jedno nebo několik období, časovou řadu vyrovnáme. A sledujeme, do jaké míry jsou dobře odhadnuty zkrácená období – pseudoprognozy známých ale nepoužitých hodnot.

V praxi se kombinuje více postupů. Např. graf nasvědčuje na exponenciálu, ale existuje logické omezení (hranice nasycení), potom použijeme s-křivku.

4.2.2. Mechanické vyrovnání

Mechanické zprůměrovávání hodnot. Používá se při výrazné sezónnosti.

Metoda vyrovnání klouzavými průměry:

M - klouzavá část – počet období - kolika členné klouzavé průměry (při roční sezónnosti se používá 12 měsíců pro měsíční data)

Klouzavé úhrny – součet prvních 7 hodnot (pro denní hodnoty s týdenní sezónností), napíšeme doprostřed období, ke 4. dnu, potom 7 hodnot od druhé do osmé, napíšeme k 5. dnu...

Klouzavé průměry - každý klouzavý úhrn vydělíme sedmi. Jejich řada už neobsahuje sezónnost, jen v malé míře nahodilé kolísání a hlavně trend (pokud je).

Centrování – Pokud je počet období sudý (12 měsíců) – vypočítáme průměr ze dvou sousedních hodnot. – Klouzavé průměry jsou mezi jednotlivými čtvrtletími, centrované klouzavé průměry jsou průměry z klouzavých průměrů.

U klouzavých průměrů nám chybí několik údajů na začátku a na konci, u centrovaných chybí ještě o jeden víc.

Vážené klouzavé průměry

Jednotlivým pozorováním přidáváme váhy – největší váhu mají hodnoty uprostřed období. Např. pro $m = 7 \dots 1/21 [-2,3,6,7,6,3,-2]$

10. přednáška

4.3. Sezónní složka

Pokud máme čtvrtletní údaje (nebo denní, měsíční, ...), se setkáváme s kolísáním během roku.

Popis sezónnosti

Sezónní očišťování časových řad

4.3.1. Model konstantní sezónnosti se schodovitým trendem

$$Y_{ij} = a_i + b_j + E_{ij}$$

i – rok

j – období (čtvrtletí, měsíc...)

a_i – průměr pro všechna období roku

b_j – sezónní výkyv (každý rok stejný)

E_{ij} – náhodná odchylka

Statgraphics: aditivní model.

Předpokládáme, že i když časová řada roste (nebo klesá), tak výkyvy jsou stejné.

4.3.2. Model proporcionální sezónnosti

Pokud výkyvy nejsou v průběhu let konstantní. Sezónní výkyv je úměrný trendové složce.

Statgraphics: multiplikační model.

Hodnota časové řady se dá vyjádřit jako součin. Viz učebnice.

Empirické sezónní indexy:

$$Y_{ij} / T_{ij} = (1 - c_j) + E_{ij}$$

Y_{ij} – skutečná hodnota

T_{ij} – teoretická hodnota trendu (často klouzavý průměr)

c_j – sezónní parametr
 E_{ij} – náhodná složka

Y_{ij} / T_{ij} = sezónní index – např. 1,45 znamená, že se vyrobilo o 45% více.
= Skutečná hodnota / vyrovnaná hodnota

Z jednotlivých sezónních indexů vypočítáme průměr (za jednotlivá čtvrtletí).

Sezónní faktor – součet průměrných indexů není 4 (kvůli průměrům) a průměr z nich není 1, tak upravíme, aby byl součet 4

Sezónní indexy vydělíme jejich součtem a vynásobíme 4.

Pokud máme velmi dlouhou řadu, můžeme používat mediány místo průměrů, tím se zbavíme výjimečných náhodných hodnot.

4.3.3. **Sezónní očišťování časové řady**

Sezónní dekompozice

Odstraníme, nevyrovnáme, sezónní složku z časové složky, zbude trendová a náhodná složka.

Hendersonovy filtry

Různé počítačové programy

Pomocí sezónních faktorů – používá statgraphics

Skutečnou hodnotu vydělíme příslušným sezónním indexem. Pokud očištěná data rostou (nebo klesají), jedná se o trend a ne jen sezónní kolísání.

4.4. **Náhodná složka**

$Epsilon_t = y_t - Y_t$

Předpoklady

Náhodná složka je opravdu náhodná – střední hodnota je 0

Rozptyl náhodné složky je konstantní – odchylky se nezvětšují

Odchylky jsou vzájemně nezávislé – žádná pravidelnost, ani nejde popsat funkcí (kdyby to šlo, tak jsme špatně vyrovnali časovou řadu)

Potom se náhodné složce říká „bílý šum“

$Epsilon_t$ – náhodná složka

e_t – reziduum – konkrétní hodnota pro určitý model (jednotlivá rezidua jsou odhadem náhodné složky)

4.4.1. **Znaménkový test**

zda jsou rezidua náhodně uspořádány v čase. Nebude v písemce.

S = počet případů, kdy $e_t - e_{t-1} > 0$

$E(S) = (n-1) / 2$ pokud je to přibližně polovina, jsou rezidua náhodně uspořádána v čase.

4.4.2. **Test bodů obratu**

viz učebnice

4.4.3. Durbin-Watsonův test auto-korelace

zda jsou náhodné odchylky nezávislé

H_0 : nezávislost odchylek

H_1 : závislost (přímá či nepřímá)

testové kritérium d nabývá hodnot od 0 do 4

přímá závislost – hodnoty kolem 0

nezávislost – hodnoty kolem 2

nepřímá závislost – hodnoty kolem 4

V tabulkách jsou dolní a horní kritické hodnoty:

Pro přímou závislost:

$d < d_l$... zamítáme H_0

$d > d_u$...nezamítáme H_0

$d_l < d < d_u$... test mlčí

Pro nepřímou závislost (kolem 4) uděláme transformaci $= 4 - d$ (dále už stejně jako u přímé)

4.5. Korelace v časových řadách

Problém je, že když srovnáme dvě rostoucí (klesající) řady, vyjde nám, že je mezi nimi přímá závislost. I když věcná závislost neexistuje.

Zdánlivá závislost

Příčinná závislost

Jednoduchá metoda:

vyrovnat každou řadu vhodnou funkcí

pro jednu i druhou určíme odchylky od trendu

zkoumáme, existuje-li závislost mezi dvěma řadami odchylek od trendu

Korelace v časových řadách může být i opožděná \Rightarrow např. zkoumáme závislost mezi investicemi v roce t a ziskem v roce $t+2$

11. přednáška

5. Indexy individuální

doplnit

12. přednáška

základ indexu = jmenovatel

individuální – stejnorodých (sčitatelných) veličin

individuální jednoduché - srovnáváme dvě hodnoty podílem = index

Pokud máme řadu hodnot, můžeme počítat řadu indexů:

Bazické - se stálým základem. např. Q_{2001}/Q_{2000} , Q_{2002}/Q_{2000} , Q_{2003}/Q_{2000}

Řetězové – s pohyblivým základem – srovnáváme hodnoty ze sousedních dvou období: např. Q_{2001}/Q_{2000} , Q_{2002}/Q_{2001} ...

Řetězové a bazické indexy můžeme navzájem přepočítávat.

podíl dvou sousedních bazických indexů = příslušný řetězový index

$$Q_{2003}/Q_{2000} / Q_{2002}/Q_{2000} = Q_{2003}/Q_{2002}$$

Rok	Bazický index v % 1998=základ	Řetězový index
1998	100	?
1999	105,1	$1,051/1 = \mathbf{1,051}$
2000	108,4	$1,084/1,051 = \mathbf{1,031}$
2001	107,1	$1,071/1,084 = \mathbf{0,988}$
2002	110,5	$1,105/1,071 = \mathbf{1,032}$
2003	116,9	$1,169/1,105 = \mathbf{1,058}$

Největší růst v roce 2003, vzrůst o 5,8% (řetězový index)

U BZ může být přepočet indexů – např. na bazické s různým základem

Irwing Fischer – (1867-1947) – statistik a ekonom, zabýval se indexy. Dílo: Tvorba indexních čísel. Vymyslel i testy:

test okružnosti:

test interkalace: $Q_m/Q_l * Q_l / Q_k = Q_m / Q_k$ Pokud platí, můžeme indexy řetězit.

5.1. Individuální složené indexy

Indexy stejnorodých veličin (produkce, dovoz, vývoz... stejných výrobků)

Provádíme shrnování (dovoz z několika zemí...)

5.1.1. Extensivní ukazatele

$$i_g = Q_1 / Q_0$$

nebo: $i_g = \text{Suma}(Q_{1,i}) / \text{Suma}(Q_{0,i})$... i v sumě se často nepíše

Absolutně: $\Delta Q = \text{Suma}(Q_1) - \text{Suma}(Q_2)$

5.1.2. Intenzitní veličiny

Nemá smysl např. sčítat ceny, musíme je průměrovat.

Např. cena, produktivita práce, hektarový výnos...

$$i_p = \text{Průměr}(p_1) / \text{Průměr}(p_0)$$

průměrná cena mandarinek v únoru / průměrná cena mandarinek v lednu

použijeme vážený průměr (váhy podle množství, které jsme dovezli)

index proměnlivého složení – jak se v průměru změnila hodnota intenzitní veličiny.

$$p * q = Q$$

$$p = Q / q$$

5.2. Rozklad indexů

důležité na zkoušku

změna může být důsledkem:

změny hodnot (např. se změnily ceny)

změny struktury (např. jsme začali nakupovat více na dražších trzích)

5.2.1. Metoda postupných změn

Rozklad indexů na základě předpokladu, že změny probíhají postupně.

Nejdříve se změní ceny a potom množství (nebo obráceně).

Index proměnlivého složení = Index stálého složení * index struktury

$$i_{ps} = i_{ss} * i_{str}$$

Index struktury říká, o kolik by se změnila veličina v důsledku samotné změny struktury. Index stálého složení říká, jak by se změnila veličina v důsledku změny cen ne změny struktury.

Příklad

Vyvážíme broušené vázy do několika zemí. Posuďte jak se změnily vývozní ceny, vyvezené množství, hodnota vývozu za jednotlivé země i za všechny země dohromady.

Země	Vývoz 2002 (ks)	Vývoz 2004 (ks)	Vývoz 2002 (USD)	Vývoz 2004 (USD)
A	1000	800	30 000	20 800
B	2000	3000	68 000	90 000
C	3000	6000	96 000	210 000
	q_0	q_1	$Q_0 = p_0 * q_0$	$Q_1 = p_1 * q_1$
	6000	9800	194 000	320 800

Stejný výrobek, několik zemí \Rightarrow použijeme individuální indexy ne souhrnné.

Veličiny: p , q , Q

$$p = Q / q$$

intenzitní veličina = podíl dvou extenzivních

intenzivní veličina – cena za jednotku (není zadána, ale můžeme ji spočítat)

cena = vývoz v USD / vývoz v kusech

$$p = Q / q$$

$$i_{ps} = \text{Průměr}(p_1) / \text{Průměr}(p_0) = 1,012 = 101,2\%$$

V roce 2004 jsme vyváželi za průměrnou cenu 32,73 USD v roce 2002 za 32,33 z toho jsme vypočítali, že průměrná cena vzrostla o 1,2%
průměrná cena vzrostla o 40 centů

Jaký vliv měla samotná změna cen a jaký vliv změna struktury vývozu?

Dopočítáme

Země	Cena v USD (2002)	Cena v USD (2004)	$p_1 * q_0$
A	30	26	26 000
B	34	30	60 000
C	32	35	105 000
Celkem	-	-	191 000

$p_1 * q_0$ = kolik bychom vydělali, kdybychom v roce 2002 prodávali za ceny roku 2004

Vyjde

index stálého složení = 98,5%

index struktury = 102,8%

Interpretace:

index stálého složení nám říká, že v důsledku samotné změny cen by poklesla průměrná cena o půl dolaru = -1,5%

index struktury nám říká, že v důsledku samotné změny struktury by vzrostla průměrná cena o 90 centů = +2,8%

součet dvou absolutních rozdílů nám dá celkový absolutní rozdíl

součin dvou analytických indexů nám dá celkový index

násobíme 0,985 * 1,028 !

13. přednáška

rozklad indexu a souhrnné indexy, dotazy

6. Indexy souhrnné

indexy nestejnorodých veličin

když není jen jeden druh zboží/výrobku, zabýváme se více druhy nejčastěji **cenové indexy**

6.1.1. Dříve

Poprvé je použil anglický biskup na univerzitě začátkem 18. století (kvůli stipendiím)

první index = $\text{Suma}(p_1) / \text{Suma}(p_0)$ = ceny dnes / ceny před 100 lety

nahradili to individuálními cenovými indexy: p_1 / p_0

z individuálních indexů počítali nevážené indexy (aritmetické) tyto metody se už dnes nepoužívají (maximálně geometrické průměry)

6.1.2. Dnes

čitatel: ceny běžného období

jmenovatel: ceny základního období

Souhrnný cenový index:

$\text{Suma}(p_1 * q) / \text{Suma}(p_0 * q)$

q – musí být pro jednotlivé druhy zboží stejné (množství)

I_p – souhrnný index
 i_p – individuální index

6.2. Cenové indexy

Různé indexy, podle toho jestli použijeme množství roku 0 nebo 1

6.2.1. Laspeyresův cenový index – q ze základního období

$$I_p = \text{Suma}(p_1 * q_0) / \text{Suma}(p_0 * q_0)$$

čitatel – stejná množství (loňská) přeceněná do letošních cen

jmenovatel – skutečné tržby za základní období

q je vahou zboží v indexu

6.2.2. Paascheho Index – q z aktuálního období

$$I_p = \text{Suma}(p_1 * q_1) / \text{Suma}(p_0 * q_1)$$

6.2.3. Loweho index – q_k

$$I_p = \text{Suma}(p_1 * q_k) / \text{Suma}(p_0 * q_k)$$

Pokud během let měníme q , nemůžeme indexy řetězit. Pokud je chceme řetězit musíme použít stejné q_k (stejně)

po dobu, co používáme stálá množství je index řetěžitelný

Jediný řetěžitelný z těchto indexů

6.2.4. Index cenový Fischerův

geometrický průměr laspeyresova a paascheho indexu

6.2.5. Montgomeryho cenový index

$$I_p^m = (I_q)^{\text{delta } Q_p / \text{delta } Q}$$

Vychází z logaritmického rozkladu indexů, je produktem snahy o objektivní cenový index

$$I_q = \text{hodnotový index} = \text{Suma}(p_1 * q_1) / \text{Suma}(p_0 * q_0)$$

6.2.6. Průměrové tvary indexů

Napsáno ve tvaru váženého průměru:

$$I_p = \text{Suma}(p_1 * q_0) / \text{Suma}(p_0 * q_0) = \text{Suma}((p_1 / p_0) * p_0 * q_0) / \text{Suma}(p_0 * q_0)$$

Jako harmonický průměr:

$$I_p = \text{Suma}(p_1 * q_1) / \text{Suma}(p_0 * q_1) = \text{Suma}(p_1 * q_1) / \text{Suma}((p_1 * q_1) / (p_1 / p_0))$$

6.3. Objemové Indexy

$$\text{Suma}(q_1 p) / \text{Suma}(q_0 * p)$$

stejně ceny působí jako souměřitel a váha, abychom mohli množství sčítat

ceny ze základního období – souhrnný objemový index Laspeyresův

ceny běžného období – souhrnný objemový index Paascheho

geometrický průměr z nich – objemový index Fischerův

ve stálých cenách – Loweho index

6.4. Hodnotový index - Rozklad

$$I_Q = \text{Suma}(p_1 * q_1 / \text{Suma}(p_0 * q_0)) = \text{tržba v běžném} / \text{tržba v základním období}$$

Říká o kolik procent se změnila tržba.

Hodnotový index můžeme rozložit – jaký vliv měla samotná změna cen, jaký vliv měla samotná změna množství.

6.4.1. Metoda postupných změn

$$\begin{aligned} I_Q &= \text{Suma}(p_1 * q_1 / \text{Suma}(p_0 * q_0)) = \\ &= \text{Suma}(p_1 * q_0 / \text{Suma}(p_0 * q_0)) * \text{Suma}(p_1 * q_1 / \text{Suma}(p_1 * q_0)) = \\ &= \text{Laspeyresův cenový index} * \text{Paascheho objemový index} \end{aligned}$$

Předpokládáme, že změny probíhají postupně.

Pokud to přehodíme (jestli se dřív změní cena a pak množství nebo obráceně), vyjde celkový hodnotový index stejný. Oba dva rozklady jsou správné. Rozložené indexy jsou ale jiné (cenový a objemový).

Ve skutečnosti ale neprobíhají změny postupně a nezávisle na sobě, proto existují i jiné metody. Přesto se nejvíce používá metoda postupných změn.

6.4.2. Rozklad se zbytkem

$$\begin{aligned} &\text{Suma}(p_1 * q_0 / \text{Suma}(p_0 * q_0)) * \text{Suma}(p_0 * q_1 / \text{Suma}(p_0 * q_0)) * \text{zbytek} \\ &= (\text{vliv cen}) * (\text{vliv množství}) * (\text{zbytek, který se nedá rozložit – společný vliv}) \end{aligned}$$

6.4.3. Logaritmický rozklad indexů

Angličan Montgomery chtěl objektivně rozložit indexy. Neujal se.

6.5. Cenové indexy v praxi

strana 380

Vychází se z Laspeyresova indexu (používá se kombinace Laspeyresova a Loweého indexu)

$$\text{Laspeyresův: } I_p = \text{Suma}(p_1 * q_0) / \text{Suma}(p_0 * q_0)$$

$$\text{CPI} = \text{Suma}((i_p) * (p_0 * q_0 / \text{Suma}(p_0 * q_0))) =$$

$$\text{CPI} = \text{Suma}(i_p * w_0)$$

i_p – individuální index

w_0 – váha = tržba z jednoho druhu zboží / celková tržba

individuální tržba - počítá se tržba za celou skupinu, do které patří reprezentant

CPI = index spotřebitelských cen

Cenové indexy se počítá každý měsíc a publikují se 6. pracovní den následujícího měsíce. Tyto indexy považujeme za řetězitelné, počítají se z toho čtvrtletní, půlroční, roční...

Ze zboží se vybere několik set reprezentantů (775, plánuje se snižování) strana 383. 10 skupin reprezentantů. Údaje se zjišťují ve 41 vybraných okresech. Ceny se zjišťují přibližně 15. dne v měsíci. U sezónního zboží se zjišťuje i několikrát měsíčně.

Spotřební koš – reprezentanti a jejich váhy. Je několik let stálý. Revize koše probíhá přibližně jednou za pět let.

6.6. *Index životních nákladů*

Vezmeme určitou skupinu obyvatelstva (např. důchodce, domácnosti s více než jedním dítětem...) a počítáme cenový index pro tuto skupinu obyvatel. Reprezentanti ve spotřebním koši jsou stejní, ale mění se váhy reprezentantů, podle potřeb skupiny.