U N I V E R Z A   V   L J U B L J A N I

E K O N O M S K A   F A K U L T E T A

Blejec Marijan:

S T A T I S T I K A   I. L E T N I K

Nadaljevanje.

Izdala Strokovna sekcija LM Ekonomske fakultete
v Ljubljani
1950.

Kazalo:

Časovne serije

Popravki glej str. 54

Časovne serije.

1. Opredelitev in naloga analize.

Časovna-dinamična serija je serija individualnih, zbirnih, srednih ali relativnih vrednosti, splošno torej serija vseh vrst statističnih količin, od katerih se vsaka nanaša na poedine sukcesivne časovne dele, oziroma momente. Členi časovne serije, katerih vsak zase da sliko samo za določen krajši razmak ali celo za moment, dajo postavljeni v časovnem zaporedju kot kompleks členov, sliko v časovnem spreminjanju določenega pojava (primerjaj filmsko kamero - zaporedje v času si sledečih momentnih slik da sliko gibanja)..

Analiza gospodarskih pojavov predpostavlja njih stalno časovno spreminjanje, ker stacionarnih gospodarskih pojavov v stvarnosti ni. Ustvarjanje fiktivnih stacionarnih pojavov problem le poenostavlja radi najenostavnejše možnosti analize.

Na spremembo določenega ekonomskega - družbenega pojava v času vplivajo najrazličnejši drugi pojavi in drugi vzroki. Naloga analize časovnih serij je, da ugotovi zvezo med posameznimi pojavi (korelacije) na eni strani, na drugi strani pa da analizira in razstavi rezultate vseh vplivov na posamezna komponenta, kar omogoča, da moremo bolj ali manj prodreti v zakonitosti socialno-ekonomskih pojavov.

2. Analiza časovnih serij v kapitalističnem gospodarstvu - konjunkturna statistika.

Kapitalističnemu gospodarstvu so lastne zakonitosti stihijskega odvijanja gospodarskih pojavov. Ker bi poznavanje teh zakonitosti moglo služiti onemu, ki bi jih poznal v njih lastno izkoriščanje je jasno, da je zanimanje za analizo časovnih serij v kapitalizmu veliko in se je pojavilo razmeroma zgodaj (že v 19. stoletju "nauk o krizah") in kasneje doseglo neverjeten napredek in se je v sklopu nauka o konjunkturah razvila specialna statistična metoda, konjunkturna statistika. Nauk o o konjunkturi išče pravilnosti in zakonitosti v gibanju gospodarskih pojavov in so njeno področje vsi gospodarski pojavi. Nauk o konjunkturi je nastal v kapitalizmu in s svojimi predpostavkami tak red tudi predstavlja, t.j.da je narodno gospodarstvo sestavljeno iz individualnih gospodarstev, in pomeni napredek za celotno gospodarstvo, ako bo vsak gospodaril tako, da zase pridobi čim več koristi). Na podlagi izsledkov o zakonitosti pojavov skuša sklepati na gospodarsko dogajanje v bodoče (prognoza). Vendar se je izkazalo, da se vendar ne da gospodarsko življenje zajeti suho v matematično - formalistične kalupe, ko je leta 1929 izbruhnila svetovna gospodarska kriza, po napovedih Harvardskega instituta pa je bila napovedana gospodarska prosperiteta. S tem je formalistična smer analiziranja doživela svoj prelom.

3. Pomen analize časovnih serij v socializmu in planskem gospodarstvu.

Za razliko od kapitalističnega sistema stopa v socializmu mesto zakona stihije plansko gospodarstvo.

Po členu 15.ustave FLRJ usmerja radibzaščite življenskih koristi ljudstva, dviga ljudske blaginje in pravilnega izkoriščanja vseh gospodarskih možnosti in sil, država gospodarsko življenje in razvoj s pomočjo splošnega gospodarskega plana; pri tem se opira na državni in zadružni gospodarski sektor ter izvaja splošno kontrolo med zasebnim sektorjem gospodarstva.

Radi tega nauk o konjunkturi za preučevanje našega gospodarstva dejansko ne pride v poštev, pač pa moramo koristiti mnoge statistične metode časovnih serij, ker je v našem gospodarstvu nebroj pojavov, ki bi jih mogli formalno analizirati s temi metodami.(vpliv naravnih pojavov, privatna gospodarstva).

4. Vrste gibanj gospodarskih pojavov.

Videli bomo, da bomo vse vrste gibanj pojavov,znane iz konjunkturne statistike, našli tudi v našem gospodarstvu.

41.Trend. Vsak pojav sledi v svojem časovnem razvoju, kljub odklonom za posamezne momente, neki osnovni smeri razvoja. To smer časovnega razvoja imenujemo "trend". Ta smer razvoja je lahko naraščajoča ali padajoča, ali tudi niha na daljše razdolje (sekularna gibanja). Plansko gospodarstvo predpostavlja stalen napredek in razvoj. Osnovna linija razvoja je postavljena z gospodarskim planom, to pa ni nič drugega kot trend. Seveda je možno, da različni vplivi odklanjajo stvarnost od te linije, vendar je osnovna smer več ali manj stalna. Trend moremo opazovati tudi pri vseh časovnih serijah socialno-ekonomskih pojavov, kjer osnovna smer ni podana s planom, ker absolutne konstance pojavov skoro da ne poznamo (specialno v meteorologiji).

42. Sezonske oscilacije pomenijo periodično odklanjanje pojava od trenda. Plansko gospodarstvo stremi za tem, da čim bolj eliminira sezonske vplive na gospodarske pojave (n.pr. gradbena dela). Vendar to v velikih primerih ni možno (n.pr. poljska dela). V takih primerih skuša sezonska oscilacija ublažiti (n.pr.turizem, kjer je perioda leto, periodični vplivi izostanki po dnevih). Iz tega je razvidno, da je važnost poznavanja sezonskih oscilacij veliko.

43. Konjunkturne oscilacije. Konjunkturne oscilacije (prosperitet - napetost - kriza - depresija) so značilne za kapitalistični red. Plansko gospodarstvo jih ne pozna, oziroma mora skrbeti, da jih eliminira. Popolna eliminacija v današnji fazi ni možna, ker imamo opravka še s privatnim sektorjem, ki je v gotovih panogah še močan (kmetijstvo). Vendar ima država možnost vplivati tudi tu z ustavo postavljeno pod kontrolo privatnega sektorja. Na drugi strani pa eliminacija konjunkturnih oscilacij ni možna radi konjunkturnih oscilacij v ostalem svetu,ki ima nujen odmev čeprav šibak tudi v državah socializma.

44. Enkratne spremembe ali enkratni vpliv. Enkratne spremembe so one, katere povzroča enkratni vpliv. Take spremembe morejo biti trajne,t.j.take, ki trenutno spremene tek,in se po spremembi nadaljuje in druge vrednosti. Vpliv spremembe je trajen (n.pr. nov izum dvigne proizvodnjo, prehod nekaj podjetij iz ene direkcije v drugo v eni zveča nivo števila delavstva v drugi pa zmanjša). Trenutne spremembe so take, ki poderejo reden tok, vendar se po prenehanju tega vzroka tek pojava nadaljuje iz prejšnjega nivoja (n.pr.vpliv štrajka na vrednost proizvodnje). Poleg te vrste trenutnih sprememb poznamo še drugo,kjer sprememba v nekem trenutku izzove v naslednjem trenutku reakcijo v obliki spremembe v obratni smeri.(n,pr.ako je tovorni transport nekega podjetja en dan prekinjen je radi tega naslednji

dan dvojen).

45. Slučajne spremembe. Vse ostale spremembe, ki so rezultat delovanja stihije oz. faktorjalnih znakov, kateri so izven možnosti uravnavanja in eliminiranja imenujemo slučajne spremembe.

5. Metode analize časovnih serij.

5o. Splošno. Težišče analize časovnih serij je na proučevanju jakosti posameznih vrst vplivov, ki vplivajo na opazovani pojav. Ena izmed tipičnih metod proučevanja časovnih serij je indeksna metoda, katere pa ne bomo na tem mestu obdelovali, ker je bila podana pri relativnih številih. V ostalem pa se poslužujemo metod, ki pomagajo, razstavljati časovno vrsto na komponente trend, sezonske oscilacije, enkratne vplive, in rezidualne slučajne spremembe. Medsebojni vpliv posameznih pojavov bomo proučevali s korelacijo, ki pa bo predmet posebnega poglavja.

51. TR E N D. Tehnično se izračunava trend kot osnovna smer razvoja na več načinov, od katerih ima vsak svoje dobre in slabe strani, odvisne bodisi od lažjega ali težjega konstruiranja, bodisi po večji ali manjši teoretični dognanosti, ker zavisi sama oblika iskanega trenda v bistveni meri od metode, ki smo jo uporabljali. Predno gremo metodo izračunavanja si moramo biti na jasnem, da bo trend dal sliko osnovne linije razvoja, da bo tekel med realnimi empiričnimi vrednostmi, da predstavlja torej neke vrste dinamično srednjo vrednost, okrog katere kolebajo empirični vrednosti in se mu več ali manj prilagajajo.

511. Prostoročno črtanje trenda. Najenostavnejši način določanja trenda je, da v diagramu, v katerega smo včrtali časovno serijo med empiričnimi vrednostmi prostoročno potegnemo črto, ki naj po občutku predstavlja osnovno smer gibanja. Ta način je najlažji, obenem pa tudi najmanj točen = natančen.

512. Polzeče povprečje. Povprečje izračunano iz členov statistične serije izraža tipično vrednost serije, ker se slučajni odkloni individualnih vrednosti kompenzirajo. Enako izraža aritmetična sredina samo nekaj zaporednih členov tipično vrednost teh členov, oziroma razdobja. Serijo povprečnih vrednosti moremo dobiti na ta način, da tvorimo povprečja iz vseh možnih zaporednih skupin členov po toliko členov, za kolikor smo se odločili, da bomo tvorili povprečja. Praktično nastane vsaka nova skupina

iz členov predidoče skupine tako, da prvi člen v tej skupini izpustimo, dodamo pa naslednji člen, ki pride v celotni seriji za zadnjim členom v skupini. Ker dobimo na ta način celo serijo povprečij in te skupine nekako polze po osnovni seriji, imenujemo to serijo serijo polzečih povprečij. Ako vzamemo n.pr. skupine po tri člene, bomo polzeča povprečja tvorili po naslednji shemi.

Osnovna serija

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6$ ----------

$p_2 = \dfrac{x_1 + x_2 + x_3}{3}$. $p_3 = \dfrac{x_2 + x_3 + x_4}{3}$ $p_4 = \dfrac{x_3 + x_4 + x_5}{3}$ v splošnem

$p_k = \dfrac{x_{k-1} + x_k + x_{k+1}}{3}$

Splošen obrazec za preračunavanje pa je:

$p_{k+1} = p_k + \dfrac{x_{k-2} - x_{k-1}}{3}$

Primer:

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

6 8 10 9 8 ----------

$p_2 = \dfrac{6 + 8 + 10}{3} = \dfrac{24}{3} = 8$

$p_3 = 8 + \dfrac{9 - 6}{3} =$

$p_4 = 9 + \dfrac{8 - 8}{3} = 9$

Če vzamemo v skupine toliko členov, da dobimo obseg periode če gre za periodičen pojav, bo polzeče povprečje eliminiralo poleg slučajnih odklonov v seriji tudi sezonske variacije radi tega, ker se bo v vsaki skupini nahajal en hrib in en dol sezonske variacije. V tem primeru bo serija polzečega povprečja služila kot trend, ker bo v glavnem vsebovala samo osnovno smer gibanja, ker je očiščena vsaj v glavnem tako slučajnih, kot periodičnih variacij. Ker pride v gospodarskih problemih največkrat v poštev lētna sezonska variacija, bomo izvedli tak primer: Izračunamo povprečje pripišemo srednjemu členu skupine. Ker je število mesecev sodo število, torej ni srednjega meseca, uporabljamo tehtano aritmetično sredino skupine trinajstih členov, da vzamemo za končna meseca 1/2 vrednosti, kjer pa so vseeno posamezni meseci enako zastopani.

$$T_{jul} = \frac{1}{1}\left(\frac{1}{1} \text{ jan} + \text{feb} + \ldots\ldots + \text{jul} + \ldots \text{dec} + \frac{1}{2} \text{ ja} \right)$$

ali v splošnem:

$$T_m = \frac{1}{12}\left(\frac{1}{2} y_{m-6} + y_{m-5} + \ldots\ldots + y_m + \ldots\ldots y_{m+5} + \frac{1}{2} y_{m+6} \right)$$

Za praktično računanje je važen postopek, kako iz izračunan povprečja za eno skupino pridemo do povprečja za posledujo skupino.

$$T_m = \frac{1}{2}\left(\frac{1}{2} y_{m-6} + y_{m-5} + \ldots + y_m + \ldots + y_{m+5} + y_{m+6} \right)$$

$$T_{m+1} = \frac{1}{12}\left(\frac{1}{2} y_{m-5} + y_{m-4} + \ldots + y_{m+1} + \ldots + y_{m+6} + \frac{1}{2} y_{m+7} \right)$$

$$T_m - T_{m+1} = \frac{1}{12}\left(\frac{1}{2} y_{m-6} + \frac{1}{2} y_{m-5} - \frac{1}{2} y_{m+6} - \frac{1}{2} y_{m+7} \right)$$

$(y_{m+6} - y_{m-6})$ oz. $(y_{m+7} - y_{m-5})$ so v stvari diference vrednosti za odgovarjajoče mesece dveh sledečih let.

Kot primer vzemimo število natovorjenih vagonov v b. Jugoslaviji po mesecih (glej primer na strani).

$$T_{avg} = T_{jul} + \frac{1}{24}\left(d_{jan} + d_{feb} \right)$$

$$T_{sept} = T_{avg} + \frac{1}{24}\left(d_{feb} + d_{jan} \right)$$

$$T_{okt} = T_{sept} \pm \frac{1}{24}\left(d_{mar} + d_{apr.} \right)$$

$T_{jul} = 115,3$

$$T_{avg} = 115,3 + \frac{1}{24}(9,9 + 7,8) = 116,0$$

$$T_{sept} = 116,9 + \frac{1}{24}(7,8 + 8,2) = 116,7$$

Postopek bo dal pričakovane rezultate v slučaju, da ima osnovno gibanje konstanten porast oz. padec, in ako ima periodično gibanje stalne amplitude. Čim bolj se empirična krivulja približuje temu idealnemu stanju, tembolj bo metoda sama zadovoljiva.

Hiba postopka je tudi ta, da ne moremo izračunati vrednosti polzečega povprečja prvih in zadnjih šest mesecev časovne serije.

513. DOLOČANJE TRENDA S POMOČJO MATEMATIČNIH KRIVULJ.

Vrednosti členov statistične serije se s časom (neposredno) spreminjajo, so torej odvisne spremenljivke, oziroma funkcije časa. Radi tega se postavlja vprašanje, dali bi mogli izraziti empirično dano statistično serijo z matematično funkcijo. E funkcija na eni strani ne sme biti prekoamplicirana, da je konitost jasna, na drugi strani pa moramo streneti za tem, da bodo razlike med empiričnimi in vrednostmi funkcije čim manjše. Ako hočemo z matematičnimi funkcijami določiti trend, ne bomo iskali takih, katere bi se skladale v vseh vrednostih z empiričnimi, kar bi bilo običajno interpoliranje skozi dane točke, temveč bomo iskali razmeroma enostavne funkcije (parabole tega reda, logaritmične krivulje in eventuelno eksponencialne funkcije) katerih vrednosti se bodo čim bolj prilegale empiričnim vrednostim. Pri izbiri tipa krivulje bomo izmed teh izbrali tisto, pri kateri bodo razlike med empiričnimi vrednostmi in vrednostmi izravnane funkcije čim manjše. Ta izbira bo izvedena na ta način, da bomo izbrali primerne paramentre, ki konkretno določajo funkcijo. Problem se v statistiki postavlja na ta način, da iščemo pri izbranem tipu funkcije tisto, za katero je vsota kvadratov odklonov empiričnih vrednosti od funkcijskih najmanjša. (Gaussova metoda najmanjših kvadratov). Ako zaznamujemo z y_K empirično vrednost K tega člena serije, z f_K vrednost funkcije prilagoditve za K, mora biti izpolnjen pogoj.

$$\sum_{K=1}^{N} \left({}_{k} - f({}_{k}) \right)^{2} = Min \quad K = 1, 2, \ldots N$$

S tem so dani paramentri funkcije. Običajno izbiramo kot funkcije, ki izražajo trend:

premico $y = a + bx$

parabolo 2 reda $y + a + bx + cx^2$

parabolo 3 reda $y + a + bx + cx^2 + dx^3$

v splošnem parabolo 12-tega reda

$$y = a + bx + cx^2 + \ldots\ldots\ldots + rx^n$$

eksponentialno funkcijo

$$y = ab^x$$

logaritmično funkcijo

$$y = a + b\lg x$$

a,b,c,itd.so paramentri, katere je treba šele od primera do primera izračunati:

Kot primer vzemimo najenostavnejši primer, da iščemo trend v obliki premice:

$y_1\ y_2\ y_3\ y_4\ y_5 \ldots$ časovna serija za vrednosti člen v

$x_1\ x_2\ x_3\ x_4\ x_5 \ldots$ vrednosti časa, za katerega veljajo y_i.

Funkcija s katero hočemo izraziti trend časovne serije y je:

$$f(x_K) = a + bx_K$$

Izbrati je treba primerne paramentre. Ti paramentri so določeni s pogojem:

$$\sum_{K=1}^{N} (y_k - a - bx_k)^2 = Min$$

Če kvadrate izračunamo, dobimo:

$$y_1^2 + a^2 + b^2x_1^2 - 2ay_1 - 2bx_1y_1 + 2abx_1 +$$

$$y_2^2 + a^2 + b^2x_2^2 - 2ay_2 - 2bx_2y_2^2 + 2abx_2 +$$

$$y_k^2 + a^2 + b^2x_k^2 - 2ay_k - 2bx_ky_k + 2abx_k +$$

$$y_N^2 + a^2 + b^2x_N^2 - 2ay_N - 2bx_Ny_N + 2abx_N$$

$$\sum_k^N y_k^2 + Na^2 + b^2\sum_k^N x_k^2 - 2a\sum_k^N y_k - 2b\sum_k^N x_k y_k + 2ab\sum_k^N x_k = Min$$

Ako vzamemo izhodišče koordinantnega sistema sredi serije, je $\sum_k x_k = 0$, ker imamo iste vrednosti enkrat pozitivne, drugič negativne. Z to okrajšavo moremo gornji izraz pisati v obliki:

$$\sum_k^N y_k^2 + N\left(a - \frac{\sum y_k}{N}\right)^2 + \sum x_k^2\left(b - \frac{\sum xy}{\sum x_k^2}\right)^2 - \left(\frac{\sum y_k}{N}\right)^2 - \frac{(\sum x_k y_k)^2}{\sum x^2} = Min$$

V tem izrazu moremo spreminjati parametra a in b. Cel izraz bo manjši (najmanjši), ako bosta drugi in tretji člen, katera s pomočjo a in b moremo še spremeniti, čim manjša. Ker ne moreta zavzeti negativne vrednosti, je njih najmanjša vrednost 0, torej je:

$$a - \frac{\sum y_k}{N} = 0 \qquad b - \frac{\sum x_k y_k}{\sum x_k^2} = 0$$

S tem sta oba parametra določena.

Primer: Povprečno mesečno število natovorjenih vagonov v bivši Jugoslaviji po letih.

Povp. mes. Prenos
c ooo vag. koord.
sist.

l	y_k	x_k	x_k^2	k	$f(x_k)$
1932	119,0	-3	9	-357,0	111,1
1933	114,5	-2	4	-229,0	116,2
1934	116,8	-1	1	-116,[illegible]	121,3
1935	121,5	0	0		126,4
1936	126,0	+1	1	126,0	131,5
1937	140,6	+2	4	28[illegible]	136,6
1938	146,4	+3	9	439,2	141,7
	885,0	0	28	142,4	

N = 7

$a = \frac{885,0}{7} = 126,4$ / $b = \frac{142,4}{28} = 5,1$

$f(x) = 126,4 + 5,1\, x$

Z vstavljanjem x = -3, -2 ....+3 v f (x) dobimo izračunane vrednosti za posamezna leta.

Postopek je podoben pri drugih tipih funkcij, le da se račun znatno komplicira. Vendar so izdelane metode, ki tudi za komplicirane tipe funkcij omogočajo razmeroma hitro rešitev problema (ortogonalni polinomi, tabele itd.).

514. ELIMINIRANJE TRENDA IZ ČASOVNE VRSTE

Trend in sezonske variacije, dve glavni komponenti gibanja v časovni vrsti, moreta vplivati na vrednost na dva načina: ali aditivno ali faktorialno. Izraženo v funkciji bi imeli za aditivno vezo ako je T (x) trend, S (x) pa periodične oscilacijo, O (x) pa ostali vplivi (slučajni itd.).

$f(x) = T(x) + P(x) + O(x)$

v slučaju faktorialne zveze pa:

$f(x) = T(x) \cdot P(x) \cdot O(x)$

Aditivno zvezo bomo imeli takrat , kadar bo jakost periodičnih oscilacij enaka ne glede na absolutni nivo pojava, faktorialna pa v primeru, da bo vplivala sprememba nivoja pojava tudi na jakost oscilacij.

V primeru aditivnih zvez bomo trend eliminirali z odštevanjem vrednosti trenda od vrednosti členov časovne serije,

$$(f(x) - T(x)) = P(x) + O(x)$$

v drugem pa z deljenjem z vrednostjo trenda

$$\frac{f(x)}{T(x)} = P(x) \cdot O(x)$$

Iz faktorialne povezave moremo priti do aditivne, ako vzamemo mesto vrednosti členov logaritme teh vrednosti.

$$\lg f(x) = \lg T(x) + \lg P(x) + \lg O(x)$$

V ekonomskih pojavih so med trendom in priodičnimi oscilacijami običajnejše faktorialne kot aditivne zveze.

N.pr. jakost sezonske variacije, števila gradbenega delavstva je enkrat, dvakrat večja, če se nivo števila gradenj enkrat, dvakrat poveča.

515. ELIMINIRANJE TRENDA. Kot primer vzemimo število natovorjenih vagonov v b. Jugoslaviji in eliminirajmo trend z odštevanjem polzečega povprečja od empirične serije. Rezultati morejo biti pozitivni ali negativni, ker so sezonske variacije pozitivne ali negativne (glej stran)

52. PERIODIČNE SPREMEMBE.

520. SPLOŠNO. Poleg trenda so glavne komponente gibanj periodične variacije. Periode morejo imeti najrazličnejše dolžine na primer teden, dekada, mesec, leto itd. Največji periodični vpliv na gospodarske pojave ima običajno leto. Letne variacije imenujemo tudi sezonske variacije. Periodične variacije vplivaj na gibanje pojavov tako, da se pojav na določenem delu periode giblje pod na drugem pa nad trendom. Učinki periodičnih variacij se v toku ene periode po pravilu izenačujejo in je zat vsota vseh periodičnih učinkov ene periode enaka 0.

Na podlagi tega sta obe metodi izračunavanja sezonskih variacij izdelani po naslednjih predpostavkah. Z izračunavanjem povprečij iz večjih členov časovne serije se kompenzirajo slučajne variacije. Z vsoto ali z izračunavanjem povprečij členov ene periode pa se poleg slučajnih variacij kompenzirajo tudi periodične (sezonske variacije). S to predpostavko smo pri metodi polzečih povprečij eliminirali slučajne in sezonske variacije, da smo dobili čisti trend.

Konkretna vrednost člena časovne serije je v glavnem sestavljena iz naslednjih komponent: trenda, periodičnih variacij in slučajnih variacij. V simbolih dobimo

$$y_{\mu\lambda} = T_{\mu\lambda} + P_{\mu} + S_{\mu\lambda}$$

Pri Bowley-Smithovi metodi najprej eliminiramo trend s tem, da ga na kateri koli način izračunamo (polzeča povprečja, metoda najmanjših kvadratov), in ga od konkretnih vrednosti časovne serije odštejemo. Tako vstaneta samo periodična komponenta in slučajna variacija. Te pa razdružimo s tem, da v seštevanju vrednosti za določen čas periode izračunamo povprečje iz vseh period in s tem odstranimo slučajne variacije. Druga metoda takozvana Personova metoda je je izdelana pod predpostavko, da je trend linearne oblike torej dan z :

$$T_{\mu\lambda} = a+bt_{\mu\lambda}$$

Z odštevanjem dveh zaporednih členov serije se komponenta trenda pojavlja v diferenci v naslednji obliki

$$T_{\mu\lambda+1} - T_{\mu\lambda} = a+bt_{\mu\lambda+1} - a-bt_{\mu\lambda} = b,$$

torej kot konstanta, ki jo moremo določiti kot kaže izpeljava metode pod 522.

521. Bowley-Smithova metoda. Po eliminaciji trenda ostanejo v časovni vrsti v glavnem sezonske oz.periodične oscilacije in slučajne oscilacije. Ako pogledamo konkretno za naš primer, (problem razdružitve obeh oscilacij) iz prometne statistike bivše Jugoslavije vrednosti za vsa leta za določen mesec, bomo videli, da te vrednosti več ali manj varirajo vendar je nivo vrednosti za vsak mesec drugo. Po podrobnejšem pregledu vidimo, da gornja variacija izhaja iz slučajnih vplivov, spreminjanje nivoja pa od periodičnih oscilacij, ki so rezultat periodičnih vplivov. Nivo gornjih vrednosti bomo dolo-

čili s tem, da bomo izračunali srednjo vrednost, ki bo podala tipično vrednost sezonskega odklona, za določen mesec in s temm merilo odklona. Ker predpostavlja, da je vsota vseh periodičnih oscilacij tekom ene periode 0 (vsota enega hriba in enega dola je 0), bomo eventuelno od 0 različno vsoto razdelili enakomerno na posamezne člene s tem dosegli gornji pogoj.

Shema računanja je naslednja:

$\lambda = \lambda$ - to leto

$n = n$ -ti mesec $\qquad y_{\lambda\mu}$ = empirična vrednost za λ-to let

ℓ število let, m štev. mesecev =12 μ -ti mesec

Potek periode $\qquad T_{\lambda\mu}$ = vrednost trenda za $\lambda\mu$

$y_{\lambda\mu} - T_{\lambda\mu} = Y_{\lambda\mu}$ $\qquad Y_{\lambda\mu}$ = empirične vrednosti očiščene trenda

$p_{\mu} = \dfrac{\sum_{\lambda=1} Y_{\lambda\mu}}{m}$ $\qquad p_{\mu}$ = sezonska variacija (nekorigirana)

$p = \dfrac{\sum_{\mu=1}^{\ell} p_{\mu}}{m}$ $\qquad p$ = korektura

$P_{\mu} = p_{\mu} - p$ $\qquad P_{\mu}$ = sezonska variacija (korigirana)

$$\sum_{\mu=1}^{m} P_{\mu} = 0$$

Gornji način uporabljamo, kadar so posamezne komponente povezane aditivno. V primeru faktorialne povezave pa bomo prešli na računanje z logaritmi ali na računske operacije, ki so za stopnjo višji od seštevanja in odštevanja, t.j. množenje in deljenje.

Pri prvem primeru je bila sezonska oscilacija dana absolutno, medtem ko je v drugem dana z indeksom in moramo za eliminacijo empirične vrednosti s temi vrednostmi deliti. Navedeni metodi se imenujeta Bowley-Smithova metoda in predpostavljata, da je v seriji trend že eliminiran.

522. <u>Persenova metoda</u> verižnih diferenc oz. verižnih indeksov. Ako nimamo trenda eliminiranega in pri predpostavki enakomernega porasta osnovne smeri moremo uporabiti Persenovo metodo verižnih diferenc pri aditivnih zvezah, oziroma verižnih in-

deksov pri faktorskih zvezah.

Metoda obstoja v tem, da izračunamo serije verižnih diferenc in poiščemo povprečne verižne diference za posamezne mesece. Kumulativna serija povprečij verižnih diferenc da nekorigirane sezonske razlike. Korektura bo obstojala v tem, da bomo verižne diference zmanjšali ali povečali za $\frac{1}{12}$ vsote verižnih diferenc, tako, da bo vsota vseh verižnih diferenc enako 0. Ker vsota sezonskih diferenc izračunanih na ta način ne bo enaka 0, iščemo pa take sezonske diference, kjer je vsota vseh členov ~~mmmm~~ periode enaka 0, bomo od vsakega člena odšteli $\frac{1}{12}$ vrednosti vsote vseh členov.

Z simboli je postopek naslednji:

$y_{\lambda.\mu}$ individualna vrednost časovne serije za λ leto in μ-ti mesec

$y_{\lambda,u+1} - y_{\lambda,u} = \vartheta_{\lambda\, u+1}$ individualna verižna diferenca

$\dfrac{\sum\limits_{\lambda=1}^{l} \vartheta_{\lambda\mu}}{l} = V_u$ povprečna verižna diferenca

$d_m = \sum\limits_{\mu=1}^{m} V_u$ nekorigirana sezonska diferenca

$\dfrac{\sum\limits_{\mu=1}^{12} V_u}{12} = \dfrac{d_{12}}{12} = V$ korekturna konstanta

$\vartheta_u - v = \vartheta_\mu'$ korigirana sezonska diferenca

$\sum\limits_{u=1}^{m} \vartheta_\mu' = d_m'$

$\dfrac{\sum\limits_{m=1}^{12} d_m}{12}$

$d_m' - d = \delta_m$ sezonska diferenca reducirana na vsoto 0 $\left[\Sigma\,\delta_m = 0\right]$

Analogen je postopek verižnih indeksov le da so operacije za eno računsko stopnjo višje; t.j. mesto verižne diference - verižni indeksi, mesto vsote produkti, mesto razlike kvocienti itd. razen pri izračunavanju povprečnih verižnih indeksov kjer ostane izračunavanje aritmetične sredine (ako ne računamo z logaritmi).

Postopek v simbolih je naslednji:

$$\frac{y_{\lambda,\mu+1}}{y_{\lambda,\mu}} = v_{\lambda\mu}; \quad \frac{\sum_{\lambda=1}^{q} v_{\lambda\mu}}{l} = v_{\mu}; \quad p_m = \prod_{\mu=1}^{m} v_{\mu};$$

$$v = \sqrt[12]{\prod_{\mu=1}^{12}} = \sqrt[12]{p_{12}}; \quad p'_m = \frac{p_m}{v}; \quad p = \frac{\sum_{1}^{12} p'_m}{12}$$

$$p'_m - p = \Pi_n$$

53. <u>Enkratne izpremembe:</u> Trajne enkratne izpremembe se očitujejo s spremembo splošnega nivoja časovne serije. Vpliv je viden iz višine prenosa nivoja. Vpliv enkratnih trenutnih sprememb eliminiramo s tem, da z običajno metodo interpolacije za trenutek take enkratne spremembe interpolirano vrednost, ki bi nastopila, če te izpremembe ne bi bilo. Razlika med empirično vrednostjo in interpolirano vrednostjo je jakost vpliva enkratne spremembe.xxx

54. <u>o s t a l e (r e z i d u a l n e) k o m p o n e n t e .</u>

Po eliminaciji trenda, periodičnih oscilacij enkratnih sprememb, ostanejo v časovni seriji še izvestne komponente, ki pa so po svoji izmeri navadno manjše, ki ki so večinoma rezultat slučajnih vplivov.

xxx Glede enkratnih sprememb naj omenimo še, da pri analizi trenotne spremembo eliminiramo na ta način, da na tisto mesto vključimo interpolirano vrednost, pri trajnih spremembah pa moramo na mestu prekinitve rednega toka razdeliti časovno serijo na dva dela in za vsak del posebno določati trend.

Izračunavanje trenda po metodi polzečih povprečih

Število natovorjenih vagonov v b.Jugoslaviji.

A: Absolutni podatki v λ μ

	1933	1934	1935	1936	1937	1938
jan.	89,0	98,9	90,5	96,5	107,1	114,6
feb.	86,6	94,4	90,0	92,5	108,4	120,1
mar.	109,7	110,1	104,3	111,6	122,2	148,0
apr.	97,1	101,1	103,8	102,4	128,4	134,8
maj	110,8	113,3	116,0	114,0	129,9	143,2
jun.	106,7	113,2	110,5	116,1	136,3	139,9
jul.	109,5	111,3	125,7	129,5	152,5	148,5
avg.	127,6	125,3	137,4	145,0	163,4	168,8
sept.	152,6	136,7	152,7	155,0	169,2	171,9
okt.	152,2	153,4	169,9	171,0	182,3	178,1
nov.	130,6	132,5	147,4	150,8	151,1	156,1
dec.	106,0	110,8	110,4	129,4	133,8	132,5

B: diferenca med leti: d λ μ

leto λ

mes. μ

	1933	1934	1935	1936	1937	1948
jan.		+ 9,9	- 8,4	+ 6,1	+ 10,5	+ 7,5
feb.		+ 7,8	- 4,4	+ 2,5	+15,9	+11,7
mar.		+ 0,4	- 5,8	+ 7,3	+10,6	+25,8
apr.		+ 4,0	+ 2,7	- 1,4	+26,0	+ 6,4
maj		+ 2,5	+ 2,7	- 2,0	+15,9	+13,3
jun		+ 6,5	- 2,7	+ 5,6	+20,2	+ 3,6
jul.		+ 1,8	+14,4	+ 3,8	+23,0	- 4,0
avg.		- 1,7	+11,5	+ 7,6	+18,4	+ 5,4
sept.		-15,9	+16,0	+ 2,3	+14,2	+ 2,7
okt.		+ 1,2	+16,5	+ 1,1	+11,3	- 4,2
nov.		+ 1,9	+14,9	+ 3,4	+ 0,3	+ 5,0
dec.		+ 4,8	- 0,4	+19,0	+ 4,4	- 1,3

C: polzeče vsote

24xTλu	1933	1934	1935	1936	1937	1938
jan.		2820,8	2785,8	2957,2	3249,0	3501,[illegible]
feb.		2820,9	2811,7	2968,6	3290,4	3503,[illegible]
mar.		2803,3	2839,2	2978,5	3323,0	3511,[illegible]
apr.		2788,5	2871,7	2981,9	3348,5	3509,[illegible]
maj		2791,7	2903,1	2986,4	3360,1	3510,[illegible]
jun.		2798,4	2917,6	3008,8	3364,8	3514,[illegible]
jul.	2766,7	2794,8	2923,3	3038,3	3376,7	
avg.	2784,4	2782,0	2931,9	3064,7	3395,9	
sep.	2792,6	2771,7	2941,7	3091,2	3433,4	
okt.	2797,0	2768,7	2947,6	3127,8	3465,6	
nov.	2803,5	2774,1	2944,2	3169,7	3485,3	
dec.	2812,5	2774,1	2947,8	3205,8	3502,2	

D: Trend izračunan s pomočjo polzečih povprečij.

Tλμ	1933	1934	1935	1936	1937	1938
jan.		117,5	116,1	123,2	135,4	145,7
feb.		117,6	117,3	123,6	137,1	145,9
mar.		116,8	118,3	124,0	138,4	146,3
apr.		116,2	119,7	124,3	139,4	145,9
maj		116,4	121,1	124,4	139,9	146,0
jun.		116,6	121,5	125,2	140,1	146,3
jul.	115,3	116,5	121,8	126,5	140,6	
avg.	116,0	116,0	122,2	127,6	141,4	
sept.	116,4	115,5	122,6	128,7	142,9	
okt.	116,5	115,3	122,8	130,2	144,2	
nov.	116,8	116,6	122,7	132,0	145,2	
dec.	117,2	115,5	122,8	133,4	145,7	

Izračunavanje sezonskih cariacij po Bowley - Smithovi metodi odklonov od trenda

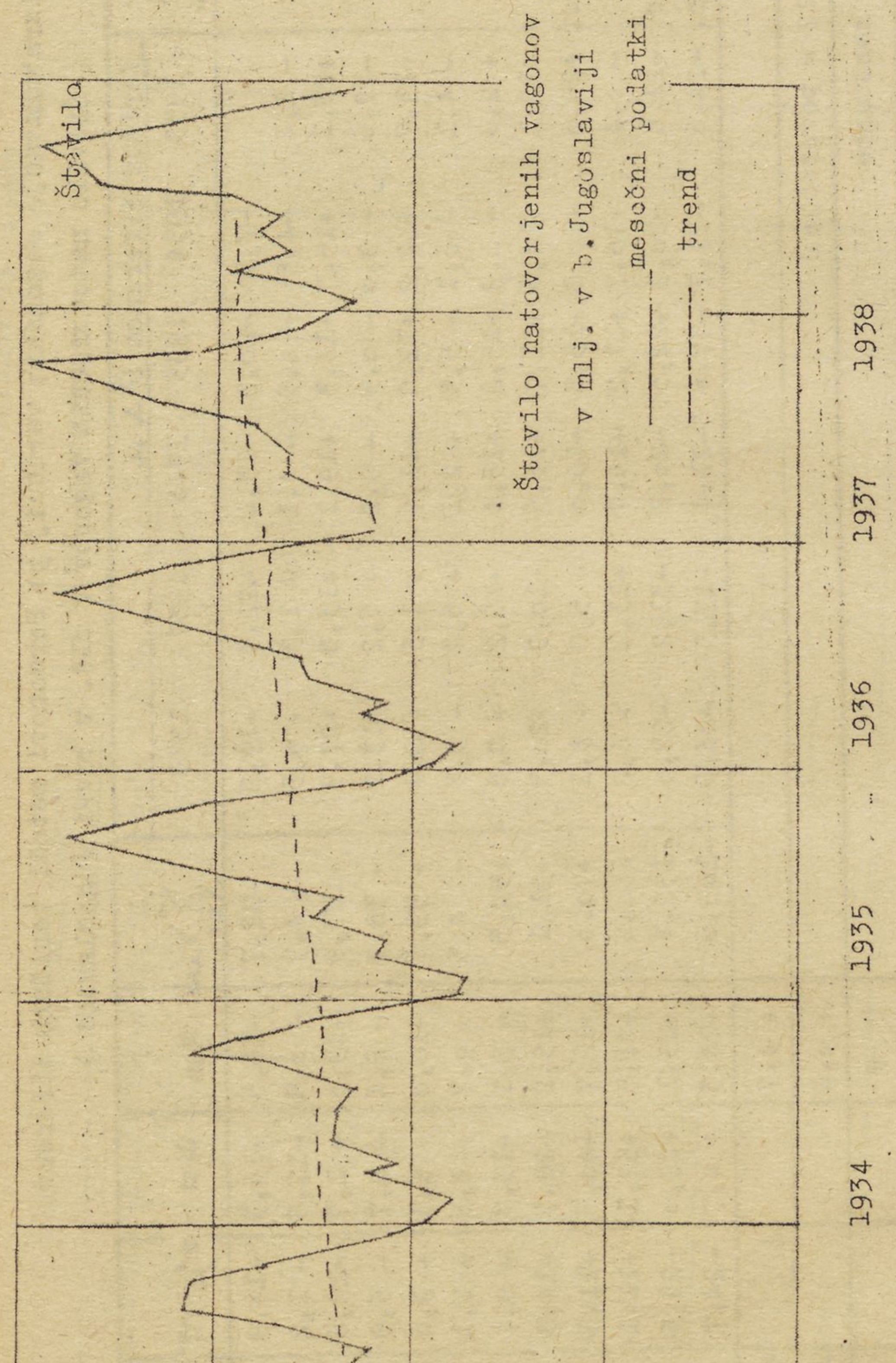
(Število natovorjenih vagonov v mlj.v bivši Jugoslaviji)

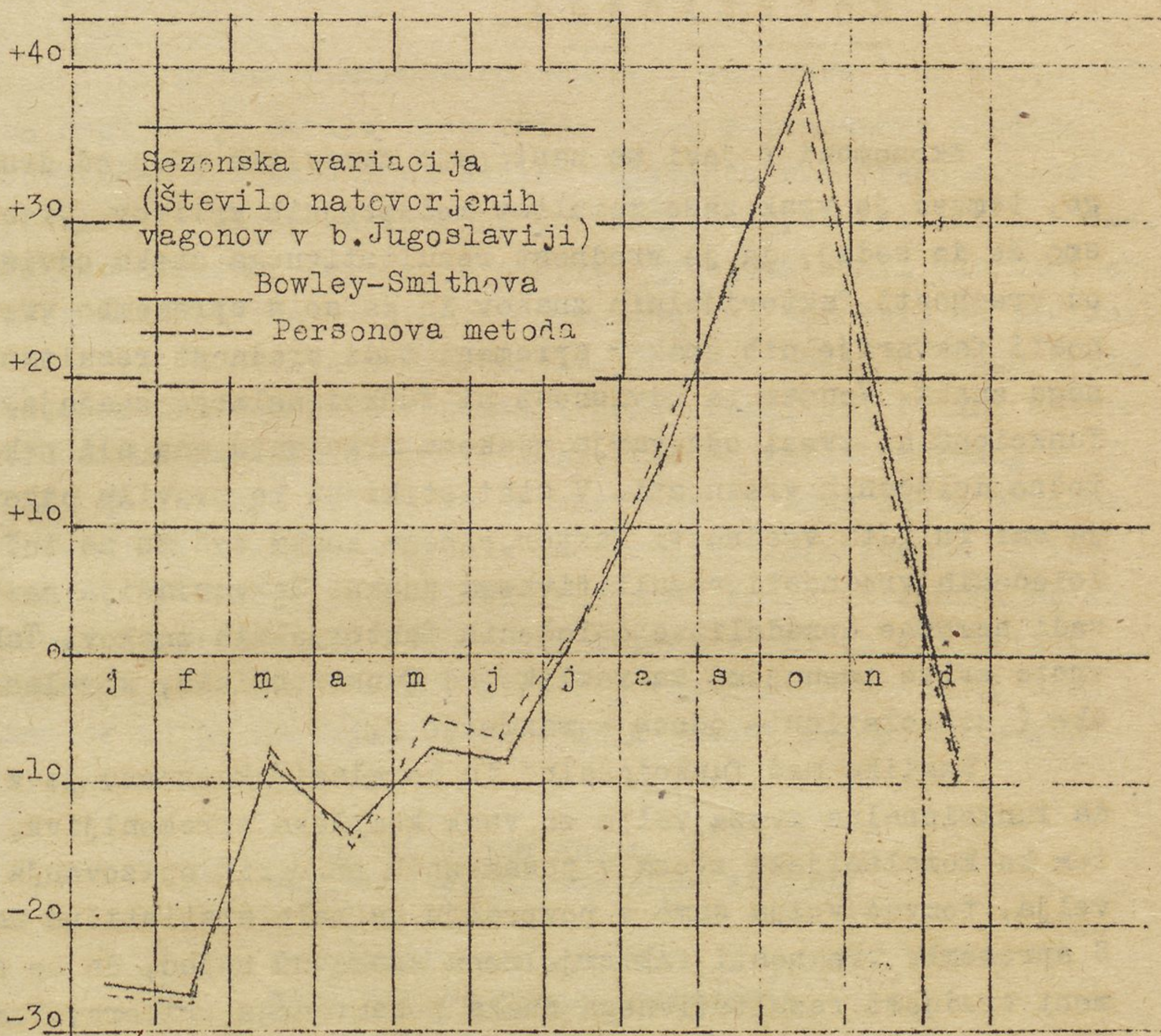
Leto λ	Odklon od trenda $y_{\lambda\mu} = y_{\lambda\mu} - T_{\lambda\mu}$						1		
mesec μ	1933	1934	1935	1936	1937	1938	$\sum_{\lambda} y_{\lambda\mu}$	p_{μ}	P_{μ}
januar	-	-18,6	-25,6	-26,6	-28,3	-31,1	-13o,2	-26,o	-25,3
februar		-23,2	-27,o	-31,1	-28,7	-25,3	-135,8	-27,2	-26,9
marec		- 6,7	-14,o	-12,4	-16,2	+ 1,7	-47,6	- 9,5	- 9,3
april		-15,1	-15,9	-21,9	-11,o	-11,1	-75,o	-15,o	-14,7
maj		- 3,1	- 5,1	-2o,4	-1o,o	- 2,8	-41,4	- 8,3	- 8,1
junij		- 3,4	-11,o	- 9,1	- 3,8	-16,4	-43,7	- 8,7	- 8,4
julij	- 5,8	- 5,2	+ 3,9	+ 3,o	+11,9		+ 7,8	+ 1,6	+ 1,8
avgust	+11,6	+ 9,9	+15,2	+17,4	+22,o		+76,1	+15,2	+15,5
september	+36,2	+21,2	+3o,1	+26,3	+26,3		+14o.1	+28,o	+28,2
oktober	+35,7	+38,1	+47,1	+4o,8	+38,1		+199,8	+4o,o	+4o,3
november	+13,8	+15,9	+24,7	+18,8	+ 5,9		+ 79,1	+15,8	+16,o
december	-11,2	- 4,8	-12,4	- 4,o	-11,9		-44,3	- 8,9	- 8,6
Vsota								-3,o	o
Vsota :12 = povprečje								-o,25	
Oznaka konstante								P	

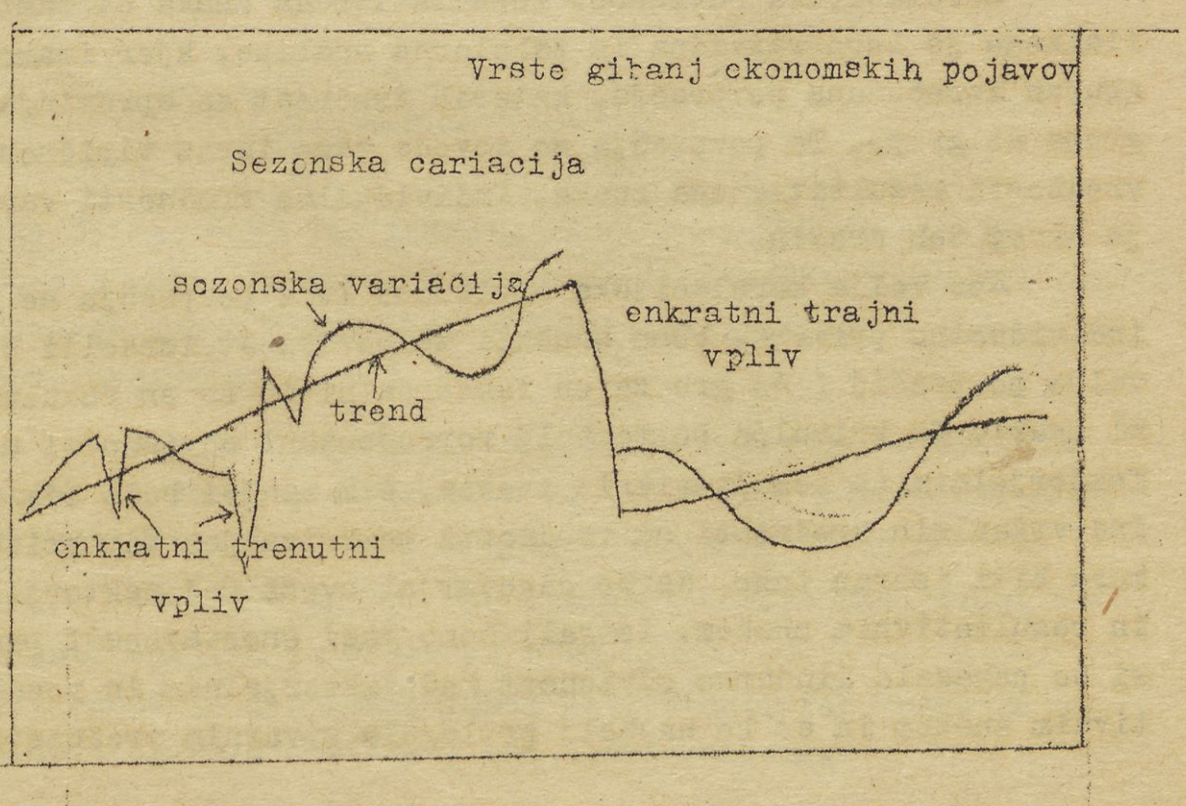
Izračunavanje sezonskih variacij po Personovi metodi verižnih diferenc

(Število natovorjenih vagonov v mlj. v bivši Jugoslaviji)

Leto λ / mesec μ	Verižna diferenca $v_{\lambda\mu}$ 1933	1934	1935	1946	1937	1938	$\sum_{\lambda=1}^{1} v_{\lambda\mu}$	v_μ	đ n	v'_μ	đ'm	Sn
januar		-7,1	-20,3	-13,8	-22,3	-19,2	-82,7	-16,5	-16,9	-16,9	-16,9	-26,1
februar	- 2,4	-4,5	- 0,5	- 4,1	+ 1,3	+ 5,5	- 4,7	- 0,8	-17,3	- 1,2	-18,1	-27,3
marec	+23,1	+15,7	+14,3	+19,1	+13,8	+27,9	+113,9	+19,0	+ 1,7	+13,6	+ 0,5	- 8,7
april	-12,6	- 9,0	- 0,5	- 9,2	+ 6,2	-13,2	- 38,3	- 6,4	- 4,7	- 5,8	- 6,3	-15,5
maj	+13,7	+12,2	+12,8	+11,6	+ 1,5	+ 8,4	+ 59,5	+10,0	+ 5,3	+ 9,6	+ 3,3	- 5,9
junij	- 4,1	- 0,1	- 5,5	+ 2,1	+ 6,4	- 3,3	- 4,5	- 0,7	+ 4,6	- 1,1	+ 2,2	- 7,0
julij	+ 2,8	- 1,9	+15,2	+13,4	+16,2	+ 8,6	+54,3	+ 9,1	+13,7	+ 8,7	+10,9	+ 1,7
avgust	+18,1	+14,6	+11,7	+15,5	+10,9	+20,3	+91,1	+15,2	+28,9	+14,8	+25,7	+16,5
september	+25,0	+10,8	+15,3	+10,0	+ 5,8	+ 3,1	+70,0	+11,7	+40,6	+11,3	+37,0	+27,8
oktober	- 0,4	+16,7	+17,2	+16,7	+13,1	+ 6,2	+68,8	+11,5	+52,1	+11,1	+48,1	+38,9
november	-21,6	-20,3	-22,5	-20,2	-31,2	-22,0	-139,4	-23,1	+29,0	-23,5	+24,5	+15,4
december	-24,6	-21,7	-37,0	-21,4	-17,3	-23,6	-145,6	-24,3	+ 4,7	-24,7	- 0,1	- 9,3
Vsota								+ 4,7			+110,7	+ 0,5
Vsota: 12 = povprečje								+ 0,4			+ 9,2	
Označba konstante								v			d	


Število
Število natovorjenih vagonov
v mlj. v b. Jugoslaviji
mesečni podatki
trend
15o
1oo
5o
1933
1934
1935
1936
1937
1938


+4o
+3o
+2o
+1o
o
-1o
-2o
-3o
Sezonska variacija
(Število natevorjenih
vagonov v b.Jugoslaviji)
Bowley-Smithova
Personova metoda
j f m a m j j a s o n d


Vrste gibanj ekonomskih pojavov
Sezonska cariacija
sezonska variacija
trend
enkratni trajni
vpliv
enkratni trenutni
vpliv

K O R E L A C I J A .

Ekonomski pojavi ne nastopajo neodvisno eden od drugega, temveč je vsak zase rezultat medsebojnih vplivov. Videli smo že do sedaj, da je vrednost rezultativnega znaka odvisna od vrednosti faktorijelnih znakov in da se s spremembo vrednosti faktorijelnih znakov spremeni tudi vrednost rezultativnega znaka. Vendar ta odvisnost ni funkcionalnega značaja. Pri funkcionalni zvezi odgovarja vsakemu argumentu ena ali nekaj točno določenih vrednosti. V statistiki pa po pravilu odgovarja eni in isti vrednosti faktorjelnega znaka več in ne točno določenih vrednosti rezultativnega znaka. Ta variacija nastane radi nemožne opredelitve določenih faktorjelnih znakov. Take vrste zveze imenujemo za razliko od funkcionalnih, korelacijske (correlation - odnos - razmerje).

Razlika med funkcionalno in korelacijsko zvezo je v tem, da funkcionalna zveza velja za vsak kompleks spremenljivk, medtem ko korelacijska zveza v posameznih primerih opazovanja ne velja, temveč velja samo v povprečju za celo statistično maso. S spremembo vrednosti faktorjelnega znaka ni nujno, da se spremeni vrednost rezultativnega znaka posameznega ~~primera, temveč~~ se spremeni nivo vrednosti rezultativnega znaka.

Korelacijska odvisnost rezultativnega znaka od faktorialnega je lepo razvidna iz tabelarne analize, kjer imamo po grupah izračunana povprečja, katerih vrednost se spreminja od grupe do grupe. Ta povprečja so seveda samo izraz tipičnosti vrednosti rezultativnega znaka. Individualne vrednosti varirajo okrog teh sredin.

Ker velja korelacijska odvisnost le v povprečju ne pa za individualne primere, bomo skušali to odvisnost izraziti s krivuljo povprečij (če gre za en faktorjelni in za en rezultativni znak). Ta krivulja bo tembolj reprezentant odvisnosti med faktorjelnim in rezultativnim znakom, čim manjši bodo odkloni individualnih vrednosti od vrednosti te krivulje. Tip krivulje mora biti izbran tako, da bo odgovarjal zvezi med faktorjelnim in rezultativnim znakom. Izbrali bomo radi enostavnosti premico, ki bo pokazala linearno odvisnost med faktorjelnim in rezultativnim znakom in se bo najbolj prilegala stvarnim vrednostim.

Povprečen kvadratičen odklon (varianca) od aritmetične sredine je merilo disperzije okrog aritmetične sredine. Povprečen k kvadratičen odklon individualnih vrednosti od premice linearne odvisnosti je merilo disperzije okrog preme. Varianca okrog premice je manjša, kvečjemu enaka varianci okrog aritmetične sredine. Tem manjša je, čim bolj se odvisnosti prilega stvarnim vrednostim. Na vsak način opisuje prema odvisnosti vrednosti rezultativnega znaka, boljša kot aritmetična sredina. To sledi iz tega, ker se varianca zmanjša. V koliko ga opisuje bolje, pa je razvidno iz zmanjšanja variance. Povezanost med faktorjelnim in rezultativnim znakom je tem večja, čim večje je rezultativno zmanjšanje variance. Relativna diferenca med variancama napram premi in aritmetični sredini more služiti kot merilo povezanosti vrednosti obeh znakov in jo imenujemo, ker določa jakost zveze, determinacijski koeficient r^2.

S^2_y = rezultativnega znaka varianca od aritmetične sredine

S^2_p = rezultativnega znaka varianca od preme odvisnosti

r^2 = determinacijski koeficient

$$r^2 = \frac{S^2_y - S^2_p}{S^2_y} = 1 - \frac{S^2_p}{S^2_y}$$

Ako je $S^2_p = S^2_y$ kar pomeni, da se varianca ni nič spremenila, moremo reči, da premica odvisnosti ne opiše vrednosti rezultativnega znaka boljšo kot aritmetična sredina. V tem primeru povezanosti ni $r^2 = 0$.

Ako je $S^2_p = 0$, pomeni, da je varianca empiričnih vrednosti od premice odvisnosti 0, kar pomeni strogo linearno odvisnost rezultativnega znaka. V tem primeru je $r^2 = 1$.

Ker je $S^2_p \leqq S^2_A$ je vrednost determinacijskega koeficienta omejena in sicer $0 \leqq r^2 \leqq 1$.

Čim večji je determinacijski koeficient, tem tesnejša je povezanost oziroma odvisnost med obema znakoma.

Ako pomnožimo r^2 z 100, dobimo odstotno determinacije, ki pove, koliko odstotkov variacije je pojasnjene z linearno povezavo.

$\frac{S_p^2}{S_A^2}$ to je $1-r^2$ pa pove, kakšen del variacije je ostal še nepojasnjen ($1-r^2$) imenujemo koeficient nedeterminacije.

$$\sqrt{1 - \frac{S_p^2}{S^2}} = r$$ pa imenujemo korelacijski koeficient.

r more biti pozitiven ali negativen. Pozitiven je v primeru preme odvisnosti (če se veča X, se veča tudi Y), negativen pa v primeru obratne odvisnosti (če se veča x, y pada). Predno preidemo k izračunavanju koeficienta determinacije, si moramo biti na jasnem, da bomo vzeli kot preme odvisnosti ono katera najbolje opisuje individualne vrednosti, za katero je torej srednji kvadratični odklon (varianca) najmanjša.

(x_k x_k) je k-ta dvojica vrednosti faktorjelnega znaka (x) in rezultativnega znaka (y).

$f = a + bx$ prema odvisnosti

$(y_k - a - bx_k)$ odklon vrednosti rezultativnega znaka od preme vrednosti

$$S_p^2 = \frac{\sum_{k=1}^{N} (y_k-a-bx_k)^2}{N} = \underset{a \text{ in } b}{\mathrm{Min}}$$ Pogoj za določitev paramentrov

Z metodo najmanjšega kvadrata je možno določiti paramentra a in b in s tem korelacijski koeficient. Pri upoštevaju gornjih predpostavk se obrazec za izračunavanje korelacijskega koeficienta glasi:

$$\frac{\dfrac{\sum_{k=}^{N} x_k \; y_k}{N} - x\,y}{\sigma_x \, \sigma_y} = r$$

Korelacijski koeficient r leži v mejah:

$-1 < r < +1$, Korelacija more torej biti ali negativna ali pozitivna.

Pozitivna korelacija obstoji npr.med velikostjo posestva in številom živine tega posestva, med uvozom in izvozom v zunanji trgovini (primer bivša Jugoslavija) med starostjo in težo živine itd.negativna pa n.pr.med količino in ceno na kapitalističnem trgu, kjer vlada zakon ponudbe in povpraševanja, med skupno površino posestva in % njiv od skupne površine itd.

Korelacijski koeficient je zelo prikladen, ker z enim številom izraža stopnjo povezanosti dveh pojavov.

Primer izračunavanja korelacijskega koeficienta

úvoz in izvoz v bivši Jugoslaviji.

x = uvoz
y = izvoz

	x	y	x^2	y^2	xy
1921	2,5	4,1	6,25	16,81	1o,25
1922	3,7	6,4	13,69	4o,96	23,68
1923	8,o	8,3	64,oo	68,89	66,4o
1924	9,5	8,2	9o,25	67,24	77,9o
1925	8,9	8,8	79,21	77,41	78,32
1926	7,8	7,6	6o,84	57,76	59,28
1927	6,4	7,3	4o,96	53,29	46,72
1928	6,4	7,8	4o,96	6o,84	49,92
1929	7,9	7,6	62,41	57,76	6o,o4
1930	6,8	7,o	46,24	49,oo	47,6o
1931	4,8	4,8	23,o4	23,o4	23,o4
1932	3,1	2,9	9,61	8,41	8,99
1933	3,4	2,9	11,56	8,41	9,86
1934	3,9	3,6	15,21	12,96	14,o4
1935	4,o	3,7	16,oo	13,69	14,8o
Σ	87,1	91,1	58o,23	616,5o	59o,84
: 15	5,81	6,07	38,68	41,1o	39,39

$$r^2 = \frac{\left(\frac{\sum xy}{N} - \frac{\sum x}{N} \cdot \frac{\sum y}{N}\right)^2}{\left[\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2\right]\left[\frac{\sum y^2}{N} - \left(\frac{\sum y}{N}\right)^2\right]} = \frac{(39{,}39 - 35{,}27)^2}{(38{,}68 - 33{,}76)(41{,}1o - 36{,}84=)}$$

$\sigma_x = 2{,}22$ $\quad \frac{(4{,}12)^2}{4{,}92 . 4{,}26} = \frac{16{,}9744}{2o{,}9592} = 0{,}81$

$\sigma_y = 2{,}o6$

$r = 0{,}9$

$y - 6{,}o7 = 0{,}834\ (x - 5{,}81)$

$y = 0{,}834x + 1{,}22$

$y - \bar{y} = \frac{\sigma_y}{\sigma_x} r\ (x - \bar{x})$ pokaže linearno odvisnost znaka in ga imenujemo regresijsko enačbo. Ako zamenjamo v tej enačbi y z x, x pa z y, dobime

$$x - \bar{x} = \frac{\sigma_x}{\sigma_y} r\ (x - \bar{x})$$

r ostane isti radi tega, ker je glede xa in ya simetričen. Ta veza kaže odvisnost x - a od y - a

To premici se ne skladata, ker koeficienta

$\frac{\sigma_y}{\sigma_x} r$ in $\frac{\sigma_x}{\sigma_y} r$ nista recipročni vrednosti.

Imamo torej 2 regresijski premi, ki gresta preko točke $\bar{x}$, $\bar{y}$, te je skozi aritmetično sredino.

Za prejšnji primer dobimo naslednji regresijski enačbi:

1. $y - 6{,}o7 = \frac{2{,}o6}{2{,}22} . 0{,}9\ (x - 5{,}81)$ $\quad y = 0{,}834x + 0{,}69$

2. $x - 5{,}81 = \frac{2{,}22}{2{,}o6} . 0{,}9\ (y - 6{,}o7)$ $\quad x = 0{,}97\ y - 0{,}73$

Korelacijska teorija razvija svoje izvajanje tudi na večje število pojavov, ki se med seboj povezujejo. V teh primerih gre za tkz. multiple in partielne korelacije.

O S N O V N E   M E T O D E   I Z B O R A .

1. ZAKON VELIKIH ŠTEVIL.

Naloga statistike ni samo opazovanje o obsegu in notranji sestavi statističnih mas, temveč tudi iskanje zvez in zakonitosti odnosov, ki se pojavljajo v socialno-ekonosmkih pojavih. Že v prejšnjih odstavkih smo videli, da je vrednost rezultativnih znakov plov vplivov faktorialnih znakov, ki se razdele v dve grupi: prva grupa tkzv.opredeljujočih znakov, katerih vrednosti moremo v vsakem konkretnem primeru določiti (n.pr. pri hektarskem odnosu vrsta žita, način gnojenja itd.) in druga grupa vplivov in pogojev, ki jih radi svoje narave ne moremo opredeliti (konkretno pogoji rasti, mikro sestav zemlje itd.) in ki jih združimo v grupo slučajnih vzrokov. Vendar moremo opazovati, da se vplivi slučajnih vzrokov, ki niso dosegljivi našemu evidentiranju in reguliranju med seboj kompenzirajo in uničujejo in izbije iz rezultata tembolj značilnost tipičnosti in zakonitosti vrednosti ali odnosov, čim večje smo vzeli število opazovanj oz. enot v statistični masi, iz katere smo dognali rezultat.N,pr.primer: Opazujemo shematični primer metanja novca, kjer je statistična enota posamezen met novca, statistična masa vsi izvršeni meti,vrednosti znaka ena ali druga stran novca (številka ali grb). Pogoji metanja niso enostranski, tako da bi bila tendenca za met ene ali druge strani, in novec je simetričen. Ako pogledamo posamezen met, oddvojeno, more pasti ali cifra ali grb. Razmerje med obema je torej pri posameznem metu 1 : 0, ali torej 0 : 1 brez zakonitosti. Ako vzamemo celo vrsto metov in pogledamo, kolikokrat je bila vržena cifra, kolikokrat grb, bomo videli, da se razmerje tembolj izenačuje in bliža vrednosti 1 : 1, čim večje število smo izvršili. Čim večje število metov izvršimo, tem bolj se očituje zakonitost, da je met tako cifre kot grba enakoverjeten, in da je ramerje med njima 1 : 1.

2. primer: Ako opazujemo spol otrok ima neka družina lahko samo sinove, druga same hčere. Ako pa opazujemo odnose v velikem številu rojstev, bomo dobili vedno bolj stalno razmerje 106 moških na 100 ženskih rojstev. Drugačno razmerje pri malo pri-

merih izhaja radi vpliva slučajnih vrzokov.

Gornja lastnost, da se tipične vrednosti (srednje vrednosti) in odnosi oz. lastnosti v homogenih statističnih masah z izvestnim povečanjem števila opazovanih enot bliža neki stalni vrednosti in da se ta vrednost z dodajanjem nadaljnjih vrednosti enot ne spremeni bistveno, imenujemo v statistiki zakon o velikih številih. Ta zakon je za iskanje zakonitosti masovnih pojavov osnovne važnosti in je bil radi tega predmet tako teoretikov, ki so izvedli njegov matematičen dokaz (Laplace, Gauss, Poisson, Čebišev) tako empirikov, ki so dokazali, da zakon v celoti velja v praksi (Buffon, ki je z 4040 meti novca dobil razmerje 49, 31 : 50, 69. Karl Pearson, ki je z 24.000 poizkusi dosegel razmerje 50,05 : 49,95). Vendar si moramo biti na jasnem, da je problematika v socialno-ekonomskih pojavih bolj zapletena, kot v teoretičnih poizkusih z novci, kockami itd.

Gornja zakonitost je napotila statistike do vprašanja, ali ne bi bilo mogoče, ako velja zakon velikih števil, priti do statističnih zaključkov že na podlagi samo delnega opazovanja statistične mase. Iz tega se je razvila celotna veja matematične statistike, ki je do podrobnosti razvila možnosti izrabe teh metod in ki se v statistiki imenuje metoda vzorca. Predno preidemo na opis same metode vzorca (izbora), moramo prej razčistiti še par pojmov, ki bodo za nadaljnja razmotrivanja potrebni.

2. VERJETNOST:

SPLOŠEN POJEM.

Ako imamo določen pojav, bodisi shematičen ali socialno ekonomski, ima vsaka enota za določen znak realizirano eno izmed možnih vrednosti. Čeprav za posamezne enote te mase ne moremo naprej vedeti kakšno vrednost ima, moremo pri poznavanju zasedenosti z vrednostmi predvidevati kakšno vrednost znaka bo imela določena enota. To predvidevanje moremo določiti pojma verjetnosti. Čim več enot mase ima neko določeno vrednost znaka, temvečja je verjetnost, da ima posamezna enota to vrednost znaka. Število enot z določeno vrednostjo v statistični masi se v tem primeru meri z relativno zasedenostjo, ki jo dobimo, da število enot z določeno vrednostjo delimo z celotnim številom enot mase.

Matematično je verjetnost limita ulomka število ugodnih primerov napram skupnemu številu primerov.

$$\lim \frac{nk}{n} = pk = \text{verjetnost.}$$

V teoriji verjetnosti imamo običajno opravka z verjetnostjo " a priori ", to pomeni, da operiramo s pojavi in shemami, za katere je možno že v naprej določiti verjetnost ponavljanja. N.pr.metanje kocke; verjetnost da vržemo številko 3 je 1/6, ker imamo 6 številk, od 1-6, od katerih za nobeno ni razloga, da bi bila verjetnost za nastop manjša, pri vsakem metu pa mora pasti ena izmed šestih številk, verjetnost za nastop katerekoli številke je 1, za eno samo številko torej 1/6. Ta verjetnost se da določiti, ne da bi predhodno izvršili eksperiment in število ugodnih slučajev delili s številom možnih slučajev in iskali limito tega ulomka.

V statistiki je prinsosialno - ekonomskih pojavih nemogoča ne enak način, kot v zgornji shemi določati verjetnost za nastop nekega dogodka ali pojava. V statistiki izvedemo najprej opazovanje poedinih pojavov in šele naknadno določimo, kolik je odnos ugodnih do možnih slučajev. Na ta način dobimo verjetnost nastopa določenega pojava šele naknadno po poiskusu, zato imenujemo tako verjetnost " a posteriori ". Ako imamo v neki statistični masi N enot, od teh pa jih ima n vrednost znaka A, je verjetnost za nastop vrednosti $P(A) = \frac{n}{N}$.

(Ako sestoji statistična masa iz 800 gospodinjstev od katerih jih je 200 z 3 člani, pravimo, da je statistična verjetnost ali relativna frekvenca (ker je v stvari to frekvenca v odnosu na celoto) $P(3) = \frac{200}{800} = 0,25$

Verjetnost 0,25 pomeni, da bi, ako bi napravili poiskus in iz 800 gospodinjstev slučajno izbrali posamezna gospodinjstva, v limiti, to je z velikim številom ponavljanj tega postopka v 0,25 slučajih od celotnega števila naleteli na gospodinjstvo z 3 člani.

Ako izračunamo relativne frekvence za celo serijo dobimo serijo relativnih frekvenc. Vsota vseh relativnih rekvenc je enaka 1, kar pomeni, da ima vsaka statistična enota mase realizirano eno izmed vrednosti znaka.

Primer diskontinuirane serije:

Razdelitev družin po številu otrok

Število otrok	Število družin	Relativna frekvence
1	12.788	0,34226
2	11,63o	0,31126
3	6,399	0,17126
4	3,721	0,09959
5	1,888	0,05053
6	71o	0,019oo
7	177	0,00484
8	4o	0,00107
9	9	0,00024
10	2	0,00005
Skupno	37.364	1,00000

V statistiki se je utrdila terminologija, da govorimo v primeru majhne statistične mase o relativnih rekvencah, v primeru velikih mas pa o verjetnostih.

Kadar je osnovni znak statistične serije zvezen, ne moremo navesti frekvenc za vsako vrednost znaka, temveč tvorimo razrede, poiščemo grekvence v teh razredih, in iz teh frekvenc izračunamo relativne frekvence.

Trajanje gorenja žarnic

Ure gorenja	štev. žar.	relativ. frekvenca
-6oo	3	0,0333
nad 6oo-8oo	1o	0,1111
nad 1ooo-12oo	32	0,3556
nad 8oo-1ooo	33	0,3667
nad 12oo-14oo	9	0,1000
nad 14oo-16oo	3	0,0333
	90	1,0000

ako vzamemoiz tega kolektiva 9o žarnic na slepo eno žarnico, je verjetnost 0,3667, da bo gorela od 8oo-1000 ur. Grafično se da relativna frekvenca pokazati z histogramom. Vsota ploščin vseh stolpcev je enaka 1. Ploščina posameznega stolpa pa je direktno enaka relativni frekvenci oz. verjetnosti za nastop pripadajoče vrednosti.

Ako bi povečali število preizkušenih žarnic, za katere smo izvršili merjenje od 9o n.pr.na 1o.ooo, razrede pa zožili na razrede z širino 1o ur, izračunali relativne frekvence in narisali frekvenčni poligon, bi dobili sliko, ki bi bila sicer prvi podobna vendar bolj precizna, ker bi pokazala relativne frekvence za manjše intervale. V tej sliki bi dobili boljši vpogled v razdelitev posameznih vrednosti. Teoretično bi prišli z večanjem števila žarnic, ki jih vključujemo v statisti masō in manjšanje širine razreda v limiti do zvezne slike razporeditve verjetnosti. Stopinčasto sliko bi zamenjala krivulja.

22. SEŠTEVANJE VERJETNOSTI oz. relativnih frekvenc.

Ako pogledamo primer razdelitve družin po številu otrok in se vprašamo po relativni frekvenci družin z 3 ali 4 otroci, moramo po gornjem pravilu deliti število družin z 3 ali 4 člani t.j. 6.399 + 3721 z skupnim številom vseh družin 37.364 torej $\frac{6399 + 3.721}{37.364} = 0,27085$

Vidimo torej, da pridemo do istega rezultata, ako relativna frekvence oz. verjetnosti seštejemo.

Verjetnost, da bi imela slučajno izbrana družina iz kolektiva 37,364 družin 3 ali 4 otroke je torej 0,27085. Vidimo, da se v primeru, da gre za verjetnost nastopa ene ali druge vrednosti znaka, verjetnosti enostavno seštevajo. N.pr. verjetnost, da ima slučajna izbrana družina manj od 5 otrok, bo enaka verjetnost, da ima družina ali enega ali dva, ali tri, ali štiri otroke, torej (1 ali 2 ali 3 ali 4) = 0,34226 + 0,31126 + 0,17126 + 0,09959 = 0,92437.

Paziti moramo seveda, da gornji stavek velja samo v primeru, da gre za vrednosti znaka, ki se izključujejo. Ne velja n.pr. v naslednjem primeru.

Verjetnost družine z 1 ali 2 otroci je 0,65352

Verjetnost družine z 2 ali 3 otroci je 0,48252

Verjetnost družin z (ali 2) ali (2 ali 3) otroci ni enaka 1,13604 kot bi veljalo po gornjem stavku, kar pa je popolen nesmisel, ker verjetnost ne more biti večja od 1, temveč manjša, ker se gornja dva pojava ne izključujeta (število otrok 2 imamo v prvi in drugi skupini). Vendar tega stavka ne bomo naprej razvijali, ker bomo v nadaljnjem imeli opravka samo z verjetnostmi, ki se izključujejo. V primeru kontinuiranega znaka se bo gornji stavek izkazal v tem. Verjetnost, da bo gorela žarnica od 600-1.000 ur je enaka vsoti verjetnosti, da bo gorela od 600-800 ur in verjetnosti, da bo gorela od 800-1.000 ur. Torej

P (600 - 1.000 n) = P (600 - 800) + P (800 - 1.000) = 0,1111 + 0,3667 = 0,4778. Na sliki bo ta verjetnost dana z vsoto ploščin obeh stolpcev med mejami 600 - 1.000 ur. Ako smo frekvenčno distribucijo izdelali do podrobnosti (velika masa, majhna širina razredov) smo v limiti prišli iz stopnjičaste distribucijske krivulje. V tem primeru je verjetnost za vrednost v poljubnem intervalu dana z vsoto ploščin stolpcev med tema mejama ali v limitiranem procesu s ploščino, omejeno z delom abscice med mejama, z ordinatama nad mejama in delom krivulje verjetnostne distribucije. Površina med celotno abscisо in krivuljo verjetnostne distribucije je v vseh primerih enaka 1. Geometrijskega oz. grafičnega podajanja verjetnosti se bomo posluževali v vseh nadaljnjih izvajanjih, ker je najlažje razumljivo.

Ostalih stavkov verjetnostnega računa se ne bomo dotikali, ker za elementarno razumevanje vsebine teorije vzorca niso potrebni.

3. NADOMESTNE METODE STATISTIČNEGA OPAZOVANJA.

31. Splošno. Statistična opazovanja so običajno v praksi akcije ogromnega obsega. Radi velikega števila enot (popis prebivalstva milijone, kmetijski popisi stotisoče) zahtevajo velika materialna sredstva, široko razpredeno organizacijo, ogromno armado kadrov, ki jih je treba šele učiti izvajati konkreten popis in obdelavo na eni strani, na drugi strani pa razmeroma dolgo dobo, predno pridemo do rezultatov. Velika finančna sredstva, nezanesljivost velikega števila kadrov in prekasna možnost koriščenja podatkov je dala že zgodaj misliti, ali bi se mogla najti metoda, ki bi gornje hibe odpravila.

S statističnimi akcijami hočemo dobiti uvid v tranji sestav statistične mase, ugotoviti zakonitosti, ki vladajo v socialno-ekonomskih pojavih s tvorjenjem grup, z izračunavanjem karakterističnih veličin (srednjih vrednosti, dispersije, korelacijskih koeficientov, relativnih števil itd.). Pod vplivom veljave zakona o velikih številih, po katerem se z večanjem števila opazovanih enot vedno bolj razkrivajo zakonitosti masovnih pojavov, se pojavlja vprašanje, dali je mogoče z opazovanjem samo dela enot statistične mase napraviti zaključke o statistični masi, oz. o socialno-ekonomskem pojavu.

Ena izmed teh nadomestnih metod je anketa, ki predstavlja deloma številčni, deloma deskriptivni opis pojava.

V statistični praksi je bila dolgo časa zelo uporabljena metoda, da se je z izbiro tipičnih pojavov oz. enot sklepalo na dogajanja v celoti. Iz te se je razvila monografska metoda. Bistvo mogografske metode je bilo, da se je podrobno proučilo za dotičen pojav nekaj tipičnih objektov oz. pojavov. Rezultate iz teh raziskovanj se je na to posplošilo na celoto. N.pr. z monografskim opisom tipične vasi se je osvetilo življenje kmetov, z študijo budžetov tipičnih družin sklepalo na budžete v splošnem. Ta metoda tipičnih pojavov se je v statistiki dolgo časa obdržala. Vendar jo je v novejšem času radi tega, ker je izbira tipičnih ~~enot oz. pojavov preveč~~ subjektivna, popolnoma izpodrinila druga metoda t.j. metoda slučajnega izbora. Izbira enot, ki jih bomo izbrali iz celote, ni več odvisna od osebe, ki izbor vrši, temveč je izbor popolnoma objektiven, ker je vsaka enota izbrana slučajno. To dejstvo omogoča, da moremo postaviti temu izboru objektivne znanstvene temelje, ki slone na verjetnostnem računu. Ker je ta metoda izmed vseh nadomestnih metod najobjektivnejša in je njena uporaba vedno večja, se bomo z njo podrobnejše pobavili. Ta metoda je matematično zelo izdelana in tvori doberšen del zanimanja matematičnih statistikov.

32. O nadomeščanju kolektiva z eno samo vrednostjo. Vzemimo na splošno neko statistično maso, v tej statistični masi opazujemo nek znak. Vrednosti tega znaka variirajo od enote do enote, za določen znak oz. maso manj za drug znak oz. maso bolj. Kot nam je že znano, označimo to variabilnost z varianco ^{2} oz. z standardno devijacijo . Tipična vrednost, ki najboljše reprezentira vrednost vseh enot, pa je aritmetična sredina A.

Postavlja se vprašanje, v koliko mora biti vrednost katere koli enote (slučajno izbrane), reprezentant vseh vrednosti oz. aritmetične sredine. V ekstremnem primeru, ko bi bile vrednosti za vse enote iste, bi bila vrednost vsake poljubne enote direkten reprezentant, ker bi bile vse vrednosti enake, in bi bilo potrebno pogledati samo vrednost ene same enote in bi s tem poznali vse ostale vrednosti z aritmetično sredino vred. V tem primeru bi bila varianca 0. Takih ekstremnih primerov pa v statistiki nimamo. V realnih statističnih masah nastopa vedno variabilnost vrednosti znakov. V tem primeru pa bo vrednost posameznega člena statistične mase temboljša reprezentirala vrednosti vseh odn. aritmetično sredino, čim manjši bo odklon od aritmetične sredine, ki je najboljši možen reprezentant. Možni odkloni poljubnega člena pa bodo tem manjši, čim manjša bo variabilnost vrednosti znaka. Vzemimo dva kolektiva ljudi, in za vsak kolektiv opazujemo višino. V prvem smo ugotovili, da je 95% višin vseh ljudi v intervalu 168 - 172 cm. in da aritmetična sredina enaka 170 cm. Ako se poslužimo stavka o verjetnosti, moremo reči, da je 95% verjetnosti, da bo višina povprečnega človeka tega kolektiva med 168 -

172 cm, oz. da je 95% verjetnosti, da odklon višine poljubne enote od aritmetične sredine ne bo večji od 2 cm.

V drugem kolektivu pa smo ugotovili, da je 95% višin vseh ljudi tega kolektiva v mejah 124 - 176 cm in da je aritmetična sredina 150 cm. Poleg tega pa smo ugotovili, da smo 30% vseh enot v intervalu 148 - 152, torej z odklonom od aritmetične sredine, ki je manjši ali enaka 2 cm.

Na prvi pogled je vidno, da je variabilnost v prvem primeru znatno manjša kot v drugem, in da je radi tega vrednost znaka poljubne enote, boljša ocena vrednosti drugih členov in aritmetične sredine kot pa kolektiv drugega primera. Čim manjša je variacija, tem več posameznih vrednosti je v neposredni aritmetične sredine.

33. NORMIRAN ODKLON.

Ako iz neke statistične mase, ki ima za določen znak povprečno vrednost $\bar{x}$, vzamemo neko enoto i, bo ta enota imela vrednost znaka x_i. Odklon $x_i - \bar{x}$ od aritmetične sredine je odklon individualne vrednosti od aritmetične sredine. Ta odklon imenujemo " konkretni odklon ". Ta odklon, čeprav absolutno enak, je, odvisno od mase lahko relativno velik ali majhen, glede na to, ali ima večje ali manjše število vrednosti mase večje konkretne odklone. Za določeno vrednost x_i, za katero je absolutna vrednost konkretni odklon ($x_i - \bar{x}$), moremo vse enote razdeliti na dva dela, v one, za katere je konkreten odklon večji in v one, za katere je konkreten odklon manjpi od tega odklona.

Ta odklon je relativno majhen, ako je odklon za veliko število enot večji ali relativno velik, ako so konkretni odkloni za veliko število vrednosti manjši. To je odvisno od variabilnosti pojava. Zato običajno odklone merimo v relativnih tkzv. normiranih odklonih, kjer so odkloni merjeni v enotah standardne devijacije celotne mase.

Ako zaznamujemo normirani odklon z t je:

$$t = \frac{x_i - \bar{x}}{\sigma_x}$$

to je neimenovano število, ker imata konkreten odklon in standardna devijacija isto enoto mere, enoto opazovanega znaka.

Ako je distribucija normalna, je verjetnost, da bo normiran odklon manjši od določenega t, za vse normalne distribucije, negledc na velikost standardne deviacije distribucije vedno enaka.

Za normalno distribucijo, ki ena osnovnih distribucij teoretične statistike veljajo naslednje vrednosti:

t	$p\left(\left\lvert\frac{x_i - \bar{x}}{\sigma_x}\right\rvert < t\right)$
0,6745	0,5000
1,0000	0,6827
1,9600	0,9500
2,0000	0,9545
3,0000	0,9973

kar pomeni, da je verjetnost P, da velja za x_i neenačba

$$\bar{x} - t\sigma_x < x_i < \bar{x} + t\sigma_x$$

Ako imamo normalno distribucijo, v kateri je $\bar{x} = 10$ $\sigma = 2$ je 95% verjetnosti, da leži individualen primer v razmaku

$$10 - 1{,}96 \cdot 2 < x_i < 10 + 1{,}96 \cdot 2$$

$$6{,}08 < x_i < 13{,}92$$

Tudi pri ostalih teoretičnih distribucijah je verjetnost za določene normirane odklone stalno neodvisna od standardne deviacije. Običajno pa nas zanima obratni problem: Kolika verjetnost je, da se v dani okolici individualnega podatka nahaja aritmetična sredina, ker na podlagi tega ocenimo interval, v katerem se z določeno verjetnostjo nahaja aritmetična sredina.

Za določen del, ki ga preistavlja $P\left(\lvert x_i - \bar{x}\rvert \quad t \quad\right)$ vseh x_i velja po prejšnjem.

$$\bar{x} - t\sigma_x < x_i < \bar{x} + t\sigma_x$$

Ako od te neenačbe odštejemo $x_i + \bar{x}$ bom še vedno veljala za iste x_i. Na ta način dobimo:

$$-x_i - t\sigma < -\bar{x} < -x_i + t\sigma_x$$

Ako neenačbi spremenimo predznak, moramo obrniti tudi znake neenačbe in dobimo:

$$x_i + t\sigma > \bar{x} > x_i - t\sigma$$

Na ta način vidimo, da je možno z poznavanjem standardne de vijacije, z določitvijo verjetnosti, za koliko enot naj ta neenačba verjetno velja (s tem dobimo t) izračunati, v kateren intervalu okrog individualne vrednosti leži aritmetična sredina.

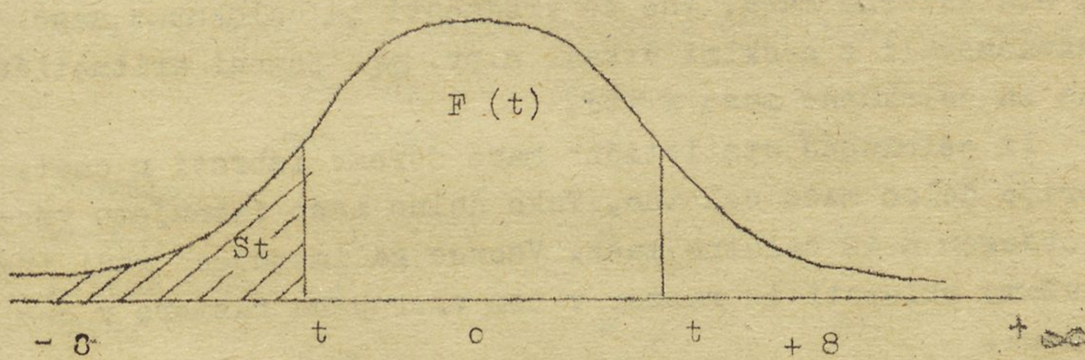
Ako ponovimo gornji primer, in vzamemo iz statistične mase za katero je distribucija normalna, poljubno enoto, za katero je vrednost opazovanega znaka xi, je 95% verjetnosti, da smo izbrali tak xi, da bo aritmetična sredina $\bar{x}$ ležala v intervalu

$$xi - 1{,}96 . 2 \leq \bar{x} \leq xi + 1{,}96 \times 2$$

$$xi - 3{,}92 \leq \bar{x} \leq xi + 3{,}92$$

Ako narišemo relativne distribucije grafično, bo verjetnost nastopa individualne vrednosti v določenem intervalu dana z ploščino lika, ki je omejen z

1. danim intervalom abscisne osi,
2. obema ordinantama nad mejama intervala,
3. z delom distribucijske krivulje nad intervalom.



Pri simetričnih distribucijah je

$$S (o) = \frac{1}{2} \qquad S (+t) + S (-t) = 1$$

in velja med S (t) naslednja zveza

$$2 S (-t) + F (t) = 1 \qquad 2 - 2 S (+t) + F (t) = 1$$

$$F (t) + 1 = 2 S (+t)$$

Iz tega prejšnjega odstavka je razvidno, da je pri znanem tipu distribucije za določen normiran odklon možno določiti verjetnost, s katero lahko pričakujemo, da bo za poljubno izbrani člen distribucije xi ležal $\bar{x}$ v intervalu.

$$xi - t\sigma < \bar{x} < x' + t\sigma$$

obratno moramo trditi, kar delamo pogosteje, da je pri dani distribuciji za določeno verjetnost F (t), možno najti normiran odklon t tak, da bo z verjetnostjo F (t) izbran poljuben xi, tak, da bo $\bar{x}$ ležal v intervalu.

$$xi - t\sigma < \bar{x} < xi + t\sigma$$

Na ta način je z intervalom, dana ocena za $\bar{x}$, z dano verjetnostjo F (t), da ta ocena velja.
Te stavke smo v nadalnjem zelo pogosto uporabljali.

34. <u>CELOKUPNA STATISTIČNA MASA, DELNA MASA ALI VZOREC.</u>

Skupnost vseh enot opazovane statistične mase imenujemo celokupno statistično maso. Število enot v celokupni statistični masi bomo vselej zaznamovali z ν . Vsaka izmed enot ima različne vrednosti znakov. Da imamo vpogled v celotno maso iz teh vrednosti izračunamo za celoto strukturo, srednje vrednosti, standardne devijacije, itd. splošno količine, s katerimi je opisana statistična masa. Vse te vrednosti za celokupno maso bomo zaznamovali z grškimi črkami n.pr. α pomeni aritmetična sredina za celokupno maso x itd.

Iz celokupne statistične mase moremo izbrati n enot, ki tvorijo delno maso celotne. Tako delno maso imenujemo <u>vzorec</u> z obsegom n iz celotne mase. Vzorec ga imenujemo radi tega, ker skušamo sklepati iz razmer v tem vzorcu na razmere v celotni masi.

Takih različnih delnih mas, oz. vzorcev z obsegom n moremo iz celotne mase z obsegom ν več, v splošnem toliko, kolikor je možnih kombinacij z n členi iz kompleksa z členi, t.j. $\binom{\nu}{n}$. Ako je $\nu = 6$, n = 3, je možnih $\binom{6}{3}$ vzorcev, t.j. $\frac{6.5.4}{1.2.3} = 20$) Število vseh možnih vzorcev z danim obsegom pri dani masi je kot vidimo veliko. Radi tega moremo smatrati celokupnost vseh možnih vzorcev kot samostojno maso vzorcev. Enota je posamezen vzorec. Kakšni so opredeljujoči znaki take enote? Da je določena statistična masa enota kolektiva vseh vzorcev, mora imeti določen obseg n, in mora biti vzeta iz iste osnovne mase.

Vedeti moramo seveda, da po vsakem " vzorcu ", ki ga vzamemo iz statistične mase, enote, ki so bile izbrane, ponovno vložimo v celokupno maso.

Vsak vzorec, če naj bo statistična enota mora imeti svoje znake in vrednosti teh znakov. Znak posamezne enote vzorca je lahko aritmetična sredina določenega znaka osnovne enote v tem vzorcu, ki jo zaznamujemo z a, ali standardna devijacija s, strukturni procent določenega znaka p, korelacijski koeficient izračunan iz tega vzorca r itd. Ker so enote posameznih vzorcev različne, se od vzorca do vzorca spreminjajo tudi gornje vrednosti. Vidimo, da imajo znaki vzorcev prave lastnosti statističnih znakov, variiranje. Vsak tak znak mase vzorcev (ker je tu masa vseh možnih vzorcev jo imenujemo statistično skupnost) ima svojo aritmetično sredino A (a) izračunano iz vrednosti vseh vzorcev, svojo standardno devijacijo S (a) in druge.

V našem odtavku o metodi izbora bomo uporabljali naslednje znake. Karakteristične količine v osnovni skupnosti zaznamujemo z grškimi črkami, karakteristične količine posameznega vzorca z malimi latinskimi črkami, karakteristične količine skupnosti vseh vzorcev z velikimi latinskimi črkami.

	Osnovna skupnost	Vzorec	Skupnost vzorcev
Število enot	ν	n	N
Aritmetična sred.	α	a	A
Varianca	σ^2	s^2	S^2
Štev.dev.	σ	s	S

351. Zakonitosti med količinami celokupne mase, delne mase in mase vzorca.

Teorija vzorcev gre za tem, da skuša dobiti zveze, v koliko bi mogli razmere v vzorcu posplošiti kot razmere v celoti. To je možno radi tega, ker veljajo med karakterističnimi količinami osnovne celokupnosti vzorcev določene zakonitosti in zveze. Mi bomo našteli od zvez samo one količine, ki se najpogosteje uporabljajo t.j. aritmetična sredina in standar-

~~na devijacija.~~

1. ~~Število vseh možnih vzorcev~~ N je enako številu kombinacij.

$\binom{\nu}{n} = N$ ako ima osnovna skupnost 10 enot je iz te mase možnih $\binom{10}{5}$ različnih vzorcev po 5 enot.

$$N = \binom{10}{5} = \frac{10.3}{1} = 210$$

V splošnem je N t.j.število vseh možnih vzorcev zelo veliko, ker se z večanjem ν - ja N zelo veča.

2. $A(a) = \alpha(x)$ Aritmetična sredina aritmetičnih sredin določenega znaka vseh možnih vzorcev je enaka aritmetični sredini znaka v osnovni celokupnosti,

3. $S^2(a) = \frac{\sigma^2(x)}{n} \cdot \frac{\nu - n}{\nu - 1}$ Varianca aritmetične sredine vseh možnih vzorcev $S^2(a)$, je z varianco znaka osnovne skupnosti ($\sigma^2(x)$).z številom enot v vzorcu (n) in številom enot osnovne skupnosti (ν) v zvezi,kot kaže enačba pod (3).

Z večanjem velikosti vzorcev, t.j.z večanjem n se varianca skupnosti vseh aritmetičnih sredin vzorcev manjša.

Ako je število enot osnovne skupnosti veliko, se zveza v limiti približa enačbi:

$$\lim S^2(a) = \frac{\sigma^2(x)}{n}$$

4. Aritmetična sredina varianc vseh možnih vzorcev je v naslednji zvezi z varianco v celokupnosti in številom enot in n.

$$A(s^2) = \sigma^2(x) \frac{n-1}{n} \cdot \frac{\nu}{\nu - 1}$$

Ako je število enot osnovne skupnosti veliko,kar običajno je, limitira $A(s^2)$

$$\lim A(s^2) = \sigma^2(x) \frac{n-1}{n}$$

5. Za velik je varianca varianc (s^2) vseh možnih vzorcev znaka

$$S^2(s) = \frac{\sigma(x)}{2n}$$

Za preizkus veljavnosti zgornjih zakonitosti izvršimo v celoti računski primer z majhnim številom enot:
Vzemimo osnovno celokupnost 6 enot.

V = 6

1 2 3 4 5 6 R.št.enote
1 1 4 4 4 4 vrednost znaka xi

Iz te skupnosti vzemimo vse možne vzorce z 3 enotami

n = 3

Možni vzorci	Vrednosti xi v vzorcih
123	114
124	114
125	114
126	114
134	144
135	144
135	144
145	144
146	144
156	144
234	144
235	144
236	144
245	144
246	144
256	144
345	444
346	444
356	444
456	444

V celoti so možni vzorci
4 krat 114
12 krat 144
4 krat 444

Že na prvi pogled je vidno, da ima največ vzorcev sličen notranji sestav kot celota t.j. vrednosti 1 in 4 v razmerju 1 : 2

	frekvenca	vzorec	arit. sredina	Varianca
1 vzorec	4	114	2	2
2 vzorec	12	1[illegible]	3	2
3 vzorec	4	[illegible]44	4	0
skupnost vseh vzorc.	20	-	3	8/5

$$N = \binom{V}{n} = \binom{6}{3} = \frac{6.5.4}{1.2.3} = 20$$

$$\alpha(x) = \frac{1+1+4+4+4+4}{6} = 3$$

$$\sigma^2(x) = \frac{4+4+1+1+1+1}{6} = 2$$

$$A(a) = \frac{4x2+12X3+4x4}{2o} = \frac{8+36+16}{2o} = \frac{6o}{2o} = 3$$

$$s^2(a) = \frac{4.(2-3)^2+12x(3-3)^2+4x(4-3)^2}{2o} = \frac{4+0+4}{2o} = 2/5$$

$$A(s^2) = \frac{4x2+12.2+4.0}{2o} = \frac{8+24}{2o} = \frac{32}{2o} = 8/5$$

$$A(a) = \alpha(x) \qquad 3 = 3$$

$$s^2(a) = \frac{\sigma^2(x)}{n} \cdot \frac{V-n}{V-1} \qquad 2/5 = \frac{2}{3} \cdot \frac{6-3}{6-1} = \frac{2}{3} \cdot \frac{3}{5} = 2/5$$

$$A(s^2) = \sigma^2(x) . \frac{n-1}{n} . \frac{V}{V-1} \qquad 8/5 = 2 , \frac{3-1}{3} . \frac{6}{6-1} = 2 . \frac{2}{3} . \frac{6}{5} = 8/5$$

352. Ocenjevanje aritmetičnih sredin.

Poleg gornjih vez veljajo pa še naslednje zakonitosti:

a) Normirani odkloni aritmetičnih sredin vseh možnih vzorcev danega obsega n, se distribuirajo v tkz."Studentove t – distribucije. Enačba za gostoto Studentove – distribucije je:

$$f(n,t) = C. \left(1 + \frac{t^2}{n-1} \right)^{\frac{n}{2}}$$

C je konstanta, n je paramenter (število enot v vzorcu), t je spremenljivka (normiran odklon)

Kot smo navedli v prejšnjih odstavkih, pomeni ploščina omejena od intervala dela abscisne osi, končni ordinant tega intervala in dela pripadajoče krivulje verjetnost s katero pričakujemo, da bo imel individualen normiran odklon vrednost med spodnjo in gornjo mejo intervala. Kot vidimo je Studentova t distribucija simetrična

$f(n_1 t) = f(n_1; -t)$ in ima modus za vrednost $t = 0$

$f(n,0) = Maks.$

Za naša nadaljnja razmotrivanja ni važna podrobna obravnava te distribucije, temveč potrebujemo samo podatke o tem, kolike so verjetnosti, da bo normiran odklon ležal v intervalu od $-t$ do $+t$. Vprašanje običajno postavljamo obratno, da izberemo verjetnosti in na podlagi teh določimo pripadajoče vrednosti t-ja.

Običajno postavljamo te verjetnosti enake:

5o %, 90 %, 95 %, 99 %, 99,9 %

Na ta način dobimo interval, v katerem se z določeno verjetnostjo nahaja normiran odklon, za individualen vzorec

$$-t < \frac{a-A(a)}{S(a)} < +t \qquad \text{xx}$$

Vendar ta ocena ne more služiti za oceno aritmetične sredine $\alpha(x)$ osnovne skupnosti, ker pove odklon od aritmetične sredine vseh sredin vzorcev.

Ker pa je po (2) $\alpha(x) = A(a)$

moremo gornjo neenačbo pisati v obliki: $-t < \frac{a-\alpha(x)}{S(a)} < +t$

Poleg tega je nemogoče izračunati standardno devijacijo aritmetičnih sredin vseh vzorcev $S(a)$.

Po stavku (3) je $S(a) = \frac{\sigma(x)}{\sqrt{n}} \sqrt{\frac{V-n}{V-1}}$

Vendar je tudi $\sigma(x)$ t.j. standardno devijacijo osnovne skupnosti nemogoče izračunati.

Radi tega jo moramo skušati oceniti na kak drug način. Po stavku 4 je aritmetična sredina vseh vzorcev S^2, t.j. varianca vseh vzorcev, če je osnovna skupnost velika, enaka

$$A(s^2) = \sigma^2(x) \cdot \frac{n-1}{n}$$

V tej enačbi ocenimo $A(s^2)$ z vrednostjo s^2 enega samega vzorca lahko radi tega, kar je varianca od s: $S^2(s)$ že pri razmeroma majhnem n zelo majhna in sicer 1/2 n-ti del variance celote, na ta način dobimo oceno da je $s^2 \approx \sigma^2(x) \frac{n-1}{n}$ ali

$$\sqrt{\frac{n}{n-1}}\, s \approx \sigma(x) \qquad s \approx \sigma(x) \sqrt{\frac{n-1}{n}} \quad \text{ali}$$

Standardno devijacijo aritmetičnih sredin vseh vzorcev imenujemo standardno pogrešk, Ako vstavimo ta izraz v neenačbo (xx) dobimo: $-t < \frac{a-\alpha}{s/\sqrt{n-1}} < +t$

Ker je t dan z izbero verjetnost a in s pa izračunamo iz enega samega vzorca je ta neenačba ocena aritmetične sredine celote, velja

Ako vstavimo ta približek v enačbo za S(a) dobimo

$$S(a) = \sqrt{\frac{n}{n-1}} \frac{s}{\sqrt{n-}} \sqrt{\frac{V-n}{V-1}} = \frac{s}{\sqrt{n-1}} \sqrt{\frac{V^2 n}{V-1}}$$

Pri velikem V limitira za S(a) (x)

$$\lim_{V \to \infty} S(a) = \frac{s}{\sqrt{n-1}}$$

Standardno deviacijo aritmetičnih sredin vseh vzorcev imenujemo standardno pogreško. Ako vstavimo ta izraz v neenačbo (xx) dobimo:

$$-\frac{t\ s}{\sqrt{n-1}} < a - \alpha < \frac{+t\ s}{\sqrt{n-1}} \quad /-a$$

$$-a - \frac{t\ s}{\sqrt{n-1}} < -\alpha < -a + \frac{t\ s}{\sqrt{n-1}}$$ spremenimo predznak

$$a + \frac{t\ s}{\sqrt{n-1}} > \alpha > a - \frac{t\ s}{\sqrt{n - 1}}$$

Ta neenačba podaja oceno za α na ta način, da daje interval, v katerem leži α, pri verjetnosti F (t), da izberemo tak vzorec. Primer: Iz produkcije žarnic smo izbrali po slučajnem izboru 10 žarnic. Dobe trajanja gorenja so naslednje:

Primer	Trajanje ur	$x_i - U$	$(x_i - U)^2$
1	1.100	0	0
2	1.050	-50	2500
3	1.080	-20	400
4	1.200	+100	10.000
5	1.150	+50	2.500
6	1.170	+70	4.900
7	1.030	-70	4.900
8	1.140	-40	1.600
9	1.120	+20	400
10	1.170	+70	4.900
Σ		+210	32.100

$$1.121 + \frac{t.52,5}{3} > \alpha > 1.121 - \frac{t.52.5}{3}$$

$$1.121 + 17,5\ t > \alpha > 1.121 - 17,5\ t$$

a = 1.121 $\quad s^2 = 3210 - 21^2 = 2769$

s = 52,5

Ocena za povprečno trajanje gorenja celotne produkcije je

1.086 < α < 1.156 z 95 % verjetnostjo

V naslednjem podajamo tabelo verjetnosti:

Verjetnosti veljajo za nastop vzorca, za katerega leži α izven gornjega intervala.

Studentova t - distribucija

n-1	P=0,05	P=0,01	P=0,001	n-1	P=0,05	P=0,01	P=0,0001
1	12,706	63,657	636,619	26	2,056	2,779	3,707
2	4,303	9,925	31,598	27	2,052	2,771	3,690
3	3,182	5,841	12,941	28	2,048	2,763	3,674
4	2,776	4,604	8,610	29	2,045	2,756	3,659
5	2,571	4,032	6,859	30	2,042	2,750	3,646
6	2,447	3,707	5,959	35	2,030	2,724	3,592
7	2,365	3,499	5,405	40	2,021	2,704	3,551
8	2,306	3,355	5,041	45	2,014	2,689	3,521
9	2,262	3,250	4,781	50	2,008	2,678	3,496
10	2,228	3,169	4,587	60	2,000	2,660	3,460
				70	1,994	2,648	3,435
11	2,201	3,106	4,337	80	1,990	2,638	3,416
12	2,179	3,055	4,318				
13	2,160	3,012	4,221	90	1,987	2,631	3,402
14	2,145	2,977	4,140	100	1,984	2,626	3,390
15	2,131	2,947	4,073	120	1,980	2,617	3,372
16	2,120	2,921	4,015	140	1,977	2,611	3,361
17	2,110	2,898	3,965	160	1,975	2,607	3,352
18	2,101	2,878	3,922	180	1,973	2,603	3,346
19	2,093	2,861	3,883				
20	2,086	2,845	3,850	200	1,972	2,601	3,340
				300	1,968	2,592	3,324
21	2,080	2,831	3,819	400	1,966	2,588	3,315
22	2,074	2,819	3,792	500	1,965	2,586	3,310
23	2,069	2,807	3,767				
24	2,064	2,797	3,745	1000	1,962	2,581	3,300
25	2,060	2,787	3,725		1,960	2,576	3,291

Pri velikih vzorcih (n velik) Studentova t - distribucija limitira proti normalni krivulji gostota normalne distribucije je dana s formulo

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

Ocena aritmetične sredine nekega pojava moremo uporabiti za ocenjevanje vrednosti določenega znaka za celoto na ta način, da pomnožimo vrednost ocene aritmetične sredine s številom enot celotne osnovne skupnosti z standardno pogreško.

Kot primer ocenjevanja celotne vrednosti vzemimo nasledDji problem:

Za statistično maso 20.550 delavcev je treba oceniti plačilni fond na 1 uro z 2 % vzorcem. 2 % od 20.550 je 411. V vzorec bomo torej vzeli 411 slučajno izbranih delavcev. Za vsakega smo poizvedovali za višino urne plače in dobili naslednjo distribucijo:

Urna plača din	Štev.del.	n_i	$n_i f_i$	$n_i^2 f_i$
18-	19	-3	-57	171
19-	20	-2	-40	80
20-	69	-1	-69	69
21-	170	0	0	0
22-	94	+1	+94	94
23-	34	+2	+68	136
24-	5	+3	+15	45
	411		+11	595

$$a = 21,5 + \frac{11}{411} = 21,527$$

$$s = \frac{595}{411} - \left(\frac{11}{411}\right)^2 = 1,205$$

Ako gre za 95 % verjetnost je t =1,966 (po tabeli)
Standardna pogreška torej enaka

$$\varepsilon = \frac{s.t}{\sqrt{n-1}} \cdot \sqrt{\frac{V-n}{V-1}} = \frac{1,205.1,966}{\sqrt{410}} \sqrt{0,98} = 0,116$$

Aritmetična sredina celote leži med $a - \varepsilon$ in $a + \varepsilon$, torej med 21,527 - 0,116 in 21,527 + 0,116. Skupni plačilni fond je:

Na = 20,550 x 21.527 = 442.380 din

Verjetni odklon plačilnega fonda pa je $N\varepsilon$ =

20.550 x 0,116 = 2.383,8 din.

Možen odklon je kakor vidimo relativno majhen,ker znaša samo

$$\frac{2.383,8 \times 100}{442.380} \% = 0,54 \%$$

Metoda izbora omogoča vnaprej določiti dopusten odklon oz. natančnost rezultatov vzorca. Na podlagi določenega odklona moremo izračunati število enot v vzorcu po formuli za realen odklon

$\varepsilon = \frac{t.s}{\sqrt{n-1}}$ iz tega obrazca moremo izračunati pri danem ε, določenem t in ocenjenem s,n

$$n = 1 + \frac{t^2.s^2}{2}$$

Število potrebnih enot je temveč je čim večja je variabilnost proučavanega znaka, in čim manjši je določen dopusten odklon.

353. OCENJEVANJE STRUKTUR:

Ker nas pri statističnem opazovanju skoro vedno zanima struktura statistične mase, je potrebno, da si ogledamo še ocenjevanje struktur po metodi slučajnega izbora. Problem je vsebinsko popolnoma analogen kot za ocenjevanje aritmetične sredine. Poiskati moramo meje zanesljivosti za posamezne verjetnosti.

Ako zaznamujemo z V_i število členov z določeno vrednostje znaka v celotni statistični masi, je $V_i/V = \pi$ i relativna frekvenca v celoti.

Analogno temu imamo v vzorcu z n enotami $\frac{n_i}{n} = p_i$

Ker te relativne frekvence v stvari niso druge, kot aritmetična sredine, ako vzamemo za n_i členov vrednost 1, za ostale pa 0

x	fi	fx	fx^2
1	n_i	n_i	n_i
0	$n-n_i$	0	0
	n	n_i	n_i

$$a = \frac{\sum fx}{n} = \frac{n_i}{n} \quad s^2 = \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2 = \left(\frac{n_i}{n}\left(\frac{n_i}{n}\right)^2 = p_i(1-p_i)$$

Zato v teh primerih veljajo vse zakonitosti, ki smo jih opazovali pri cenitvah aritmetičnih sredin s to izjemo, da relativna frekvenčna distribucija ni $\frac{n_i}{n}$ vseh možnih vzorcev ni Studentova t - distribucija, temveč tkzv. binomialna - Bernoullijeva distribucija. Relativna frekvenca za $\left(\frac{n_i}{n}\right)$ je dana za binomialne distribucijo po obrazcu

$$P\left(\frac{n_i}{n}\right) = \binom{n}{n_i}\pi_i^{n_i}(1-\pi_i)^{n-n_i}$$

distribucija ima dva parametra n in π_i, t.j. velikost vzorca in verjetnost za nastop vrednosti x_i.

Ta distribucija v splošnem ni simetrična razen kadar je $\pi_i = \frac{1}{2}$

V limiti preide binomialna distribucija z $n \to \infty$ pri $\frac{V_i}{V} = \frac{1}{2}$ v normalno distribucijo.

Običajno uporabljamo normalno distribucijo kot približek že za končne n in pri $\pi_i \approx 1/2$, ker napaka ni velika, pač pa se zelo olajša računska operacija.

Meja zanesljivosti ocene so pri predpisani verjetnosti F (x) dane za aritmetično sredino v tem primeru za

$$\frac{n_i}{n} - \frac{x.s}{\sqrt{n-1}} < \pi_i < \frac{n_i}{n} + \frac{x\,s}{\sqrt{n-1}} \quad \text{ali}$$

ker je $s \pm \sqrt{p_1 (1-p_1)}$

$$\frac{n_i}{n} - x\sqrt{\frac{p_i(1-p_i)}{n-1}} < \pi_i < \frac{n_i}{n} + x\sqrt{\frac{p_i(1-p_i)}{n-1}}$$

F(x) za normalno distribucijo

x	f(x)	x	F(x)	x	F(x)	x	F (x)
0,0	0,000	1,0	0,683	2,0	0,955	3,0	0,9973
0,1	0,080	1,1	0,729	2,1	0,964	3,5	0,99953
0,2	0,159	1,2	0,770	2,2	0,972	4,00	0,99993
0,3	0,236	1,3	0,806	2,3	0,979		
0,4	0,311	1,4	0,838	2,4	0,984		
0,5	0,383	1,5	0,866	2,5	0,986		
0,6	0,451	1,6	0,89[illegible]	2,6	0,991		
0,7	0,516	1,7	0,911	2,7	0,993		
0,8	0,576	1,8	0,928	2,8	0,995		
0,9	0,632	1,9	0,943	2,9	0,996		

Primer:

V nekem okraju je bilo rojenih v določenem razdobju 1.352 otrok. Iz te statistične mase izberemo 5 % enot tj. 67 otrok in ugotovimo, da je od teh 37 dečkov. Določiti interval v katerem predvidoma leži odstotek dečkov celote z 95 % verjetnostjo.

$n_i = 37$

$n = 67 \quad \varepsilon = 196.\sqrt{\frac{0,552.0,448}{66}} \cdot \frac{1352 - 67}{1351} =$

$p_i = 0,552 \quad = 196.0,091.0,97 =$

$1 - p_i = 0,048 \quad = 0,173$

$$0,552 - 0,173 \leq \pi_i \leq 0,552 + 0,173$$

$$0,379 \leq \pi_i \leq 0,725$$

Kot vidimo, je rezultat radi premajhnega števila primerov zelo nezanesljiv, verjetno je z 95 % verjetnosti vrednost odstotka dečkov v celoti od 37,9 % % do 72, 5 %.

36. NAČINI SLUČAJNEGA IZBIRANJA. Že uvodoma o metodi izbora smo naglasili, da je slučajen izbor edina nadomestna metoda, ki ima teoretično izgrajeno osnovo, to je verjetnostni račun. Shematično dobimo slučajen izbor, da iz žare z raznobarvnimi krogljicami vlečemo na slepo kroglje. Pri stvarnem statističnem opazovanju pa je treba iz danega materiala, spiskov itd. izbrati slučajno določeno število enot. Ker je to v gotovih primerih težko, ne da bi vnesli subjektiven moment v sam izbor, so postopki izbora že v naprej izdelani. Najenostavnejši način izbora je mehanični izbor. Pri mehaničnem izboru izbereno posamezne enote na ta način, da v spisku enot celote, kjer imamo oštevilčene posamezne note z tekočimi številkami izberemo n.pr. pri 95 % izboru vsako 20 enoto n.pr. Z redno štev. 1, 21, 41, 61 itd. Ta način je zelo enostaven, treba pa je paziti, da ta delna masa ni enostranska, t.j, da se enota z določenimi značilnosti ne kopičijo ravno v teh številkah, kar more vplivati na kakovost rezultatov.

Mehanični izbor v regionalnem smislu pa pomeni, da so enote , vzete v izbor v naravi krajevno enako oddaljene druga od druge (n.pr. pri netraži).
Ta način ni čisti slučajni izbor in more imeti radi tega pod nepravimi pogoji škodljive vplivna rezultate. Radi tega je treba pri uporabi mehaničnega izbora predhodno preštudirati ako ima sistematičnega odklona.

Pravi slučajni izbor bi imeli, ako bi iz žare z dobro premešanimi oštevilčenimi listki z vsemi tekočimi številkami celote izvlekli na slepo številu enot odgovarjajoče število listkov in bi iz spiska enot vzeli enote z tistimi tekočimi številkami, ki so napisane na izvlečenih listkih. Da ni treba od primera do primera konstruirati takih pomožnih žar, so izdelane tabele " slučajnih številk ". To su nanizane številke, ki bi jih z vlečenjem listkov dobili iz žar. Radi nazornosti prinašamo izrez iz tabele " slučajnih števil" iz " Romanovskega" dela "Uporaba matematične statistike v poskusih.

Tablica slučajnih števil

8574	5490	4069_x	0163^x	1241^2	1270^2
4575	6276	2709^x	4732_x	0301^x	8730
4999	8826	0291^x	0258^x	9430	
7627	6090	9572_x			
4315	5778	1508^x			
6987	8055				
0387^x					

Ako je število vseh enot trištevilčno vzamemo v obzir v vsakem štirištevilčnem številu samo prve ali zadnje tri številke. Ako je število vseh enot 378, bodo prišla v poštev samo ona števila, ki so manjša od 378 (v tabeli zaznamovane z x). V izbor bi iz teh števil prišle v poštev redne številke

38, 270, 29, 150, 16, 25, 124, 30, 127.

Slučajni izbor v splošnem ni reprezentativen, vendar je verjetno reprezentativen, t.j. obstoja večja možnost da je, kot da ni.

Pogreško v izboru, ki jih povzroči lahko slučajen vzorec, če ni reprezentativen, so lahko velike in povzročajo lahko velike motnje v uporabnosti vzorca. Raziskovalci so spoznali, da se napake morejo zelo zmanjšati, če imamo možnost, da izberemo tak vzorec, da je podoben celotni tj. da ima slične značilnosti kot celota. Izvedba tega zahteva nekaj predhodnih informacij o celoti, da raziskovalec vsaj približno ve, kaj naj bi predstavljal reprezentativen vzorec.

Reprezentativen vzorec dosežemo z izborom v legah (stratificiran vzorec). Zanesljivost ocene v odločilni meri zavisi od velikosti variacije znakov. Če celotno maso razdelimo na homogene dele, bo variacija homogenih delov znatno padla. Variacija v grupi bo majhna, variacija med grupami pa velika. Ako bomo jemali iz posameznih homogenih delov celote slučajne vzorce, bodo ocene za te delne mase radi manjših variacij bolj zanesljive. Iz teh ocen se naknadno sestavi rezultat za celoto. Moremo pa vzeti samo izbiranje enot po legah, obdelavo podatkov pa napravimo skupno.

Primer za stratificiran izbor je predhodna razdelitev na okraje, ako ima regionalna razlika vpliv na pojav. (n.pr. pri vzorčenju zemljiških gospodarstev). Pri vzorcu zemljiških gospodarstev je dobro kot lega vzeti poljedelske rajone, ali razdelitev na nižinska in višinska gospodarstva, na majhna, srednja, velika.

Ako je število grup - leg veliko, moremo izvor vršiti tako, da ne izbiramo iz vseh leg, temveč samo iz nekaterih, ki jih pa zopet izberemo slučajno. V teh slučajno izbranih grupah pa nato izvršimo slučajno izbiranje. Ta način je zelo priporočljiv v aplikaciji na regionalno razdelitev in ga imenujemo stratificiran podizbor. Pri dosedanjih vzorcih, ki jih je vršil zvezni statistični urad FLRJ n.pr. o vezanih cenah ali materah z več kot 5 otroci, je bila kot lega KLO (v LRS jih je 1266). Iz teh KLO-jev je bilo po slučajnem izboru izbranih določeno število. Nadaljnji izbor se je vršil samo v teh KLO-jih. Podizbor uporabljamo v glavnem radi administrativnih razlogov. Popisovanje samo v nekaterih KLO-jih je zvezano z manjšo organizacijo, manjšimi finančnimi stroški in manjšo obremenitvijo terena. Kajti ni vseeno, ako imamo popisati 10.000 enot, katere so raztresene v 1.300 KLO-jih in je v vsakem KLO-ju 5 - 10 enot, ali so vse te enote v 50 KLO-jih in je v vsaki povprečno 200 enot. V praksi se tudi mnogo uporablja dvojni izbor. Določene podatke posebno iz ekonomskega področja posameznikov je zelo težko zbrati, ker naletimo na upor in tendenco utaje in neresničnih podatkov. V takih primerih najprej podrobno opazujemo drug pojav, ki je z iskanim v korelacijski zvezi. Korelacijo med obema pa poiščemo z manjšim vzorcem, tako da o kočljivih vprašanjih vprašujemo zelo malo oseb.

Večji izbor dvojnega izbora moremo uporabljati tudi za določevanje tehtanega izbora v legah, ker z njim dobimo relativno razmerje velikosti posameznih grup.

Tako moremo pri zemljiških popisih z velikim vzorcem ugotoviti razmerje med številom posestnikov po velikostnih skupinah, z manjšim vzorcem znotraj vsake skupine za medsebojne odnose ostalih znakov.

37. UPORABA METODE IZBORA. Uporaba metode izbora je vsestranska. Povsod kjer imamo opravka z statističnim opazovanjem, moremo s pravilno izpeljanimi postopki metode vzorca priti do rezultatov, ki imajo predpisano natančnost. Posebej se je metoda slučajnega izbora utrla pot v tista področja, v katerih ne moremo izvajati kompletna opazovanja kot je n.pr. kontrola kvalitete proizvodnje. Statistična kontrola kvalitete proizvodnje je po-

stala posebna statistična disciplina in je v nekaterih državah vedno bolj uveljavlja. (ZDA, SSSR itd.).

Poleg tega se metoda izbora uporablja s pridom tudi v poljedelski statistiki, ravno radi tega, ker imamo opravka zvzelo številnimi pojavi. Zanesljivo ugotavljanje hektarskega donosa posameznih kultur za večja področja, kjer ne more priti v pošte v kompletno opazovanje je možno le s pomočjo na metodi slučajnega izbora bazirani metodi metraže. S pomočjo vzorcev izbranih z mehaničnim izborom, se more s poljubno natančnostjo oceniti hektarski donos določene kulture. Za žita sestoji enota vzorca iz poskušne parcele v izmeri 1 m2. Na tej parceli preštejemo število klasov, iz določenega števila klasov znotraj tega kvadratnega metra (običajno ca 25) preštejemo zrna posameznih klasov in stehtamo težo 1.000 zrn. Iz skupnosti vseh enot vzorca (poskusnih parcel vzamemo več(moremo napraviti oceno hektarskega donosa celote. Oceno moremo napraviti poljubno natančno s tem da dvigamo število enot vzorca. Za homogena področja zadošča že razmeroma majhno število poskusnih parcelic za dosego zanesljivega rezultata za nehemogena pa si običajno pomagamo z stratificiranimi izbori.

38. <u>PO METODI IZBORA IZVRŠENE STATISTIČNE AKCIJE V LRS</u>. Do sedaj je statistični urad LR Slovenije bodisi v zveznem merilu, bodisi samostojno izvedel s pomočjo metode izbora naslednje statistične akcije:

<u>10 % vzorec popisa zemljiških gospodarstev v letu 1947.</u>
Ker bi razmeroma komplicirana obdelava podatkov popisa zemljiških gospodarstev z leta 1947 po velikostnih skupinah zahtevala preveč časa in finančnih sredstev je bil popis sumarno obdelan kompletno, po velikostnih skupinah pa z 10 % izborom. Izbor je bil mehaničen in se je vzelo v obzir vsako deseto gospodarstvo po obstoječem spisku. Na ta način je bila obdelana struktura po kategorijah kultur po velikostnih skupinah.

<u>Anketa o vezanih cenah</u>. Akcija je bila zveznega merila in izvršena v letu 1948. Izbor je bil izvršen na podlagi stratificiranega podizbora s tem, da je bil popis izvršen v omejenem številu KLO-jev, kar je znatno pocenilo in skrajšalo celotno delo. Akcija je bila uspešno izvršena (glej govor maršala Tita na novo leto 1949), hitro in z razmeroma majhnim številom kvalificiranih kadrov.

<u>Vzorec popisa živine leta 1948.</u> Poleg rednega popisa živine, ki se vrši vsako leto decembra meseca je bil izveden leta 19.. vzorec popisa živine ob času najvišjega stanja živine. Izbor bil stratifiiciran podizbor in sicer 5%. Izbor KLO-jev je bil slučajen v okviru OLO-ja in sicer na ta način, da ni bilo vzetih 5 % KLO-jev temveč toliko KLO-jevq da je skupnoštevilo zneslo 5%.

<u>Poskusna metraža v LRS jeseni leta 1948.</u> Izvršena je bila po vseh OLO-jih Slovenije. Material služi za študij za uvedbo znanstveno osnovane metraže v LRS. Proučeval se je hektarski donos pšenice.

<u>Anketa o materah z nad 5 živimi otroci v letu 1949.</u> Akcija je bila zveznega obsega in sicer po metodi stratificiranega podizbora. Obdelava je bila izvršena v zveznem obsegu. Rezultati so služili za osnutek vladne uredbe.

<u>2 % vzorec popisa prebivalstva v LRS.</u> Ker bo kompletna obdelava popisa prebivalstva z dne 15.marca 1948 trajala še vse leto 1949, je statistični urad Slovenije za hiter pregled na podlagi osnovnega materiala izvedel 2 % izbor. Kot enota izbora je bilo vzeto gospodinjstvo, obdelane pa vse osebe, ki so spadale v to gospodinjstvo. Gradivo je bilo obdelano na Powers strojih in je dal zadovoljive rezultate. Izbor je bil regionalno stratificiran, ker so bila gospodinjstva izbrana tako, da so bila enakomerno razdeljena po vseh KLO-jih, na vsakih 100 gospodinjstev 2 primera.

<u>Ocena socialne strukture iz popisa prebivalstva.</u> Med obdelavo popisa prebivalstva je bila na hitro obdelana po metodi vzorca socialna struktura in sicer na podlagi mehaničnega izbora (vsak 61 slučaj iz kontrolnika).

<u>Anketa o živalski proizvodnji.</u> Istočasno s popisom živine, ki je bil kompletna statistična akcija je bil izvršen vzorec o živalski proizvodnji za vsako dvajseto posestvo. Izbor je bil mehaničen in so bila gospodarstva izbrana na podlagi obstoječih spiskov. Za ta gospodarstva so bila izbrana potrebna vprašanja o živalski proizvodnji. Obdelava je bila izvršena v zveznem obsegu.

<u>Kontrola popisa živine leta 1949.</u> Za ocenitev zanesljivosti podatkov popisa živine je bila tik za popisom izvedena za vsaki 54 gospodarstvo kontrola glavnih podatkov. Na podlagi podatkov te kontrole je bil za posamezne vrste živine po metodi vzorca izračunan % pravilnosti popisa.

Popravki.

stran	vrsta	namesto	vstavi
5	1o	$p_2=x_1+x_2+x_3$	$p_2=\frac{x_1+x_2+x_3}{3}$
5	18	$p_3=8+\frac{9-6}{3}$	$p_3=8+\frac{9-6}{3}=9$
6	7	$T_{jul}=\frac{1}{1}(\frac{1}{1}\,jan+feb+\quad)$	$T_{jul}=\frac{1}{12}(\frac{1}{2}jan+feb+\quad)$
6	13	$T_m=\frac{1}{2}(\frac{1}{2}\,y_{m-6}+\quad)$	$T_m=\frac{1}{12}(\frac{1}{2}\,y_{m-6}+\quad)$
8	2	$\overset{N}{\underset{K=1}{}}\ K-f(\ _k)^2=Min$	$\sum_{k=1}^{N}\left[y_k-f(x_k)\right]^2=Min$
8	3	paramentri	parametri
povsod	dalje	paramenter	parameter
8	8	parabolo 12-tega	parabolo n-tega
8	25	$N\ (y_k-\quad-bx_k)^2=Min$	$\sum_{k=1}^{N}(y_k-a-b\,x_k)^2=Min$
13	13	$y_{\lambda\mu}=T_{\lambda\mu}=$	$y_{\lambda\mu}-T_{\lambda\mu}=y_{\lambda\mu}$
14	18	$v_\mu-v=v_\mu$	$v_\mu-v=v'_\mu$
14	2o	$\frac{\sum_{m=1}^{12} d_m}{12}\quad d'_m-d=\delta_m$	$\frac{\sum_{m=1}^{12} d_m}{12}=d$ $d'_m-d=\delta_m$
19	3	$\vert\ n\ \ v'u\ \vert\ d'm\ \vert\ \delta u\ \vert$	$\vert\ d_m\ \vert\ 'u\ \vert\ d'm\ \vert\ \delta m\ \vert$
24	22	$\frac{\frac{\sum_{k=}^{N} X_k\,Y_k - XY}{N}}{\sigma\ \sigma}=\gamma$	$\frac{\frac{\sum_{k=1}^{N} X_kY_k}{N}-\bar{X}\,\bar{Y}}{\sigma_x\sigma_y}=\gamma$

stran	vrsta	namesto	vstavi
26	3	$(4,12)^2$ 16,9744 4,92.4,26 20,9592	$\frac{(4,12)^2=16,9744}{4,92.4,26\ \ 20,9592}$
26	7	$x-\bar{x} = \frac{\sigma_x}{\sigma_y} r(x-\bar{x})$	$(x-\bar{x}) = \frac{\sigma_x}{\sigma_y} r(y-\bar{y})$
39	5	mora	more
39	19	enake	znake
40	31	za velik je	za velik V je
42	6	$\underline{4x2+12x2+4.0}$	$\frac{4.2+12.2+4.0}{20}$
45	29	$\frac{1,960 \quad 2,576}{2}$	$\frac{\infty \mid 1,960 \mid 2,576}{2}$
45	34	$f(x)=\frac{1}{\sqrt{2}} e^{-\frac{x^2}{2}}$	$f(x)=\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
47	7	$n = 1+\frac{t^2 S^2}{2}$	$n = 1 + \frac{t^2 . s^2}{E^2}$
47	29	$P(\frac{ni}{n}) = \binom{n}{ni} \pi_2^{ni} (1-\pi_i)^{n-ni}$	$P(\frac{ni}{n}) = \binom{n}{ni} \pi_i^{ni} (1-\pi_i)^{n-ni}$

Pravilna razmestitev koncem strani 43 in v začetku 44 je naslednja:

Namesto predzadnjega odstavka na strani 43 začni na strani 44 zgoraj!:

" Alo vstavimo...... do stavka :Ako vstavimo ta izraz v neenačbo (xx) dobimo:"(Ta odstavek je že na strani 43, kjer nadaljuješ.)

Zadnjemu odstavku [illegible] sledijo neenačbe na strani 44. (po odstavku, ki se prične:" S[illegible]....)