EduardasVakrina

MATEMATINĖS STATISTIKOS PRADMENYS. STATISTINIŲ DUOMENŲ ANALIZĖ NAUDOJANT MS **EXCEL**

METODINIAI NURODYMAI NEAKIVAIZDININKAMS

2007m

T u r i n y s

1 Įvadas

Šis metodinis darbas skirtas susipažinimui su MS EXCEL statistinių funkcijų panaudojimu atliekant paprasčiausią statistinę analizę.

Statistika (*lot. status* – buklė) reiškia: 1) kiekybinę masinių reiškinių apskaitą; 2) mokslą, kuris tiria kiekybinius pokyčius visuomenės ir ūkio vystymesi ir apdoroja tų tyrimų duomenis mokslo ir praktikos tikslams.

Jei reiškinius, stebimus įvairiose mokslo srityse (fizikoje, chemijoje , biologijoje, medicinoje) ar visuomenės gyvenime, vertinsime kaip tam tikrus eksperimentus, tai pastebėsime kad jų rezultatus lemia daugybė atsitiktinių faktorių, todėl eksperimento rezultatas paprastai yra atsitiktinis dydis arba įvykis. Tyrėjo uždavinys – už atsitiktinių svyravimų pamatyti priežastinio faktoriaus veikimą ir surasti dėsningumus.

Tikimybių teorijoje įvedama eilė svarbių sąvokų atsititinių įvykių ir atsitiktinių dydžių apibūdinimui: tikimybės, pasiskirstymo funkcijos, teorinio vidurkio, dispersijos, koreliacijos koeficiento, regresijos lygties ir kt. Praktikoje teorinius modelius konkrečioms tikimybinėms situacijoms galime priskirti tik remdamiesi eksperimentiniais duomenimis.

Matematinės statistikos turinį sudaro statistinių eksperimentų planavimas, statistinių duomenų grupavimas ir jų analizė. Čia taikomi tyrimo metodai gali būti bendri analizuojant įvairių mokslo sričių ir visuomeninių reiškinių dėsningumus.

2 Generalinė aibė ir imtis

Dažniausisi tenka spręsti tokius uždavinius: parenkama tiriamoji aibė, kurios objektai (elementai) turi vieną ar keletą tyrėją dominančių požymių. Pavyzdžiui, sociologą domina kandidatų į prezidento postą reitingai. Čia tiriamoji aibė – visi potencialūs rinkėjai; tyrėją dominantis požymis – nuomonė apie vieną ar kitą kandidatą. Energetikai planuoja pajamas, surenkamas iš daugiabučių namų gyventojų už komunalines paslaugas. Tiriamoji aibė – daugiabučių gyventojai, tiriamas požymis – gyventojų, laiku sumokančių komunalinius mokesčius, skaičius. Tiriamoji aibė - visi gamyklos vieno tipo gaminiai; požymis – gaminio atitikimas standartų reikalavimams, t.y., gaminio kokybė.

Statistinių tyrimų nagrinėjamų objektų aibė vadinama *generaline aibe (populiacija).*

Pilniausius tyrimo duomenis gautume, jei galėtume ištirti visus generalinės aibės elementus. Praktikoje dažniausisi tai padaryti neįmanoma (objektų labai daug; tyrimas susijęs su didžiulėmis lėšų ar laiko sąnaudomis, su tiriamojo objekto sunaikinimu ir kt.). Todėl dažniausiai tiriama tik aibės dalis, o apie visų aibės elementų savybes sprendžiama iš šios dalies savybių.

Generalinės aibės tiriamų objektų dalį vadiname *imtimi*. Imties elementų skaičių vadiname *imties tūriu*. Imties elementų tiriamo požymio reikšmes vadiname *duomenimis*.

Vienas iš svarbiausių reikalavimų – imtis turi būti *reprezentatyvi*, t.y., ji turi teisingai atspindėti tiriamo požymio galimų reikšmių proporcijas generalinėje aibėje. Būtent reprezentatyvumas lemia, ar ištyrus imtį galime daryti patikimas išvadas apie visą generalinę aibę. Ši sąlyga yra išpildyta, jei imtis sudaroma *atsitiktiniu būdu*, t.y., jei kiekvienas generalinės aibės elementas su vienoda tikimybe gali patekti į imtį. Reprezentatyvumas priklauso ir nuo imties dydžio..

Matematinės statistikos metodais nagrinėjant imties elementų *tiriamojo požymio X reikšmių seką*, sudaromas *empyrinis* (statistinis, imties) tiriamojo požymio pasiskirstymas, apskaičiuojamos empyrinio pasiskirstymo skaitinės charakteristikos. Dėl imties atsitiktinumo šios charakteristikos yra atsitiktiniai dydžiai, tuo besiskiriantys nuo tikrųjų generalinės aibės pasiskirstymo skaitinių charakteristikų.

Pagrindiniai matematinės statistikos uždaviniai yra:

1) statistinių duomenų grupavimas;

2) nežinomų teorinio pasiskirstymo parametrų taškinių ir intervalinių įverčių radimas;

3) hipotezių apie teorinį pasiskirstymą ir jo parametrus tikrinimas;

4) regresinė ir koreliacinė analizė, leidžianti tirti priklausomybės tarp atsitiktinių dydžių pobūdį ir stiprumą.

3 Duomenų grupavimas

Į vienos generalinės aibės elementų požymio X n stebėjimų rezultatus galime žiūrėti kaip į n nepriklausomų vienodai pasiskirsčiusių atsitiktinių dydžių arba kaip į vieno atsitiktinio dydžio X n nepriklausomų reikšmių ir jas nagrinėti jų pasirodymo tvarka, pagal jų didumą arba atsitiktine tvarka. Stebėjimų rezultatai $x_1,\ x_2,\ x_3,\ \dots,\ x_i,\ x_{i+1},\ \dots,\ x_n$ paprastai taip pat vadinami imtimi.

Tarkime, kad tiriant generalinės aibės požymį X, gauta imtis

$$x_1,\ x_2,\ x_3,\ \dots,\ x_i,\ x_{i+1},\ \dots,\ x_n\ .$$

Kai kurios stebėtos reikšmės gali būti vienodos, tarkime x_1 pasikartoja n_1 kartą, x_2, - n_2 kartų, … , x_k - n_k kartų , čia

$$n_1 + n_2 + \dots + n_k = n.$$

Skaičius n_1 , n_2 , ..., n_k vadiname reikšmių x_i *dažniais*, o santykius $v_i = \frac{n_i}{n}$ (i =1, 2,…, k) - *santykiniais dažniais*

Reikšmes $x_1,\ x_2,\ x_3,\ \dots,\ x_k$ išdėstome didėjimo tvarka ir sudarome lentelę

X	x_1	x_2	. . .	x_k
n_i	n_1	n_2	. . .	n_k

Šią lentelę vadiname *variacine eilute.* Galime sudaryti lentelę, kurios pirmojoje eilutėje yra imties $x_1,\ x_2,\ x_3,\ \dots,\ x_k$ reikšmės o antrojoje – šių reikšmių santykiniai dažniai $v_1 = \frac{n_1}{n}$, $v_2 = \frac{n_2}{n}$, … , $v_k = \frac{n_k}{n}$:

X	x_1	x_2	. . .	x_k
$v_i = \frac{n_i}{n}$	$v_1 = \frac{n_1}{n}$	$v_2 = \frac{n_2}{n}$	. . .	$v_k = \frac{n_k}{n}$

čia

$$v_1 + v_2 + ... + v_k = 1.$$

Gausime požymio X *empirinį skirstinį*, arba *statistinę eilutę.*

Esant tolydiems dydžiams arba didelėms imtims, variacinėje eilutėje vietoj variantų $x_1, x_2, x_3, ..., x_k$ rašomi intervalai. Jei visos požymio X stebėtos reikšmės patenka į intervalą $[a; b]$, čia a yra mažiausia imties reikšmė, o b didžiausia imties reikšmė, tai šį intervalą taškais $a = a_o < a_1 < a_2 < ... < a_{k-1} < a_k = b$ padaliname į k lygių dalių. Dalijimo intervalo ilgis $h = \dfrac{b-a}{k}$ (kad žingsnis būtų patogesnis skaičius, kartais reikšmę a truputį sumažinam, o reikšmę b – padidinam).

Tarkime, n_i yra skaičius imties reikšmių, priklausančių intervalui $[a_{i-1}; a_i)$, $i = 1, 2, ..., k$. Sudarome intervalinę statistinę eilutę .

Intervalai	Dažniai n_i	Santykiniai dažniai $v_i = n_i / n$	v_i / h
$[a_0; a_1)$	n_1	v_1	v_1 / h
$[a_1; a_2)$	n_2	v_2	v_2 / h
. . .	. . .	. . .	
$[a_{k-1}; a_k]$	n_k	v_k	v_k / h
Σ	n	1	$1 / h$

Pastaba. Paprastai sudaromi 5 – 6, iki 10 intervalų. Didesnį intervalų skaičių imti netikslinga, nes labai padidėja tyrimo sąnaudos, o gaunamos informacijos patikimumas padidėja nežymiai.

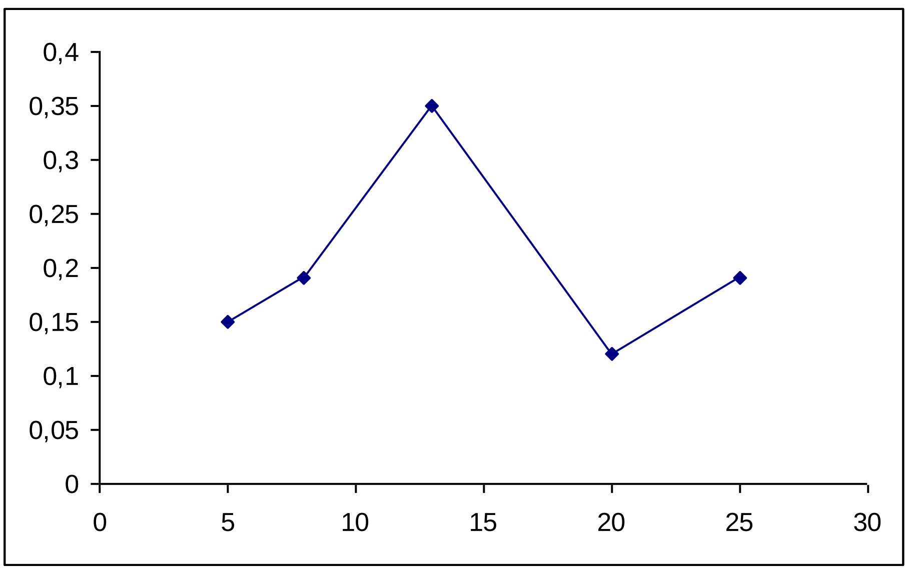
Empirinį skirstinį grafiškai galime pavaizduoti daugiakampiu. Abscisų ašyje atidedame X reikšmes $x_1, x_2, x_3, ..., x_k$, o ordinačių ašyje – atitinkamas santykinių dažnių reikšmes $v_1, v_2, v_3, ..., v_k$. Sujungę gautus plokštumos taškus atkarpomis, turėsime *empirinio skirstinio santykinių dažnių daugiakampį.*

Norėdami grafiškai pavaizduoti intervalinę statistinę eilutę, abscisų ašyje atidedame kiekvieno intervalo vidurio taškus, o ordinačių ašyje - dažnius n_i arba santykinius dažnius ν_i , $i = 1, 2, \ldots, k$. Gautus plokštumos taškus sujungiame laužtine linija.

Dažniausiai intervalinės eilutės vaizduojamos *histogramomis*. Histograma sudaroma iš stačiakampių, kurių pagrindai – intervalai $[\,a_{i-1}\,;\,a_i\,)$, $i = 1, 2, \ldots, k$, o aukštinės - n_i arba $\nu_i\ /\ h$, $i = 1, 2, \ldots, k$. Pirmuoju atveju gaunama dažnių histograma (visas jos ribojamas plotas lygus n), antruoju – santykinių dažnių histograma (visas jos ribojamas plotas lygus 1). Santykinių dažnių histograma yra tolydaus atsitiktinio dydžio tankio funkcijos grafiko statistinis analogas.

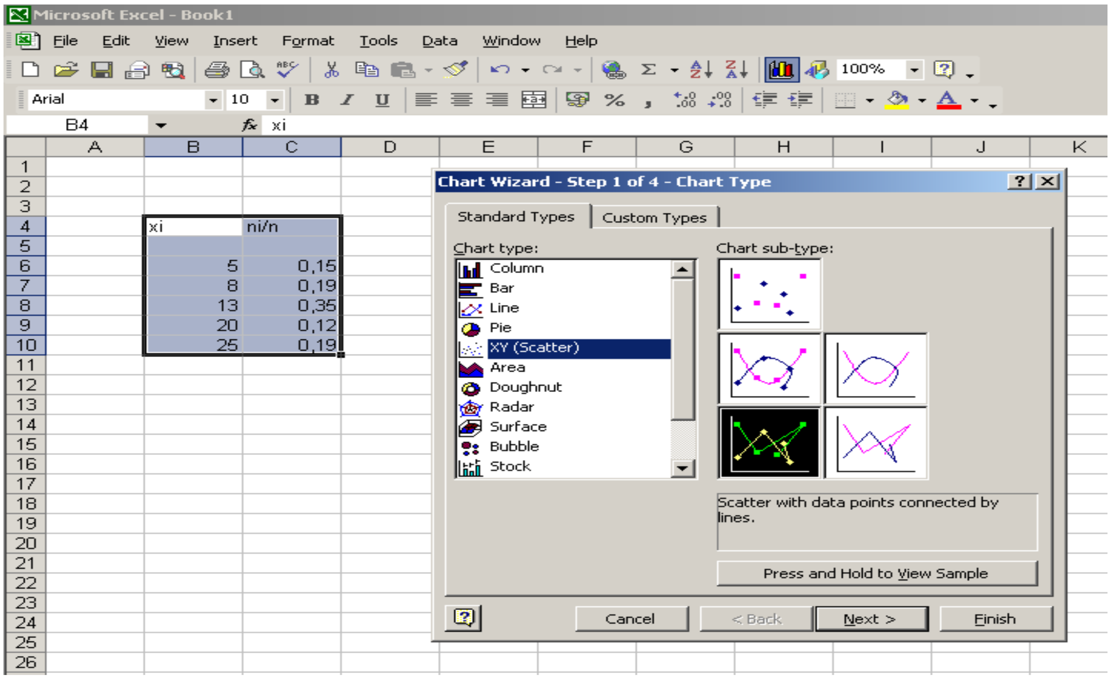
1 pavyzdys. Atsitiktinio dydžio X imties reikšmės x_i, tų reikšmių dažniai n_i ir santykiniai dažniai $\nu_i = \dfrac{n_i}{n}$ duoti lentelėje:

x_i	5	8	13	20	25
n_i	4	5	9	3	5
ν_i	0,15	0,19	0,35	0,12	0,19



1 pav.

Šį grafiką galime gauti naudodami MS EXCEL. Lentelėje įvedame statistinės eilutės duomenis, t. y. imties reikšmes ir santykinius dažnius. Tada **l**entelės viršuje esančioje simbolių eilutėje paspaudę simboliu  pažymėtą “klavišą”, iškviečiame langą **Chart Wizard ,** pasirenkame nuorodas, kurios pav.2 pažymėtos ir paspaudžiame klavišą **Finish.**



2 pav.

2 pavyzdys. Atsitiktinio dydžio *X* imties reikšmės x_i duotos lentelėje:

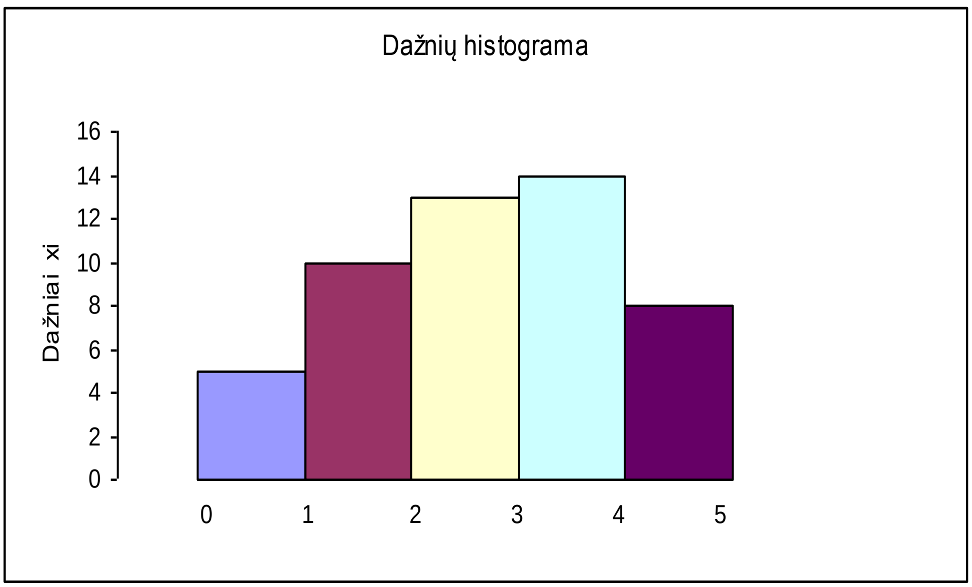
2,3	5,0	3,7	4,0	1,70	2,7	4,5	3,5	1,80	2,2
3,1	3,6	2,4	2,0	2,5	1,8	3,7	1,7	2,4	2,7
2,9	4,4	2,7	1,0	0,9	2,3	3,8	3,7	1,4	1,7
3,8	2,2	3,7	4,4	3,2	1,5	2,5	0,0	2,9	0,1
3,1	1,7	1,9	3,6	0,4	4,6	4,1	4,4	5,0	3,3

Kadangi imties tūris didelis (50 reikšmių), tai sudarysime intervalinę statistinę eilutę. Imties plotis yra 5,0 – 0,0 = 5. Visą imties plotį padalinsime į 5 intervalus, kurių ilgiai h = 1.

Intervalai	Dažniai n_i	Santykiniai dažniai	ν_i / h

		$v_i = n_i / n$	
[0; 1)	5	0,10	0,10
[1; 2)	10	0,20	0,20
[2; 3)	13.	0,26	0,26
[3; 4)	14	0,28	0,28
[4; 5]	8	0,16	0,16
Sumos	50	1,00	1,00

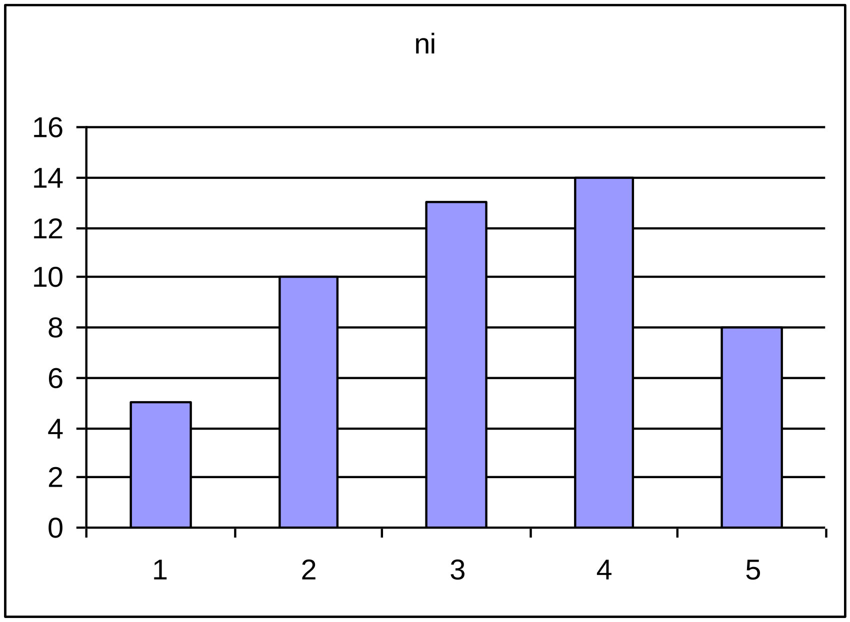
Žemiau patalpinta šios intervalinės eilutės dažnių histograma (3 pav.).



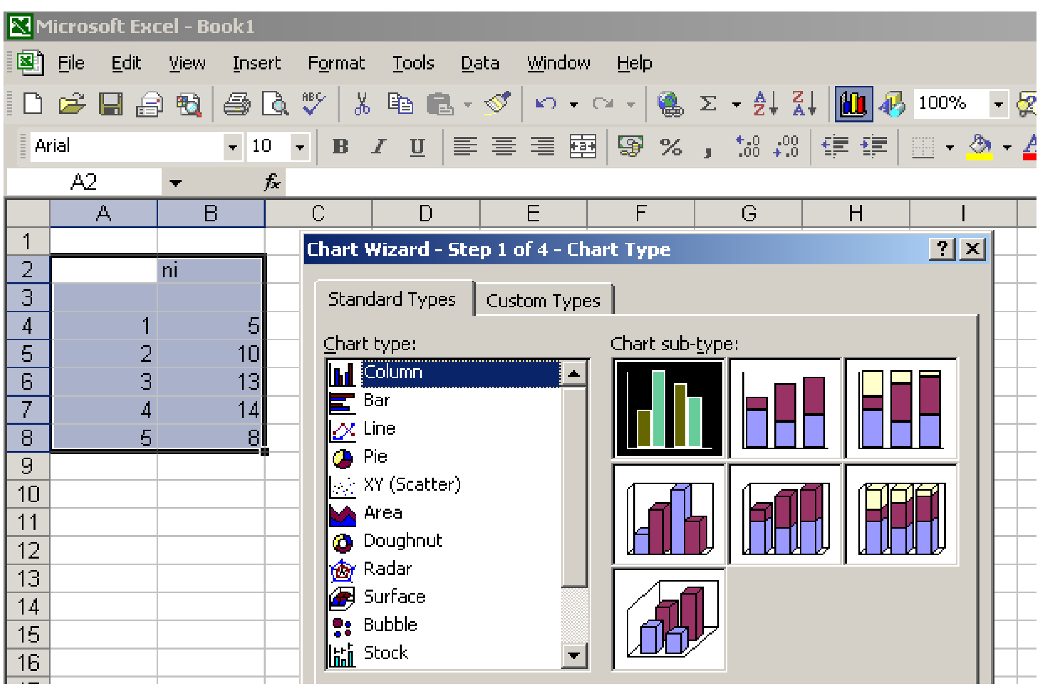
3 pav.

Šią diagramą galime gauti grafinės funkcijos **Chart Wizard** pagalba , į du EXCEL lentelės stulpelius įvedę intervalų dešiniųjų galų ir dažnių (arba santykinių dažnių) stulpelius ir juos pažymėję (užtamsinę), po to iškviestame **Chart Wizard** lange pasirinkę diagramos tipą, kurį matome 5 paveikslėlyje.

Turėdami 4 paveikslėlyje matomą diagramą, jos stulpelius galime suglaudinti aktyvavę **Series □Rows.** Tokiu būdu gausime 3 paveikslėlyje matomą histogramą**.**

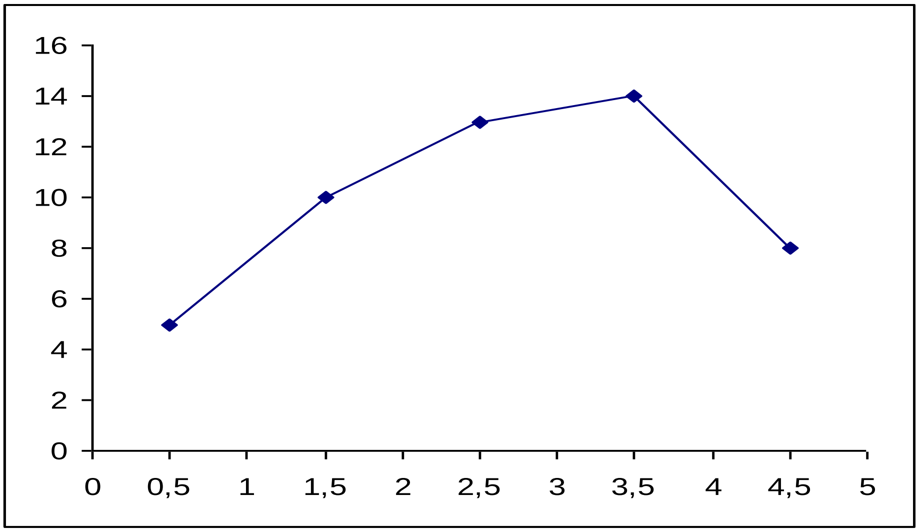


4 pav.



5 pav.

Intervalinės eilutės poligoną gausime laužtine linija sujungę taškus, kurių abscisės yra intervalų vidurio taškai, o ordinatės – tų intervalų reikšmių dažniai (6 pav.).



6 pav.

Analogiškai galime gauti santykinių dažnių histogramą ir poligoną (6 pav.)

4 Imties skaitinės charakteristikos

4.1 Imties vidurkis

Požymio *X empiriniu vidurkiu* $\overline{X}$ vadiname skaičių

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

arba

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i \,. \tag{2}$$

Pavyzdys.Tarkime, turime imtį 1; 3; 5; 11; 4; 2; 7; 6; 9. Empirinį vidurkį apskaičiuojame pagal formulę (A):

$$n = 9, \qquad \sum_{i=1}^{n} x_i = 1 + 3 + 5 + 11 + 4 + 2 + 7 + 6 + 9 = 48.$$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{9} . 48 = 5{,}33.$$

Pavyzdys. Požymio *X* variacinė eilutė tokia:

x_i	1	3	4	5	7	8
n_i	2	1	2	3	1	4

Empirinį vidurkį apskaičiuojame pagal formulę (2):

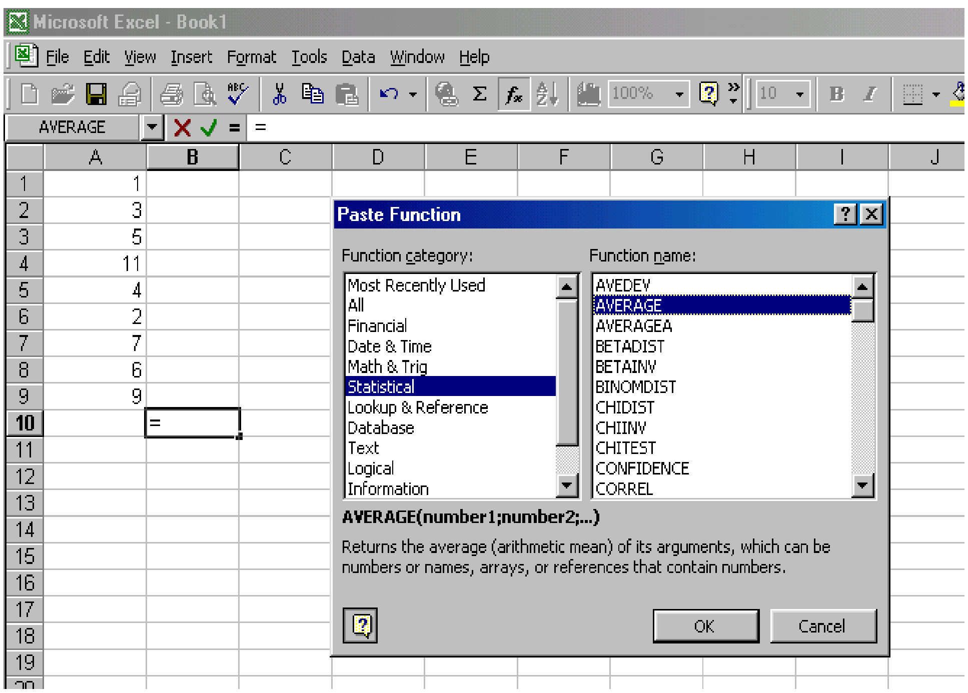
$$n = \sum_{i=1}^{k} n_i = 2 + 1 + 2 + 3 + 1 + 4 = 13;$$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i = \frac{1}{13}(1.2 + 3.1 + 4.2 + 5.3 + 7.1 + 8.4) = \frac{67}{13} = 5{,}15.$$

4.2 *Imties vidurkio radimas naudojant MS Excel*

Vidurkio radimo būdą pailiustruosime pavyzdžiu. Tarkime, turime imtį 1; 3; 5; 11; 4; 2; 7; 6; 9. Šios imties reikšmes *Excel* lentelėje patalpiname į kurio nors stulpelio (eilutės) langelius, suformuodami skaičių masyvą, pvz., A1:A9, ir pažymime langelį (mūsų pavyzdyje B10), kuriame norime gauti ieškomąjį rezultatą.

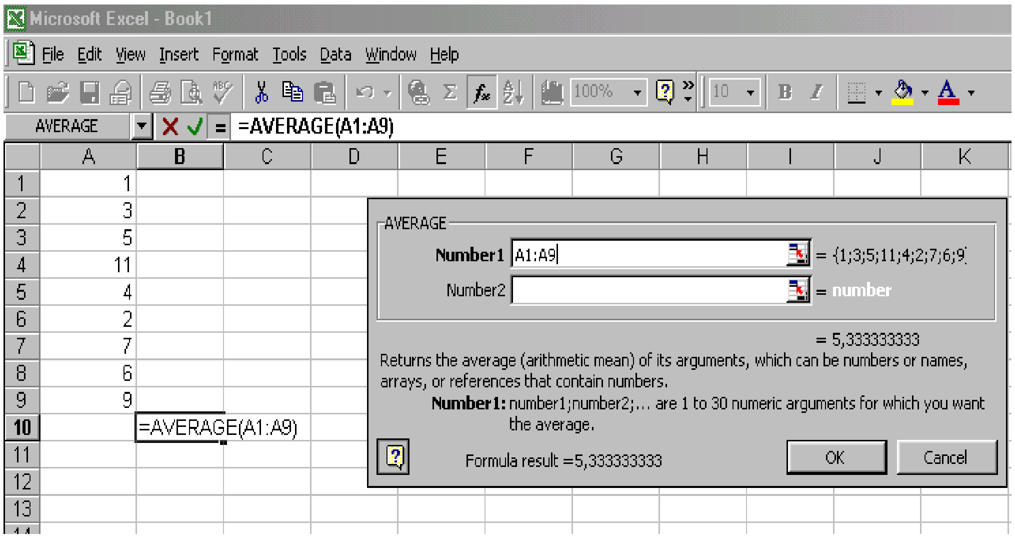
Lentelės viršuje esančioje simbolių eilutėje paspaudę simboliu  pažymėtą “klavišą”, iškviečiame langą **Paste Function** (7 pav.):



7 pav.

Kairėje lango dalyje stulpelyje **Function category** pažymime eilutę **Statistical,** dešiniajame **Function name** stulpelyje pažymime funkciją AVERAGE (vidurkis). Paspaudę OK, ekrane matome langą AVERAGE, kuriame, į langelį **Number 1** įrašius masyvo pavadinimą A1:A9 (EXCEL lange pažymėjus (užtamsinus) imties reikšmių stulpelį ir žymeklį nuvedus į AVERAGE lango **Number 1** langelį, jame atsiranda masyvo pavadinimas), iškart gauname vidurkio reikšmę 5,333333333 (žiūr.8 pav.).

Paspaudus OK, langas išnyks, o vidurkio reikšmė atsiras anksčiau pažymėtame langelyje B10.



8 pav.

Tą patį rezultatą turėsime, jei langelyje **Number 1** išvardinsime visus imties elementus, atskirdami juos vieną nuo kito taško kablelio ženklu. Šiuo atveju imties masyvą preliminariai įvesti į *Excel* lentelę nereikia.

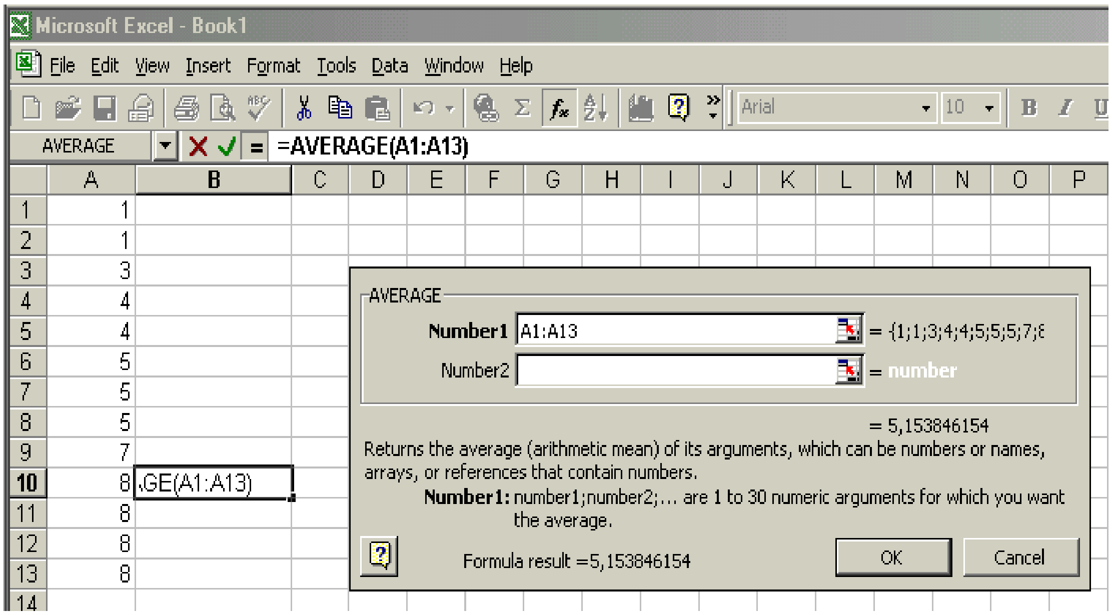
Imties vidurkį galime surasti nenaudodami AVERAGE lango, o tiesiog langelyje prie funkcijos simbolio  įvesdami komandą **=AVERAGE(A1:A9)**

Pavyzdys. Požymio X empirinis skirstinys duotas variacine eilute:

x_i	1	3	4	5	7	8
n_i	2	1	2	3	1	4

Rasime empirinį vidurkį.

Excel lentelėje į kurio nors stulpelio (eilutės) langelius patalpiname *visas* imties reikšmes, pakartodami jas tiek kartų, kiek nurodyta variacinės eilutės dažnių eilutėje: 1; 1; 3; 4; 4; 5; 5; 5; 7; 8; 8; 8; 8, ir suformuojame skaičių masyvą, pvz., A1:A13. Pažymime langelį (mūsų pavyzdyje B10), kuriame norime gauti ieškomąjį rezultatą. Paspaudę OK, B10 langelyje gausime vidurkio reikšmę 5,153846154.



9 pav.

Pastaba. Kai variacinėje eilutėje n_i eikšmės didelės, aukščiau aprašytas vidurkio radimo būdas nepatogus. Patogiau būtų skaičiavimus atlikti tiesiogiai *Excel* lentelėje:

x_i	n_i	x_i*n_i
1	2	2
3	1	3
4	2	8
5	3	15
7	1	7
8	4	32
Suma:	13	67
$\bar{X}$ =	67:13 =	5,153846

(Darbas su *Excel* lentelėmis aptariamas visose su šia programa supažindinančiose knygose.)

4.3 Imties dispersija

Požymio X *empirine dispersija* vadiname skaičių

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 \tag{3}$$

arba

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{X})^2 n_i \tag{4}$$

Formulėje (C), išskleidę $(x_i - \overline{X})^2$ ir pasinaudoję vidurkio $\overline{X}$ apibrėžimu, lengvai gauname patogesnę praktiniam skaičiavimui formulę:

$$S^2 = \overline{X^2} - \overline{X}^2 \tag{5}$$

Pavyzdys.Tarkime, turime imtį 1; 3; 5; 11; 4; 2; 7; 6; 9. Apskaičiuosime imties dispersiją. Empirinį vidurkį $\overline{X}$ apskaičiuojame pagal formulę (1):

$$n = 9, \qquad \sum_{i=1}^{n} x_i = 48.$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{9}.\ 48 = 5{,}33.$$

$$\sum_{i=1}^{n} x_i^2 = 342,$$

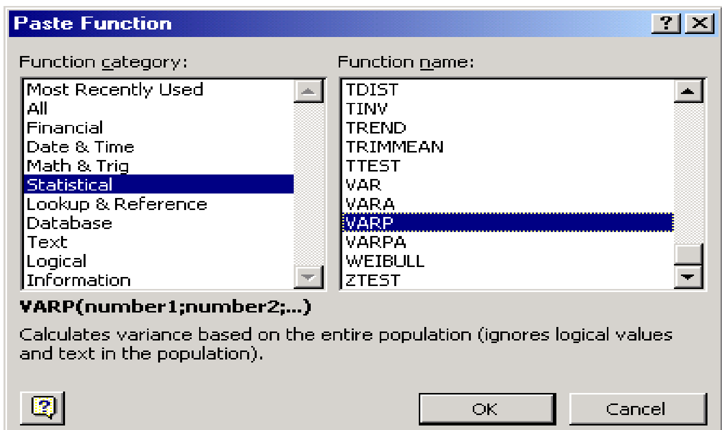
$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{9}.\ 342 = 38;$$

$$S^2 = \overline{X^2} - \overline{X}^2 = 38 - 5{,}33^2 = 9{,}5555\ldots$$

Įrašius duotosios imties reikšmių masyvą A1:A9, iškart gauname nuokrypių nuo vidurkio kvadratų sumą 86 . Paspaudus OK, langas išnyks, o minėta suma atsiras anksčiau pažymėtame langelyje.

4.4 *Imties dispersijos skaičiavimas naudojant MS Excel*

Kaip ir anksčiau aptartais atvejais, imties masyvas užrašomas *Excel* lentelėje ir iškviečiamas langas **Paste Function** , kuriame pasirenkama **Statistical** →VARP (10 pav.):



10 pav.

Su atidarytu VARP langu elgiamės taip pat, kaip ir skaičiuojant imties vidurkį $\overline{X}$ ir nuokrypį .

Imties dispersiją galime surasti nenaudodami VARP lango, o tiesiog langelyje prie funkcijos simbolio  įvesdami komandą **=VARP(**1;3;5;11;4;2;7;6;9**)**

<u>*Pastaba*</u>.Imties dispersiją S^2 padauginę iš $\frac{n}{n-1}$, čia n – imties tūris, gauname pataisytą imties dispersiją $S_1^2 = \frac{n}{n-1} S^2$

Pavyzdys. Požymio X empirinis skirstinys duotas variacine eilute:

x_i	1	3	4	5	7	8
n_i	2	1	2	3	1	4

Rasime empirinę dispersiją, panaudodami *Excel* lentelę.

	n_i	$\overline{X}$	$(x_i - \overline{X})$	$(x_i - \overline{X})^2$	$(x_i - \overline{X})^2 n_i$
1	2	5,153846	-4,153846	17,25443659	34,50887318
3	1	5,153846	-2,153846	4,639052592	4,639052592
4	2	5,153846	-1,153846	1,331360592	2,662721183
5	3	5,153846	-0,153846	0,023668592	0,071005775
7	1	5,153846	1,846154	3,408284592	3,408284592
8	4	5,153846	2,846154	8,100592592	32,40237037
Sumos:	13			34,75739555	77,69230769

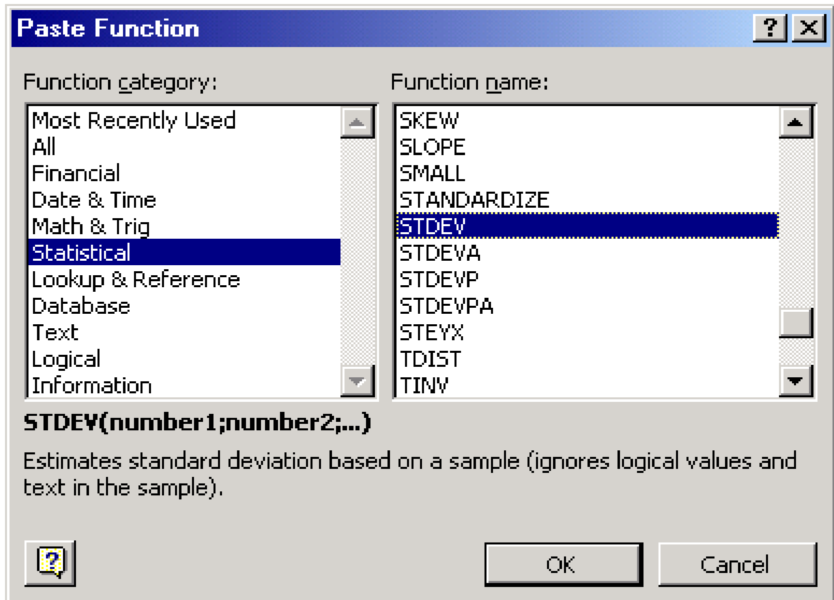
$$\boxed{S^2 = \quad 77{,}6923 :13 = \quad 5{,}976331361}$$

4.5 Pataisyto imties vidutinio kvadratinio nuokrypio radimas

Empirinės imties S^2 *vidutiniu kvadratiniu nuokrypiu* vadiname kvadratinę šaknį iš empirinės dispersijos $\sqrt{S^2}$; *pataisytu kvadratiniu nuokrypiu* - kvadratinę šaknį iš pataisytos dispersijos $\sqrt{S_1^2}$.

$$S_1 = \sqrt{S_1^2} = \sqrt{\frac{n}{n-1} S^2}$$

Norėdami rasti $S_1 = \sqrt{S_1^2}$, imties masyvą užrašome *Excel* lentelėje ir iškviečiame langą **Paste Function .** **Statistical** kategorijoje pažymime funkciją STDEV (11 pav.):



11 pav.

Tolimesnė darbo eiga analogiška aukščiau aptartiems atvejams

4.6 Imties asimetrijos koeficientas

Centriniu empiriniu k-osios eilės momentu vadinamas

$$m_k = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^k ,$$

čia $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$ .yra imties vidurkis.

Imties asimetrijos koeficientas $g_1 = \dfrac{m_3}{S_1^{\,3}}$,

čia $S_1 = \sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{X})^2}$ yra imties standartinis nuokrypis (šaknis iš pataisytos dispersijos).

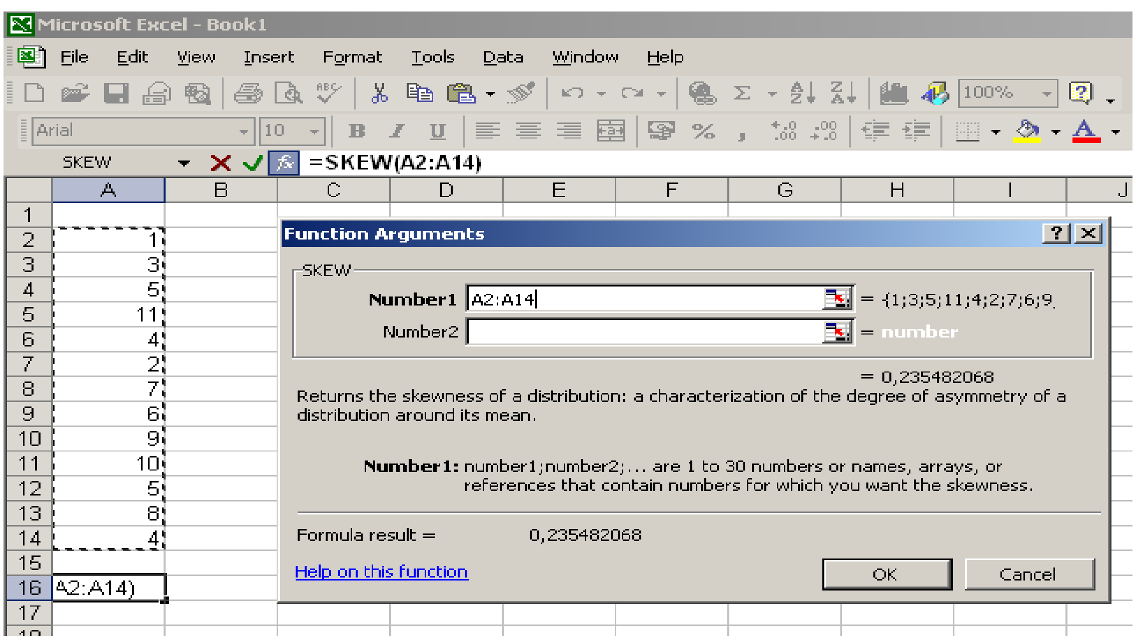
Asimetrijos koeficientas yra statistinių dažnių skirstinio simetrijos matas arba histogramos simetrijos matas. Histograma simetriška, kai $g_1 = 0$. Kai $g_1 < 0$, imties vidurkis $\overline{X}$ mažesnis už medianą . Kai $g_1 > 0$, imties vidurkis $\overline{X}$ yra didesnis už medianą .

Imties mediana yra skaičius , už kurį 50% variacinės eilutės narių yra ne didesnės ir 50% ne mažesnės, (Variacinė eilutė yra imties reikšmės išdėstytos nemažėjimo tvarka).

Asimetrijos koeficiento radimą naudojant MS EXCEL pailiustruosime pavyzdžiu. Tarkime, turime imtį 1; 3; 5; 11; 4; 2; 7; 6; 9, 10, 5, 8, 4. Šios imties reikšmes *Excel* lentelėje patalpiname į kurio nors stulpelio (eilutės) langelius ir pažymime langelį (mūsų pavyzdyje B10), kuriame norime gauti ieškomąjį rezultatą.

Lentelės viršuje esančioje simbolių eilutėje paspaudę simboliu  pažymėtą “klavišą”, iškviečiame langą **Paste Function.** Stulpelyje **Function category** pažymime eilutę **Statistical,** dešiniajame **Function name** stulpelyje pažymime funkciją SKEW. Paspaudę OK, ekrane matome langą SKEW, žymeklį nuvedę į AVERAGE lango **Number 1** langelį EXCEL lange pažymėję (užtamsinę) imties reikšmių stulpelį, iškart gauname asimetrijos koeficiento reikšmę 0,235482 (žiūr.12 pav.). Paspaudus OK, langas išnyks, o asimetrijos koeficiento reikšmė atsiras anksčiau pažymėtame langelyje B16.

Imties asimetrijos koeficientą galime surasti nenaudodami **SKEW** lango, o tiesiog langelyje prie funkcijos simbolio  įvesdami komandą **=SKEW (A2:A14),** prieš tai pažymėję langelį, kuriame norime gauti asimetrijos koeficiento reikšmę.



12 pav.

4.7 *Imties eksceso koeficientas*

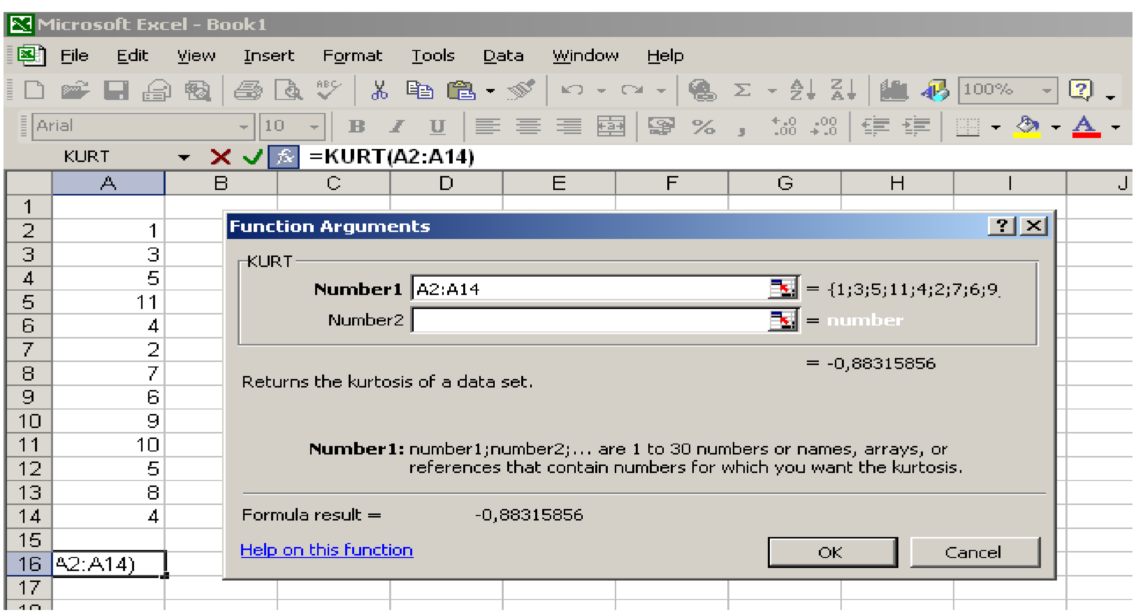
Imties eksceso koeficientas $g_2 = \dfrac{m_4}{S_1^{\,4}} - 3$ yra statistinio skirstinio histogramos lėkštumo (arba smailumo) matas. Jeigu $g_2 > 0$, histograma smaila , t.y. duomenų sklaida apie vidurkį mažesnė nei normaliosios (Gauso) kreivės. Jeigu $g_2 < 0$, histograma lėkšta, t.y. duomenų sklaida apie vidurkį didesnė nei normaliosios kreivės atveju.

Kai empiriniai asimetrijos ir eksceso koeficientai artimi nuliui, galima laikyti, kad histograma panaši į normalijo skirstinio tankio funkcijos grafiką.

Eksceso koeficiento radimą naudojant MS EXCEL pailiustruosime tuo pačiu pavyzdžiu. Apskaičiuosime imties 1; 3; 5; 11; 4; 2; 7; 6; 9, 10, 5, 8, 4 eksceso koeficientą.

Paste Function lange pasirenkame **Statistical** □ **KURT.** Funkcijos **KURT** lange žymeklį nuvedę į lango **Number 1** langelį ir EXCEL lange pažymėję (užtamsinę) imties reikšmių stulpelį, iškart gauname eksceso koeficiento reikšmę - 0,883159 (žiūr. 13 pav.). Paspaudus OK, langas išnyks, o eksceso koeficiento reikšmė atsiras anksčiau pažymėtame langelyje B16.

Imties eksceso koeficientą galime surasti nenaudodami **KURT** lango, o tiesiog langelyje prie funkcijos simbolio *fx* įvesdami komandą **=KURT (A2:A14),** prieš tai pažymėję langelį, kuriame norime gauti eksceso koeficiento reikšmę.



13 pav.

5 Nežinomų pasiskirstymo parametrų statistinis įvertinimas

5.1 Taškiniai įverčiai

Tarkime, tiriant generalinės aibės požymį X, sudaryta imtis $x_1,\ x_2,\ x_3,\ \ldots,\ x_i,\ x_{i+1},\ \ldots,\ x_n$.

Jeigu kai kurios imties reikšmės kartojasi, sudaroma variacinė eilutė

X	x_1	x_2	. . .	x_k
n_i	n_1	n_2	. . .	n_k

čia $\sum_{i=1}^{k} n_i = n$. Poligonas ir histograma yra atsitiktinio dydžio X, jei šis yra tolydusis, teorinės tankio funkcijos grafiko statistiniai analogai. Pagal poligono, histogramos formą ar kokių nors sudėtingesnių samprotavimų pagalba parenkamas hipotetinis požymio X skirstinys (tikimybinis pasiskirstymas)

Tiriant tolydžius atsitiktinius dydžius, matematiniu modeliu dažnai parenkamas normalusis pasiskirstymas. Šio pasiskirstymo funkcijų klase laikysime aibę funkcijų

$$F\,(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\,dt\,,$$

kurios viena nuo kitos skiriasi bent vienu parametru (a arba σ), arba abiem. Jeigu pasirinktas kitas, pavyzdžiui, Puasono pasiskirstymas, tai reikia įvertinti tik vieną parametrą λ.

Tarkime, bendru atveju, turime pasiskirstymo funkciją $F(\ x,\ \theta\ )$, čia θ – nežinomas parametras. Nagrinėsime šio parametro statistinį įvertinimą (taškinį įvertį) $\hat{\theta}$, kuris yra tam tikra imties reikšmių funkcija $\hat{\theta}(\ x_1,\ x_2,\ \ldots,\ x_n\ )$. Suprantama, kad paėmę kitą imtį, gausime kitą $\hat{\theta}(\ x_1,\ x_2,\ \ldots,\ x_n\ )$ reikšmę, todėl taškinis įvertis $\hat{\theta}$ yra atsitiktinis dydis. Vienos imties atveju turime vieną šio atsitiktinio dydžio realizaciją $\hat{\theta}$ ir ją vadiname nežinomo parametro θ taškiniu įverčiu.

“Geras” taškinis įvertis turi būti artimas tikrajai vertinamo parametro reikšmei, todėl jam keliami toki reikalavimai:

a) įvertis $\hat{\theta}\ (x_1,\ x_2,\ x_3,\ \dots,\ x_n)$ turi būti *pagrįstas,* t.y., jis turi konverguoti pagal tikimybę į vertinamą teorinę charakteristiką θ, kai stebėjimų skaičius neribotai didėja:

$$\lim_{n\to\infty} P\{|\ \hat{\theta}\ (x_1,\ x_2,\ x_3,\ \dots,\ x_n) - \theta| < \varepsilon\} = 1$$

esant bet kokiam teigiamam ε ; kitaip sakant, didėjant imčiai, įvertis turi būti tikslesnis;

b) įvertis $\hat{\theta}\ (x_1,\ x_2,\ x_3,\ \dots,\ x_n)$ turi būti *nepaslinktas,* t.y., jo teorinis vidurkis turi būti lygus vertinamai charakteristikai θ nepriklausomai nuo stebėjimų skaičiaus:

$$M[\hat{\theta}\ (x_1,\ x_2,\ x_3,\ \dots,\ x_n)] = \theta;$$

c) įvertis turi būti *efektyvus*, t.y., turėti mažiausią galimą dispersiją.

Pateiksime keletą taškinių įverčių **pavyzdžių**

Požymio *X empiriniu vidurkiu* $\overline{X}$ vadiname skaičių

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{6}$$

arba

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i\,. \tag{7}$$

Platesniame matematinės statistikos kurse įrodoma, kad normaliojo skirstinio atveju empirinis imties vidurkis yra suderintasis, nepaslinktas ir efektyvus nežinomo parametro (teorinio vidurkio) a įvertis. T. y. $\widehat{a} = \overline{X}$.

Požymio X *empirine dispersija* vadiname skaičių

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^2 \tag{8}$$

arba

$$S^2 = \frac{1}{n} \sum_{i=1}^{k} (x_i - \overline{X})^2 n_i = \overline{X^2} - \overline{X}^2, \tag{9}$$

čia $\overline{X^2}$ - atsitiktinio dydžio X kvadrato vidurkis, $\overline{X}^2$ - šio dydžio vidurkio kvadratas Įrodoma, kad normaliojo skirstinio atveju empirinė dispersija S^2 yra paslinktasis teorinės dispersijos σ^2 įvertis, todel dažnai empirinė dispersija S^2 pakeičiama nepaslinktuoju dispersijos D(X) įverčiu - pataisytąja imties dispersija

$$S_1^2 = \frac{n}{n-1} S^2$$

Taigi, geras parametro σ^2 taškinis įvertis yra $\sigma^2 = S_1^2$.

(Kai n didelis, skirtumas tarp S^2 ir S_1^2 praktiškai išnyksta.)

Atlikus n bandymų, įvykio A pasirodymų santykinis dažnis W(A) = $\frac{m}{n}$ yra suderintasis, nepaslinktas ir efektyvus binominio skirstinio parametro p įvertis: $\hat{p} = \frac{m}{n}$.

Čia n yra atliktų bandymų skaičius, o m skaičius bandymų, kurius atliekant įvykis A pasirodė (įvyko).

Eksponentinio pasiskirstymo atveju, dydis $1/\overline{X}$ yra suderintasis, nepaslinktas ir efektyvus parametro λ įvertis: $\hat{\lambda} = 1/\overline{X}$.

Empyrinis imties vidurkis yra taip pat geras taškinis įvertis $\hat{\lambda} = \overline{X}$ nežinomam teoriniam Puassono skirstinio vidurkiui M(X) = λ ir dispersijai D(X) = λ.

5.2 Pasikliautinųjų intervalų (intervalinių įverčių) radimas

Vietoje nežinomo pasiskirstymo parametro θ naudodami jo taškinį įvertį $\hat{\theta}(x_1, x_2, \ldots, x_n)$, niekada nežinome, kokio dydžio paklaidą darome, todėl daugeliu atvejų patogesnis yra *intervalinis* įvertis, apibrėžiantis intervalą, kuriame su tam tikra tikimybe yra parametro θ reikšmė.

Tarkime, kad pagal imties reikšmes surastas nežinomo parametro θ taškinis įvertis $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$. Įverčio $\hat{\theta}$ *patikimumu* (arba pasikliovimo lygmeniu) vadinsime nelygybės $|\theta - \hat{\theta}| < \varepsilon$ galiojimo tikimybę. Žymėsime

$$P\,(\,|\,\theta - \hat{\theta}\,| < \varepsilon\,) = 1 - \alpha$$

arba

$$P\,(\hat{\theta} - \varepsilon < \theta < \hat{\theta} + \varepsilon\,) = 1 - \alpha.$$

Intervalas $(\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon)$, su tikimybe $1 - \alpha$ uždengiantis nežinomą parametrą θ, vadinamas *pasikliautiniu intervalu*. Kuo mažesnis šio intervalo ilgis 2ε, tuo didesnis tikslumas.

ε dydis priklauso nuo imties tūrio ir nuo patikimumo, t. y., dydžiai ε, n ir $1 - \alpha$ yra tarpusavy susiję - žinodami du iš jų, galime surasti trečią.

Bendra pasikliautinojo intervalo sudarymo schema yra tokia:

a) iš generalinės aibės, kurios pasiskirstymo funkcija yra $F(x, \theta)$, sudaroma n tūrio imtis ir iš jos gaunamas nežinomo parametro θ taškinis įvertis $\hat{\theta}$;

b) sudaromas atsitiktinis dydis $Y(\theta)$, susietas su parametru θ ir turintis žinomą tankio funkciją $f(y, \theta)$;

c) parenkamas reikiamas patikimumas $1 - \alpha$ (paprastai 0,95 arba 0,99);

d) pasinaudojant Y pasiskirstymo tankiu, surandami du skaičiai c_1 ir c_2, tokie, kad galiotų lygybė

$$P\,(c_1 < Y(\theta) < c_2) = \int_{c_1}^{c_2} f(y,\theta)dy = 1 - \alpha\,.$$

Skaičiai c_1 ir c_2 paprastai parenkami taip, kad būtų teisingos lygybės

$$P\,(Y(\theta) < c_1) = \frac{\alpha}{2} \quad \text{ir} \quad P\,(Y(\theta) > c_2) = \frac{\alpha}{2}\,,$$

t.y., kad plotas, apribotas tankio funkcijos $f(y, \theta)$ grafiku iš viršaus, y ašimi iš apačios ir tiesėmis $y = c_1$, $y = c_2$ būtų lygus $1 - \alpha$, o plotai, esantys tiesės $y = c_1$ kairėje ir tiesės $y = c_2$ dešinėje kiekvienas būtų lygus $\frac{\alpha}{2}$.

5.2.1 Normaliai pasiskirsčiusio atsitiktinio dydžio X teorinio vidurkio a pasikliautinio intervalo radimas, kai žinomas σ

Tarkime, eksperimentas aprašomas atsitiktiniu dydžiu X, ir šio eksperimento dėsningumų analizei sudaromas normalusis modelis su pasiskirstymo funkcija

$$F(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt .$$

Laikysime, kad šiame modelyje σ žinomas, o a – nežinomas. Šio parametro nustatymui iš generalinės aibės paimta imtis $x_1, x_2, \dots, x_n$ ir surastas taškinis nežinomo teorinio vidurkio įvertis

$$\hat{a} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Į imtį $x_1, x_2, \dots, x_n$ galime žiūrėti kaip į n vienodai (normaliai) pasiskirsčiusių nepriklausomų atsitiktinių dydžių $X_1, X_2, \dots, X_n$ su pasiskirstymo funkcija $F(x, a, \sigma)$. Remiantis šiomis prielaidomis yra įrodyta, kad atsitiktinis dydis

$$u = \frac{\overline{X} - a}{\sigma / \sqrt{n}}$$

. yra pasiskirstęs pagal normalųjį dėsnį su parametrais 0 ir 1 (t.y. $N_0(0, 1)$).

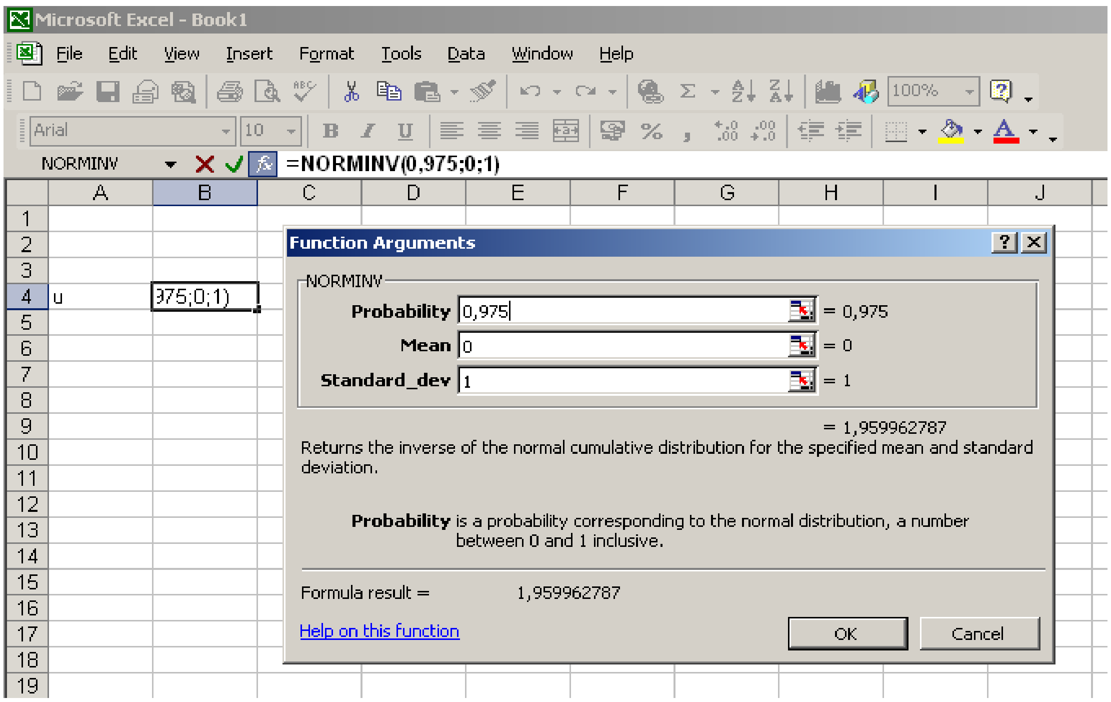
Tikimybė, kad šis dydis nukryps nuo savo teorinio vidurkio dydžiu $u_{\frac{\alpha}{2}}$ randama pagal formulę:

$$P(|u - M(u)| < u_{\frac{\alpha}{2}}) = P(|u| < u_{\frac{\alpha}{2}}) = P(-u_{\frac{\alpha}{2}} < \frac{\overline{X} - a}{\sigma / \sqrt{n}} < u_{\frac{\alpha}{2}}) =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-u_{\frac{\alpha}{2}}}^{u_{\frac{\alpha}{2}}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{u_{\frac{\alpha}{2}}} e^{-\frac{t^2}{2}} dt = \underline{2\Phi(u_{\frac{\alpha}{2}}) = 1 - \alpha} .$$

Iš pabrauktos lygybės, naudodami Laplaso funkcijos lenteles, surandame skirstinio N(0;1) kritinę reikšmę $u_{\frac{\alpha}{2}}$.

Kritinėms reikšmėms $u_{\frac{\alpha}{2}}$ surasti galime panaudoti EXCEL statistinę funkciją NORMINV. Atsidarę NORMINV langą **Probability** eilutėje turime įvesti tikimybę $1 - \frac{\alpha}{2} = P(u > u_{\frac{\alpha}{2}})$. Mūsų uždavinyje $1 - 0{,}025 = 0{,}975$



14 pav.

Kritinę reikšmę galime surasti nenaudodami NORMINV lango, o tiesiog langelyje prie funkcijos simbolio  įvesdami komandą **=NORMINV(0,975;0;1)**

Suradę reikalingą kritinę reikšmę, pertvarkome skliaustuose esančią nelygybę:

$$P(-u_{\frac{\alpha}{2}} < \frac{\overline{X} - a}{\sigma/\sqrt{n}} < u_{\frac{\alpha}{2}}) = P(-u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \overline{X} - a < u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) =$$

$$= P(\overline{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < a < \overline{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha = \gamma .$$

Gavome pasikliautinį intervalą nežinomam normaliai pasiskirsčiusio atsitiktinio dydžio teoriniam vidurkiui a :

$$\bar{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < a < \bar{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Pažymėkime $u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \delta$. Nesunku pastebėti, kad didėjant imties tūriui n, dydis δ mažėja, t.y., didėja įverčio tikslumas.

Didinant patikimumą $\gamma = 1 - \alpha = 2\Phi(u_{\frac{\alpha}{2}})$ didėja δ, nes $\Phi(u_{\frac{\alpha}{2}})$ – didėjanti funkcija, todėl mažėja įverčio tikslumas.

Pavyzdys. Normaliai pasiskirsčiusio atsitiktinio dydžio X empirinis vidurkis $\bar{X} = 5$, o vidutinis kvadratinis nuokrypis $\sigma = 3$. Su pasikliovimo lygmeniu $\gamma = 0.95$ raskime teorinio vidurkio (skirstinio parametro a) pasikliautinąjį intervalą kai $n = 36$.

Sprendimas. $\gamma = 1 - \alpha = 2\Phi(u_{\frac{\alpha}{2}}) = 0{,}95$. Aukčiau pateiktoje lentelėje arba EXCEL statistinės funkcijos pagalba **NORMINV** surandame: $u_{\frac{\alpha}{2}} = 1.96$, $u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} =$

$$1.96.\frac{3}{\sqrt{36}} = 1{,}96.0{,}5 = 0{,}98 \,,$$

todėl pasikliautinasi intervalas yra $(\bar{X} - 0{,}98 ; \bar{X} + 0{,}98)$, čia $\bar{X}$ - imties vidurkis.

Arba P(4,02< a <5,98)=0.95.

5.2.2 Normaliai pasiskirsčiusio atsitiktinio dydžio X teorinio vidurkio apasikliautinojo intervalo radimas, kai σ nežinomas

Sudaromas atsitiktinis dydis:

$$T = \frac{\bar{X} - a}{S_1}\sqrt{n}\,,$$

čia $\bar{X}$ - imties vidurkis, n – imties tūris, S_1 – “pataisytas” imties vidutinis kvadratinis nuokrypis, $S_1 = \sqrt{\frac{n}{n-1} S^2}$, $S_1 = S^2$ - imties dispersija.

Platesniame matematinės statistikos kurse parodoma, kad atsitiktinis dydis T yra pasiskirstęs pagal *Stjudento dėsnį su (n – 1) laisvės laipsniu* ir patogus tuo , kad priklauso nuo vienintelio parametro n – imties tūrio, t.y., nepriklauso nei nuo vidurkio a, nei nuo vidutinio kvadratinio nuokrypio σ.

Iš lygybės

$$P\left(\left| \frac{\bar{X}-a}{S_1}\sqrt{n} \right| < t_{\frac{\alpha}{2};n-1} \right) = \gamma = 1-\alpha$$

gauname:

$$P\left(-t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} < \bar{X}-a < t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} \right) = \gamma$$

arba

$$P\left(\bar{X} - t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} < a < \bar{X} + t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} \right) = \gamma .$$

Iš šios lygybės turime, kad $\left(\bar{X} - t_{\frac{\alpha}{2};n-1\,\gamma}\frac{S_1}{\sqrt{n}} < a < \bar{X} + t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} \right)$ yra intervalas, su patikimumu γ dengiantis teorinį vidurkį a.

Žinodami $\gamma = 1 - \alpha$ ir n, $t_{\frac{\alpha}{2};n-1}$ reikšmes surandame iš Stjudento skirstinio kritinių reikšmių lentelių.

Pavyzdys. Iš normaliai pasiskirsčiusio atsitiktinio dydžio X $n = 16$ tūrio imties surastas imties vidurkis $\bar{X} = 20.2$ ir imties (empirinė) dispersija $S^2 = 0,6$. Raskime teorinio vidurkio a pasikliautinąjį intervalą kai $\gamma = 0,95$.

Sprendimas. Surandame $S_1 = \sqrt{\frac{n}{n-1}S^2} = \sqrt{\frac{16.0,6}{15}} = 0,8$. Iš Stjudento skirstinio lentelių, kai $\gamma = 0,95$ $(1-\alpha = 0,05)$ ir $n - 1 = 15$, randame $t_{\frac{\alpha}{2};n-1} = 2,13$.

Tada

$$\bar{X} - t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} = 20,2 - 2,13.\frac{0,8}{\sqrt{16}} = 20,2 - 0,426 = 19.774;$$

$$\bar{X} + t_{\frac{\alpha}{2};n-1}\frac{S_1}{\sqrt{n}} = 20,2 + 2.13.\frac{0,8}{\sqrt{16}} = 20,2 + 0,426 = 20.626.$$

Gavome, kad su patikimumu 0,95 intervalas (19,774; 20,626) dengia atsitiktinio dydžio X teorinį vidurkį a.

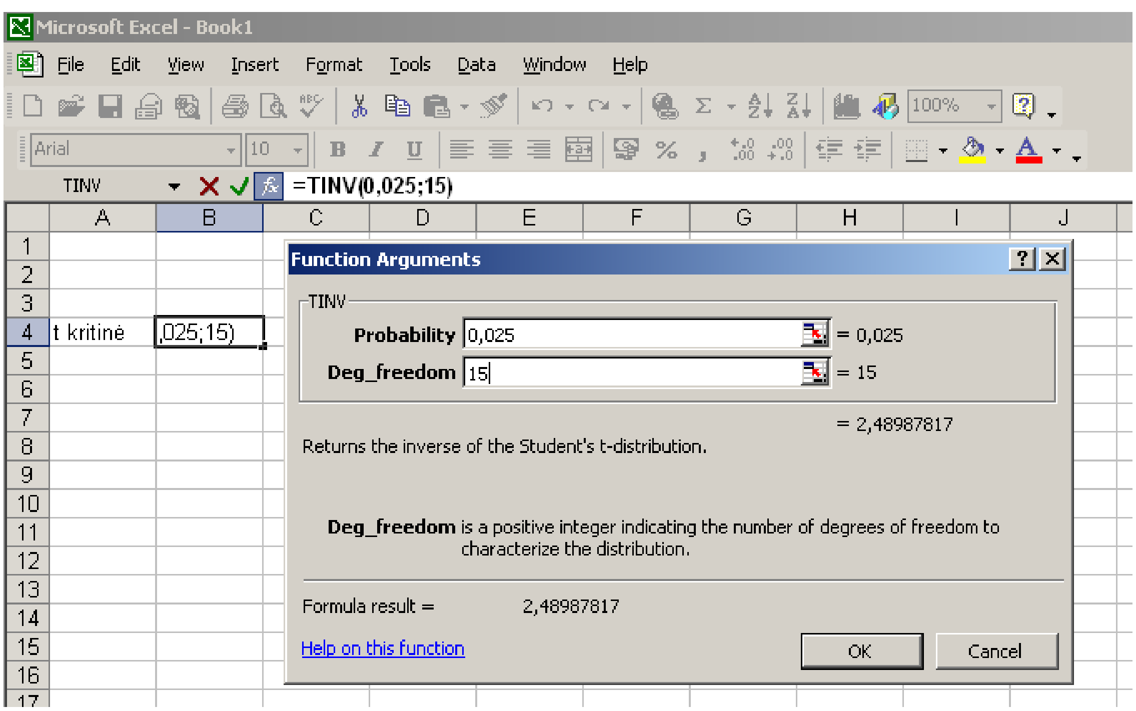
Pastaba. Pasinaudodami lygybe $S_1 = \sqrt{\frac{n}{n-1} S^2}$ galime parašyti:

$$t_{\frac{\alpha}{2};n-1} \frac{S_1}{\sqrt{n}} = t_{\frac{\alpha}{2};n-1} \frac{\sqrt{\frac{n}{n-1} S^2}}{\sqrt{n}} = t_{\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n-1}},$$

čia S = $\sqrt{S^2} = \sqrt{\overline{X^2} - \overline{X}^2}$ = "nepataisytas" imties vidutinis kvadratinis nuokrypis.

Kritinėms reikšmėms $t_{\frac{\alpha}{2};n-1}$ surasti galime panaudoti EXCEL programos statistinę funkciją

TINV. Atsidarę TINV langą **Probability** eilutėje turime įvesti tikimybę $1 - \frac{\alpha}{2} = P$ $(u > u_{\frac{\alpha}{2}})$. Mūsų uždavinyje 1 - 0,025 = 0,975. **Deg_freedom** langelyje - laisvės laipsnių skaičių



15 pav.

Kritinę reikšmę $t_{\frac{\alpha}{2};n-1}$ galime surasti nenaudodami TINV lango, o tiesiog langelyje prie funkcijos simbolio įvesdami komandą **=TINV(0,975;15)**

5.2.3 Pasikliautinasis intervalas normaliai pasiskirsčiusio atsitiktinio dydžio X vidutiniam kvadratiniam nuokrypiui σ

Atsitiktinis dydis

$$\chi^2 = \frac{(n-1)S_1^2}{\sigma^2}$$

yra pasiskirstęs pagal χ^2 dėsnį su n-1 laisvės laipsniu, todėl teisinga lygybė;

$$\mathrm{P}\left(\chi_{1-\frac{\alpha}{2};n-1} < \frac{(n-1)S_1^2}{\sigma^2} < \chi_{\frac{\alpha}{2};n-1} \right) = \gamma = 1-\alpha ,$$

kurioje $\chi_{1-\frac{\alpha}{2};n-1}$ ir $\chi_{\frac{\alpha}{2};n-1}$ yra χ^2 skirstinio kritinės reikšmės $1-\frac{\alpha}{2}$ ir $\frac{\alpha}{2}$ eilės atitinkamai.

Pertvarkę skliaustuose esančias nelygybes, gauname pasikliautinąjį intervalą normaliojo atsitiktinio dydžio dispersijai σ^2:

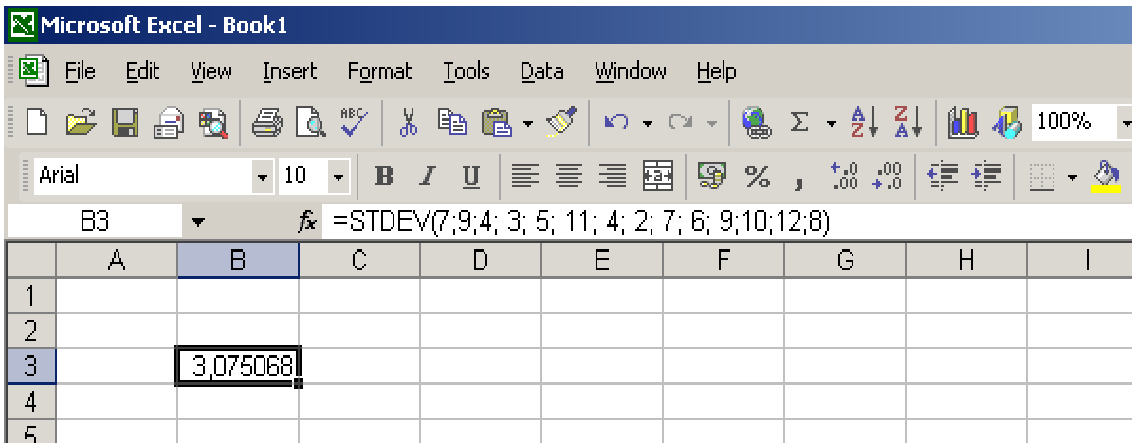
$$\mathrm{P}\left(S_1^2 \frac{n-1}{\chi_{\frac{\alpha}{2};n-1}} < \sigma^2 < S_1^2 \frac{n-1}{\chi_{1-\frac{\alpha}{2};n-1}} \right) = \gamma = 1-\alpha$$

Ištraukę kvadratinę šaknį iš visų skliaustuose esančias nelygybes sudarančių reiškinių, turėsime pasikliautinąjį intervalą normaliojo skirstinio parametrui σ.

$$\mathrm{P}\left(S_1 \sqrt{\frac{n-1}{\chi_{\frac{\alpha}{2};n-1}}} < \sigma < S_1 \sqrt{\frac{n-1}{\chi_{1-\frac{\alpha}{2};n-1}}} \right) = \gamma = 1-\alpha$$

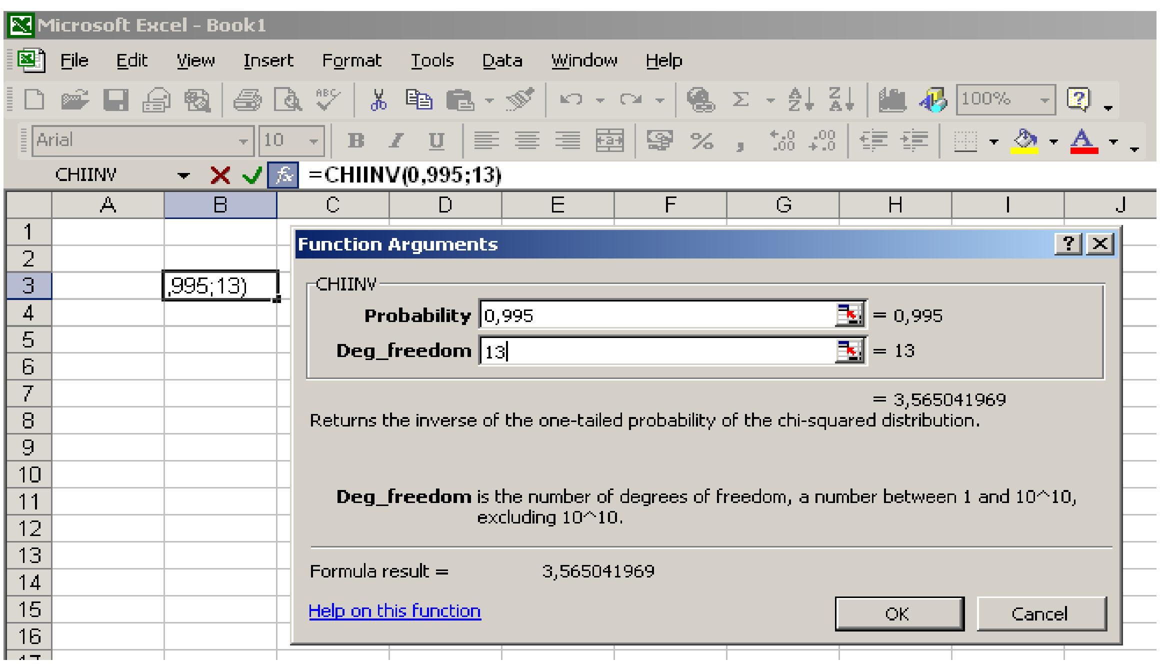
Pavyzdys.Tarkime, turime imtį 7;9;4; 3; 5; 11; 4; 2; 7; 6; 9;10;12;8 paimtą iš normaliosios generalinės aibės. Su pasikliovimo lygmeniu 0,99 suraskime pasikliautinąjį intervalą parametrui σ.

Pataisytam imties vidutiniam kvadratiniam nuokrypiui rasti pasinaudokime, kaip buvo paaiškinta aukščiau, EXCEL statistinę funkciją STDEV:



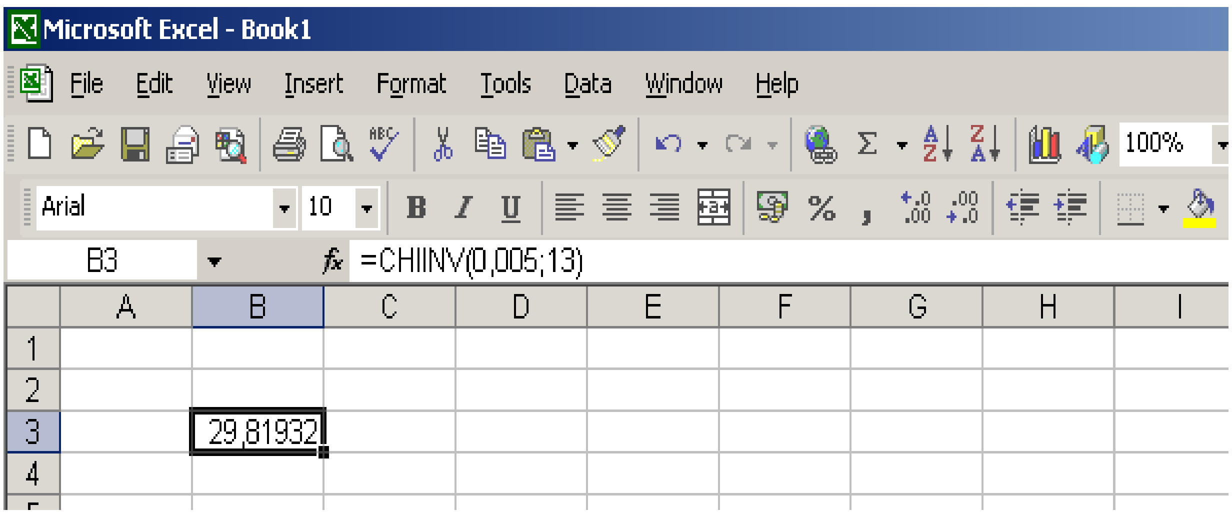
16 pav

χ^2 skirstinio kritinėms reikšmėms $\chi_{1-0.005;14-1}$ ir $\chi_{0.005;14-1}$ panaudosime EXCEL statistinę funkciją CHIINV.



17 pav.

Kritines reikšmes, pavyzdžiui $\chi_{0.005;14-1}$, galime surasti nenaudodami CHIINV lango, o tiesiog langelyje prie funkcijos simbolio f_x įvesdami komandą **=CHIINV(0,005;13)**



18 pav.

$S_1 = 3{,}075068$, $\chi_{1-0.005;14-1} = 3{,}56504$, $\chi_{0.005;14-1} = 29{,}81932$

$$P\left(3{,}075068\sqrt{\frac{13}{29{,}81932}} < \sigma < 3{,}075068\sqrt{\frac{13}{3{,}56504}}\right) = \gamma = 1 - 0{,}01$$

$$P(3{,}075068.0{,}6602718 < \sigma < 3{,}075068.1{,}909587) = \gamma = 1 - 0{,}01$$

$$P(2{,}03038 < \sigma < 5{,}87211) = 0{,}99.$$

6 Koreliacijos teorijos elementai

Paprasčiausia ryšio tarp dydžių forma yra funkcinė priklausomybė. Ji išreiškia tokį ryšį tarp dviejų kintamų dydžių, kai kiekvieną vieno iš jų reikšmę x atitinka viena griežtai apibrėžta kito dydžio y reikšmė:

$y = f(x)$.

Gamtos ir visuomenės reišiniuose funkciniai ryšiai sutinkami retai. Dažniau sutinkame ryšius tarp atsitiktinių dydžių, kai kiekvieną vieno kintamojo reikšmę atitinka ne viena , o kelios kito dydžio reikšmės.

Pavyzdžiai.

1.Gaminio savikaina susijusi su darbo našumu, bet ši atitiktis nėra griežta: savikainą sąlygoja ir eilė kitų faktorių, todėl esant tokiam pat darbo našumui, gaminių savikaina gali svyruoti, įgydama skirtingas skaitines reikšmes.

2.Derlius priklauso nuo trąšų kiekio, tačiau, esant tam pačiam išbertų trąšų kiekiui ir kokybei, derliai gali būti skirtingi.

3.Ryšys tarp ruošimosi egzaminui laiko ir egzamino įvertinimo.

Jei tarp dviejų atsitiktinių dydžių X ir Y egzistuoja toks ryšys, kad kiekvieną dydžio X reikšmę atitinka apibrėžtas dydžio Y skirstinys, dėsningai besikeičiantis kintant X reikšmei, tai tokį ryšį tarp X ir Y vadiname **statistiniu**.

Jei keičiantis vieam atsitiktiniam dydžiui keičiasi kito atsitiktinio dydžio v i d u r k i s, tai tokį s t a t i s t i n į ryšį vadiname **koreliaciniu**.

Tarkime, atsitiktinio dydžio X stebimos reikšmės yra $x_1, x_2, \ldots, x_k$, *o atsitiktinio dydžio Y stebimos reikšmės yra* $y_1, y_2, \ldots, y_n$. *Esant statistiniam ryšiui tarp X ir Y , kiekvieną dydžio X reikšmę* x_i *atitinka dydžio Y skirstinys*

Y	y_1	y_2	...	y_j	...	y_n
m	m_{i1}	m_{i2}		m_{ij}		m_{in}

(10)

$$\sum_{j=1}^{n} m_{ij} = m_{x_i}$$

arba kiekvieną dydžio Y reikšmę y_j atitinka dydžio X skirstinys

X	x_1	x_2	...	x_i	...	x_k
m	m_{1j}	m_{2j}	...	m_{ij}		m_{kj}

(11)

$$\sum_{i=1}^{k} m_{ij} = m_{y_j}$$

Taigi, stebėjimų rezultatus galime surašyti lentelėje:

X / Y	x_1	x_2	...	x_i	...	x_k	
y_1	m_{11}	m_{21}	...	m_{i1}	...	m_{k1}	$m_{y_1} = \sum_{i=1}^{k} m_{i1}$
y_2	m_{12}	m_{22}	...	m_{i2}	...	m_{k2}	$m_{y_2} = \sum_{i=1}^{k} m_{i2}$
...	...	...	...	...	...	...	...
y_j	m_{1j}	m_{2j}	...	m_{ij}	...	m_{kj}	$m_{y_j} = \sum_{i=1}^{k} m_{ij}$
...	...	...	...	...	...	...	...
y_n	m_{1n}	m_{2n}	...	m_{in}	...	m_{kn}	$m_{y_n} = \sum_{i=1}^{k} m_{in}$
	$m_{x_1} = \sum_{j=1}^{n} m_{1j}$	$m_{x_2} = \sum_{j=1}^{n} m_{2j}$	...	$m_{x_i} = \sum_{j=1}^{n} m_{ij}$	...	$m_{x_k} = \sum_{j=1}^{n} m_{kj}$	N

Šią lentelę vadiname *koreliacine lentele*. Ji yra statistinės priklausomybės tyrinėjimo pagrindas.

Lentelės analizė:

1. $x_1, x_2, \ldots, x_k$ - atsitiktinio dydžio X reikšmės;

$y_1, y_2, \ldots, y_n$ - atsitiktinio dydžio Y reikšmės.

2. Eilutės ir stulpelio susikirtime esantis skaičius m_{ij} parodo, kiek kartų stebėta reikšmių pora (x_i, y_j). m_{ij} vadinamas dažniu.

3. Paskutinėj eilutėj esantys skaičiai m_{x_1}, $m_{x_2}, \ldots, m_{x_k}$ parodo, kiek kartų visuose stebėjimuose pasirodė reikšmės $x_1, x_2, \ldots, x_k$ atitinkamai .

$$m_{x_i} = \sum_{j=1}^{n} m_{ij} = m_{i1} + m_{i2} + \ldots + m_{in} .$$

4.Paskutiniame stulpelyje skaičiai m_{y_1}, $m_{y_2}, \ldots, m_{y_n}$ parodo, kiek kartų visuose stebėjimuose pasirodė reikšmės $y_1, y_2, \ldots, y_n$ atitinkamai.

$$m_{y_j} = \sum_{i=1}^{k} m_{ij} = m_{1j} + m_{2j} + \ldots + m_{kj} .$$

5. Visų skaičių m_{y_j} suma lygi N; visų skaičių m_{x_i} suma lygi N, t.y.,

$\sum_{i=1}^{k} m_{x_i} = \sum_{j=1}^{n} m_{y_j}$ = N – visų stebėjimų skaičiui.

6. Statistiniai skirstiniai (10) ir (11) vadinami sąlyginiais atsitiktinio dydžio Y (dydžio X) skirstiniais, atitinkančiais X reikšmę x_i ($Y = y_j$).

7. Lentelės pirmoji ir paskutinė eilutės

X	x_1	x_2	...	x_i	...	x_k
m_X	$m_{x_1} = \sum_{j=1}^{n} m_{1j}$	$m_{x_2} = \sum_{j=1}^{n} m_{2j}$	...	$m_{x_i} = \sum_{j=1}^{n} m_{ij}$	...	$m_{x_k} = \sum_{j=1}^{n} m_{kj}$

sudaro požymio X besąlyginį pasiskirstymą; pirmasis ir paskutinis stulpeliai

Y	y_1	y_2	...	y_j	...	y_n
m_y	$m_{y_1} = \sum_{i=1}^{k} m_{i1}$	$m_{y_2} = \sum_{i=1}^{k} m_{i2}$	...	$m_{y_j} = \sum_{i=1}^{k} m_{ij}$	...	$m_{y_n} = \sum_{i=1}^{k} m_{in}$

sudaro požymio Y besąlyginį pasiskirstym

6.1 *Koreliacinio ryšio reiškimas regresijos lygtimi*

Tarkime, turime atsitiktinio dydžio Y sąlyginį pasiskirstymą, atitinkantį X reikšmę x_i:

Y	y_1	y_2	...	y_j	...	y_n
m_i	m_{i1}	m_{i2}		m_{ij}		m_{in}

Simboliu $\overline{Y}_{x_i}$ pažymėkime sąlyginį atsitiktinio dydžio Y vidurkį, atitinkantį atsitiktinio dydžio X reikšmę x_i:

$$\overline{Y}_{x_i} = \frac{m_{i1}y_1 + m_{i2}y_2 + ... + m_{in}y_n}{m_{i1} + m_{i2} + ... + m_{in}} = \frac{\sum_{j=1}^{n} m_{ij}y_j}{m_{x_i}}$$

Suradę sąlyginius Y vidurkius visoms X reikšmėms, gausime lentelę

X	x_1	x_2	...	x_i	...	x_k
$\overline{Y}_x$	$\overline{Y}_{x_1}$	$\overline{Y}_{x_2}$	$\cdots$	$\overline{Y}_{x_i}$	$\cdots$	$\overline{Y}_{x_k}$

Kiekvieną X reikšmę x atitinka pilnai apibrėžta sąlyginio vidurkio $\overline{Y}_x$ reikšmė, todėl $\overline{Y}_x$ yra reikšmių x funkcija, t.y.,

$$\overline{Y}_x = f(x) \qquad (12)$$

Analogiškai

$$\overline{X}_{y_j} = \frac{m_{1j}x_1 + ... + m_{ij}x_i + ... + m_{kj}x_k}{m_{1j} + ... + m_{ij} + ... + m_{kj}} = \frac{\sum_{i=1}^{k} m_{ij}x_i}{m_{y_j}}$$

ir sąlyginių vidurkių lentelė:

Y	y_1	y_2	...	y_j	...	y_n
$\overline{X}_y$	$\overline{X}_{y_1}$	$\overline{X}_{y_2}$	$\cdots$	$\overline{X}_{y_j}$	$\cdots$	$\overline{X}_{y_n}$

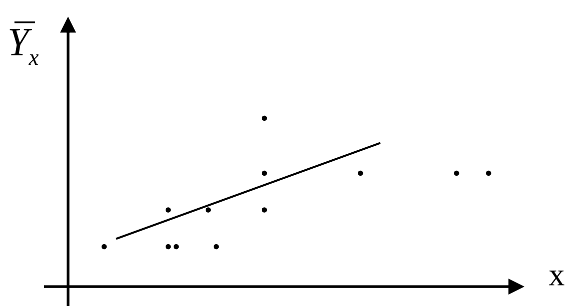
Iš jos:

$$\overline{X}_y = g(y) \qquad (13)$$

Lygybė (12) vadinama *koreliacine lygtimi arba Y regresijos lygtimi X atžvilgiu*, o (13) - *koreliacine lygtimi arba X regresijos lygtimi Y atžvilgiu*. Šių lygybių grafikus vadiname regresijos linijomis. Jos gali būti tiesės arba kreivės. Jei grafikai tiesės, tai turime *tiesinę regresiją*; jei grafikai kreivės – (parabolė, hiperbolė, eksponentė ir kt.) – *kreivinę regresiją*.

Vienas iš koreliacijos teorijos uždavinių – nustatyti regresinės priklausomybės tarp duotųjų dydžių formą, t.y., nustatyti regresijos lygties pavidalą ir tos lygties parametrus.

Atidėję plokštumoje taškus su koordinatėmis $(x_i , \overline{Y}_{x_i})$ $i = 1, 2, \dots , k,$ gauname vaizdą, iš kurio sprendžiame apie ryšio tarp X ir $\overline{Y}_x$ formą.



$\overline{Y}_x = aX + b$

Dabar reikia rasti spėjamos lygties parametrus.

6.2 Tiesinė regresijos lygtis

Kai regresijos lygtys $\overline{Y}_x = f(x)$ ir $\overline{X}_y =$ g (y) yra tiesinės, t. y., išreiškiamos pavidalu $\overline{Y}_x = ax + b$ ir $\overline{X}_y = cy + d$, tai koreliacinė priklausomybė tarp X ir Y vadinama tiesine. Šios lygtys vadinamos *tiesinės regresijos lygtimis*, o jų grafikai – *tiesinės regresijos tiesėmis*.

Sudarydami lygtį $\overline{Y}_x = ax + b$ koeficientus a ir b parenkame taip, kad regresijos tiesė būtų arčiausiai prie taškų $(x_i, \overline{Y}_{x_i})$. Taškų atstumus nuo tiesės $\overline{Y}_x = ax + b$ matuosime jų nuokrypiais nuo tiesės Oy ašies kryptimi.

Tegul $\tilde{Y}_{x_i}$ - tiesės $\overline{Y}_x = ax + b$ taško su abscise x_i ordinatė, o $\tilde{Y}_{x_i}$ - taško $(x_i, \overline{Y}_{x_i})$ ordinatė. Tada

$$\tilde{Y}_{x_i} - \overline{Y}_{x_i} = a x_i + b - \overline{Y}_{x_i}, \qquad i = 1, 2, \dots, k.$$

Sudarome šių nuokrypių kvadratų, padaugintų iš dažnių, sumą:

$$S(a, b) = \sum_{i=1}^{k} m_{x_i} (ax_i + b - \overline{Y}_{x_i})^2 .$$

Reikalaujame, kad nuokrypių kvadratų suma būtų mažiausia, t. y., iešome S(a, b) minimumo:

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum_{i=1}^{k} m_{x_i} (ax_i + b - \overline{Y}_{x_i}) x_i = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum_{i=1}^{k} m_{x_i} (ax_i + b - \overline{Y}_{x_i}) = 0. \end{cases} \qquad (14)$$

Atskliaudę reiškinius, esančius po sumų ženklais įvedame tokius pažymėjimus:

$$\sum m_{x_i} x_i = n\frac{\sum m_{x_i} x_i}{n} = n\overline{X}, \qquad \sum m_{x_i} x_i^2 = n\overline{X^2}, \qquad \sum m_{x_i} \overline{Y}_{x_i} = n\overline{Y},$$

$$\sum m_{x_i} x_i \overline{Y}_{x_i} = n\overline{XY}, \qquad \sum m_{x_i} = n .$$

Dabar sistema (14) virsta tokia:

$$\begin{cases} an\overline{X^2} + bn\overline{X} = n\overline{XY}, \\ an\overline{X} + bn = n\overline{Y} \end{cases} \qquad (15)$$

Tai dviejų tiesinių lygčių sistema ieškomų koeficientų a ir b atžvilgiu. Iš (15) sistemos antrosios lygties turime: $b = \overline{Y} - a\overline{X}$, tada $a = \dfrac{\overline{XY} - b\overline{X}}{\overline{X^2}} =$

$= \dfrac{\overline{XY} - \overline{X}(\overline{Y} - a\overline{X})}{\overline{X^2}}$, iš čia

$$a\overline{X^2} - a(\overline{X})^2 = \overline{XY} - \overline{X}\overline{Y} \qquad \text{ir} \qquad a = \frac{\overline{XY} - \overline{X}\overline{Y}}{\overline{X^2} - \overline{X}^2} = \frac{\overline{XY} - \overline{X}\overline{Y}}{S_x^2}$$

Į regresijos lygtį $\overline{Y}_x = aX + b$ įstatę $b = \overline{Y} - a\overline{X}$, galime užrašyti:

$$\overline{Y}_x - \overline{Y} = a\,(X - \overline{X}),$$

Analogiškai gautume:

$$\overline{X}_y - \overline{X} = c\,(Y - \overline{Y}).$$

Matome, kad abi tiesės eina per tą patį tašką ($\overline{X}, \overline{Y}$). Šis taškas yra atsitiktinių dydžių X ir Y pasiskirstymo centras.

Dydžiai X ir Y paprastai yra skirtingų dimensijų (pvz., X – ilgis, Y – svoris), todėl, pakeitus matavimo vienetus, keisis ir tiesių krypties koeficientas . Kad taip neįvyktų, nuokrypių matavimo vienetu imamas vidutinis kvadratinis nuokrypis. Lygtį

$$\overline{Y}_x - \overline{Y} = a\,(X - \overline{X})$$

pertvarkome:

$$\frac{\overline{Y}_x - \overline{Y}}{S_y} = a\frac{S_x}{S_y}\,\frac{X - \overline{X}}{S_x}.$$

Pažymėję $a\dfrac{S_x}{S_y} = r$, gauname $\dfrac{\overline{Y}_x - \overline{Y}}{S_y} = r\,\dfrac{X - \overline{X}}{S_x}$,

arba

$$\overline{Y}_x - \overline{Y} = r\,\frac{S_y}{S_x}\,(X - \overline{X}).$$

Koeficientas $r = a\frac{S_x}{S_y}$ nepriklauso nuo matavimo vienetų ir vadinamas *koreliacijos koeficientu.*

Jei tiesinės koreliacijos koeficientas

$$r = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{S_x S_y}$$

lygus nuliui, tai tarp X ir Y nėra tiesinio koreliacinio ryšio (nors netiesinis ryšys gali būti).

Jeigu $|r| = 1$, tai tarp X ir Y yra funkcinis ryšys.

Kuo $|r|$ artimesnis vienetui, tuo stipresnis ryšys tarp X ir Y .

Jeigu r teigiamas, tai X didėjant Y taip pat didėja; jei neigiamas, tai X didėjant Y mažėja (ir atvirkščiai).

Pavyzdys. **B**andymo metu stebėtos tokios X ir Y reikšmės:

Nr.	1	2	3	4	5	6	7	8	9	10
X	1	1	1	2	2	2	3	3	3	4
Y	3	3	3	4	4	5	5	5	6	7

Rasime Y regresijos lygtį X atžvilgiu (*ir X regresijos lygtį Y atžvilgiu.)*

1. Sudarome lentelę

y_j \ x_i	1	2	3	4	m_y
3	3				3
4		2			2
5		1	2		3
6			1		1
7				1	1
m_x	3	3	3	1	10

2. Apskaičiuojame sąlyginius vidurkius $\overline{Y}_{x_i} = \dfrac{\sum_{j=1}^{n} m_{ij} y_j}{m_{x_i}}$:

$$\overline{Y}_{x_1} = \frac{3.3}{3} = 3; \quad \overline{Y}_{x_2} = \frac{2.4+1.5}{3} = 4{,}33; \quad \overline{Y}_{x_3} = \frac{2.5+1.6}{3} = 5{,}33;$$

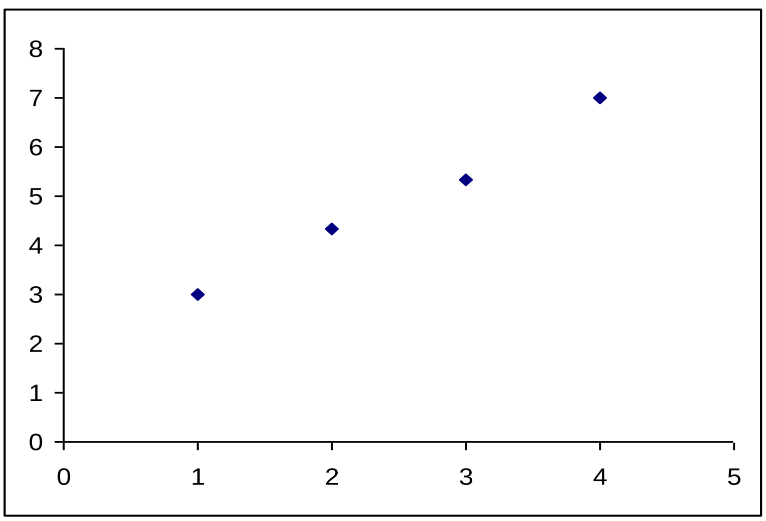
$$\overline{Y}_{x_4} = \frac{1.7}{1} = 7.$$

Sudarome lentelę:

x_i	1	2	3	4
$\overline{Y}_{x_i}$	3	4,33	5,33	7

Atidėję taškus (x_i , $\overline{Y}_{x_i}$) koordinačių sistemoje matom, kad jie išsidėstę beveik tiesėje 19 pav.), todėl turime tiesinės regresijos atvejį.

Užpildome dar tris lenteles:



19 pav.

x_i	m_{x_i}	$m_{x_i} x_i$	$m_{x_i} x_i^2$

1	3	3	3
2	3	6	12
3	3	9	27
4	1	4	16
$\sum$	10	22	58

$$\overline{X} = \frac{\sum m_{x_i} x_i}{n} = \frac{22}{10} = 2{,}2; \quad \overline{X^2} = \frac{\sum m_{x_i} x_i^2}{n} = \frac{58}{10} = 5{,}8;$$

$$S_x^2 = \overline{X^2} - \overline{X}^2 = 5{,}8 - (2.2)^2 = 0{,}96; \quad S_x = 0{,}979796;$$

y_j	m_{y_j}	$m_{y_j} y_j$	$m_{y_j} y_j^2$
3	3	9	27
4	2	8	32
5	3	15	75
6	1	6	36
7	1	7	49
$\sum$	10	45	322

$$\overline{Y} = \frac{\sum m_{y_j} y_j}{n} = \frac{45}{10} = 4{,}5; \quad \overline{Y^2} = \frac{\sum m_{y_j} y_j^2}{n} = \frac{219}{10} = 21{,}9;$$

$$S_y^2 = \overline{Y^2} - \overline{Y}^2 = 21{,}9 - (4{,}5)^2 = 1{,}65; \quad S_y = 1{,}284523$$

x_i	y_j	$x_i\ y_j$	m_{ij}	$x_i\ y_j\ m_{ij}$
1	3	3	3	9

2	4	8	2	16
2	5	10	1	10
3	5	15	2	30
3	6	18	1	18
4	7	28	1	28
				111

$$\overline{XY} = \frac{\sum\sum x_i y_j m_{ij}}{n} = \frac{111}{10} = 11{,}1; \quad r = \frac{11{,}1 - 2{,}2.4{,}5}{0{,}979796.1{,}284523} = 0{,}9534626;$$

$$a = 0{,}9534626 \frac{1{,}284523}{0.979796} = 1{,}249999;$$

Empirinė tiesinės regresijos lygtis yra tokia:

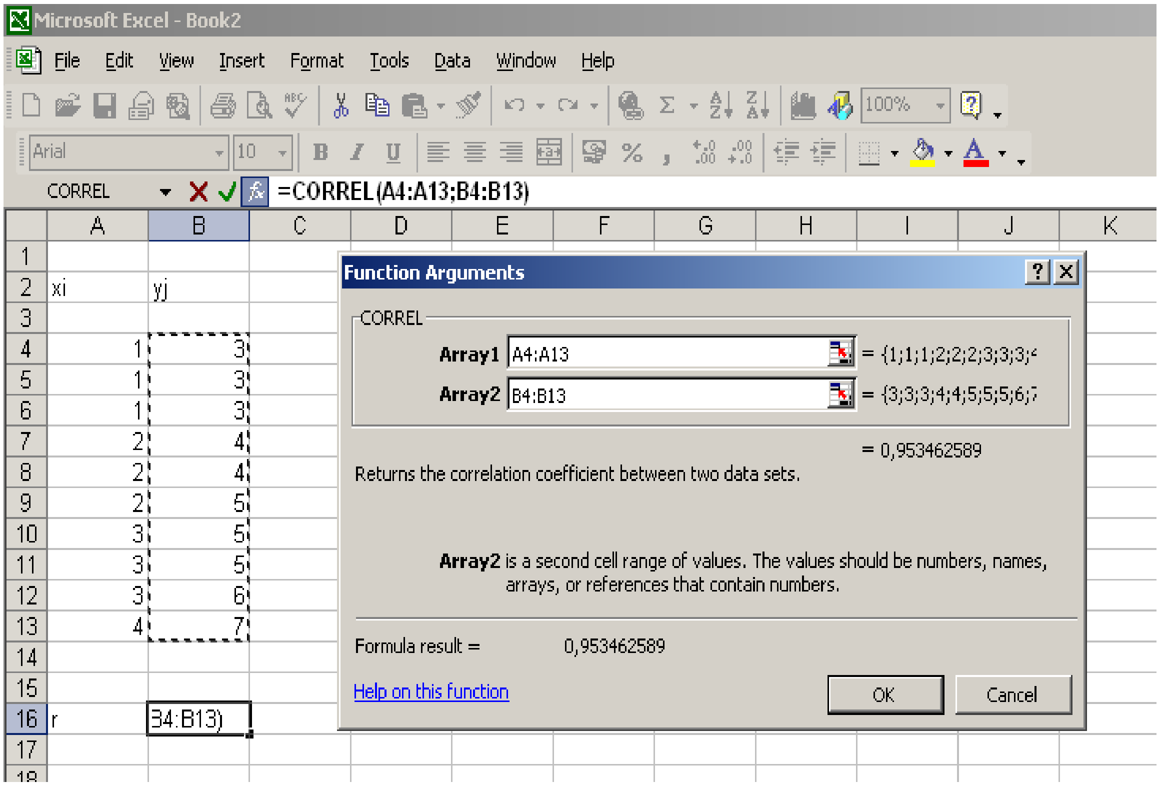
$$\overline{Y}_x - 4{,}5 = 1{,}25\,(x - 2{,}2) \quad \text{arba} \quad \underline{\overline{Y}_x = 1{,}25\,x + 1{,}75}$$

6.3 Empirinio koreliacijos koeficiento ir empirinės tiesinės regresijos lygties radimas su MS EXCEL

Norėdami gauti empirinį koreliacijos koeficientą su EXCEL programa stebėtas dvimačio atsitiktinio dydžio (X,Y) reikšmių poras patalpiname *Excel* lentelėje į kuriuos nors du stulpelius (eilutes) suformuodami skaičių masyvą, pvz., A1:A9. Kiekvieną reikšmių porą įvedame tiek kartų, koks yra šios poros dažnis m_{ij} .Pažymime langelį, kuriame norime gauti ieškomąjį rezultatą.

Lentelės viršuje esančioje simbolių eilutėje paspaudę simboliu  pažymėtą “klavišą”, iškviečiame langą **Paste Function**

Kairėje lango dalyje stulpelyje **Function category** pažymime eilutę **Statistical,** dešiniajame **Function name** stulpelyje pažymime funkciją CORREL. Paspaudę OK, ekrane matome langą CORREL , kuriame, į langelį **Number 1** įrašome masyvą A4:A13, o į langelį **Number 2** masyvą . B4:B13 iškart matome empirinio koreliacijos koeficiento reikšmę 0,953463 (žiūr.20 pav.). Paspaudus OK, langas išnyks, o vidurkio reikšmė atsiras anksčiau pažymėtame langelyje B16.



20pav.

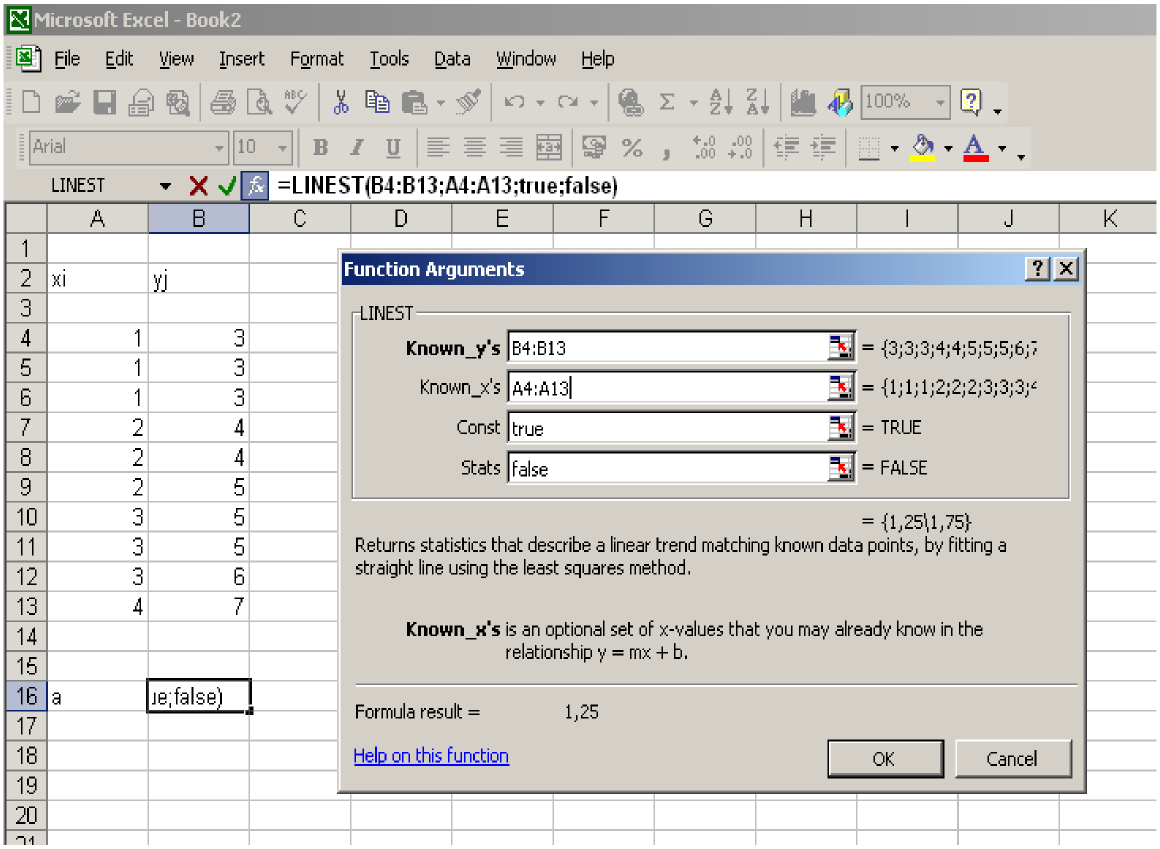
Galima nekviesti CORREL lango, o tiesiog simbolio  eilutėje įvesti komandą **=CORREL(A4:A13;B4:B13**) ir paspausti *Enter.*

Empirinės tiesinės regresijos lygties koeficientams a ir b surasti taip pat užtenka panaudoti komandas

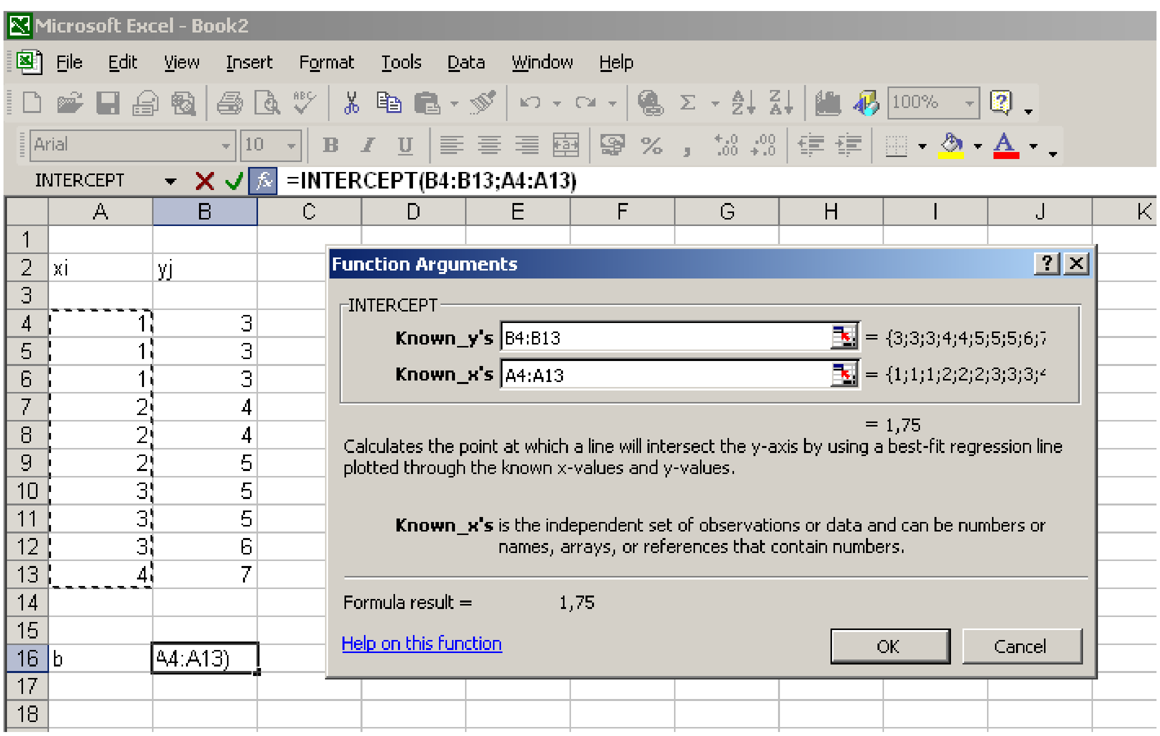
=LINEST(B4:B13; A4:A13;true;false) ir **=INTERCEPT(B4:B13; A4:A13).**

Kitas būdas yra išsikviesti LINEST langą , kad pažymėtame langelyje gautume regresijos lygties koeficientą a. Pav. 21 LINEST lange matome abu regresijos lygties koeficientus. Kai koeficiento a reikšmė turi daugiau skaitmenų , laisvojo nario b gali ir nesimatyti.

Koeficientui b pažymėtame langelyje gauti naudojame INTERCEPT langą (22 pav.)


Microsoft Excel - Book2
File Edit View Insert Format Tools Data Window Help
LINEST
=LINEST(B4:B13;A4:A13;true;false)
xi
yj
a
Function Arguments
LINEST
Known_y's B4:B13 = {3;3;3;4;4;5;5;5;6;7
Known_x's A4:A13 = {1;1;1;2;2;2;3;3;3;4
Const true = TRUE
Stats false = FALSE
= {1,25\1,75}
Returns statistics that describe a linear trend matching known data points, by fitting a straight line using the least squares method.
Known_x's is an optional set of x-values that you may already know in the relationship y = mx + b.
Formula result = 1,25
Help on this function
OK
Cancel

21 pav.


Microsoft Excel - Book2
File Edit View Insert Format Tools Data Window Help
INTERCEPT
=INTERCEPT(B4:B13;A4:A13)
xi
yj
b
Function Arguments
INTERCEPT
Known_y's B4:B13 = {3;3;3;4;4;5;5;5;6;7
Known_x's A4:A13 = {1;1;1;2;2;2;3;3;3;4
= 1,75
Calculates the point at which a line will intersect the y-axis by using a best-fit regression line plotted through the known x-values and y-values.
Known_x's is the independent set of observations or data and can be numbers or names, arrays, or references that contain numbers.
Formula result = 1,75
Help on this function
OK
Cancel

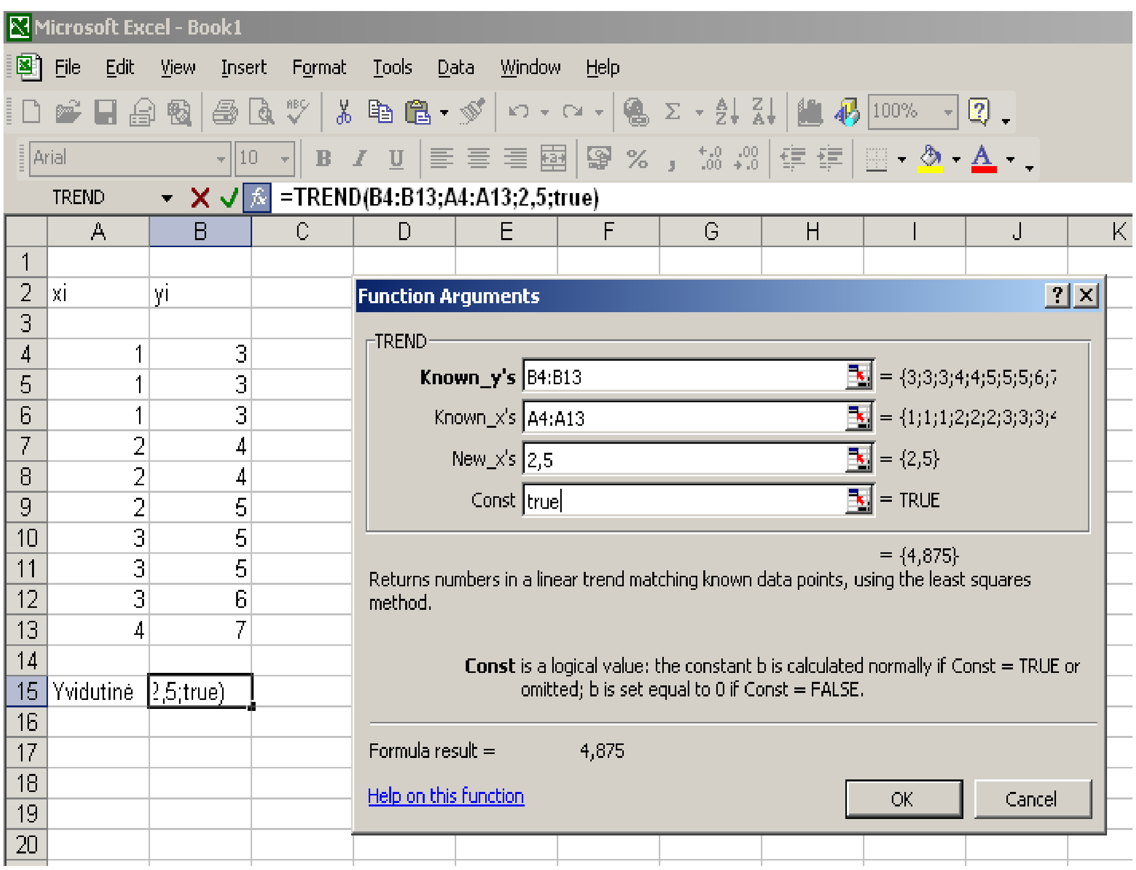
22 pav.

6.4 *Vidutinės Y reikšmės prognozavimas naudojant tiesinį trendą, kai žinoma x reikšmė*

Panaudodami empyrinę regresijos lygtį galime prognozuoti vidutinę Y reikšmę, kai X reikšmė x žinoma arba pasirenkama. EXCEL statistinės funkcijos TREND pagalba galima atlikti šią prognozę nesuradę prieš tai regresijos lygties.

Past Function lange **Statistical** kategorijoje pažymime funkciją TREND (23pav.). Į pirmus du langelius įvedame y_i ir x_i reikšmių masyvus, į trečią langelį *New_x's* įrašome laisvai pasirinktą *x* reikšmę (mūsų pavyzdyje $x = 2{,}5$). Langelyje *Const* įrašius loginio kintamojo reikšmę *true*, gauname vidutinės *Y* reikšmės prognozę pagal tiesinį trendą $y = ax+b$, o parinkę reikšmę *false* – pagal tiesinį trendą $y = ax$.

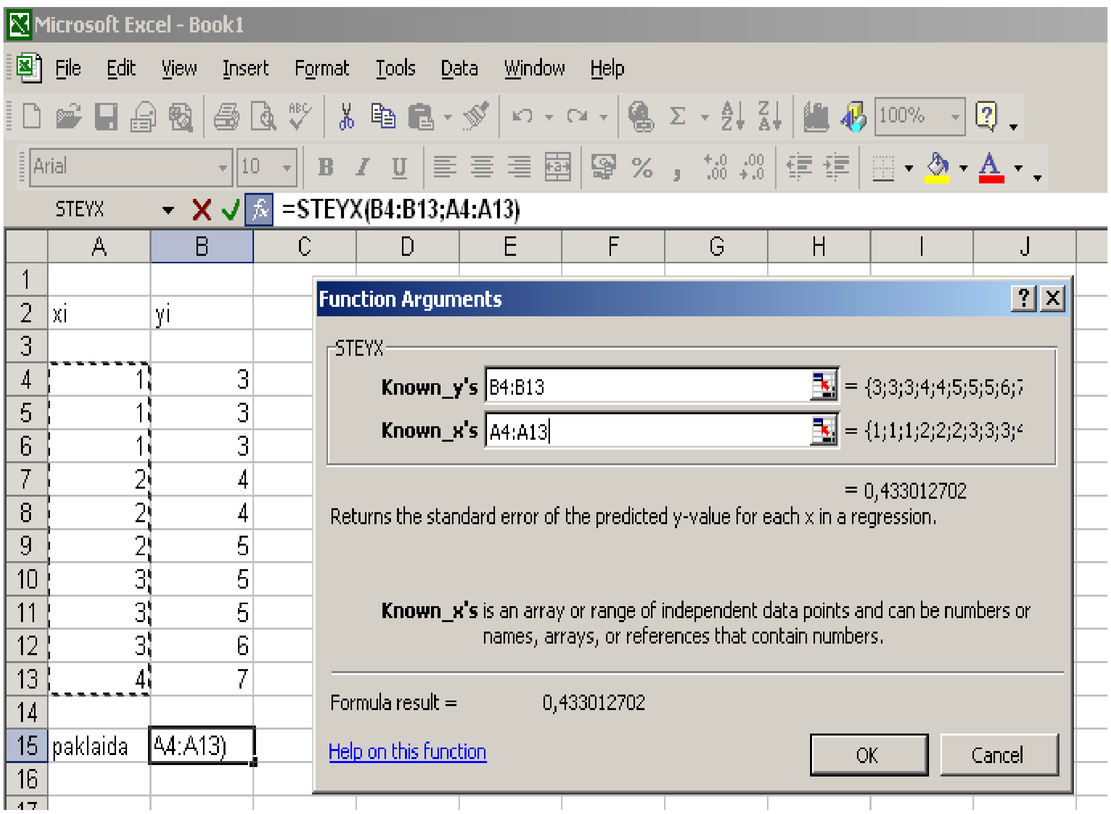
Paspaudę OK, prognozuojamą vidutinę Y reikšmę 4,875 gausime iš anksto parinktame *Excel* lentelės langelyje (mūsų pavyzdyje – langelyje B15).



23 pav

6.5 *Vidutinė kvadratinė paklaida tiesinės regresijos lygčiai* $y = ax + b$

Suformavę y_i ir x_i reikšmių masyvus, **Paste Function** lange **Statistical** kategorijoje pažymime funkciją STEYX (24pav.). Į langelius įvedę y_i ir x_i reikšmių masyvus, matome vidutinę kvadratinę paklaidą s = $\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - mx_i - b)^2}{n-2}}$, daromą prognozuojant pagal tiesinį trendą $y = ax + b$. Paspaudę OK, vidutinės kvadratinės paklaidos reikšmę 0,433013 gausime iš anksto parinktame langelyje (mūsų pavyzdyje – langelyje B15).



24 pav.

Literatūra

1. A.Žemaitis. Trumpas tikimybių teorijos ir matematinės statistikos kursas. Vilnius: Technika. 2001.
2. F.Mišeikis. Statistika ir ekonometrija. Vilnius: Technika. 1997.
3. J.Raulynaitis,V.Podvezko, S.Vakrinienė, J.Daunoravičius. Matematinė statistika. Vilnius: Technika. 1997.
4. A.Apynis, E.Stankus. Matematika. Vilnius: TEV. 2000.