

VILNIAUS UNIVERSITETO

KAUNO FAKULTETAS

SOCIALINIŲ MOKSLŲ IR TAIKOMOSIOS INFORMATIKOS INSTITUTAS

Informacinių sistemų ir kibernetinės saugos studijų programa

IRMANTAS VASILJEVAS

STATISTIKOS STUDIJŲ DALYKAS

SAVARANKIŠKAS DARBAS NR. 1

KAUNAS 2019

TURINYS

PAVEIKSLŲ SĄRAŠAS.....	4
IŽANGA.....	5
1. PIRMOJI UŽDUOTIS	6
1.1 Duomenų pasirinkimo situacija	6
1.2 Geriausio pasiskirstymo dažnių lentelė	6
1.3 Dažnių lentelė pagal vietą sąraše	8
1.4 Dažnių lentelė pagal standartinį nuokrypį	9
1.5 Duomenų sklaidos bei padėties charakteristikos.....	10
1.6 Ūselinė diagrama.....	10
2. ANTROJI UŽDUOTIS	12
2.1 Apklausos tema	12
2.2 Apklausos klausimai	12
2.3 Apklausos duomenų analizė.....	14
IŠVADOS	18
LITERATŪRA.....	19

PAVEIKSLŲ SĄRAŠAS

pav 1 dažnių lentelė su geriausiu pasiskirstymu	7
pav 2 histograma pagal geriausią pasiskirstymą	7
pav 3 dažnių lentelė pagal vietą sąraše.....	8
pav 4 histograma pagal vietą sąraše	8
pav 5 dažnių lentelė pagal standartinį nuokrypį.....	9
pav 6 histograma pagal standartinį nuokrypį	9
pav 7 sklaidos bei padėties charakteristikos.....	10
pav 8 sklaidos bei padėties charakteristikos.....	10
pav 9 ūselinė diagrama	11
pav 10 apklausos klausimai. Pirma dalis.....	12
pav 11 apklausos klausimai. Antra dalis	13
pav 12 apklausos klausimai. Trečia dalis	13
pav 13 ėjimo miegoti laikas fakultetuose	15
pav 14 moterų ir vyrų miego trukmės skirtumai	15
pav 15 miego kokybė	16
pav 16 savijautos priklausomybė nuo miego trukmės	16
pav 17 nemigos priklausomybė nuo miego trukmės	17
pav 18 veikla prieš miegą.....	17

ĮŽANGA

Šio darbo tikslas yra pademonstruoti Statistikos studijų dalyko praktinių ir teorinių užsiėmimų metu įgytas žinias bei įgūdžius.

Šį darbą sudaro dvi dalys. Pirmoji yra atliekama su statistinės analizės įrankiu STATISTICA. Užduoties pagrindiniai tikslai yra normalaus pasiskirstymo atsitiktinių skaičių lentelės generavimas, tų duomenų skirstymas į dažnių lenteles, duomenų vaizdavimas histogramose ir ūselinėje diagramoje bei duomenų padėties ir sklaidos charakteristikų generavimas. Antroji užduotis atliekama su EXCEL duomenų apdorojimo įrankiu. Jos metu reikėjo sukurti apklausos anketą, jos rezultatus pavaizduoti įvairiausiais grafikai ir diagramomis bei surasti tų duomenų padėties ir sklaidos charakteristikas.

1. PIRMOJI UŽDUOTIS

Pirmojoje užduotyje duomenys bus analizuojami su statistinės analizės programa STATISTICA.

1.1 Duomenų pasirinkimo situacija

Buvo sugeneruoti 504 skaičiai. Šie duomenys atspindėjo trijų kambarių butų kainas Vilniaus mieste. Prieš generuojant buvo peržiūrėti butų skelbimai tokiose skelbimų svetainėse kaip *aruodas.lt*, *skeliu.lt*, *alio.lt* bei buvo įvertintos mažiausios ir didžiausios galimos butų kainos, butų kiekio pasiskirstymas įvairiuose intervaluose. Pagal tai buvo pasirinktas tikėtinas butų kainų vidurkis - 380 000 bei standartinis nuokrypis 160 000. Statistinės analizės progamoje 504 reikšmės buvo sugeneruotos pagal normaliojo pasiskirstymo formulę - $RNDNORMAL(\text{standartinis.nuokrypis}) + \text{vidurkis}$.

1.2 Geriausio pasiskirstymo dažnių lentelė

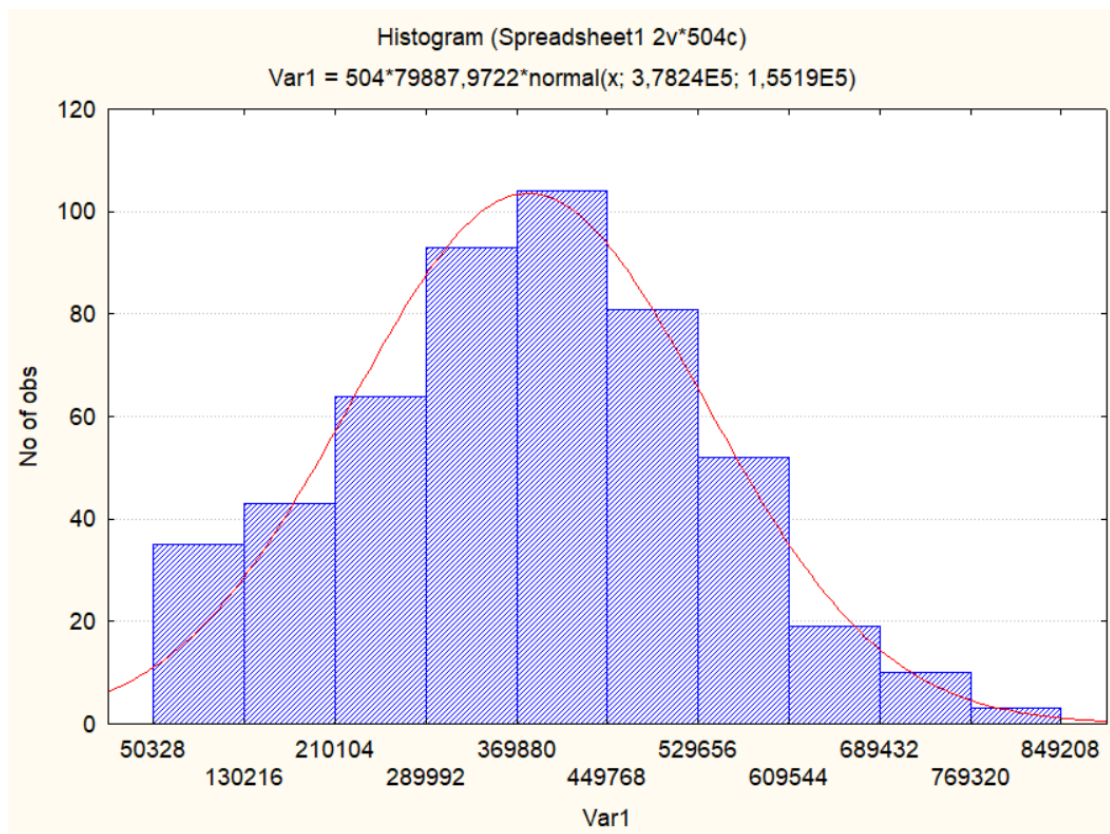
Geriausio pasiskirstymo dažnių lentelei generuoti buvo pasirinkta 10 intervalų. Norint sukurti dažnių lentelę pagrindiniame meniu reikia pasirinkti *Statistics* punktą, po to *Basic Statistics/Tables*. Iššokusiam dialogo lauke tarp įvairių duomenų vaizdavimo galimybių reikia pasirinkti *Frequency tables* bei paspausti mygtuką *OK*. Po to, sklityje *Advanced* nusirodome intervalų kiekį pasirinkdami punktą *No. Of exact intervals*. Galiausiai spaudžiame mygtuką *Summary: Frequency Tables* ir naujame puslapyje gauname dažnių lentelę.

Nubrėžti histogramą galima keliais scenarijais. Pirmasis būdas butų pagrindiniame meniu pasirinkus punktą *Graphs* ir po to *Histograms...*, kuriame galima pasirinkti histogramos nustatymus. Kitas būdas yra panašus į dažnių lentelių generavimo scenarijų. Tik vietoj mygtuko *Summary: Frequency Tables* derėtų spausti mygtuką *Histograms*.

		Frequency table: Var1: =RndNormal(160000) + 380000 (Spreadsheet1)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
6 000	<x<=94 000	19	19	3,76984	3,7698
94 000	<x<=182 000	41	60	8,13492	11,9048
183 475	<x<=272 000	69	129	13,69048	25,5952
272 000	<x<=358 000	96	225	19,04762	44,6429
358 000	<x<=446 000	114	339	22,61905	67,2619
446 000	<x<=534 000	89	428	17,65873	84,9206
534 000	<x<=622 000	52	480	10,31746	95,2381
622 000	<x<=710 000	16	496	3,17460	98,4127
710 000	<x<=798 000	5	501	0,99206	99,4048
798 000	<x<= 886 000	3	504	0,59524	100,0000

pav 1 dažnių lentelė su geriausiu pasiskirstymu

Iš dažnių lentelės matome, jog intervalai pasiskirstę tolygiai, tad dėl noramalaus pasiskirstymo galime neabejoti.



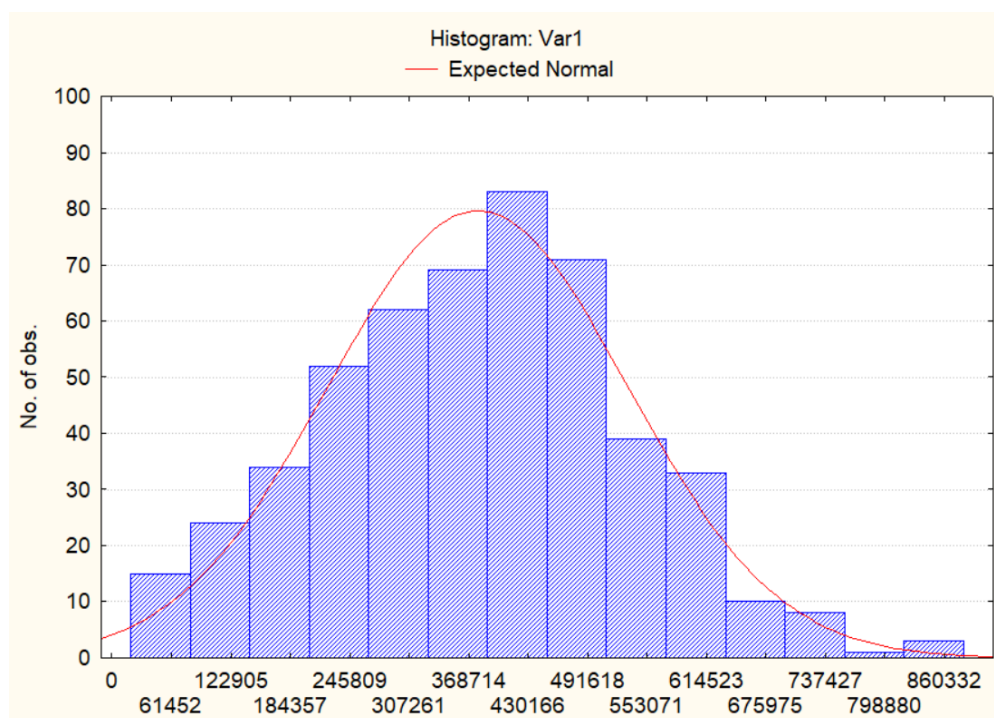
pav 2 histograma pagal geriausią pasiskirstymą

1.3 Dažnių lentelė pagal vietą sąrašė

Pagal vietą sąrašė šią dažnių lentelę reikia suskirstyti į 14 intervalų. Scenarijus toks pat kaip ir prieš tai buvusiame punkte 1.1 Geriausio pasiskirstymo dažnių lentelė. Tik vietoj dešimties intervalų punkte *No. Of exact intervals* derėtų pasirinkti keturioliką.

		Frequency table: Var1: =RndNormal(160000) + 380000 (Spreadsheet1)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
19 602	<x<=81 055	15	15	2,97619	2,9762
81 055	<x<=142 507	24	39	4,76190	7,7381
142 507	<x<=203 959	34	73	6,74603	14,4841
203 959	<x<=265 411	52	125	10,31746	24,8016
265 411	<x<=326 864	62	187	12,30159	37,1032
326 864	<x<=388 316	69	256	13,69048	50,7937
388 316	<x<=449 768	83	339	16,46825	67,2619
449 768	<x<=511 221	71	410	14,08730	81,3492
511 221	<x<=572 673	39	449	7,73810	89,0873
572 673	<x<=634 125	33	482	6,54762	95,6349
634 125	<x<=695 577	10	492	1,98413	97,6190
695 577	<x<=757 030	8	500	1,58730	99,2063
757 030	<x<=818 482	1	501	0,19841	99,4048
818 482	<x<=879 934	3	504	0,59524	100,0000
Missing		0	504	0,00000	100,0000

pav 3 dažnių lentelė pagal vietą sąrašė



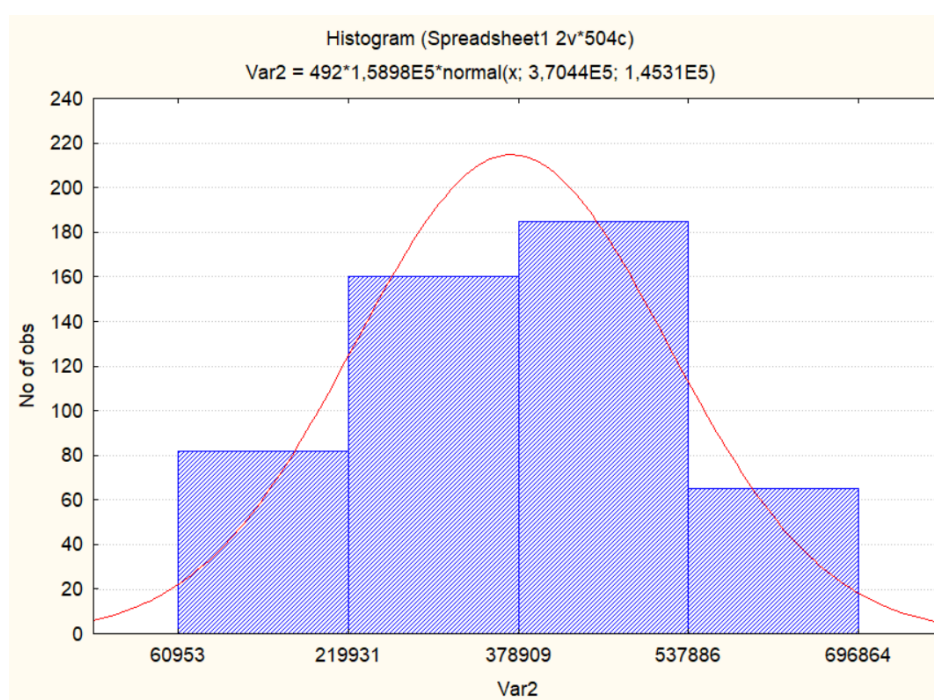
pav 4 histograma pagal vietą sąrašė

1.4 Dažnių lentelė pagal standartinį nuokrypį

Šiame punkte reikėjo sugeneruoti dažnių lentelę, pagal reikšmes besiskiriančias nuo vidurkio ne daugiau kaip per 2 standartinius nuokrypius. Tai reiškia, kad šioje lentelėje neturi būti mažiausių ir didžiausių būtų kainų reikšmių. Lentelėje yra numatomi keturi intervalai su šiomis ribinėmis reikšmėmis: po du ir po vieną standartinį nuokrypį nuo vidurkio į didesnę ir mažesnę puses bei pats vidurkis. Šias reikšmes atitinka šie skaičiai: 60 000, 220 000, 380 000, 540 000, 700 000. Mažesnės reikšmės už 60 000 bei didesnės už 700 000 negali patekti į grafikus, todėl tas reikšmes panaikiname, tad reikšmių skaičius sumažėja iki 492. Dabar su likusiais 492 skaičiais generuojame dažnių lentelę bei histogramą su keturiais intervalais.

		Frequency table: Var2: =RndNormal(160000) + 380000 (Spreadsheet1)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
60 000	<=x<220 000	82	82	16,26984	16,2698
220 000	<=x<380 000	162	244	32,14286	48,4127
380 000	<=x<540 000	184	428	36,50794	84,9206
540 000	<=x<700 000	64	492	12,69841	97,6190
700 000	<=x<860 000	0	492	0,00000	97,6190
Missing		12	504	2,38095	100,0000

pav 5 dažnių lentelė pagal standartinį nuokrypį



pav 6 histograma pagal standartinį nuokrypį

1.5 Duomenų sklaidos bei padėties charakteristikos

Duomenų sklaidos bei padėties charakteristikos sugeneruojamos pagal šį scenarijų. Rekamės meniu juostos punktą *Statistics*, po to *Basic Statistics/Tables*. Iššokusiame dialogo lauke tarp įvairių duomenų vaizdavimo galimybių reikia pasirinkti *Descriptive Statistics* bei paspausti mygtuką *OK*. Po to, sklityje *Advanced* nusirodome paspaudžiame mygtuką *Select all stats*. Paspaudus mygtuką *Summary* naujame lange sugeneruojami sklaidos ir padėties charakteristikos.

Descriptive Statistics (Spreadsheet1)												
Variable	Valid N	Mean	Confidence -95.000%	Confidence +95.000%	Geometric Mean	Harmonic Mean	Median	Mode	Frequency of Mode	Sum	Minimum	Maximum
Var1	504	378244,0	364662,4	391825,7	338395,4	285447,4	385511,6	Multiple	1	190634983	50328,44	849208,2

pav 7 sklaidos bei padėties charakteristikos

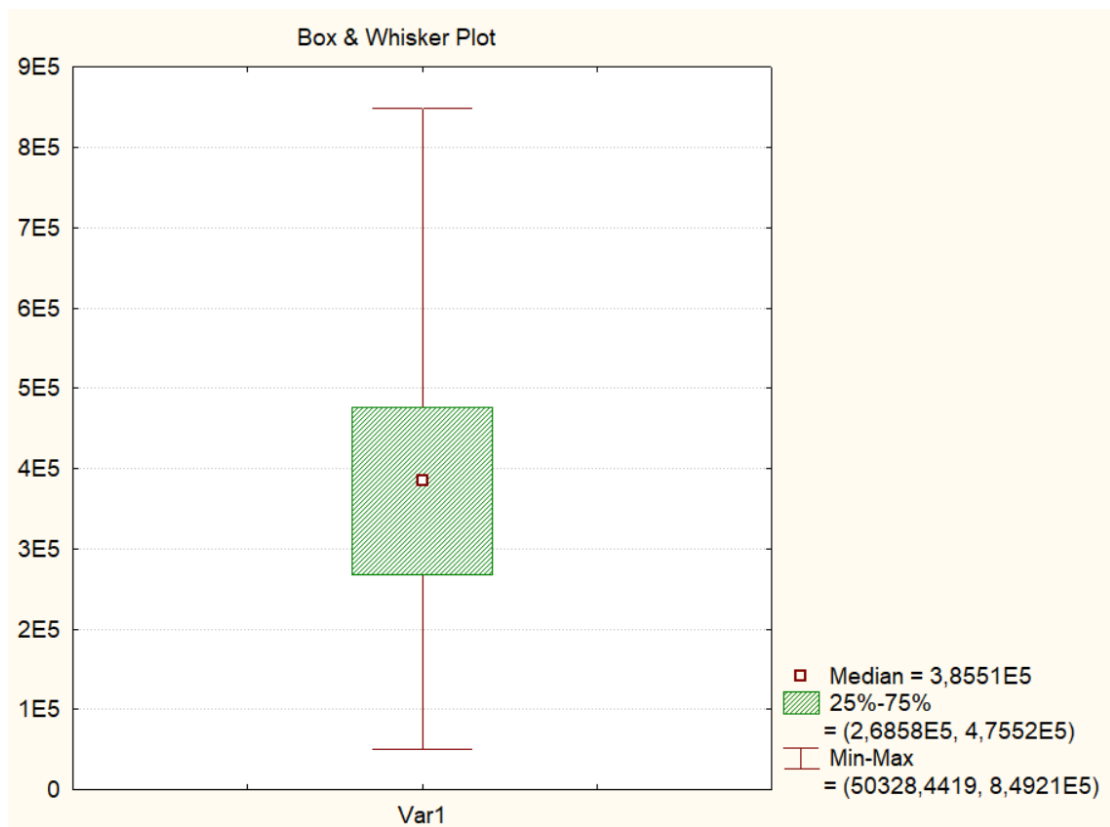
Descriptive Statistics (Spreadsheet1)													
Variable	Lower Quartile	Upper Quartile	Percentile 10.00000	Percentile 90.00000	Range	Quartile Range	Variance	Std.Dev.	Standard Error	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
Var1	268578,6	475520,7	168747,0	580908,8	798879,7	206942,1	2,408504E+10	155193,6	6912,871	0,086442	0,108786	-0,214588	0,217148

pav 8 sklaidos bei padėties charakteristikos

1.6 Ūselinė diagrama

Norint sukurti ūselinę diagramą pagrindiniame meniu reikia pasirinkti *Statistics* punktą, po to *Basic Statistics/Tables*. Iššokusiame dialogo lauke tarp įvairių duomenų vaizdavimo galimybių reikia pasirinkti *Frequency tables* bei paspausti mygtuką *OK*. Po to, pasirinkti skiltį *Descr.* bei paspausti mygtuką *Box & whisker plot for variables (1)*. Atsidaro langas, kuriame reikia pasirinkti kokią informaciją norime matyti ūselinėje diagramoje. Renakmės punktą *Median/Quart./Range*, kuris grafike pavaizduoja medianą, kvartilius bei ribines skaičių sekos reikšmes - mažiausią ir didžiausią.

Žemiau pavaizduota ūselinė diagrama vaizduoja sugeneruotos skaičių sekos intervalą, medianą ir reikšmių sritį tarp pirmo ir trečio kvartilių.



pav 9 ūselinē diagrama

2. ANTROJI UŽDUOTIS

Antrajau užduočiai atlikti buvo būtina sukurti apklausą, pasirinkta tema, svetainėje *apklausa.lt* bei grafiškai pavaizduoti surinktus duomenis.

2.1 Apklausos tema

Apklausa buvo orientuota į jaunų žmonių miego kokybės tyrimą. Respondentai turėjo atsakyti, kokiam Vilniaus Universiteto fakultete mokosi, kuriuo metu eina miegoti, kiek laiko miega, kaip dažnai prabunda naktį, kaip jaučiasi atsikėlę, ar išlieka energingi visos dienos metu, ką veikia prieš einant miegoti. Šiais klausimais buvo siekta išsiaiškinti jaunų žmonių, besimokančių Vilniaus universitete poilsio kokybę bei pastebėti dėsningumus tarp miego kokybės ir jai įtaką darančių veiksnių.

2.2 Apklausos klausimai

Apklausą sudaro 8 klausimai, iš kurių du yra kiekybiniai.

pav 10 apklausos klausimai.
Pirma dalis

Kiek laiko miegate?

Kiek laiko per dieną būnate maksimaliai darbingi?

Ar esate patenkinti savo miego kokybe?

- ☐ Taip
☐ Ne

Kokia Jūsų lytis?

- ☐ Vyras
☐ Moteris

Kokiame fakultete mokotės?

- ☐ Chemijos ir geomokslų fakultetas
☐ Ekonomikos ir verslo administravimo fakultetas
☐ Filologijos fakultetas
☐ Filosofijos fakultetas
☐ Fizikos fakultetas
☐ Gyvybės mokslų centras
☐ Istorijos fakultetas
☐ Kauno fakultetas
☐ Komunikacijos fakultetas
☐ Matematikos ir informatikos fakultetas
☐ Medicinos fakultetas
☐ Tarptautinių santykių ir politikos mokslų institutas
☐ Teisės fakultetas
☐ Verslo mokykla

Ar jus kankina nemiga?

- ☐ Taip
☐ Ne
☐ Tik prieš atsiskaitymus

Kaip jaučiatės dienos metu?

- ☐ Esu energingas(s) iki vidurdienio
☐ Esu energingas(s) iki vakaro
☐ Esu darbingas(s) visą dieną
☐ Visą laiką pavargęs ir be nuotaikos
☐ Tampu energingas tik vakare
☐ Kita

pav 11 apklausos klausimai. Antra dalis

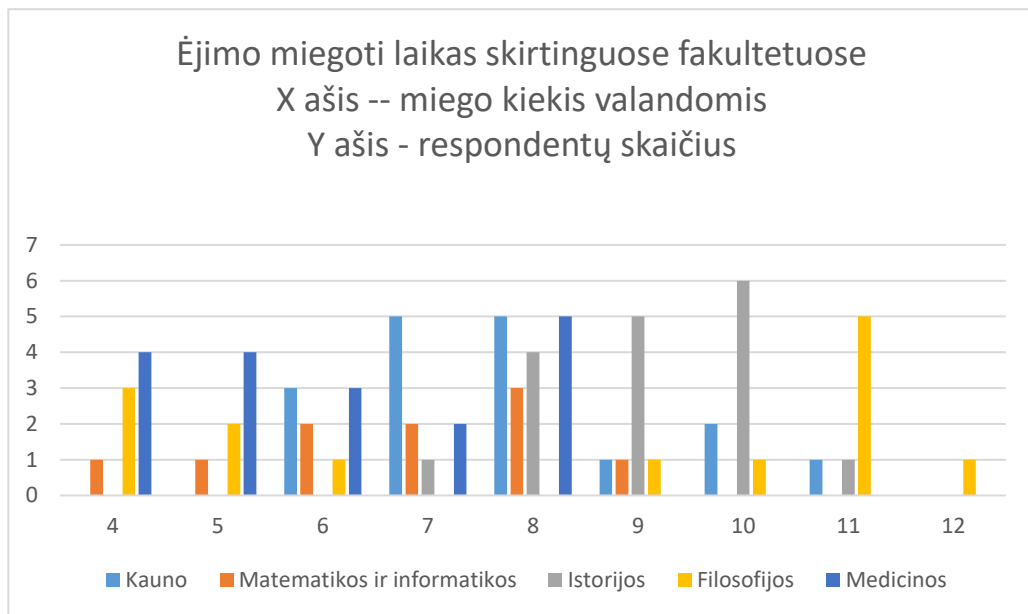
Ką veikiate prieš miegą?

- ☐ Naršau internete arba žiūriu filmus
☐ Dirbu
☐ Skaitau knygas
☐ Medituoju
☐ Bendrauju su namiškiais arba draugais
☐ Aktyviai sportuoju
☐ Einu pasivaikščioti gryname ore
☐ Kita

pav 12 apklausos klausimai. Trečia dalis

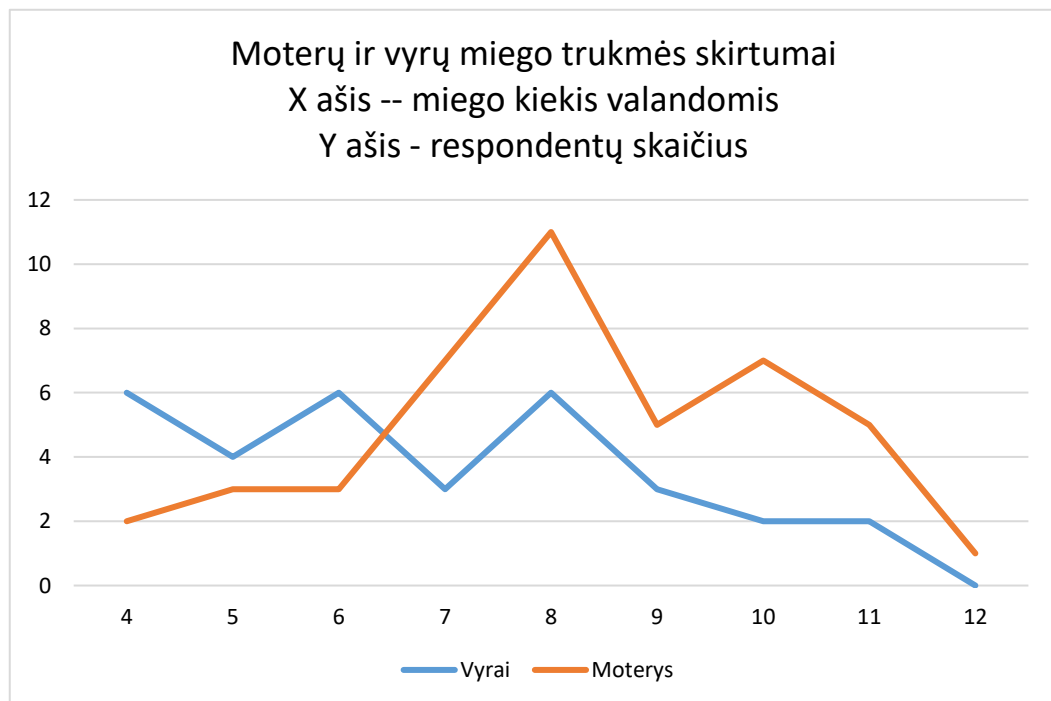
2.3 Apklausos duomenų analizė

Trumpiausiai miega filosofai ir medikai. Tarp ilgiausiai miegančių daugiausiai yra filosofų ir istorikų. Optimaliausią valandų skaičių tarp 7 ir 9 valandų miega Kauno fakulteto studentai. Tai vaizduoja žemiau esantis paveikslas.



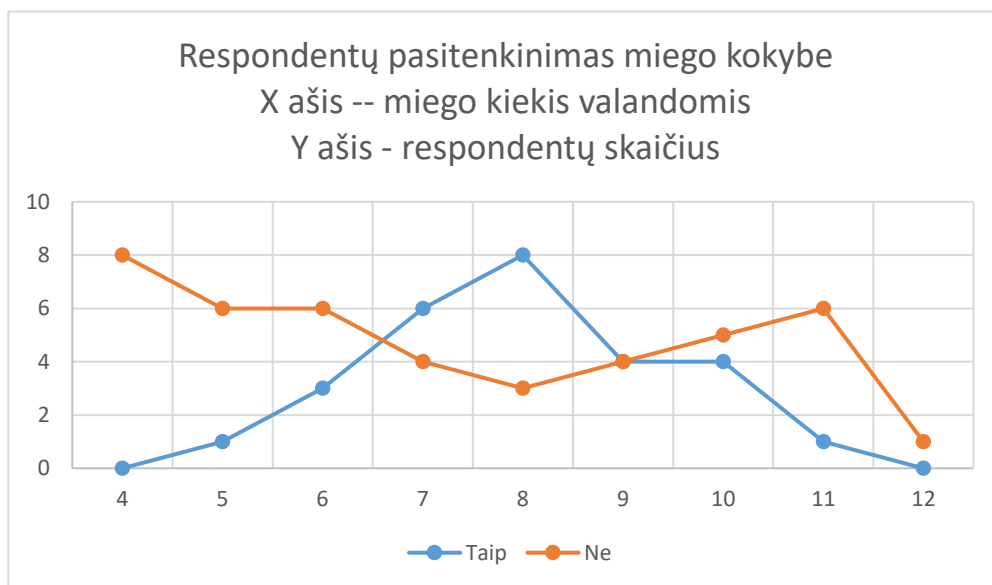
pav 13 ėjimo miegoti laikas fakultetuose

Apklausa parodė, kad moterys atsakingiau žiūri į miegą, nes dauguma jų miega daugiau nei 7 valandas. Tuo tarpu vyrai yra linkę miegoti ir mažiau nei 7 valandas, bet rečiau renkasi ilgą miegą tarp 9 ir 12 valandų.



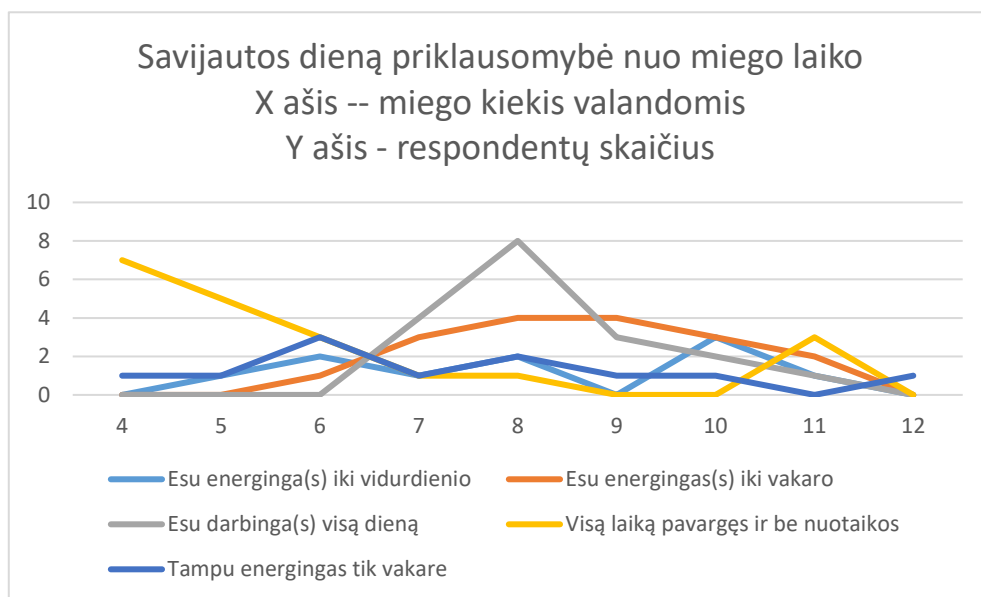
pav 14 moterų ir vyrų miego trukmės skirtumai

Didžiausią nepasitenkinimą miego kokybe jaučia miegantys trumpai apie 4 ar 5 valandas, taip pat per ilgai miegantys žmonės. Miegantys tarp 7 ir 8 valandų pasiilsį geriausiai.



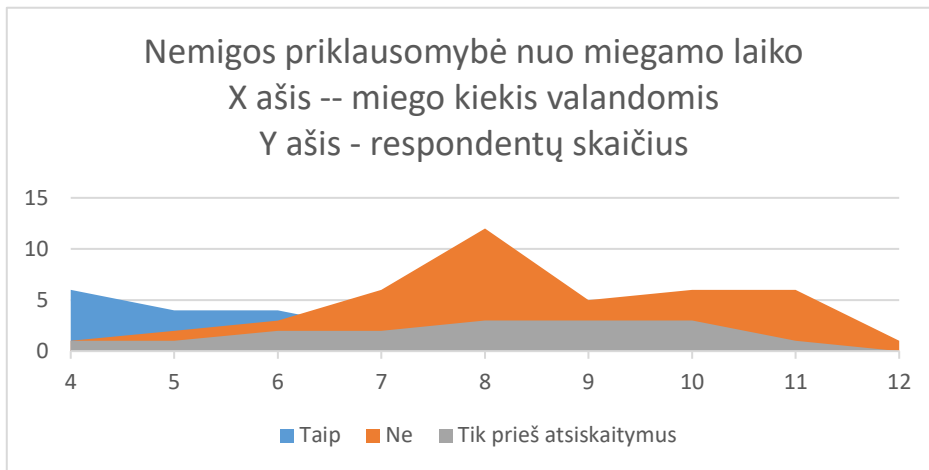
pav 15 miego kokybė

Visą dieną energingi jaučiasi žmonės, kurie miega tarp 7 ir 9 valandų. Dalis šių žmonių jaučiasi energingi iki vakaro. Visą dieną pavargę bei be nuotaikos yra žmonės, kurie miega mažiau nei 7 valandas arba tie, kurie miega per daug.



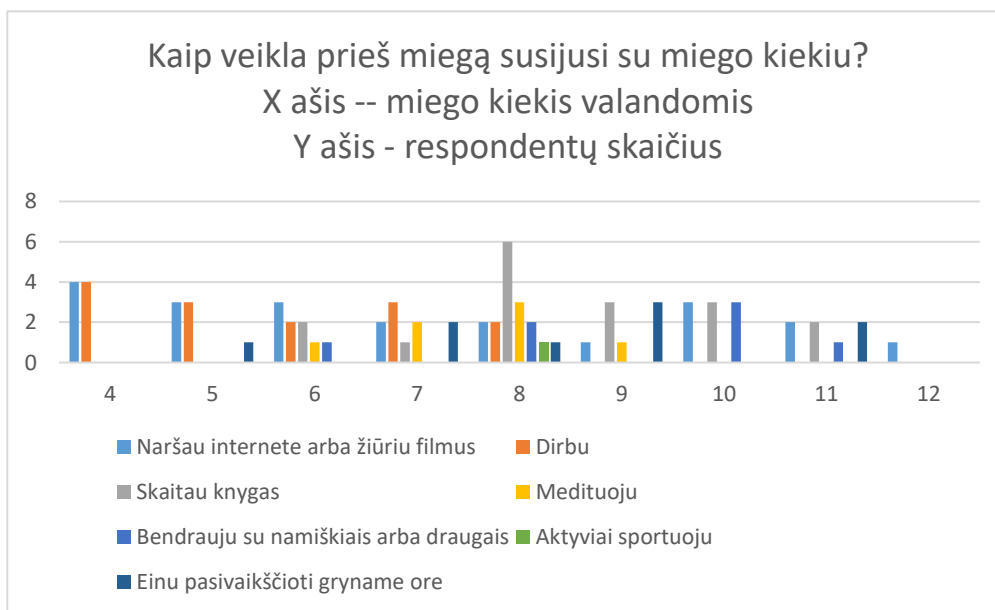
pav 16 savijautos priklausomybė nuo miego trukmės

Nemigą patiria žmonės, kurie miega per trumpai. Dalis studentų kenčia nuo nemigos prieš atsiskaitymus nepriklausomai nuo to, kiek jie vidutiniškai miega. Miegantys daugiau nei 7 valandas pagal apklausos duomenis nėra linkę į nemigą.



pav 17 nemigos priklausomybė nuo miego trukmės

Trumpiausiai miegantys studentai yra linkę prieš miegą užsiimti psichiką aktyvinančiomis veiklomis, tokiomis kaip darbas, filmų žiūrėjimas ar naršymas internete. Studentai miegantys ilgai prieš miegą yra linkę į atsipūtusį laiko leidimą - bendravimą, filmų žiūrėjimą, naršymą internete bei kartais pasivaikščiojimus gryname ore. Tarp miegančių 8 valandas yra daugiausiai tų, kurie skaito knygas bei užsiima harmonizuojančiomis veiklomis.



pav 18 veikla prieš miegą

IŠVADOS

Duomenų analizės įrankių, tokių kaip STATISTICA ir EXCEL naudojimas padeda efektyviai atlikti statistinę duomenų analizę bei atrasti įžvalgų apie duomenis.

LITERATŪRA

1. Virgilijus Sakalauskas. *Statistikos paskaitų medžiaga*. Žiūrėta 2019-10-08 per internetą: VU informacijos šaltiniai.
2. Nekilnojamojo turto portalas. *Trijų kambarių butų kainų paieška Vilniaus mieste*. Žiūrėta 2019-09-29 per internetą: <<https://www.aruodas.lt/>>.