

## Laboratorinis darbas nr.5

**Tema:** Duomenų analizė

### Užduotys

1. R duomenų rinkinį *mtcars* prijunkite prie R paieškos kelio naudodami *data()*. R duomenų rinkinio *mtcars* kintamuosius prijunkite prie R paieškos kelio naudodami *attach()*.
  - a. Nustatykite *mtcars* kintamojo *am* pradinę klasę ir pakeiskite jo klasę į faktorių *class()*, *factor()*.
  - b. Nustatykite duomenų rinkinio *mtcars* kintamųjų tipus, apskaičiuokite pagrindines skaitines charakteristikas.
  - c. Sudarykite kintamojo *cyl* santykinių dažnių lentelę ir ją pavaizduokite stulpeline diagrama;
  - d. Sudarykite kintamųjų *cyl* ir *am* kryžminę dažnių lentelę ir ją pavaizduokite stulpeline diagrama;
  - e. Nubrėžkite kintamojo *carb* skritulinę diagramą.
  - f. Tegul visos šio uždavinio diagramos bus viename grafiniame lange išsaugotos pavadinimu *mtcars.pdf*
2. Naudodami komandą *with()* apskaičiuokite duomenų rinkinio *mtcars*, kintamojo *qsec* medianą.
3. Taikydami komandą *tapply()* apskaičiuokite kiek vidutiniškai mylių/galonui nuvažiuoja automobiliai su automatine ir su mechanine pavarų dėže.
4. Išrikiuokite duomenų rinkinį *mtcars* pagal karbiuratorių skaičių *carb* didėjimo tvarka *order()*.
5. Taikant komandą *split()* suskirstykite kintamojo *carb* reikšmes pagal faktorių *am*.
6. Padalinkite duomenų rinkinio *mtcars*, kintamojo *qsec* reikšmes pagal pavaros tipą į grupes. Gautus duomenų vektorius (grupes) pavaizduokite viename grafike stačiakampių diagramomis. Grafike nurodykite kiekvieno faktoriaus pavadinimą statmenai ašiai (*las = 2*), *split()*, *boxplot()*.
7. Gauti matematinės statistikos testo rezultatai:  
52, 54, 57, 49, 63, 54, 38, 46, 49, 33, 43, 40, 29, 43, 60, 69, 54, 64, 41, 63, 44, 55, 58, 41, 37, 49, 36, 43, 36, 44, 35, 54, 57, 55, 56, 56, 56, 41, 49, 63, 41, 46, 45, 55, 45, 49, 47, 37, 62, 48, 44, 45, 48, 62.
  - a. Nubraižykite santykinių dažnių histogramą.
  - b. Apskaičiuokite formos charakteristikas: asimetrijos koeficientą (angl. *skewness*); eksceso koeficientą (angl. *kurtosis*) ir pakomentuokite gautus rezultatus.
  - c. <https://www.spssanalyze.lt/dazniu-skirstiniu-formos-charakteristikos/>
8. Importuokite (nuskaitykite, įkelkite) duomenų rinkinį *Duomenys\_lab5* į Rstudio. Atsitiktinių reikšmių generavimui nustatykite *set.seed(1126)*. Naudodami komandą *sample()* iš šio duomenų rinkinio atrinkite atsitiktinę 250 eilučių imtį. Naudodami imties duomenis:
  - a. apskaičiuokite kintamojo *year* padėties charakteristikas *summary()* ir nubrėžkite stačiakampių diagramą *boxplot()*, grafiniam langui uždėkite groteles (31 × 31) *grid()*.
  - b. apskaičiuokite koreliacijos koeficientus visoms nagrinėjamų kintamųjų poroms. Nubrėžkite sklaidos diagramas visoms kintamųjų poroms viename grafiniame lange.
  - c. didžiausią koreliacijos koeficientą turinčiai kintamųjų porai sudarykite tiesinės regresijos lygtį. Regresijos tiesę nupieškite ant sklaidos diagramos.

- d. komentaruose užrašykite determinacijos koeficiento reikšmę.
- e. Patikrinkite hipotezę apie koreliacijos koeficiento  $\rho(\text{area}, \text{price})$  reikšmingumą, kai reikšmingumo lygmuo 0.05.
9. Sugeneruokite 500 kintamojo  $X$  reikšmių, jeigu  $X \sim N(0.5, 2)$ . Naudodami šios imties duomenis:
- Raskite populiacijos vidurkio pasikliautinąjį intervalą, kai  $\sigma = 2$  ( $\sigma$  žinoma), su pasiklivimo lygmeniu  $Q = 0.99$ ;
  - Raskite populiacijos vidurkio pasikliautinąjį intervalą, kai  $\sigma$  nežinomas, su pasiklivimo lygmeniu  $Q = 0.95$ ;
  - Esant reikšmingumo lygmeniui  $\alpha = 0.05$  patikrinkite parametrines hipotezes:
    - $\begin{cases} H_0: \mu = 1, \\ H_1: \mu \neq 1. \end{cases}$
    - $\begin{cases} H_0: \mu = 1, \\ H_1: \mu > 1. \end{cases}$
    - $\begin{cases} H_0: \mu = 1, \\ H_1: \mu < 1. \end{cases}$

**PASTABA.** Violetine spalva pažymėtos užduotys nebūtinės. Jų laboratorinio darbo nr.2 atsiskaityme nebus.

### Komandos

Komanda	Paaishkinimas
<b>factor</b> ( $x$ )	Vektorius $x$ koduojamas faktoriumi. Iš diskretaus kiekybinio kintamojo padaromas faktorius (kokybinis kintamasis) <b>Pvz.:</b> <code>factor(am, labels = c("automatic", "manual"))</code>
<b>class</b> ( $x$ )	Objekto $x$ klasei nustatyti
<b>tapply</b> ( $vector$ , $index$ , $function$ )	Apskaičiuojamos <i>funkcijos</i> reikšmės <i>vektoriaus</i> duomenims, kiekvieno <i>indekso</i> atveju. <b>Pvz.:</b> <code>tapply(alga, lytis, vidurkis)</code>
<b>order</b> ( $x$ , $decreasing=FALSE$ )	Pateikia vektoriaus $x$ elementų indeksų vektorių, jeigu vektoriaus $x$ elementai būtų rikiuojami didėjimo tvarka. Duomenų rinkinio <i>mtcars</i> rikiavimo pagal cilindų skaičių <i>cyl</i> didėjimo tvarka. <b>Pvz.:</b> <code>mtcars[order(cyl, decreasing=FALSE),]</code>
<b>split</b> ( $x$ , $f$ )	Padalija vektoriaus $x$ reikšmes į faktoriaus $f$ grupes. <b>Pvz.:</b> <code>split(alga, lytis)</code>
<b>boxplot</b> ( $x_1, x_2, \dots, x_n$ )	Brėžiama kintamųjų $x_1, x_2, \dots, x_n$ stačiakampių diagrama.
<b>axis</b> ( $side=1, at=c(), label=c(), las=2$ )	Jau sukurtam grafikui leidžia pridėti stulpelių pavadinimus vienoje iš 4-ių grafiko pusių ( <i>side</i> ), nurodant koku eiliškumu bus surašyti pavadinimai ( <i>at</i> ), su <i>label</i> nurodant pavadinimus. Parametras <i>las</i> nurodo kintamųjų vardų padėtį ašies atžvilgiu: <i>las = 0</i> (rašoma lygiagračiai ašiai); <i>las = 2</i> (rašoma statmenai ašiai). <b>Pvz.:</b> <code>axis(side=1, at=c(1,2), labels=c("Automatine", "Rankine"), las=2)</code>

<b>grid</b> (n)	Grafiniam langui uždeda groteles iš $n \times n$ langelių.
<b>with</b> (data,f(x))	Apskaičiuojama duomenų rinkinio <b>data</b> , kintamojo <b>x</b> , funkcijos <b>f</b> reikšmė.  <b>Pvz.:</b> <b>with</b> (mtcars,mean(mpg))
<b>install.packages</b> ("e1071") <b>library</b> (e1071) <b>skewness</b> (x)	<b>Asimetrijos koeficientas</b>  $g_1 = \frac{m_3}{s^3} \qquad m_k = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^k$
<b>install.packages</b> ("e1071") <b>library</b> (e1071) <b>kurtosis</b> (x)	<b>Eksceso koeficientas</b>  $g_2 = \frac{m_4}{s^4} - 3$
<b>sample</b> (x, n, replace = FALSE, prob = NULL)	Komanda <b>sample</b> () atsitiktine tvarka, iš vektoriaus <b>x</b> atrenka <b>n</b> elementų be grąžinimo. Galima nurodyti tikimybės reikšmę <b>prob</b> .
<b>cor</b> (x,y)	Apskaičiuojamas kintamųjų x ir y koreliacijos koeficientas
<b>cor.test</b> (x,y)	Apskaičiuojamas kintamųjų x ir y koreliacijos koeficientas ir patikrinama hipotezė apie jo lygybę nuliui.
<b>lm</b> (y~x)	Apskaičiuojami tiesinės regresijos lygties koeficientai.
<b>abline</b> (lm(y~x))	Brėžiama regresijos tiesė.
<b>t.test</b> (x, mu = 5, alternative = c("two.sided", "less", "greater"), conf.level = 0.95)	Tikrinama parametrinė hipotezė jog populiacijos vidurkis = 5; < 5; > 5.  Kai pasiklovimo lygmuo Q = 0,95.
<b>t.test</b> (x, y, alternative = "two.sided", paired = FALSE, var.equal = TRUE, conf.level = 0.95)	Tikrinama parametrinė hipotezė apie kintamųjų x ir y vidurkių lygybę, nepriklausomoms imtims, kai dispersijos lygios, esant pasiklovimo lygmeniui Q = 0,95.
<b>density</b> (x)	neparametrinis histogramos glodinimas
<b>boxplot</b> ()	Brėžiama stačiakamių (dėžučių) diagrama
<b>attach</b> ()	Duomenų sistema (kintamieji, jų struktūra rinkinyje) prijungiama prie R paieškos kelio
<b>data</b> ()	R duomenų rinkiniui prijungti, bei R duomenų rinkinius pristatyti.
<b>data</b> ( <b>dat</b> )	Prie R paieškos kelio prijungiamas R duomenų rinkinys <b>dat</b>
<b>table</b> (cut(d3,int))	...