

Tartu Ülikool

Loodus- ja täppisteaduste valdkond

Matemaatika ja statistika instituut

Argo Rihkrand

**Õppeaine “Tõenäosusteooria ja matemaatiline
statistika” uurimine**

Matemaatilise statistika eriala

Bakalaureusetöö (9 EAP)

Juhendaja Natalja Lepik

Tartu 2018

Õppeaine “Tõenäosusteooria ja matemaatiline statistika” uurimine

Bakalaureusetöö

Argo Rihkrand

Lühikokkuvõte. Käesolevas bakalaureusetöös on vaadeldud õppeaasta 2016/2017 kevadsemestri ja 2017/2018 sügissemestri aine “Tõenäosusteooria ja matemaatiline statistika” osavõtnute tulemusi, ning analüüsitud, millest võisid tuleneda erinevused semestritel. On uuritud ka, mida lisaõppevahendi kasutuselevõtt sügissemestril võis kaasa tuua hinnete jaotusele, ning kas see võis parandada tudengite oskusi. Peale selle on töös analüüsitud küsitluse tulemusi, et selgitada välja tudengite arvamus lisaõppevahendisse.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: tõenäosusteooria, matemaatiline statistika, statistiline test, andmeanalüüs

Study of the course “Probability theory and mathematical statistics”

Bachelor’s thesis

Argo Rihkrand

Abstract. This thesis reviewed the results of students, who during the 2016/2017 spring or 2017/2018 autumn semester participated in the course “Probability theory and mathematical statistics” and analyzed possible reasons why differences occurred between two terms. In addition, it was viewed how additional learning tool introduced during the autumn semester affected grade distribution and whether it improved students’ skills. Lastly, based on the conducted survey, students’ opinion over additional learning tool was analyzed.

CERCS research specialisation: P160 Statistics, operations research, programming, actuarial mathematics

Keywords: probability theory, mathematical statistics, statistical test, data analysis

Sisukord

Sissejuhatus	5
1. Taust	6
2. Hinnete jaotus	9
2.1 Hinnete jaotuse võrdlus	9
3. Korrelatsioonid tulemuste vahel	11
3.1 Kevad 2016/2017 andmestik	12
3.2 Sügis 2017/2018 andmestik	13
4. Kontrolltööde keskmiste tulemuste võrdlus	15
4.1 Kontrolltööde keskmiste tulemuste võrdlus semestri siseselt	17
4.2 Kontrolltööde keskmiste tulemuste võrdlus semestrite lõikes	17
5. Kodu testide analüüs	19
5.1 Kodu testide tulemused ja tudengite osakaalud testide lahendamisel	19
5.2 Kahe katse erinevuste võrdlus	21
6. Küsitlus <i>DataCamp</i> lisapraktikumide kohta	23
6.1 Küsitluse läbiviimine	24
6.2 Küsitluse tulemused	24
Kokkuvõte	28
Kasutatud kirjandus	29
Lisad	30
Lisa 1. Hiiruut testi väljund hinnete jaotuse võrdlemiseks	30
Lisa 2. T-testi väljund kodu testide katsete tulemuste võrdlemiseks	31

Lisa 3. T-testi väljund kodu testide katsete aegade võrdluseks	38
Lisa 4. Küsitlus lisapraktikumide kohta <i>DataCamp</i> keskkonnas	45

Sissejuhatus

Bakalaureusetöös on vaadeldud õppeaine “Tõenäosusteooria ja matemaatiline statistika” õppetulemusi kahe semestri andmetelt, mis on kogutud 2016.-2018. aastal. Töö põhieesmärgiks on välja selgitada, millistest tunnustest sõltub aine hinne ja millised tunnused on omavahel seotud. Vaadeldavateks tunnusteks on võetud kõik aine lõplikku punktisummat kujundavad tegurid, nagu näiteks eksami punktisumma, kontrolltööde punktisumma jne.

Lähemalt on vaadeldud ka aine lõpphindeid semestritel, ning kontrolltööde tulemuste erinevusi. Lisaks on töös uuritud ka tudengite kodu testide lahendamise harjumusi. Analüüsitud on ka tudengitele loodud lisapunktide kogumise võimalust.

Antud bakalaureusetöö on jaotatud kuueks peatükiks, millest esimeses, on toodud ülevaade aine korralduse kohta. Järgnevalt on uuritud kahe töös vaatluse all oleva semestri hinnete jaotusi. Kolmandas peatükis on leitud korrelatsioonid erinevate vaadeldud tunnuste põhjal. Neljandas peatükis on uuritud, kas aines korraldatud kahe kontrolltöö tulemused on erinevad. Järgmises peatükis analüüsitakse tudengite igapäevase kodu teste. Viimaseks on kirjeldatud küsitluse tulemusi lisapraktikumide kohta.

Andmete analüüsimisel on kasutatud tarkvara *R* ning töö on vormistatud *R Markdown* paketi abil.

1. Taust

Antud töös on keskendutud õppeaasta 2016/2017 kevadsemestri ja 2017/2018 sügissemestri uurimisele. Sellel perioodil on olnud aine maht 6 ainepunkti. Mõlemal semestril oli aine korraldus sama. Igal nädalal toimus üks loeng ning üks praktikum. Loeng oli ühine kõikidele tudengitele, kuid praktikumid toimusid väiksemates, kuni 30 liikmelistes gruppides. Peale selle oli võimalus vabatahtlikult osaleda ka lisapraktikumis, mis toimus samuti iganädalaselt.

Aine lõppes kirjaliku eksamiga, millele eelnesid kaks kontrolltööd, üks semestri keskel ja teine lõpus. Eksam ja mõlemad kontrolltööd sisaldasid lävendit 50% ning olid kohustuslikud positiivse ainehinde saamiseks. Nendel töödel oli üks järeletegemise võimalus. Lisaks, tuli üliõpilastel peaaegu iga nädal sooritada test *Moodle* keskkonnas, kuid see polnud kohustuslik.

Kokkuvõttes kujunesid aine eest saadud punktid summana, mille liikmed olid järgnevad:

- kontrolltöö 1 (max 20p),
- kontrolltöö 2 (max 20p),
- eksam (max 50p),
- kodutööde keskmine (max 10p),
- lisapunktid.

Et hinnata, kuidas on tunnused omavahel seotud ning millised neist mõjutavad enim aine lõpphinnet, on välja valitud nn põhilised tunnused. Kevadsemestri andmestikus oli algselt 32 tunnust, millest on siin kasutatud 7 ning sügissemestri andmestikus 36 tunnust, millest vaadeldud 8. Andmestikud on saadud *Moodle* õppekeskkonna kaudu, kus vastavad tabelid on automaatselt genereeritud õppejõu poolt sisestatud andmete põhjal. Andmestikus on igal real ühe kursusest osavõtnud tudengi õpitulemused. Vaadeldud on järgmisi põhilisi tunnuseid:

- Hinne - aine lõplik punktisumma,
- Eksam - lõplik eksami tulemus, ehk parim tulemus eksami ja järeleksami seast,
- KT1 - esimese kontrolltöö lõplik tulemus,

- KT2 - teise kontrolltöö lõplik tulemus,
- Kodutööd - 12 *Moodle* testi punktide keskmine,
- Praktikumid - praktikumides saadud lisapunktid ülesannete lahendamise eest (iga praktikumi eest 0.5 punkti, kokku maksimaalselt 7 punkti),
- Lisapunktid - lisapunktide summa, praktikumide eest ja loengukonspekti näpuvigadest teatamise eest.

Täpsustuseks tuleb mainida, et aine lõplik punktisumma võib teoreetiliselt ületada ka 100 punkti lisapunktide tõttu.

Aine hinne sõltus saadud summaarsest punktiskoorist.

- A ehk suurepärane, kui oli saadud vähemalt 90 punkti,
- B ehk väga hea, kui vähemalt 80, kuid vähem kui 90 punkti,
- C ehk hea, kui vähemalt 70, kuid vähem kui 80 punkti,
- D ehk rahuldav, kui vähemalt 60, kuid vähem kui 70 punkti,
- E ehk kasin, kui vähemalt 50, kuid vähem kui 60 punkti,
- F ehk mitteametustatud, kui kogutud vähem kui 50 punkti.

Sügissemestril õppeaastal 2017/2018 võttis õppejõud kasutusele lisa õppevahendi veebis, keskkonnas *DataCamp* [1]. See pakkus tudengitele võimaluse teenida veel rohkem lisapunkte. Tunnus *DataCamp* on kuue veebi testi tulemuse summa, millest iga test andis maksimaalselt ühe punkti.

DataCamp on veebipõhine platvorm, mis pakub erinevaid andmeteadusele spetsialiseeritud kursusi ning õppevideoid. Kasutajad saavad harjutada programmeerimist ning andmeanalüüsi läbi oma veebilehitseja just neile sobival ajal. Põhiliselt pakutakse tarkvara *R* kursusi programmeerimise, visualiseerimise ning andmetöötlus oskuste arendamiseks, kuid on ka treeningvideoid ja õpetusi tarkvara *Python*, *SQL*, *Git* jaoks.

Platvormis on rakendatud interaktiivset õppemeetodit, kus kasutaja õpib ning programmeerib

samaaegselt. Ülesanded koosnevad praktilistest programmeerimisosadest, mille kõrval on toodud juhised lahendamiseks. Õppides antakse vahetult tagasisidet tulemustele ning soovi korral on võimalik vihjete abil ülesannet lihtsamaks muuta. Õppeaine “Tõenäosusteooria ja matemaatiline statistika” raames kasutasid tudengid õppejõu poolt tekitatud spetsiaalselt kohandatud materjale.

DataCamp on integreeritav *Moodle* keskkonda ja seetõttu sai praktikumidele ligi otse *Moodle*-st. Samuti läks praktikumi eest saadud tulemus otse *Moodle*-sse hinnete plokki.

2. Hinnete jaotus

Kevadsemestril 2016/2017 osales kursusel 283 tudengit. Järgnevas sagedustabelis on näha, kuidas jaotusid kursuse lõpphinded.

Tabel 1: 2016/2017 aine kuulajate hinnete jaotus

A	B	C	D	E	F
0.14	0.15	0.18	0.19	0.1	0.23

Kursuse sooritas positiivsele tulemusele 218 tudengit 283 seast, ning kursuse sooritas tulemusele F 65 tudengit, mis moodustab 23 % aine kuulajaskonnast.

Järgnevalt on iseloomustatud ainet, mida loeti sügissemestril 2017/2018. Kõikidest osavõtnust, keda oli kokku 144, sooritas kursuse positiivsele hindele 106 õpilast. Mittesooritanute protsent oli 26. Paneme tähele, et eelneval semestril osales pea poole rohkem tudengeid.

Tabel 2: 2017/2018 aine kuulajate hinnete jaotus

A	B	C	D	E	F
0.27	0.16	0.14	0.12	0.04	0.26

2.1 Hinnete jaotuse võrdlus

Kuna sügisel 2017 oli esmakordselt võetud kasutusele *DataCamp* lisapraktikum, siis pakkus huvi võimalik hinnete jaotuse muutus. Sel juhul vaadeldi kahe erineva semestri andmeid valimitena ja püstitati järgmine hüpoteeside paar:

H_0 : kevadsemestri ja sügissemestri hinded on sama jaotusega,

H_1 : kevadsemestri ja sügissemestri hinded on erineva jaotusega.

Semestrite tulemuste jaotuste võrdlemiseks kasutatakse χ^2 statistikut, juhul kui $x_1^{(1)}, x_2^{(1)}, \dots, x_k^{(1)}$ ja $x_1^{(2)}, x_2^{(2)}, \dots, x_k^{(2)}$ on sagedused esimeses ja teises valimis, k-ndas erinevas klassis, siis χ^2 statistik leitakse järgneva valemi abil:

$$\chi^2 = \sum_{i=1}^k \frac{(x_i^{(1)} - e_i^{(1)})^2}{e_i^{(1)}} + \sum_{i=1}^k \frac{(x_i^{(2)} - e_i^{(2)})^2}{e_i^{(2)}}, \quad (1)$$

kus

$$e_i^{(1)} = n_1 \frac{x_i^{(1)} + x_i^{(2)}}{n_1 + n_2},$$

$$e_i^{(2)} = n_2 \frac{x_i^{(1)} + x_i^{(2)}}{n_1 + n_2},$$

ja n_1 ning n_2 on vastavad valimimahud ([2], lk 167-168).

Kõikide töös uuritud testide juures on valitud olulisuse nivooks $\alpha = 0.05$. Eelnevalt kirjeldatud χ^2 testi tulemus on toodud lisas 1, mille kohaselt on statistikule $\chi^2 = 16,38$ vastav olulisuse tõenäosus $p = 0.005839$.

See ei tõesta *DataCamp*-i mõju hinnete jaotusele. Erinevus võib olla põhjustatud kuulajaskonna suurest hulgast ühel semestril või ka sellest, et aine õpetamisse olid kaasatud kahel semestril erinevad õppejõud või muudest põhjustest. Järgnevas peatükis on uuritud erinevate näitajate seoseid aine hindega.

3. Korrelatsioonid tulemuste vahel

Kuna kõik antud töös uuritud tunnused olid pidevad, siis tunnustevahelise korrelatsiooni kirjeldamiseks on kasutatud korrelatsioonimaatriksit. Kahjuks sisaldasid tulemuste andmetikud puuduvaid väärtusi, milleks võisid olla näiteks eksami ja teise kontrolltöö tulemus, kui tudeng ei ületanud vajalike lävendeid. Korrelatsioonimaatriksi leidmisel võib kasutada kahte lähenemisviisi: kas välja jätta uuringust kõikide nende üliõpilaste andmed, kellel on vähemalt üks tulemus puudu, või välja jätta andmed paarikaupa, viimane tähendab seda, et iga korrelatsiooni väärtuse leidmisel vaadeldakse vaid kahte tunnust korraga. See variant võimaldab kaasata rohkem andmeid ning andmed on omavahel paariviisiliselt täielikud [3]. Korrelatsiooni tugevust on hinnatud Pearsoni korrelatsioonikordaja väärtuse abil, kus lähtuti järgmisest reeglist. Kui korrelatsioonikordaja absoluutväärtus on

- suurem kui 0.7, siis on tugev korrelatsioon,
- suurem kui 0.5, väiksem kui 0.7, siis on keskmine korrelatsioon,
- suurem kui 0.3, väiksem kui 0.5, siis on nõrk korrelatsioon ([2], lk 190).

Kui korrelatsioonikordaja on positiivne, siis on tegemist samasuunalise, ehk positiivse korrelatsiooniga, kui kordaja on negatiivne, siis on tegemist vastassuunalise, ehk negatiivse korrelatsiooniga. Pearsoni korrelatsioonikordaja r_{xy} leitakse järgmise valemi abil:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2)$$

Valemis (2) vastab n valimimahule, kus tunnuste x ja y korral kõik puuduvate väärtustega read on välja jäetud ([2], lk 187).

3.1 Kevad 2016/2017 andmestik

Korrelatsiooni tabelist (vt tabel 3) näeme, millised tunnused on omavahel seotud rohkem kui teised, kevadsemestri andmete põhjal. Esmalt paneme tähele, et kõik vaadeldavad tunnused on omavahel positiivselt korreleeritud. Lõpliku hindega on kõige nõrgemini seotud praktikumide eest saadud punktid (0.38), keskmise tugevusega on seotud kodutööd, lisapunktid ja esimese kontrolltöö tulemus. Tugevat positiivset seost hindega näeme teise kontrolltöö hinde (0.78) ja eksami tulemusega (0.83).

Tabel 3: Tunnustevahelised korrelatsioonid aasta 2016/2017 kevadsemestri põhjal

	Hinne	Eksam	KT1	KT2	Kodutööd	Praktikumid	Lisap.
Hinne	1.00	0.83	0.59	0.78	0.62	0.38	0.63
Eksam	0.83	1.00	0.64	0.66	0.30	0.31	0.43
KT1	0.59	0.64	1.00	0.65	0.34	0.42	0.50
KT2	0.78	0.66	0.65	1.00	0.52	0.41	0.58
Kodutööd	0.62	0.30	0.34	0.52	1.00	0.23	0.60
Praktikumid	0.38	0.31	0.42	0.41	0.23	1.00	0.88
Lisap.	0.63	0.43	0.50	0.58	0.60	0.88	1.00

Kuna eksam omab lõpphindest 50% osakaalu, siis on tugev korrelatsioon mõnevõrra oodatav. Paneme tähele, et teise kontrolltöö tulemus on oluliselt suuremas korrelatsioonis aine hindega, kui esimese kontrolltöö tulemus. Teist kontrolltööd võis sooritada alles siis, kui esimene kontrolltöö oli positiivsele tulemusele sooritatud. See sõelus välja need üliõpilased, kes ainega hakkama ei saanud ja see võib põhjendada teise kontrolltöö tugevamat seost aine hindega. Lisapunktid on väga tugevalt positiivselt korreleeritud praktikumide punktidega (0.88), sest praktikumide punktid moodustavad suure osa lisapunktidest. Seega on üllatav, et lisapunktid on keskmiselt korreleeritud hindega (0.63), kuid praktikumid vaid nõrgalt positiivselt seotud

hindegaga (0.38).

Kirjaliku eksami tulemus on keskmise korrelatsiooniga mõlema kontrolltööga (vastavalt 0.64 ja 0.66). Kontrolltööd on omavahel seoses korrelatsioonikordajaga 0.65, mis näitab keskmist positiivset lineaarset korrelatsiooni. Tulemustest oleks oodanud veidi tugevamat positiivset korrelatsiooni kodutööde ning kontrolltööde tulemuste vahel (vastavalt 0.34 ja 0.52). Ka praktikumide ning kontrolltööde tulemuste vahel on vaid nõrk positiivne korrelatsioon (0.42 ja 0.41), kuigi head tulemused kodutöodes ning aktiivsus praktikumides võiksid anda eeldused sooritada ka kontrolltööd kõrgetele punktidele.

3.2 Sügis 2017/2018 andmestik

Tabelist 4 näeme, et aine lõpphinne on kõige tugevamalt korreleeritud teise kontrolltöö tulemusega (0.89) ning kodutööde tulemusega (0.73), tugev korrelatsioon on ka hinde ja esimese kontrolltöö (0.7) ning eksami vahel (0.71). Sarnaselt eelmise semestri tulemustele on jällegi teise kontrolltöö ja hinde vahel suurem korrelatsioon, kui esimese kontrolltöö ja hinde vahel. Mõõdukalt tugev korrelatsioon on aine hindel, tunnuse *DataCamp*-iga (0.66) ja lisapunktidega (0.45). Nõrgalt positiivselt korreleeritud on hinne praktikumide punktidega (0.29).

Eksami tulemus on keskmiselt positiivset korreleeritud kontrolltööde tulemustega (0.54 ja 0.47). Kontrolltööd on antud korrelatsioonitabelis seotud omavahel korrelatsioonikordajaga 0.63, mis näitab keskmist korrelatsiooni. Huvipakkuv tunnus *DataCamp* ei ole ühegi teise tunnusega tugevas korrelatsioonis. Tunnused *DataCamp* ja kodutööd on omavahel seotud keskmise positiivse korrelatsioonikordajaga 0.65. Huvitav on asjaolu, et tunnuse *DataCamp* ja eksami tulemuse vahel korrelatsioon peaaegu puudub (0.07), kuid hinde ja tunnuse *DataCamp* vahel keskmise tugevusega seos kordajaga 0.66.

Tabel 4: Tunnustevahelised korrelatsioonid aasta 2017/2018 sügissemestri põhjal

	Hinne	Eksam	KT1	KT2	Kodutööd	Praktikumid	Lisap.	DataCamp
Hinne	1.00	0.71	0.70	0.89	0.73	0.29	0.45	0.66
Eksam	0.71	1.00	0.54	0.47	0.17	0.10	0.29	0.07
KT1	0.70	0.54	1.00	0.63	0.47	0.13	0.36	0.36
KT2	0.89	0.47	0.63	1.00	0.70	0.19	0.37	0.55
Kodutööd	0.73	0.17	0.47	0.70	1.00	0.26	0.44	0.65
Praktikumid	0.29	0.10	0.13	0.19	0.26	1.00	0.14	0.38
Lisap.	0.45	0.29	0.36	0.37	0.44	0.14	1.00	0.38
DataCamp	0.66	0.07	0.36	0.55	0.65	0.38	0.38	1.00

Mõlemast andmestikust tuli välja, et aine hinne on tugevalt seotud eksamitulemusega, ning ka teise kontrolltööga. Esimese kontrolltööga oli seos nõrgem. Ka kodutööde ja hinde vahel oli märgatav seos.

4. Kontrolltööde keskmiste tulemuste võrdlus

Kontrolltööde keskmiste erinevus annab aimu sellest, kas kontrolltööd võisid olla erineva raskusega ja kas tudengitel võis olla mõnest teemast arusaamisega rohkem probleeme. Kui kontrolltööde keskmistes ei oleks erinevusi, siis võiks sellest välja lugeda, et tudengite tase oli mõlema kontrolltöö puhul sarnane. Ühe aine piires 1. ja 2. kontrolltöö keskmiste uurimiseks on kasutatud sõltuvate valimitega t-testi.

Kuna teemad kontrolltöödes on aastast aastasse samad, siis on vaadeldud sügis- ja kevadsemestri andmeid kui valimeid. Ning üldistame t-testi tulemusi kõikidele nendele semestritele, kus teemad esimese ja teise kontrolltöö juures on olnud samad, mis töös uurimise all olevatel semestritel.

Järgnevalt on sõnastatud hüpoteeside paar kontrollimaks, kas õnnestub näidata kontrolltööde erinevust olulisuse nivool 0.05:

$$H_0 : \mu = 0,$$

$$H_1 : \mu \neq 0,$$

kus μ on esimese ja teise kontrolltöö vahe keskmine. Statistika t-testi korral leitakse järgneva valemi abil,

$$t = \frac{\bar{x}_D}{\frac{s_D}{\sqrt{n}}}, \quad (3)$$

kus \bar{x}_D esimese ja teise kontrolltöö keskmine vahe valimis ja s_D on esimese ja teise kontrolltöö vahede standardhälve valimis [4].

Et kontrollida, kas kontrolltööde keskmised on erinevad kahel eraldi semestril, on kasutatud sõltumatute valimite t-testi. Selle abil on kontrollitud kahte hüpoteeside paar, esiteks uurimaks

kas 1. kontrolltöö keskmised erinevad sügis- ja kevadsemestril,

$$H_0 : \mu_{1,1} = \mu_{1,2},$$

$$H_1 : \mu_{1,1} \neq \mu_{1,2}.$$

Ning teiseks kas 2. kontrolltöö keskmised erinevad sügis- ja kevadsemestril,

$$H_0 : \mu_{2,1} = \mu_{2,2},$$

$$H_1 : \mu_{2,1} \neq \mu_{2,2}.$$

Siin tähistab esimene indeks kontrolltöö numbrit (1. või 2.) ja teine indeks vaadeldavat semestrit, kus 1 tähistab sügissemestrit ja 2 kevadsemestrit.

Kuna kaks semestrit erinevad selle poolest, et sügissemestril on kasutusel *DataCamp* lisapraktikumid, kuid kevadsemestril mitte. Seega üldistame t-testi tulemusi kogumitele, kus on lisamaterjalina kasutatud *DataCamp* lisapraktikume, ning juhule kus neid ei ole kasutatud. On vaadeldud seega kas lisapraktikumide kasutuselevõtmine on parandanud kontrolltööde tulemusi.

Kahe semestri keskmiste võrdlemiseks on kasutatud sõltumatute valimite t-testi eeldades võrdseid dispersioone vastavates üldkogumites. T-testi statistik on arvutatud valemiga,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{u*}}, \quad (4)$$

kus \bar{x}_1 ja \bar{x}_2 on kahe valimi, mahtudega n_1 ja n_2 keskmised ja s_{u*} on leitav järgnevalt ([2], lk 133):

$$s_{u*} = s \sqrt{\frac{n_1 + n_2}{n_1 n_2}},$$
$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

4.1 Kontrolltööde keskmiste tulemuste võrdlus semestri siseselt

On tehtud kaks t-testi sõltuvate andmete korral iga semestri jaoks eraldi. T-testi väljundi põhjal (vt tabel 5) saame vastu võtta sisuka hüpoteesi ning, seega sügissemestril 2017/2018 erinevad kontrolltööde keskmised, ning teise kontrolltöö tulemus on 0.8634 punkti võrra suurem esimesest kontrolltööst.

Tabel 5: 2017/2018 aine kontrolltööde t-test

	t
Test statistik	-4.40816
p väärtus	1.535e-05
Keskmiste erinevus	-0.8634241
Alternatiivne hüpotees	kahepoolne

Tehes sama testi kevadsemestri 2016/2017 kontrolltööde tulemuste põhjal, peame jääma nullhüpoteesi juurde, ning ei suuda tõestada kontrolltööde keskmiste erinevust (vt tabel 6).

Tabel 6: 2016/2017 aine kontrolltööde t-test

	t
Test statistik	-0.6996249
p väärtus	0.4855385
Keskmiste erinevus	-0.2151261
Alternatiivne hüpotees	kahepoolne

Kahe semestri teised kontrolltööd olid küll keskmiselt suurema punktiskooriga, kuid vaid sügissemestri tulemuste põhjal saaksime üldistada erinevust laiemalt ja väita, et teised kontrolltööd on suurema keskmisega kui esimesed kontrolltööd.

4.2 Kontrolltööde keskmiste tulemuste võrdlus semestrite lõikes

Uurides 1. kontrolltöö tulemuste keskmist vastavalt kevad- ja sügissemestril (vt tabel 7), saab vastu võtta sisuka hüpoteesi, ning on tõestatud, et 1. kontrolltöö keskmine tulemus

sügissemestril on 0.757 punkti võrra suurem, kui kevadsemestri keskmine tulemus.

Tabel 7: Kevad- ja sügissemestri 1. kontrolltöö t-test

	t
Test statistik	-2.0336
p väärtus	0.04309
Keskmete erinevus	-0.75745
Alternatiivne hüpotees	kahepoolne

Teise kontrolltöö juures ei saa vastu võtta sisukat hüpoteesi, et teise kontrolltöö tulemused erinesid (vt tabel 8). Kuid keskmete erinevus vaatluse all olevates tulemustes, annab teada, et sügissemestri 2. kontrolltöö keskmine on kevadsemestri keskmisest 0.52 punkti võrra suurem.

Tabel 8: Kevad- ja sügissemestri 2. kontrolltöö t-test

	t
Test statistik	-1.2443
p väärtus	0.2145
Keskmete erinevus	-0.52132
Alternatiivne hüpotees	kahepoolne

Võimalik et *DataCamp* lisapraktikumid annavad tõepoolest esimese kontrolltöö keskmise tulemuse punktiskoori juurde, kuid sellest hoolimata, ei ole võimalik tõestada, et lisapraktikumidel on põhjuslik seos kontrolltööde tulemuste edukamaks muutmisel. Seda enam, et kevad- ja sügissemestri 2. kontrolltöö t-test ei suutnud tõestada statistilist erinevust.

5. Kodu testide analüüs

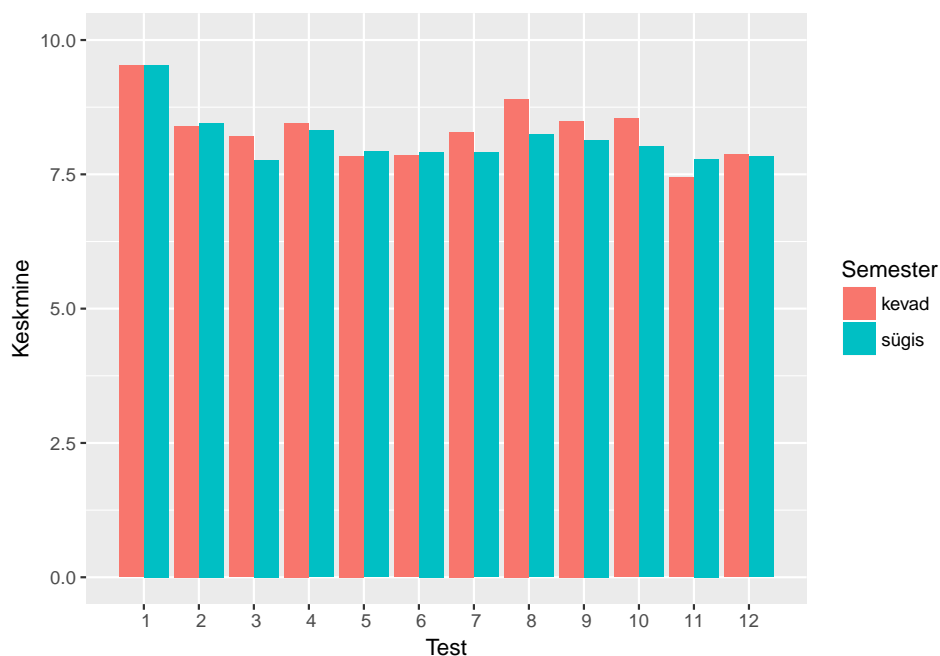
Järgnevalt on soovitud uurida, kuidas lahendavad tudengid kodu teste keskkonnas *Moodle*. Iga test oli üliõpilasele *Moodle* poolt genereeritud kord nädalas (kokku 12 testi), ning iga test andis 10 punkti, mille pealt aine lõplikku skoori läheb arvesse testide aritmeetiline keskmine. Kuna iga testi korral oli kasutada kaks katset, siis on võimalik vaadata, kui palju erineb esimese katse tulemus teisest. Testi tulemusena läks kirja kahest katsest parim.

Antud peatükis on uuritud, kumma katse peale kulus keskmiselt rohkem aega, ning kumb tulemus oli keskmiselt edukam. Lisaks on ka vaadeldud testide lõikes keskmisi tulemusi ja kursusest osavõtnud tudengite osakaalu testide lahendamisel kahel semestril.

5.1 Kodu testide tulemused ja tudengite osakaalud testide lahendamisel

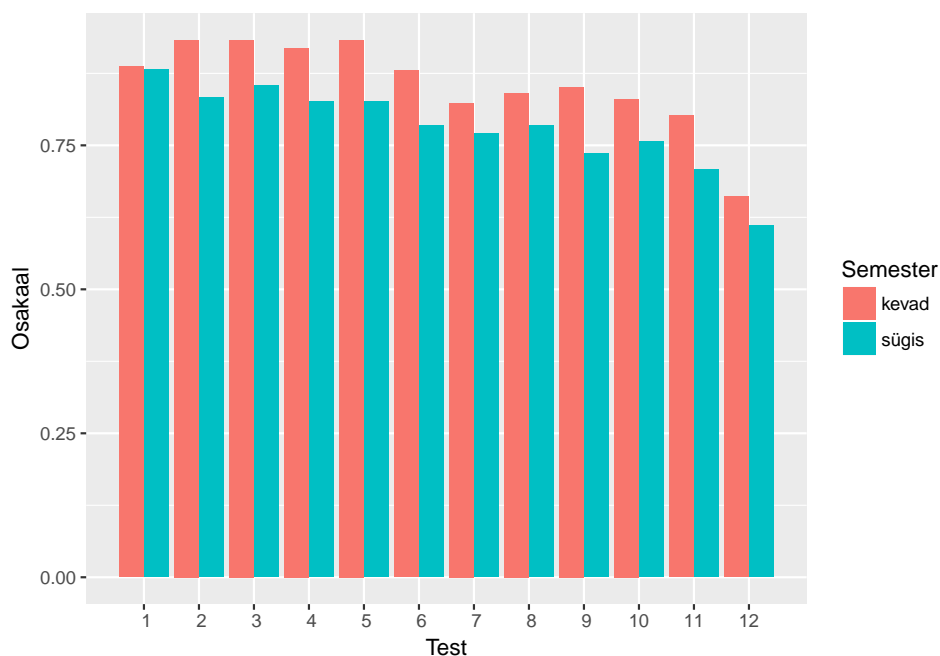
Esmalt on vaadeldud kahe semestri kodu testide keskmisi tulemusi testide kaupa. Iga testi juures on arvatatud selle testi keskmine testist osavõtnud tudengite tulemuste pealt, seega on arvestatud neid tulemusi, kus tudeng kogus rohkem kui 0 punkti (vt joonis 1). On näha, et mõlemal semestril lahendavad tudengid esimese testi keskmiselt kõige suuremale punktiskoorile. Siiski jooniselt 1 ei tule välja trendi, kus tulemused oleksid ühel semestril suuremad kui teisel, ning ka semestri alguse ja lõpu poole ei ole näha suuri erinevusi. Pigem on näha, millised testid on tudengite jaoks raskemad.

Joonis 1: Kodu testide keskmised tulemused kevad- ja sügissemestril



Hindamaks kursusest osavõtnud tudengeid testide lahendamisel, on moodustatud joonis 2, kus näeb teste lahendanud tudengite osakaalu kursusele registreerunutest. Tuleb välja, et kevadsemestril lahendas suurem osakaal registreerunutest teste, kui sügissemestril, üks põhjus sellele võib tulla ka sellest, et kevadsemestril oli pea poole rohkem registreerunuid, kui sügissemestril ning seega võis kevadisele kursusele sattuda veidi väiksem osakaal selliseid tudengeid, kes ei lahendanud teste piisava huviga. Teisalt võisid ka *DataCamp* lisapraktikumid pakkuda osadele tudengitele sobilikumat võimalust punktide teenimiseks, kui kodu testid. Peale selle on näha, et üldiselt semestri lõpupoole väheneb tudengite osakaal testide lahendamisel (vt joonis 2).

Joonis 2: Kursusest osavõtnute osakaal kodu testide lahendamisel kevad- ja sügissemestril



5.2 Kahe katse erinevuste võrdlus

Tulemuste analüüsimiseks on iga testi kohta eraldi andmestik, mis sisaldab igal real ühe tudengi ühe testi katse kohta infot, kokku 20 veergu ning igas andmestikus erinev arv ridu sõltuvalt sellest, kui mitu tudengit mitut katset sooritas. Vaadeldud on vaid õppeaasta 2016/2017 kevadsemestri andmeid. Huvipakkuvad tunnused on,

- hinne, mis näitab katse tulemust (maksimaalselt 10 punkti),
- ajakulu, mis näitab kui palju aega kulus katse peale minutites (teoreetiliselt maksimaalselt väärtuseks üks nädal).

Kuna andmestikus on igal real info ühe katse kohta, siis andmetabelit tuli esmalt teisendada kujule, kus igas reas on ühe tudengi kahe võimaliku katse tulemused. Lisaks tuli töödelda tunnust ajakulu, kuna *Moodle* väljastatud andmetabelis on antud testile kulutatud aeg sõnena, mis sisaldab päevi, tunde, minuteid ja sekundeid, seda on teisendatud numbriliseks tunnuseks, mis väljendab testi lahendamiseks kulunud aega minutites. Kuna testide lahendamisel oli

ajapiirang üks nädal, siis oli osa tudengitest jätnud testi lahti mitmeks päevaks, ilma, et nad oleksid seda testi mitu päeva lahendanud. Seega usaldusväärsemate tulemuste saamiseks on ajakulu uurimisel eemaldatud read, kus ajakulu on suurem või võrdne kuue tunniga.

Viies läbi t-testi (vt valem 3) võrdlemaks 12 testi esimese ja teise katse keskmisi, on leitud, et kõigi 12 testi teise katse keskmine oli suurem, kui esimese katse keskmine (vt lisa 2). Ning võrreldes 12 testi katsete aegade keskmisi, saime kõikide testide, välja arvatud 9. testi juures vastu võtta hüpoteesi, mis ütles, et esimese katse keskmine ajakulu on suurem, kui teise katse oma (vt lisa 3).

Selle põhjal oleks loogiline väita, et tudengid kasutavad esimest katset pigem harjutamiseks ning arvestavad, et neil on kasutada veel ka teine katse. Küllap on esimesest katsest abi, sest vastamisel näidatakse ka tudengitele õiget vastust. Seega teist katset, kus võivad olla sarnased ülesanded nagu esimesel katsel, lahendatakse kiiremini ja kogutakse juba rohkem punkte.

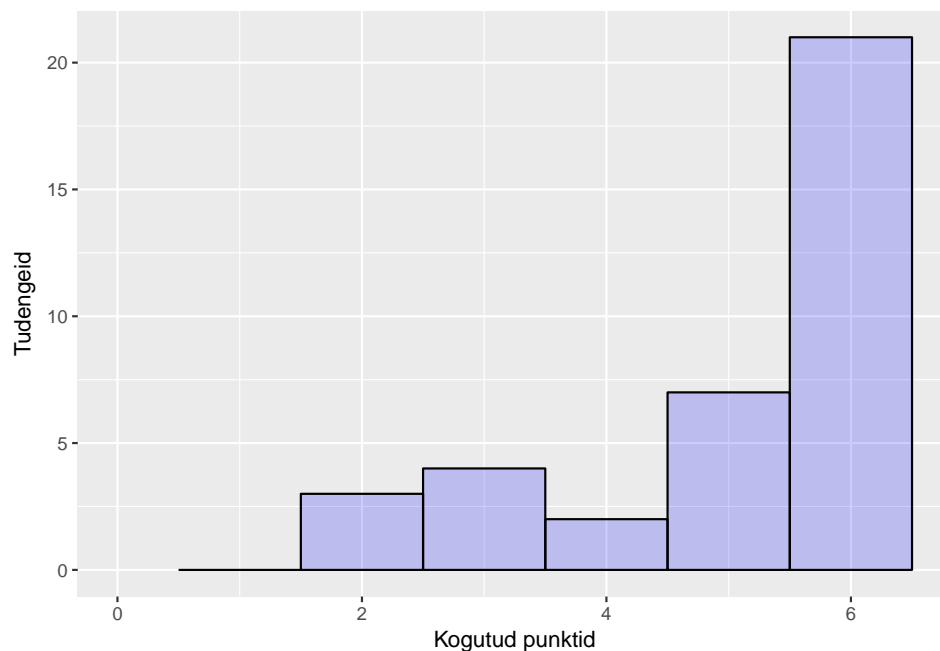
6. Küsitlus *DataCamp* lisapraktikumide kohta

Sügissemestril 2017/2018, oli tudengitel võimalus teenida lisapunkte *DataCamp* veebilehel lahendades õppejõu poolt loodud praktilisi ülesandeid *R*'i vahenditega. Et hinnata, kui kasulikud tundusid üliõpilastele *DataCamp* lisapraktikumid, viidi läbi ankeetküsitlus tudengite seas (vt lisa 4).

Küsitluse eesmärgiks oli välja selgitada, kuidas hindasid tudengid *DataCamp*-i, lisaks kui palju kulus lisapunktide teenimiseks aega ning, kas tudengite arvates võiksid sellised lisapraktikumid olla kohustuslikud.

Kuna tabelitest 1 ja 2 oli näha, et tudengite osakaal, kes said hinde A oli sügissemestril 0.13 võrra suurem kui kevadsemestril, siis on vaadeldud kui palju punkte andsid *DataCamp* lisapraktikumid nendele 39 tudengile juurde, kes said hindeks A. On näha, et 21 tudengit kogusid tänu lisapraktikumidele üle 5.5 lisapunkti. See moodustab ligi 54% hindeg A tudengitest (vt joonis 3).

Joonis 3: 2017/2018 aine hindeg A tudengite *DataCamp* lisapunktide histogramm



Ilma *DataCamp* lisapunktideta oleks hinnete jaotus sarnasem kevadsemestri tulemustele (vt tabel 9), ning χ^2 testi tulemusena ei saaks tõestada, et kevadsemestri ja sügissemestri hinded on erineva jaotusega (vt lisa 1).

Tabel 9: 2017/2018 aine kuulajate hinnete jaotus ilma *DataCamp* lisapunktideta

A	B	C	D	E	F
0.16	0.19	0.15	0.15	0.08	0.27

6.1 Küsitluse läbiviimine

Küsitlus viidi läbi internetis ning oli vabatahtlik, vastuseid laekus 29-lt tudengilt. Küsitlus oli mõeldud aine üliõpilastele, kes olid proovinud lahendada vähemalt ühte lisapraktikumi. Et analüüsida tudengite huvi *DataCamp*-i vastu on uuritud töös järgnevaid küsimusi:

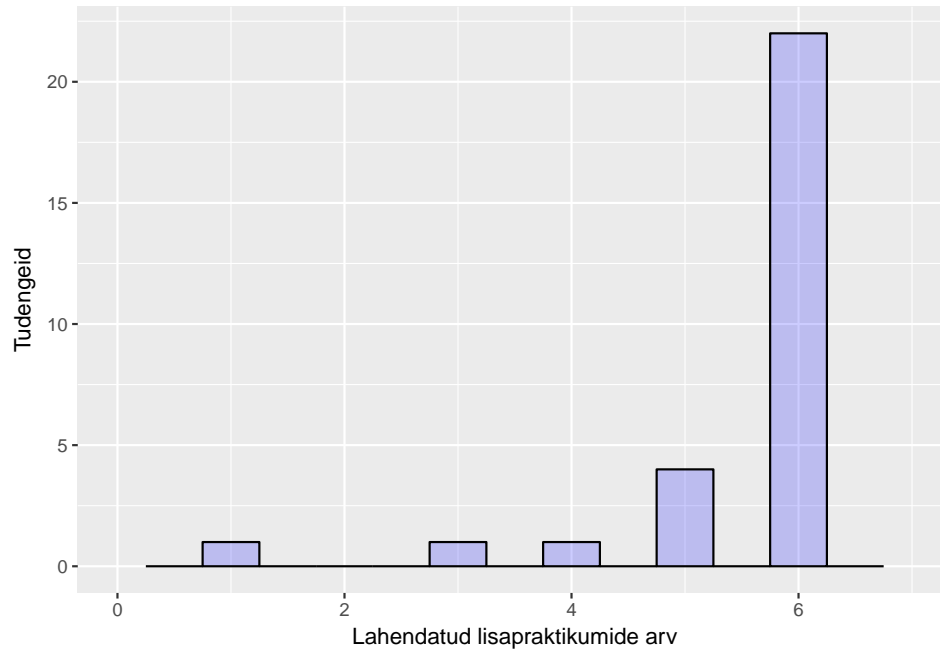
- Mitu lisapraktikumi *DataCamp* is oled lahendanud?
- Kas oled varem *R*-ga kokku puutunud (st enne ainet)?
- Kui kaua läks sul aega viimasena tehtud lisapraktikumi lahendamiseks? (vali kõige sobivam variant)
- Kas harjutused olid sõnastatud selgelt ja arusaadavalt?
- Kas on juhtunud, et sinu vastus oli õige, aga tehnilistel põhjustel luges *DataCamp* vastuse valeks?
- Kas need praktikumid võiksid kohustuslikud olla?

6.2 Küsitluse tulemused

Küsimusele, mitu lisapraktikumi *DataCamp*-is oled lahendanud, saab leida ka huvipakkuva kogumi reaalsed tulemused sügissemestri 2017/2018 andmestikust. Seega saab võrrelda, kuidas jaotusid vastanute lahendatud praktikumide arvud ja kui palju kõik kursusest osavõtnud

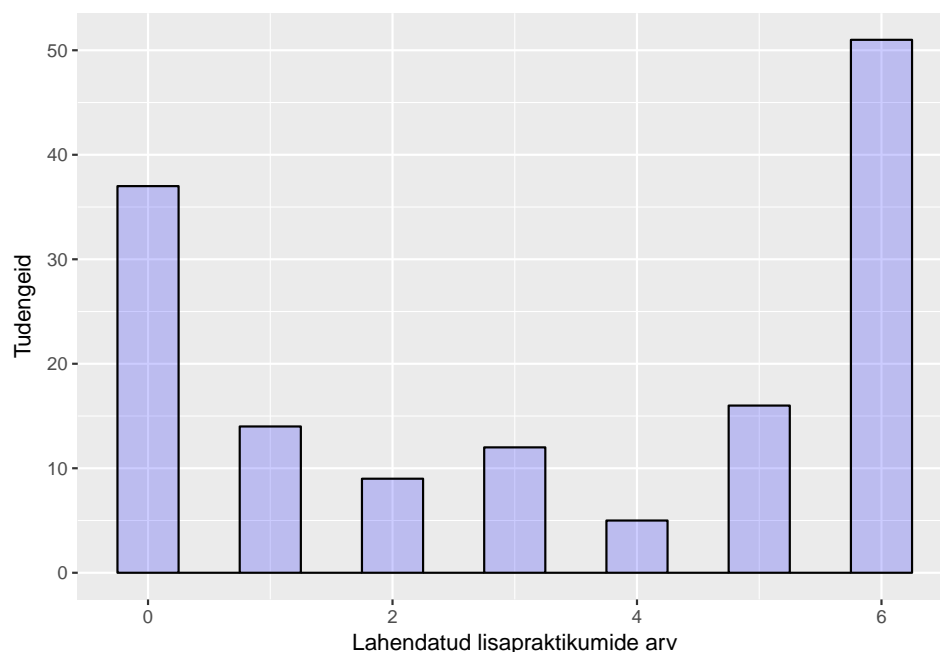
tudengid praktikume tegelikult lahendasid. Küsitlusest tuli välja, et 29-st vastanust 22 lahendasid ära kõik 6 praktikumi, moodustades ligi 76% küsitluses osalenutest (vt joonis 4).

Joonis 4: Vastanute lahendatud praktikumide arvu histogramm



Selgub, et lahendatud lisapraktikumide jaotus on veidi teistsugune kui valimi põhjal leitud, kuna tegelikkuses lahendas 6 praktikumi ära 51 tudengit, mis moodustab ligi 48% tudengitest, kes on lahendanud ära vähemalt ühe praktikumi (vt joonis 5). Peale selle 37 tudengit ei lahendanud ühtegi praktikumi, mis moodustab ligi 26% ainele registreerunutest. Seega võib väita, et küsitlusest osavõtnute seas oli suur osakaal aktiivsematel lisapraktikumi lahendajatel, kes võivad olla meelestatud vastama positiivselt.

Joonis 5: Kursusest osavõtnud tudengite lahendatud praktikumide arvu histogramm



Selgub, et vaid 7% küsitluses osalenust oli ka enne lisapraktikumide lahendamist kokku puutunud tarkvaraga R . Tulemus on üsna oodatav, sest suur hulk aine kuulajatest on esmakursuslased, kes ei ole ühegi teise aine raames R -iga kokku puutunud. Seejuures on positiivne, et paljud tudengid saavad esmase kokkupuute vajaliku tarkvaraga kohe 1. aastal.

Et saada ülevaade, kui palju aega kulub tudengitel lisapunktide teenimisele, uuriti täpsemalt, kui palju aega kulus viimase lisapraktikumi lahendamiseks aega. Valikvariantide vastustest on näha, et kõige suurema osakaaluga, tervelt 62% tudengitest, kulutas viimase lisapraktikumi sooritamiseks 30 minutit kuni 1 tunni. Lisaks 24% tudengitel kulus veelgi rohkem aega ning 14% vähem aega (vt tabel 10).

Tabel 10: Viimase praktikumi lahendamisele kulunud aegade osakaal valimis

Vähem kui 0.5 tundi	0.5 - 1 tund	1 - 1.5 tundi	1.5 - 2 tundi
0.14	0.62	0.17	0.07

Õppejõule on oluline teada, kas ülesanded on püstitatud tudengite arvates selgelt. Tuleb välja, et 90% tudengitest said enamasti ülesannetest aru, kuid 10% tudengitele jäid umbes pool ülesannetest arusaamatuks. Kahjuks ei jätnud tudengid konkreetset tagasisidet selle kohta, mis jäi neile arusaamatuks.

Küsimusele, kas on juhtunud, et sinu vastus oli õige, aga tehnilistel põhjustel luges *DataCamp* vastuse valeks, vastas 83% tudengitest, et neil ei ole ette tulnud tehnilist probleemi, kuid 14% tudengitest on loetud tehnilistel põhjustel paar korda õige vastus valeks, ning 3% tudengitest on see juhtunud üle kahe korra (vt tabel 11).

Tabel 11: Tehniliste probleemide esinemiste osakaal valimis

Ei	Paar korda	Rohkem kui 2 korda
0.83	0.14	0.03

Viimaseks on kirjeldatud seda, kuidas tudengid hindavad, kas *DataCamp* lisapraktikumid võiksid olla kohustuslikud. Selgus, et suurem osa vastanutest 79% arvab, et lisapraktikumid ei peaks tulevikus olema siiski kohustuslikud, ning ülejäänud 21% vastanud tudengitest arvab, et lisapraktikumid võiksid olla kohustuslikud.

Kokkuvõte

Käesolevas töös analüüsiti ja kirjeldati õppeaasta 2016/2017 kevadsemestri ja 2017/2018 sügissemestri tulemusi. Lõpphinnete jaotus oli semestritel erinev, ning sügissemestril oli suurem osakaal hindel A. Mõlema semestri puhul leiti, et aine hinne on tugevalt seotud eksamitulemusega, ning ka teise kontrolltööga, kusjuures esimesel kontrolltööl oli seos hindega nõrgem. Ka kodutööde ja lõpliku hinde vahel oli märkimisväärne seos.

Kontrolltööde keskmiste erinevusi uurides selgus, et teise kontrolltöö keskmine tulemus oli mõlemal semestril suurem, kui esimese kontrolltöö keskmine, kuid tulemusi ei saa üldistada kõikide loetud aine semestrite jaoks. Kõige suurem erinevus oli sügissemestril 2017/2018. Lisaks ei saa ka väita, et *DataCamp* lisapraktikumide kasutuselevõtt parandaks kontrolltööde keskmisi tulemusi.

Kodu testide analüüsis selgus, et mõlemal semestril langeb pea igal järgneval testil tudengite arv, kes teste lahendasid. Uurides kevadsemestril ka katsete erinevusi leiti, et tudengid kasutavad esimest katset pigem harjutamiseks ning teist katset lahendatakse kiiremini kui esimest ja kogutakse ka rohkem punkte.

Üliõpilaste tagasiside lisapraktikumide kohta oli valdavalt positiivne, sest ülesannete lahendamine mõõdukas ilma tehniliste ja sisuliste probleemideta. Kuid vastanute valim erines kursuses osavõtnutest kogutud lisapunktide jaotuse poolest. Sellegipoolest valdav enamus küsitlusele vastanutest arvab, et praktikumid ei peaks tulevikus olema kohustuslikud.

Autor loodab, et töö panustab tulevikus õppeaine “Tõenäosusteooria ja matemaatiline statistika” kvaliteeti. Lisaks annab ülevaate tudengite tulemustest ning aitab mõista detailsemalt, kuidas kujuneb lõpphinne.

Kasutatud kirjandus

- [1] *DataCamp* (2018). *Platformi DataCamp avaleht*, <https://www.datacamp.com/>, kasutatud (15.05.2018).
- [2] A.-M. Parring, M. Vähi, E. Käärrik (1997). *Statistilise andmetöötuse algõpetus*. Tartu. Tartu Ülikooli kirjastus.
- [3] T. J. Leeper (2013). *Missing data handling*. <https://thomasleeper.com/Rcourse/Tutorials/NAhandling.html>, kasutatud (15.05.2018).
- [4] I. Traat (2006). *Matemaatilise statistika põhikursus*. Tartu. Tartu Ülikooli kirjastus, lk 128.

Lisad

Lisa 1. Hiiruut testi väljund hinnete jaotuse võrdlemiseks

```
## Hiiruut test kevad- ja sügissemestri hinnete jaotuse võrdluseks
##
## Pearson's Chi-squared test
##
## data:  cbind(jaotus_kevad, jaotus_sygis)
## X-squared = 16.38, df = 5, p-value = 0.005839
##
## Hiiruut test kevad- ja sügissemestri hinnete jaotuse võrdluseks,
##           kui sügissemestril poleks saanud DataCamp lisapunke
##
## Pearson's Chi-squared test
##
## data:  cbind(jaotus_kevad, jaotus_sygis_ilmaDC)
## X-squared = 3.6804, df = 5, p-value = 0.5963
```

Lisa 2. T-testi väljund kodu testide katsete tulemuste võrdlemiseks

```
## #####  
## Testi nr 1  
## 1. ja 2. katse tulemuse võrdlus  
## Paired t-test  
##  
## data: Katse1 and Katse2  
## t = -10.108, df = 159, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##      -Inf -1.003419  
## sample estimates:  
## mean of the differences  
##      -1.199812  
##  
## #####  
## Testi nr 2  
## 1. ja 2. katse tulemuse võrdlus  
## Paired t-test  
##  
## data: Katse1 and Katse2  
## t = -7.5366, df = 180, p-value = 1.136e-12  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##      -Inf -1.095459  
## sample estimates:
```

```

## mean of the differences
##          -1.403315
##
## #####
## Testi nr 3
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -7.2086, df = 210, p-value = 5.016e-12
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.033506
## sample estimates:
## mean of the differences
##          -1.340806
##
## #####
## Testi nr 4
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -3.8466, df = 179, p-value = 8.318e-05
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.5948732

```



```

## sample estimates:
## mean of the differences
##          -1.043333
##
## #####
## Testi nr 5
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -8.1724, df = 198, p-value = 1.769e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.56879
## sample estimates:
## mean of the differences
##          -1.966432
##
## #####
## Testi nr 6
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -6.6834, df = 187, p-value = 1.303e-10
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:

```

```

##      -Inf -1.34555
## sample estimates:
## mean of the differences
##           -1.787713
##
## #####
## Testi nr 7
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -4.0201, df = 157, p-value = 4.504e-05
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.7644904
## sample estimates:
## mean of the differences
##           -1.299241
##
## #####
## Testi nr 8
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -3.9723, df = 135, p-value = 5.764e-05
## alternative hypothesis: true difference in means is less than 0

```

```

## 95 percent confidence interval:
##          -Inf -0.8295719
## sample estimates:
## mean of the differences
##          -1.422794
##
## #####
## Testi nr 9
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -3.0098, df = 158, p-value = 0.001522
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##          -Inf -0.4134363
## sample estimates:
## mean of the differences
##          -0.9181761
##
## #####
## Testi nr 10
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -4.1556, df = 136, p-value = 2.852e-05

```

```

## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.7982889
## sample estimates:
## mean of the differences
##      -1.327226
##
## #####
## Testi nr 11
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2
## t = -8.6134, df = 164, p-value = 2.818e-15
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -2.172504
## sample estimates:
## mean of the differences
##      -2.688909
##
## #####
## Testi nr 12
## 1. ja 2. katse tulemuse võrdlus
## Paired t-test
##
## data:  Katse1 and Katse2

```

```
## t = -4.2557, df = 125, p-value = 2.026e-05
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.047581
## sample estimates:
## mean of the differences
##      -1.715635
```

Lisa 3. T-testi väljund kodu testide katsete aegade võrdluseks

```
## #####  
## Testi nr 1  
## 1. ja 2. katse aja võrdlus  
## Welch Two Sample t-test  
##  
## data: Katse1 and Katse2  
## t = 5.5624, df = 311.29, p-value = 2.867e-08  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 11.21233      Inf  
## sample estimates:  
## mean of x mean of y  
## 26.73186 10.79191  
##  
## #####  
## Testi nr 2  
## 1. ja 2. katse aja võrdlus  
## Welch Two Sample t-test  
##  
## data: Katse1 and Katse2  
## t = 5.9749, df = 379.43, p-value = 2.653e-09  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 22.25456      Inf  
## sample estimates:
```

```

## mean of x mean of y
## 66.79907 36.06213
##
## #####
## Testi nr 3
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data: Katse1 and Katse2
## t = 4.8582, df = 381.76, p-value = 8.662e-07
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 15.44188      Inf
## sample estimates:
## mean of x mean of y
## 59.53048 36.15500
##
## #####
## Testi nr 4
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data: Katse1 and Katse2
## t = 3.658, df = 351.73, p-value = 0.0001465
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 10.28231      Inf

```

```

## sample estimates:
## mean of x mean of y
## 62.19889 43.47514
##
## #####
## Testi nr 5
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data: Katse1 and Katse2
## t = 5.5942, df = 402.27, p-value = 2.052e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 21.81884      Inf
## sample estimates:
## mean of x mean of y
## 75.13532 44.19942
##
## #####
## Testi nr 6
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data: Katse1 and Katse2
## t = 4.8529, df = 384.41, p-value = 8.86e-07
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:

```



```

## 17.60417      Inf
## sample estimates:
## mean of x mean of y
## 68.42390 41.76052
##
## #####
## Testi nr 7
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2
## t = 3.7532, df = 328.36, p-value = 0.0001032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 12.94306      Inf
## sample estimates:
## mean of x mean of y
## 69.72275 46.63116
##
## #####
## Testi nr 8
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2
## t = 3.9398, df = 302.63, p-value = 5.066e-05
## alternative hypothesis: true difference in means is greater than 0

```

```

## 95 percent confidence interval:
## 13.37102      Inf
## sample estimates:
## mean of x mean of y
## 56.41471 33.40955
##
## #####
## Testi nr 9
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2
## t = 0.70023, df = 215.56, p-value = 0.2423
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -5.543416      Inf
## sample estimates:
## mean of x mean of y
## 52.28592 48.20739
##
## #####
## Testi nr 10
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2
## t = 2.8102, df = 255.09, p-value = 0.002667

```

```

## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 6.586911      Inf
## sample estimates:
## mean of x mean of y
## 52.19286 36.22672
##
## #####
## Testi nr 11
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2
## t = 3.1345, df = 318.31, p-value = 0.000941
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 9.178338      Inf
## sample estimates:
## mean of x mean of y
## 62.54841 43.17291
##
## #####
## Testi nr 12
## 1. ja 2. katse aja võrdlus
## Welch Two Sample t-test
##
## data:  Katse1 and Katse2

```

```
## t = 2.0216, df = 173.53, p-value = 0.02238
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 3.500825      Inf
## sample estimates:
## mean of x mean of y
## 66.90462 47.66679
```

Lisa 4. Küsitlus lisapraktikumide kohta *DataCamp* keskkonnas

Mitu lisapraktikumi *DataCampis* oled lahendanud? (Kirjuta arv) (0 - 6)*

Kas oled varem *R*-ga kokku puutunud (st enne ainet)?*

- Jah
- Ei

Kui kaua läks Sul aega viimasena tehtud lisapraktikumi lahendamise? Kui tegid natuke, aga erinevatel päevadel, siis leia summaarne aeg, mis kulus kogu lisapraktikumi lahendamisele.

Vali kõige sobivam variant:*

- < 0,5 tundi
- 0,5 t kuni 1 t
- 1 t kuni 1,5 t
- 1,5 t kuni 2 t
- 2 t kuni 2,5 t
- 2,5 t kuni 3 t
- Üle 3 tunni (kirjuta oma aeg järgmise küsimuse vastuseks)

Kui vastasid eelmises küsimuses, et Sul läks viimasena tehtud lisapraktikumi lahendamise aega üle 3 tunni, siis kirjuta siia, mitu tundi Sul kulus:

Kas harjutused olid sõnastatud selgelt ja arusaadavalt?*

- Enamasti jah.
- Umbes pooled olid sõnastatud selgelt ja pooled ebaselgelt.
- Enamasti olid sõnastatud ebaselgelt.
- Raske oli harjutustest aru saada.

Kui vastasid eelmises küsimuses, et harjutustest oli raske aru saada, siis täpsusta, miks:

Kas on juhtunud, et sinu vastus oli õige, aga tehnilistel põhjustel luges *DataCamp* vastuse

valeks?*

- Ei, ei ole juhtunud.
- Jah, paar korda.
- Jah, rohkem kui 2 korda.

Kas *DataCamp* praktikumid aitasid Sul teemast paremini aru saada? Vali kõik sobivad variandid.*

- Pigem jah, mõistan paremini rakendamise võimalusi.
- Pigem jah, aitas teoreetilisi mõisteid paremini ära õppida.
- Pigem jah, sest ... (täpsusta allpool).
- Pigem ei, tekitas segadust juurde.
- Pigem ei, see võttis liiga palju aega.
- Pigem ei, sest ... (täpsusta allpool)

Palun täpsusta, miks praktikumid aitasid / ei aidanud Sul teemast paremini aru saada:

Kui oled teinud esimesi lisapraktikume ja enam ei tee, siis mis oli Sinu loobumise põhjuseks?

Kas need praktikumid võiksid kohustuslikud olla?*

- Jah, võiksid
- Ei ole vajalik

Palun täpsusta, miks praktikumid võiksid /ei võiks olla kohustuslikud:

Vorm sisaldab kohustuslikke välju, mis on tähistatud märgiga *.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Argo Rihkrand,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Õppeaine “Tõenäosusteooria ja matemaatiline statistika” uurimine”, mille juhendaja on Natalja Lepik,
 - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 06.06.2018