# Andmeanalüüsi kodused ülesanded 2019, I osa

1. *Seda peab tegema käsitsi, tee kodus ja ära raiska praksi aega*. Saad paberil viie metsast püütud karu kaalud. Leia oma valimi (aritmeetiline) keskmine ning hinnang kaalu dispersioonile karupopulatsioonis. Tee oma arvutused taskuarvutil ja esita lisaks lõpptulemusele ka mõni vahetulemus. Teiseks võrdlesid miski metsa mustikate kaalu naabermetsa mustikate kaaluga, vastavas t-testis df=..... ja t=.... (paberil). Mitme mustika andmed olid analüüsis? Leia tabelist p. Kuidas tulemust tõlgendad? Mille tõenäosus on p (vihje: *valim, üldkogum, erinevus, juhuslik*)?

2. Mõtle välja jaotus, kus (aritmeetiline) keskmine on vahemikus 2*a* kuni 2*a+2* ja mediaan on (vähemalt 1,25 korda) suurem kui valimi keskmine, esita nad mõlemad numbriliselt. Tee Exceliga sagedusjaotuse pilt (histogramm), märgi käega joonisele mediaan ja keskmine. Arvuta ja esita ka dispersioon, variatsioonikoefitsient ja 25% kvartiil.

3. Mõtle välja olukord (seleta, mis andmed need on!), kus peaks kasutama t-testi ja t-testi eeldused on täidetud (silma järgi, testida pole vaja) ja vii analüüs Excelis läbi. Ühe rühma valimi keskmine peab olema vahemikus 2*a* kuni 2*a+2*. Esita joonisena sagedusjaotus võrreldavate rühmade kaupa ja t-testi tulemus sisuka lausena nii, nagu teeksid seda teadusartiklis. Teiseks mõtle välja ja esita kahe sellise rühma sagedusjaotused, mille puhul mõni t-testi eeldus (ütle, milline?) ei ole täidetud. Vii läbi logaritmteisendus ja esita uus sagedusjaotus. Kas t-testi eeldused on peale logaritmteisendust täidetud? (Ei ole häda, kui ei ole.)

4. Mõtle välja andmestik, kus kolme rühma keskmisi võrdleva ANOVA läbiviimisel saame R-ruudu väärtuse vahemikku *a/10* kuni *a/10+0,1.* ANOVAeeldused ei tohi ilmselt rikutud olla. Mis andmed need on? Vii analüüs Excelis läbi. Esita andmestik graafiliselt (sagedusjaotused rühmade kaupa) ja ANOVA tulemused nii Exceli toortabelina kui ka jutustava lausena nii, nagu teeksid seda artiklis.

5. Mõtle välja andmestik (selgita, mis andmed need võiksid olla), kus võrdleksid kahte rühma mitteparameetrilise ANOVA’ga. Rühmade mediaanid peavad olema vahemikus 2*a* kuni *3a*. Miks ei sobi seekord parameetrilised meetodid (vasta lausega)? Analüüsi ei ole vaja läbi viia, kuid esita 1) tabelina ja 2) graafiliselt karpdiagrammidena (*box plot*), mis peavad olema nõuetekohaselt lahti seletatud. Joonista karpdiagrammid käega, kuid siiski nii, et nad on andmetega kvantitatiivses kooskõlas (päris millimeetri pealt ei pea olema). Kas ja kuidas näeme jooniselt, et parameetrilist ANOVAt ei tohi siin kasutada?

6. Mõtle välja andmestik (koos bioloogiliselt mõtteka tõlgendusega), milles Pearsoni korrelatsiooni eeldused pole rikutud (silmaga vaadates) ning kus korrelatsioonikordaja väärtus on vahemikus +0,6 kuni +0,7. Esita vastav graafik ja korrelatsioonanalüüsi tulemus lausena. Pane tähele, et mingid jooned vms ei kuulu korrelatsioonanalüüsi pildi juurde. Nihuta joonisel ühte või rohkemat punkti nii, et parameetrilist korrelatsiooni ei tohiks enam rakendada. Selgita, miks. Esita graafik.

7. Mõtle välja viiest vaatlusest koosnev andmestik, milles on küll silmaga märgatav negatiivne korrelatsioon kahe muutuja väärtuste vahel (esita joonis), kuid see ei ole statistiliselt oluline. Suurenda valimit 2, 3, 4 ja 5 korda (*copy-paste* samade arvudega, ära jooniseid esita). Jälgi r ja p väärtuse muutumist valimi suurenedes ja esita need seosed (st r ja p sõltuvus N-st) graafiliselt, võta tulemus kokku ühe sisuka lausega. Pea meeles, et sõltuva muutuja paneme püstteljele ja sõltumatu rõhtteljele.

8. Mõtle välja andmestik, kus regressioonanalüüsi eeldused pole rikutud (silma järgi vaadates) ning sirge tõus on vahemikus *-2a* kuni *-2,2a* ning R-ruut vahemikus 0,6 kuni 0,7. Esita graafik kahe muutuja seose kohta, seose võrrand nii ja R-ruut nii, nagu programm selle ise trükib ja lisaks toimetatud kujul.

9. Mõtle välja kaks erinevat andmestikku, kus tasakaalulise kahefaktorilise ANOVA puhul:

 9A faktorite koosmõju puudub, kuid üks (ja ainult üks) peamõju on oluline;

 9B mõlemad peamõjud on olulised ning esineb “suuna muutusega” koosmõju;

Vii läbi analüüs tõestamaks, et asi on just nii. Esita visandlik joonis (või siis vähemalt rühmade keskmised), mis veenaks lugejat, et nii see asi tõesti on. Esita vastavad ANOVA tabelid 1) toorkujul, st nagu Excel nad annab ja 2) nii, nagu seda teadusartiklis teeksid, st toimetatud kujul.

10. Mõtle välja realistlik taimeökoloogilises uurimistöös ette tulla võiv olukord, kus kasutaksid sellist kovariatsioonanalüüsi, milles kaasaksid mudelisse (lisaks faktorile, mille mõju sisuliselt huvitab) kaks kovariaati (mille mõju sisuliselt ei huvita). Kirjelda olukorda sisuliselt. Millist kasu loodaksid saada kovariaatide kaasamisest? Analüüsi ei ole vaja reaalselt läbi viia, aga esita hüpoteetilised ANOVA tabelid olukorra jaoks, kus a) kovariaadid ei ole kaasatud ja b) kovariaadid on kaasatud. Kuidas tulemust tõlgendad? Palun väldi olukordi, kus faktor mõjutab ka kovariaadi väärtust (selliste olukordade tõlgendus on keerukam ja see on kursuse teise poole teema).