

Kommentaariid

Iseseisvate tööde lahendustest võiksite tekitada tekstidokumendi (näit. Wordis), kuhu on kopeeritud küsimus ja selle järgi vastus (koos vajalike tabelite ja/või joonistega ning kommentaaridega).

Kuigi küsimused on koostatud praktikumides R-i abil uuritud andmestike ja teostatud analüüside baasil, võib analüüsimiseks kasutada ka mõnd teist tarkvara (miskit saab Exceliski ära teha, või siis Statistica-s või SPSS-s või ...).

Iseseisev töö 1

Andmestik: http://www.eau.ee/~ktanel/DK_0007/studentsR.csv

või Exceli failina: http://www.eau.ee/~ktanel/DK_0007/studentsR.xls

Ülesanded

- 1.1. Leia keskmine ja mediaan tunnusele peaümberrõõrt sõltuvalt matemaatika hindest.
- 1.2. Kas mannaputru söövate tudengite kehakaalude varieeruvus on suurem, kui mannaputru mitte söövatel tudengitel?
- 1.3. Kuidas tudengid jaotuvad matemaatika hinde lõikes – tee sagedustabel ja tulpdiaagramm.
- 1.4. Kas erialati on matemaatika hinnete suhtelised sagedused erinevad?
- 1.5. Konstrueeri sektordiagramm tunnusele sugu ja kirjuta sektoritele juurde "naised" ja "mehed".
- 1.6. Joonista mannaputru söövate ja mitesöövate tudengite kehakaalude histogrammid ja tihedusfunktsioonid.
- 1.7. Konstrueeri sagedustabel tunnusele pikkus, jagades pikkused 6 klassi.
- 1.8. Konstrueeri karp-vurrud diagrammid tunnusele peaümberrõõrt sõltuvalt matemaatika hindest.
- 1.9. Leidke naistudengite keskmise kaalu 95%-usalduspiirid. Kas naistudengite keskmine kaal erineb statistiliselt oluliselt 60 kg-st?
- 1.10. Kas nais- ja meestudengite keskmised kehamassiindeksid on statistiliselt oluliselt erinevad?
- 1.11. Kas tudengite kehamassiindeksite jaotus erineb normaaljaotusest?

Iseseisev töö 2

Ülesanded baseeruvad osal Mariann Nõlvaku poolt aastail 2004-2006 kogutud Eesti kalade andmebaasist.

Andmestik: http://www.eau.ee/~ktanel/DK_0007/kala.xls

Andmestik sisaldab järgmisi andmeid:

- kala number (lihtsalt identifitseerimiseks);
- liik (6 liiki: haug, särg, latikas, luts, ahven ja koha);
- rühm: röövkala või lepiskala;
- 5 püügikohta (Võrtsjärv, Kärevere, Kastre, Praaga ja Peipsi järv);
- püügisesoon (kevad-suvi või sügis-talv);
- kaal ja pikkus;
- sugu;
- laiussiga (*Diphyllobothrium latum*) nakatumine ('diphyl' = 0 või 1);
- laiussi leidude arv kalal (diph_arv).

Ülesanded

Võtame uurimise alla üksnes haugid.

2.1. Kas püügisesoon ja laiussiga nakatumine (on nakatunud / ei ole nakatunud) on seotud?

a) Konstrueerige 2-mõõtmeline sagedustabel nii absoluutsete kui ka suhteliste sagedustega. Kommentaarid?

b) Testige seose statistilist olulisust nii χ^2 -testiga kui ka Fisheri täpse testiga.

c) Arvutage šansside suhe (OR) ja selle 95% usalduspiirid võrdlemaks nakatumist sügis-talvisel sesoonil püütud haugidel kevad-suvisel sesoonil püütutega. Järeldused?

2.2. Konstrueerige karp-vurrud diagramm illustreerimaks erinevatest kohtadest püütud haugide kaalude erinevust (või sarnasust). Kas suudate *R*-i *Help*'st välja lugeda, millal loetakse mingi kaal erindiks ja tähistatakse joonisel eraldi punktina?

2.3.

a) Leidke haugide pikkuse ja kaalu vaheline lineaarne (Pearsoni) korrelatsioonikordaja ning kaalu ja laiussi leidude arvu vahelised Spearmani and Kendalli korrelatsioonikordajad. Kas need seosed on statistiliselt olulised?

b) Leidke lineaarne regressioonivõrrand prognoosimaks haugide kaalu nende pikkuse järgi. Kas leiud võrrand on statistiliselt oluline? Illustreerige seost hajuvusdiagrammiga, kuhu kandke peale ka regressioonisirge ja viimase 95% usaldusintervall. Kui palju võiks leitud võrrandi alusel kaaluda 60 cm pikkune haug?

Iseseisev töö 3

Ülesanded baseeruvad osal Mariann Nõlvaku poolt aastail 2004-2006 kogutud Eesti kalade andmebaasist.

Andmestik: http://www.eau.ee/~ktanel/DK_0007/kala.xls

Ülesanded

Võtame uurimise alla üksnes haugid.

3.1. Uurige (ükshaaval) püügikoha, soo ja sesooni mõju haugide pikkusele. Kas kõigi nimetatud faktorite mõju on statistiliselt oluline?

3.2. Kas Peipsi ja Võrtsjärve haugid on erineva pikkusega (kui suur see erinevus on)? Kas see erinevus on statistiliselt oluline?

3.3. Kui pikad on keskmiselt Võrtsjärvest sügis-talvisel sesoonil püütud isased ja emased haugid? Leidke 95%-usalduspiirid sügis-talvisel sesoonil Võrtsjärvest püütud isaste ja emaste haugide mudeli abil prognoositud keskmistele pikkustele.

Iseseisev töö 4

Andmestik *R*-i andmefailina: http://www.eau.ee/~ktanel/DK_0007/puud.rda

või pakitud Exceli failina: http://www.eau.ee/~ktanel/DK_0007/puud.rar

Ülesanded

4.1. Linnamehest metsaomanik ei tea, kui vana on tema mets. Küll aga teab ta seda, et tal on metsas valdavalt kuused ja samuti teab ta puude diameetreid.

a) Leidke logistilise regressiooni mudel, mis prognoosib, kui tõenäoliselt on mingi diameetriga kuusk raieküps (andmebaasis sisalduvad 0-lise diameetriga puud jätke välja).

Pange kirja logistilise regressiooni võrrand ning illustreerige tulemust joonisega.

b) Leidke optimaalne raieküpsuse tõenäosuse väärtus, millest alates maksab kuusk lugeda raieküpsuks. Kui suur on sellise otsustusreegli korral testi tundlikkus ja spetsiifilisus?

Kommentaar. Praktikum nr 7 loodi puude andmestikku uus tunnus, mis omab väärtust 1 raieküpsu puu puhul ja väärtust 0 mitteraieküpsu puu puhul. Järgnevalt hinnati logistilise regressiooni võrrand prognoosimaks puu raieküpsuse tõenäosust vanuse alusel. Iseseisev ülesanne palub sama teha puu diameetri järgi.

4.2. Võtke puuliikide klasterdamisel arvesse ka puuliigi maksimaalne kõrgus. Kuidas muudab selline täiendus puuliikide klasterdamise tulemusi?

Kommentaar. Praktikum nr 8 loodi uus puuliikide kohta käiv andmestik, kuhu iga puuliigi kohta arvatati keskmised diameetrid 20 ja 50 aasta vanuses ning diameetri juurdekasv vanusevahemikus 60-80 (viimane võeti 0-ga võrdseks liikidel, kelle maksimaalne vanus < 80). Ja edasi jagati puuliigid klasteritesse leitud 3 tunnuse alusel. Iseseisev ülesanne eeldab ühe täiendava tunnuse väärtuste arvutamist igale puuliigile ja selle kaasamist klasteranalüüsi.

