

Eksamist

Kirjalik eksam koosneb kahest osast. Esimene osa testib mõisteid ja teadmisi, mida te peaksite peast teadma. Antud osa ülesanded/küsimused peaksid olema lihtsad, kui te antud mõistet/definitsiooni teate ja olete tema tähendust õieti mõistnud. Eksami esimene osa toimub väga tempokalt – aega vastamiseks antakse umbes 1-2 minutit iga küsimuse kohta.

Eksami teises osas võib materiale kasutada. Teise osa põhiosa moodustavad ülesanded – tuleb otsustada küsimusega kaasa pandud material põhjal, kas tegemist on normaaljaotusega; arvutada ise usaldusintervall keskväärtusele vms. Ka eksami teine osa saab olema tempokas – umbes 10 min iga ülesande kohta. Seega ei tasu lootma jääda, et te suudate eksami ajal loengumaterialiga tutvuda – loengumaterial peaks ikka eelnevalt olema läbi uuritud, muidu jääte eksamil lihtsalt ajahätta.

Kordamisküsimused

(*) - tähendab teemat, mille kohta esitatud küsimustele tuleb osata vastata ilma konspekti kasutamata.

1. ANDMETE ESITAMINE; PÕHISTATISTIKUD

- keskmine, mediaan, mood (*)
- miinimum, maksimum, dispersioon ja standardhälve (*)
- sagedustabel (*)
- tulpdiaagramm, karp-vurrud diagramm (*)
- tunnuste tüübid (*)

2. POPULATSIOON JA VALIM

- populatsioon ja valim (*)
- tunnuse jaotus populatsioonis
- tõenäosusfunktsioon, tihedusfunktsioon
- binoomjaotus
- normaaljaotus, normaaljaotuse kvantiilid
- normaaljaotuse kontrollimine
- keskväärtuse omadused (*)
- dispersiooni omadused (*)

3. VAHEMIKHINNANGUD

- prognoosiintervall (normaaljaotusega juhuslikule suurusele) (*)
- usaldusintervall keskväärtusele

4. HÜPOTEESIDE KONTROLLIMINE

- nullhüpotees/alternatiivne hüpotees (*)
- esimest liiki viga/teist liiki viga (*)
- olulisuse nivoo (*)
- olulisustõenäosus (*)
- testi võimsus
- t-test keskväärtuse kohta käivate hüpoteeside kontrollimiseks
- hii-ruut test

(jätkub)

5. REGRESSIOONANALÜÜS

- lineaarse regressiooni mudel
- regressioonanalüüsi juures kontrollitavad hüpoteesid
- determinatsioonikordaja, lineaarne (Pearsoni) korrelatsioonikordaja, tema tähendus

6. DISPERSIOONANALÜÜS

- mitmese võrdluse probleem, Bonferroni meetod (*)
- dispersioonanalüüsi mudel
- dispersioonanalüüsi eeldused

Näidisülesandeid

1. Sooviti uurida kirjatuvide lennukiirust. 15 lindu viidi tuvilast 200 km kaugusele ja lasti vabaks. Kahe esimese päeva jooksul jõudis tuvilasse tagasi 10 lindu, neil kulus saabumiseks 18, 24, 28, 29, 31, 35, 36, 37, 39 ja 43 lennutundi. Kas kahe päeva jooksul saabunud tuvide andmestiku põhjal on võimalik iseloomustada kirjatuvide keskmist lennukiirust? Kas selleks on sobilikum kasutada valimikeskmist või valimimediaani? Miks?

Vastus: Valimikeskmist ei ole võimalik arvutada, sest meil pole teada viie linnu lennuaeg. Seevastu valimimediaani on võimalik arvutada (variatsioonrea 8. element on meil teada – 37 tundi – sest me teame, et ülejäänud viiel linnul kulus lennu peale rohkem aega kui teistel). Seega on antud näite puhul ainus võimalus iseloomustada kirjatuvide lennukiirust valimimediaani abil (mis hindab üldiselt kirjatuvidel antud distantsti läbimiseks kuluva aja mediaani)

2. Kääbikupaar toob igal kevadel ilmale 5 pisikest ja karvast kääbikulast. Kui nii kääbikuema kui ka kääbikuisa kannavad retsessiivset pikakõrvalisuse geeni (st. nende laps omandab tõenäosusega 1/4 mõlemalt vanemalt pikakõrvalisuse geeni ja kasvatab seetõttu suureks saades jäneslikult kikkis kõrvad), kui suur on siis tõenäosus, et taolise vanematepaari viiest lapsest pole ükski pikkkõrvaline? Aga kui suure tõenäosusega on viiest lapsest just kaks pikkkõrvalised? (Vihje: ehk aitab binoomjaotus?)

Vastus:

Antud juhul saab tõepoolest kasutada binoomjaotust - võttes ühe “katse” “õnnestumise” tõenäosuseks $1/4=0,25$ ja “katsete” arvuks 5, saame (võttes arvesse, et $0!=1$ ja $x^0 = 1$):

a) leida tõenäosuse, et lastest pole keegi pikakõrvaline:

$$\begin{aligned} P(X = 0) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{5!}{0!(5-0)!} 0,25^0 0,75^{5-0} \\ &= 0,75^5 \quad (= 0,237) \end{aligned}$$

b) tõenäosus, et lastest täpselt kaks on pikakõrvalised:

$$\begin{aligned}P(X = 0) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{5!}{2!3!} 0,25^2 0,75^3 \\ &= 10 \cdot 0,25^2 \cdot 0,75^3 \quad (= 0,2636\dots)\end{aligned}$$

3. Kümnes erinevas Eestimaa paigas mõõdeti 4.aprillil õietolmu sisaldust õhus. Tulemuseks saadi (mõõtmistäpsuse piirides) järgmised arvud:

3, 4, 8, 5, 4, 2, 6, 4, 6, 5

Leia antud valimi keskvärtus, dispersioon, standardhälve ja mediaan. Leia usaldusintervall 4.aprillil keskmisele õietolmu sisaldusele õhus.

Vastus:

keskmine on 4,7

dispersioon $s^2=2,9$

standardhälve $s=1,7029\dots$

mediaan=4,5

95% usaldusintervall oleks (3,48...5,92)

(t-jaotuse vastav täiendkvantiil oli tabeli andmetel 2,26)

4. Rakubioloogia instituuti kraamides leiti kokku 9 töökorras ATDP mõõturit. Taibukas tudeng otsustas mõõta ATDP sisaldust oma proovis kõigi üheksa mõõturiga, ja sai tulemuseks järgmised arvud:

1,8 2,6 2,1 2,4 2,1 1,9 2,0 2,5 2,4

Eeldades, et mõõtmistulemused erinevad tegelikust väärtusest vaid juhusliku mõõtmisvea tõttu (mille jaotus võiks olla üsna lähedane normaaljaotusele), anna usaldusintervall tegelikule ATDP sisaldusele proovis.

Lahendus:

$$\bar{x} = 2,2$$

$$Dx = 0,08$$

$$s = \sqrt{0,08} = 0,2828$$

Usaldusintervall tegelikule ATDP sisaldusele oleks

$$\left[\bar{x} - t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}} ; \bar{x} + t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}} \right]$$

$$\left[2,2 - 2,31 \cdot \frac{\sqrt{0,08}}{3} ; 2,2 + 2,31 \cdot \frac{\sqrt{0,08}}{3} \right]$$

$$\left[2,2 - 2,31 \cdot \frac{\sqrt{0,08}}{3} ; 2,2 + 2,31 \cdot \frac{\sqrt{0,08}}{3} \right]$$

$$[1,98 ; 2,42]$$

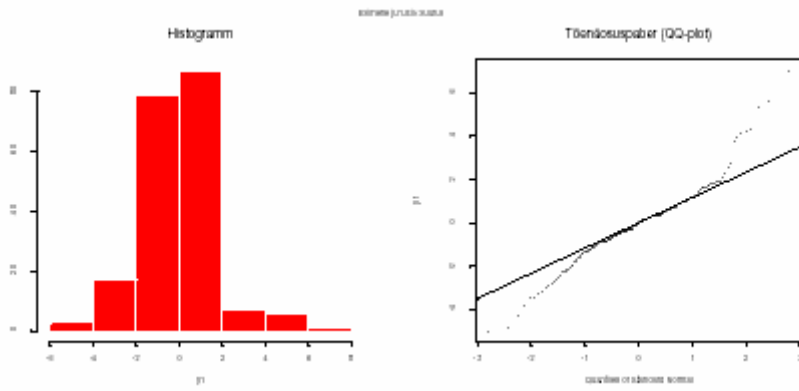
Vastus: 95% usaldusintervall proovi tegelikule ATDP sisaldusele on [1,98; 2,42]

5. Loodusuurija Harry Hoolas on oma elu jooksul leidnud 2000 linnuliigi keskmisele sabapikkusele 97%-lised usaldusintervallid. Kui teil palutakse hinnata, mitme linnuliigi tegelik keskmine sabapikkus ei asu prof. H.Hoolas poolt leitud usaldusvahemikes, siis millist vastust pakute?

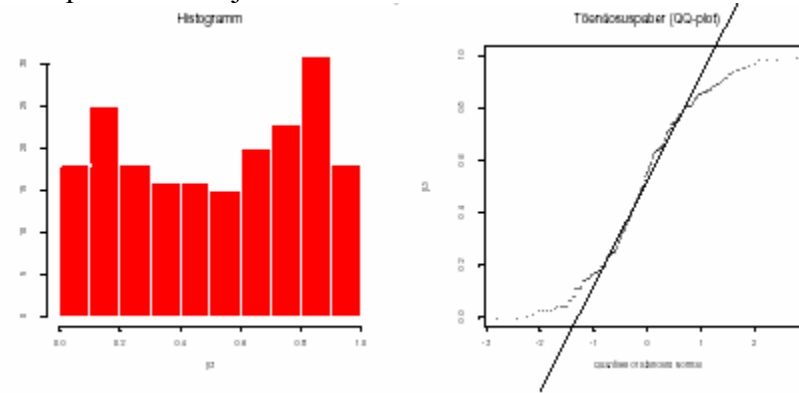
Vastus:

Tegelik keskmine ei sattu 97%-lisse usaldusintervalli keskmiselt 3% juhtudest ehk tõenäosusega $p=0,03$. 2000 usaldusintervalli seas oleks seega ligikaudu $0,03 \cdot 2000 = 60$ tegelikku keskmist sabapikkust mittesisaldavat usaldusintervalli.

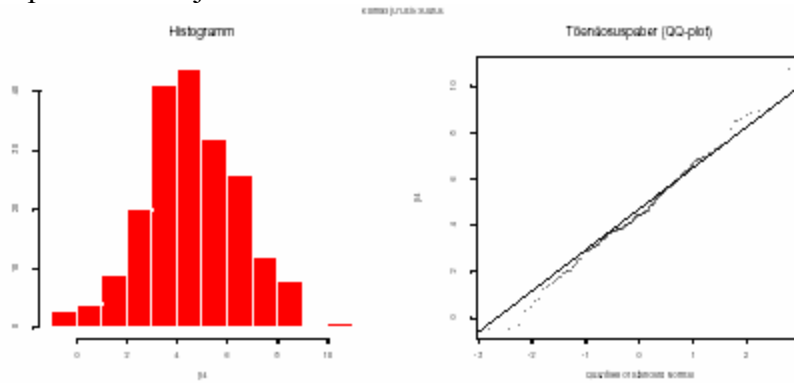
6. Ütle, milliste alljärgnevat juhulike suuruste jaotus võiks olla normaaljaotus?



Vastus: pole normaaljaotus



Vastus: pole normaaljaotus



Vastus: võib olla normaaljaotus

6. Loodusuurija tundis huvi, kas põtrade oksatoiduperioodi keskmine pikkus on Eesti erinevates piirkondades erinev. Tema uurimistöö käigus leidis kinnitust dispersioonanalüüsi mudel kujul

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Hinnatud parameetrite väärtused olid:

$$\hat{\mu} = 112$$

$$\alpha_{Põhja-Eesti} = 0$$

$$\alpha_{Lõuna-Eesti} = 12$$

- a) Mitu päeva kestab keskmiselt põtrade oksatoiduperiood Lõuna-Eestis?
- b) Mitme päeva jagu erineb oksatoiduperioodi pikkus Lõuna- ja Põhja-Eestis?

Vastus:

- a) Põtrade oksatoiduperiood Lõuna-Eestis kestab keskmiselt $112+12=124$ päeva.
- b) Põtrade oksatoiduperiood on Põhja-Eestis keskmiselt 12 päeva lühem kui Lõuna-Eestis.

8. Mis on laiem, kas 90%-usaldusintervall või 95%-usaldusintervall? Miks?

9. Biotoobi kirjeldamisel kasutati tunnust "sademete hulk" mõõdetuna ühikuis mm/kuus. Mis tüüpi tunnusega on tegemist?

10. Populatsioonis leidub väike grupp isendeid kelle uuritav tunnuse väärtused on järsult suuremad kui ülejäänutel. Kui arvutada välja nii populatsiooni mediaan kui ka keskvärtus, siis kumb neist võiks tulla suurem? Miks?

11. Too kolm näidet bioloogias uuritavatest tunnustest, mille jaotus on binoomjaotus. Seleta binoomjaotuse parameetrite (p ja n) tähendust antud näidete kontekstis.

12. Väliseksperdid uurisid eesti vaenulindude sabalaiust X ja leidsid, et $EX=4$, $DX=0,25$. Mõõtmisühikuks oli neil toll. Millised oleksid tulnud keskvärtus ja dispersioon kui uurijad oleksid kasutanud sentimeetreid mõõtmisühikutena (üks toll on ligikaudu 2,5cm)?

Näide ühest varasemast eksamist

Biomeetria eksam. A osa. (Sooritada ilma lisamaterjalideta)

1. Mis on olulisusnivoo?
2. Kala pikkust on mõõdetud sentimeetrites. Kala pikkuse dispersiooniks saadi 4. Milline oleks dispersioon siis, kui pikkust oleks mõõdetud meetrites?
3. Mis on esimest liiki viga?
4. Tahetakse teada, milline/sed toidulisandid soodustavad pleektatsudel liimja lima tootmist. Katsetati kümmet eri toitu ja tehti 10 t-testi (lima hulk ilma toidulisandita vs toidulisandiga toitmisel). Saadi järgmised olulisustõenäosused:
1. 0,896 2. 0,042 3. 0,969 4. 0,002 5. 0.444
6. 0.044 7. 0.500 8. 0.051 9. 0.555 10. 0.712
Millistel juhtudel võib vastu võtta alternatiivse hüpoteesi ja öelda, et toidulisand mõjutab lima teket?
5. Millised toodud väidetest peavad paika (märgi õiged väited)?
 - a) Kui olulisustõenäosus on suurem olulisusnivoost, siis oleme tõestanud nullhüpoteesi;
 - b) Kui olulisustõenäosus on väiksem olulisusnivoost, siis oleme tõestanud alternatiivse hüpoteesi;
 - c) Kui tegelikult kehtib nullhüpotees, siis me ei saa teha esimest liiki viga;
 - d) Kui tegelikult kehtib nullhüpotees, siis on olulisuse tõenäosus alati suurem olulisuse nivoo;
6. On teada, et 95% aastatest raiutakse Tartumaal metsa 6000..8900 ha. Kas antud vahemik on
 - a) kvartiilihaare
 - b) 95%-usaldusintervall boniteedile
 - c) 95%-line tolerantsiintervall
 - d) 95%-usaldusintervall keskmisele?
7. Sooviti uurida, milline toit kõige enam karudele meeldib. Uuringu käigus anti äsja loomaaeda sattunud karumammile erinevaid roogasid ja mõõdeti karumemme naeratuselaiust (naeratus puudub; kerge muhelus; naeratus; lai naeratus). Mis tüüpi tunnusega on tegemist?

Näide ühest varasemast eksamist

Biomeetria eksam. B osa.

1. Eksperimendis kasutatakse lisaainet X, mida saadakse tablettide kujul (tablett lahustatakse ja manustatakse katseloomale). Ühes tabletis on toimeainet 2mg, standardhälve 0,03. Eeldades, et toimeaine kogus tabletis on normaaljaotusega juhuslik suurus, siis millisesse vahemikku jääb 95% tablettide puhul toimeaine kogus?

2. Kolm haruldase linnu muna sattus uurijate kätte. On teada, et antud linnuliigi puhul on emaslinnu koorumise tõenäosus (0,6) veidi suurem kui isaslinnu koorumise tõenäosus (0,4). Milline on tõenäosus, et nendest kolmest munast kooruvad linnulapsed selliselt, et tekiks vähemalt üks linnupaar (kooruks vähemalt üks emaslind ja vähemalt üks isaslind)?

3. Kolm korda on õnnestunud mõõta haruldase linnuliigi pesitsusperioodi pikkust: 27, 31 ja 29 päeva. Eeldades, et pesitsusperioodi pikkus on normaaljaotusega juhuslik suurus, leia 95%-usaldusintervall pesitsusperioodi pikkuse keskväärtusele.

4. Genotüübiga AA madagaskari kärnkonnadel kasvavad teistest sama liigi kärnkonnadest lühemad tagajalad. Genotüüpidega Aa ja aa isenditel on koibade pikkus normaalne. Loodusuurijad uurisid 1000 juhuslikult valitud konna, kellest 42 olid genotüübiga AA, 316 genotüübiga Aa ja 642 genotüübiga aa. Geenialleelide jaotus sarnaneb olukorrale, kus looduslikku valiku survet märgatavalt ei esine (populatsioon näib olevat Hardy-Weimbergi tasakaalus). Loodusuurijaid huvitab, mis võiks kompenseerida lühemate (ja seetõttu halvema hüppevõimega) jalgade omamist. Hüpoteesiks on, et lühemate tagajalgadega konnade keskmine reaktsioonikiirus on suurem. Kõigil 1000 konnal mõõdeti reaktsioonikiirused (ms). Millise testi abil saaks mainitud hüpoteesi kontrollida?

5. Mõõdeti ühe loomaliigi kurameerimisperioodi pikkust (päevades) ja saadi järgmised tulemused: 5, 7, 5, 5, 7, 7, 5, 2, 9, 8. Õpik ütleb, et antud loomaliigi puhul on keskmine kurameerimisperioodi pikkus 5 päeva. Kas antud väide on vaatlustulemustega kooskõlas? Püstita hüpoteesid, maini, millist testi kasutad selle ülesande lahendamiseks ja näita, kuidas jõudsid tulemuseni. Millise järelduse teed?

6. Otsusta, kas eksisteerib seos tunnuste X ja Y vahel. Kas see seos on tugev? Milline võiks olla Y-i väärtus, kui X=20? Põhjenda oma arvamusi!

```
> summary(lm(y~x))
Residuals:
    Min       1Q   Median       3Q      Max
-14.75012  -3.03364   0.09066   3.00279  12.53587

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.1505     2.0831   4.393 2.84e-05 ***
x            0.5372     0.1164   4.616 1.19e-05 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.913 on 98 degrees of freedom
Multiple R-Squared: 0.1786,    Adjusted R-squared: 0.1702
F-statistic: 21.31 on 1 and 98 DF,  p-value: 1.186e-05
```