

Biomeetria bioloogidele
Hüpoteeside testimine. T-test.

OSA I. Tunnetuse tekitamine

Esmalt teeme läbi ühe eksperimendi mis võiks aidata mõista hüpoteeside statistilisel kontrollimisel kasutatavat loogikat. Eksperimendi käigus kontrollime, kui hästi suudab statistiline test (t -test) ära aimata õige, tegelikult kehtiva vastuse.

Võtame esmalt valimi ($n=10$ ehk kümme vaatlust) populatsioonist, mille tegelikuks keskväärtuseks on 3:

```
> valim=rnorm(10, mean=3)
```

Meie vaatlustulemused:

```
> round(valim, 2)
[1] 3.56 3.28 2.75 1.75 3.20 4.28 3.36 3.95 3.42 2.85
```

Meie valimi keskväärtus tuleb arvatavasti veidi erinev kolmest:

```
> mean(valim)
[1] 3.239976
```

Üritame nüüd kontrollida hüpoteesipaari:

$H_0: \mu = 3$

$H_1: \mu \neq 3$

Meie teame, et nullhüpotees on õige, aga arvuti seda ei tea! Vaatame, kas ta teeb õige otsuse! Leiame t -statistiku väärtuse:

```
> t=(mean(valim)-3)/(sd(valim)/sqrt(10))
> t
[1] 1.089526
```

$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ ←

Millist olulisuse nivood kasutame? Oletame, et lubame halvimal võimalikul juhul teha I-liiki viga 5%-tõenäosusega, ehk kasutame olulisuse nivood $\alpha=0,05$.

Nullhüpoteesi kehtides peaks leitud t -statistiku jaotus olema t -jaotusega vabadusastmete arvuga 9, ehk 95% valimite korral peaks tema väärtus jääma $\alpha/2=0,025$ ja $1-\alpha/2=0,975$ kvantiilide vahele. Leiame need kvantiilid:

```
> qt(0.025, df=9)
[1] -2.262157
> qt(0.975, df=9)
[1] 2.262157
```

```
> if (t>qt(0.025, df=9) & t<qt(0.975, df=9)) otsus="H0" else otsus="H1"
> otsus
[1] "H0"
```

Minul jäi t -statistiku väärtus $-2,26$ ja $2,26$ vahele, seega mina peaksin jääma nullhüpoteesi juurde. Enamikul teist peaks tulemus olema samalaadne, kuigi eriti halva õnne korral (5% kõikmõeldavaist katsetest) võib juhtuda, et võtsite ekslikult vastu alternatiivse hüpoteesi.

Teeme nüüd läbi ühe eksperimendi – võtame palju valimeid (näiteks 1000), kõik olukorras kus nullhüpotees tegelikult kehtib. Vaatame, kui sageli arvuti oma t -testi põhjal tehtud otsustes eksib. Salvestame arvuti otsused tunnusesse otsus:

```
katseid=1000
otsus=rep(NA, katseid)

for (i in 1:katseid){
  valim=rnorm(10, mean=3)
  t=(mean(valim)-3) / (sd(valim)/sqrt(10))
  if (t>qt(0.025, df=9) & t<qt(0.975, df=9)) otsus[i]="H0" else otsus[i]="H1"
}
table(otsus)
```

Lubasime arvutil teha rasket viga, ekslikult vastu võtta H_1 5% tõenäosusega, seega peaks 1000-st katsest ligikaudu 50 lõppema vale otsusega. Kas on nii?

Proovi, mis juhtuks otsustega, kui

- a) tõstaksime olulisuse nivood 0,2 peale;
- b) tegelikult kehtiks alternatiivne hüpotees (populatsiooni keskväärtus oleks näiteks 4). Testi võimsuseks kutsutakse tõenäosust vastu võtta alternatiivset hüpoteesi (juhul, kui alternatiivne hüpotees on õige). Milline on praegu testi võimsus (kui kasutad olulisuse nivood 0,05)?
- c) tegelikult kehtiks alternatiivne hüpotees, kuid tegelik keskväärtus on väga lähedal meie poolt oletatud väärtusele (tegelik keskväärtus 3,1, meie väide aga oleks $H_0: EX=3$);
- d) Olgu jätkuvalt tegelik keskväärtus 3,1; meie testime hüpoteesi $H_0: EX=3$, aga seekord kasutades suurt valimit – 10000 vaatlust. Milline on testi võimsus nüüd?
- e) Ehk ajab lihtsalt suur valim testil pea sassi? Kontrollime igaks juhuks üle, ega suure valimi korral (10000 mõõtmistulemust) 1.liiki vea tegemise tõenäosus liiga suureks ei lähe. Muuda populatsiooni keskväärtus tagasi 3-ks ja vaata, kui sageli otsustatakse ekslikult H_1 kasuks.

Võime ka graafikult kiigata, milliseid t-statistiku väärtuseid võime saada nullhüpoteesi kehtides (salvestame seekord iga valimi jaoks t-statistiku väärtuse):

```
katseid=1000
t=rep(NA, katseid)

for (i in 1:katseid){
  valim=rnorm(10000, mean=3)
  t[i]=(mean(valim)-3)/(sd(valim)/sqrt(10000))
}

hist(t)
```

Laias laastus öeldes: oodatavad väärtused peaksid olema vahemikus -2... 2.

Nüüd, kui loodetavasti on tekkinud mingi sisetunne hüpoteeside statistilise kontrollimise loogikast, võime edasi liikuda praktikas esineva juhu juurde, kus me tegelikku vastust oma küsimustele ei tea.

OSA II. Praktilisem pool

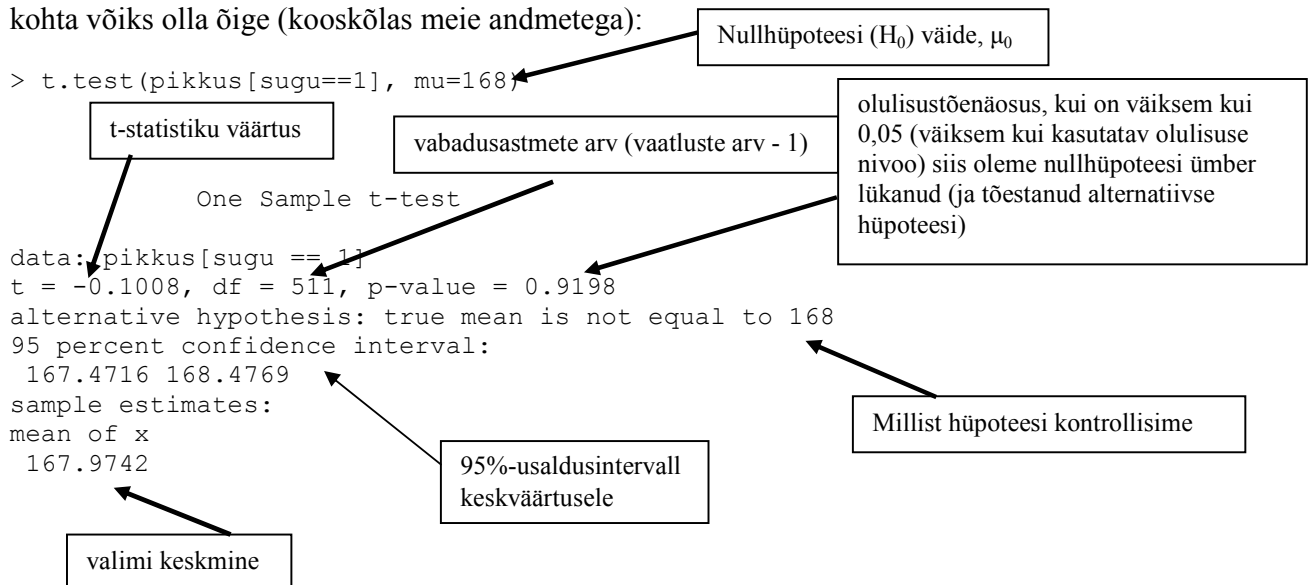
Loeme sisse tudengite andmestiku:

```
load(url("http://www.ms.ut.ee/mart/biomeetria2012/andmefail.RData"))
```

Väidetavalt olla Eesti meeste keskmine pikkus 1800mm ja eestlannade keskmine pikkus 1680mm (väide võetud:

<http://innomet.ttu.ee/oppetoo/Inimenemasinsuhted/Antropomeetria.doc>, lk 11).

Kontrollime, kas tudengite andmed on kooskõlas väidetuga. Naised olid tudengite andmestikus kodeeritud kui sugu=1, mehed kui sugu=2. Vaatame esmalt, kas väide naiste kohta võiks olla õige (kooskõlas meie andmetega):



Näeme, et naiste puhul on meie kasutuses olevad andmed kooskõlas väidetuga – kui tegelik keskvärtus oleks 168cm, siis võiksime sellise valimi, nagu on meie kasutuses, saada küll.

Kontrolli iseseisvalt väidet eesti meeste keskmise pikkuse kohta. Kommenteeri.

Proovime veel.

Ühe allika väitel on eesti meeste keskmine kaal 78kg. Teise allika väitel on meeste keskmine kaal 75kg. Kontrolli mõlemat hüpoteesi (tudengite andmestiku põhjal. Meestel on tunnus sugu väärtuseks 2). Millise tulemuseni jõuad?

Erinevad t-testid

Algset t-testi on mitut moodi kohandatud, kolm neist variantidest on väga sageli kasutatud:

- T-test hüpoteeside kontrollimiseks keskvärtuse kohta – kontrollib mingit väidet ühe populatsiooni (ühe tunnuse) keskvärtuse kohta. Kontrollitavaks hüpoteesiks on näiteks väide $H_0: EX=10$; $H_1: EX \neq 10$.
- T-test keskvärtuste võrdlemiseks, sõltumatud valimid. Sobib kasutamiseks, kui soovime võrrelda, kas kahe uuritava populatsiooni keskvärtused on võrdsed või mitte. Näiteks tahame teada saada, kas naiste ja meeste keskmine palk on sama või mitte. Või kui tahame uurida, kas genotüübiga AA isenditel on keskmine kolesteroolitase samasugune kui genotüübiga aa isenditel (või mitte). Kontrollitavaks hüpoteesideks on väited $H_0: EX = EY$; $H_1: EX \neq EY$
- T-test keskvärtuste võrdlemiseks, sõltuvad valimid. Kontrollitakse ikka väidet, kas kahes grupis on uuritava tunnuse keskvärtus sama või mitte, näiteks uuritakse, kas alkoholi tarbinute keskmine reaktsioonikiirus on sama, mis alkoholi mittetarbinutel. Eelmisest olukorrast on erinev vaid see, et samad objektid kuuluvad mõlemasse gruppi – samal inimesel mõõdetakse reaktsioonikiirust enne ja pärast alkoholi tarbimist. Või mõõdetakse mingi ensüümi aktiivsust igal loomal 2 ja 12 päeva peale sündi ja tahetakse teada, kas antud ensüümi (keskmine) aktiivsustase on muutunud - ka sellisel juhul on tegemist sõltuvate valimitega (2. päeva mõõtmised ja 12. päeva mõõtmised on tehtud samadel loomadel).

Vaatamegi alljärgnevalt, kuidas teha t-testi kahe populatsiooni keskvärtuste võrdlemiseks (nii sõltuvate kui ka sõltumatute valimite korral)


T-test sõltumatute valimite jaoks

Võimaldab kontrollida, kas kahe populatsiooni keskvaartused on võrdsed või mitte:

$$H_0: EX = EY$$

$$H_1: EX \neq EY$$

Näiteks kontrollime, kas meestel ja naistel on süstoolse vererõhu keskvaartused samad või mitte.



```
> t.test(SVR[sugu==1], SVR[sugu==2])

Welch Two Sample t-test

data:  SVR[sugu == 1] and SVR[sugu == 2]
t = -10.975, df = 179.329, p-value < 2.2e-16
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -16.11443 -11.20286
sample estimates:
mean of x mean of y
 114.5841  128.2427
```

Milline on t-testi otsus? Mille põhjal otsuseni jõudsid?

Muuseas, sama tulemuseni oleksime jõudnud ka andes R'ile käsu

```
t.test(SVR~sugu)
```

Kontrolli veel järgmiseid hüpoteese:

- Kas haiglaravi või kiirabi vajanutel oli suurem keskmine kaal?
- Kas viina tarbivatel inimestel on suurem keskmine kaal kui viina mittetarbivatel inimestel ($viin=1$)?

T-test sõltuvate valimite jaoks

Loe R'i andmestik "seeme.txt". Selleks:

```
seeme=read.table("http://www.ms.ut.ee/mart/biomeetria2012/seeme.txt", header=TRUE)
```

Soovitakse teada, kas kuivatis kuivatatud seeme annab parema saagi kui põllupeal kuivanud vili (tegemist on Inglismaal kogutud andmetega). Väikesed katsepõllud jagati kaheks, ühele poolele külvati põllul kuivanud vilja, teisele poole aga kuivatis kuivanud vilja. Tunnus HAR iseloomustab põllul kuivanud vilja saagikust (lbs/aaker, nael on u 0,45 kg; aaker on u. 0,4 ha; korruta 1,12 saamaks kg/hektarilt); tunnus KUIV kuivatis kuivatatud vilja saagikust.

Moodustame uue tunnuse, mis näitab erinevalt kuivatatud seemnete viljakuse erinevust samal põllulapil:

$VAHE = KUIV - HAR$

Kui viljakus ei sõltu seemnete kuivatamise viisist, siis saagikus peaks olema ligikaudu samasuur. Erinevused saagikuses oleksid siis tingitud pigem sellest, et sama põllu kaks poolt ei pruugi olla täiesti samaväärsed kasvutingimuste poolest, samuti võib ühele põllupoolele ka muidu mõni hakkajam taim sattuda, mis samuti tasakaalu rikub. Aga üle kõigi erinevate juhuslikult poolitatud põldude peaks keskmine erinevus tulema 0, ehk tunnuse VAHE keskväärtus peaks sellisel juhul olema 0. Matemaatiliselt kirja pandult:

$H_0: E(VAHE) = 0$ (kuivatamisviis ei mõjuta keskmist saagikust)

$H_1: E(VAHE) \neq 0$ (kuivatamisviis mõjutab keskmist saagikust)

Selliseid hüpoteese saab kontrollida t-testi abil:

```
> t.test(VAHE)
```

```
One Sample t-test
```

```
data: VAHE
t = 1.6905, df = 10, p-value = 0.1218
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -10.72710 78.18164
sample estimates:
mean of x
 33.72727
```

Kus $t = 1.6905$ on t-statistiku väärtus, $df=10$ näitab kasutatud vabadusastmete arvu, p-value on olulisustõenäosus.

Kas olulisuse nivool 0,05 saab järeldada, et põllu viljakus sõltub seemnete kuivatamise viisist?

Kas võib tõestatuks lugeda, et viljakus ei sõltu seemnete kuivatamise viisist?

Üks lugupeetud teadlane on ütelnud, et kuivatis kuivatatud seemnete saagikus on keskmiselt 100 naela/aakri kohta suurem kui põllul kuivanud vilja viljakus. Kas meie andmed kinnitavad selle maineka teadlase arvamust? Püstitame hüpoteesid:

$H_0: E(VAHE) = 100$ (keskmine erinevus saagikuses on 100naela/aaker)

$H_1: E(VAHE) \neq 100$

Kontrollime neid hüpoteese:

```
> t.test(VAHE, mu=100)
```

Kas saame tõestada, et teadlasel oli õigus?

Kas saame tõestada, et teadlane eksis?

Milliste hüpoteeside kontrollimiseks tudengite andmestiku peal sobiks t-test sõltuvate valimite jaoks?

R' is koos lisamoodulitega on võimalik kasutada üle 100 erineva statistilise testi, millistest paljud on välja mõeldud üsna spetsiifiliste hüpoteeside kontrollimiseks. Alljärgnevalt mainime paari testi nende sajakonna seast:

Shapiro-Wilk'i test (normaaljaotuse eelduse testimine)

Kui valim on väike, siis tagab t-testi õiged arvutustulemused vaid siis, kui uuritav tunnus on normaaljaotusega. Järvest on aga välja püütud vaid 6 siiga, tegemist on selgelt väikese valimiga. Peaksime kontrollima, kas siigade kaalude jaotus võiks olla normaaljaotusega. Võisime seda teha graafiliselt, aga vastavaid hüpoteese on võimalik kontrollida ka statistilise hüpoteeside testimise protseduuri abil. Shapiro-Wilk'i testi võib kasutada kontrollimaks, kas pideva tunnuse jaotuseks võiks olla mõni normaaljaotus. Kontrollime näiteks, kas tudengite pikkuste jaotus on normaaljaotus:

H_0 : Tunnuse *Pikkus* jaotus on (mingi) normaaljaotus, $Pikkus \sim N(\mu; \sigma^2)$
 H_1 : Tunnuse *Pikkus* jaotus pole normaaljaotus

```

Tunnus, mille jaotust kontrollime
↓
> shapiro.test(pikkus)

Shapiro-Wilk normality test

data:  Weight
W = 0.9729, p-value = 1.035e-09
↑
olulisustõenäosus on 0,000000001035
```

Antud juhul näeme, et olulisustõenäosus tuleb väiksem kui 0,05, seega oleme nullhüpoteesi kummutanud (nullhüpotees ei saa paika pidada – antud tunnuse jaotuseks pole küll normaaljaotus).

Kas meeste pikkus on normaaljaotusega? Aga naiste pikkus?

Märkus: tasub teada, et eriti suuremate valimite korral võib Shapiro-Wilk'i test olla mõttetult range. Halvimal juhul võib ta normaaljaotuse eelduse kummutada (suure valimi korral) imeväikeste probleemide pärast (näiteks mõõtmistulemuste ümmardamine). Sellised imepisikesed probleemid enamasti ei muuda mistahes normaaljaotust eeldava meetodi tulemusi kuigivõrd. Sestap vahel soovitatakse eelistada suuremate valimite korral ikka normaaljaotuse graafilist kontrollimist – inimsilmal on nagunii raske selliseid tähtsusetuid kõrvalekaldeid normaaljaotusest tuvastada ja see ongi hea. Enamasti on normaaljaotust eeldavad meetodid robustsed – kui vaatlused on peaaegu normaaljaotusega, siis kõlbab neid meetodeid ikka veel kasutada. Matemaatilist täpsust pole antud eelduse kontrollimisel niiväga vajagi ☺.

Uuritava tunnuse hajuvust kontrollivad testid

Uurivad tunnuse hajuvust võrdlevad testid võiksid näiteks kasulikud olla sama liigi asurkondade uurimisel – kui antud loomad ühes asurkonnas on väga sarnased võrreldes teiste asurkondadega, siis võib põhjuseks olla hiljutine „pudelikaela“ läbimine, mis geneetilist varieeruvust on tugevalt vähendanud. Samuti eeldavad mõned statistilised meetodid, et uuritava tunnuse hajuvus kõigis uuritavates populatsioonides on samasuur – ka selle eelduse paikapidavust saab kontrollida mainitud testide abil.

Bartlett' test

bartlett.test – (F-testi edasiarendus) – kontrollib, kas uuritava tunnuse hajuvus on vaadeldavates populatsioonides samasugune või mitte:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Eeldab, et uuritavad tunnuse jaotus (kõigis populatsioonides) on normaaljaotus. Kasutusnäiteid:

```
bartlett.test(pikkus, olu)
bartlett.test(kaal, sport)
```

Viimase testi jaoks kontrollime ka eelduse täidetust (kas kõigi sportimisintensiivsuste korral võiks kaalu jaotuseks olla normaaljaotus):

```
par(mfrow=c(3,2))
by(kaal, sport, qqnorm)
```

Fligner-Killeen'i test

fligner.test - kontrollib, kas uuritava tunnuse hajuvus on vaadeldavates populatsioonides samasugune või mitte. Ei nõua, et uuritava tunnuse jaotus oleks normaaljaotus. Kasutusnäide (kalade andmestiku puhul):

```
fligner.test(kaal, sport)
```

Sündmuse toimumise tõenäosuse kohta käivate hüpoteeside kontrollimine

Suure valimi korral võib hüpoteeside kontrollimisel tõenäosuse(te) kohta kasutada ka t-testi. Väikese valimi korral seda teha ei tohi (uuritava tunnuse jaotus pole ju normaaljaotus). Sestap on olemas spetsiaalne test hüpoteeside kontrollimiseks tõenäosuste kohta, mida võib kasutada väikeste valimite korral ja mis annab täpsemaid tulemusi kui t-test ka suurte valimite korral. Selle testi nimeks on binom.test. Näide binom.test-i abiinfost:

```
## Conover (1971), p. 97f.
## Under (the assumption of) simple Mendelian inheritance, a cross between plants of two particular
## genotypes produces progeny 1/4 of which are "dwarf" and 3/4 of which are "giant", respectively.
## In an experiment to determine if this assumption is reasonable, a cross results in progeny having
## 243 dwarf and 682 giant plants. If "giant" is taken as success, the null hypothesis is that
## p = 3/4 and the alternative that p != 3/4.
binom.test(c(682, 243), p = 3/4)
binom.test(682, 682 + 243, p = 3/4) # The same.
## => Data are in agreement with the null hypothesis.
```

Sama test võrdluseks t-testi abil tehtuna (eelistata päriselus binom.test-i!):

```
t.test(rep(c(1,0),c(682, 243)), mu = 3/4)
```