

Peatükk 4

Hüpoteeside statistiline kontrollimine

Jälgides enda ümber toimuvat, elusloodust vaadeldes või aretusprotsessi juhtides tekib mõtleval inimesel paratamatult oletusi või hüpoteese ümbritseva maailma kohta. Oletus või kahtlus pole veel teadmine. Kuidas leida kinnitust oma kahtlustele? Kas kogutud andmed kinnitavad või hoopis kummutavad meie oletuse/hüpoteesi?

4.1 Hüpoteeside kontrollimise filosoofiast

Kuidas saaks kontrollida, kas mingi oletus, teooria või hüpotees peab paika? Üks võimalus on järgmine. Oletatakse, et kontrollitav teooria peab paika. Arutluste abil leitakse, mis sellisel juhul (kui kontrollitav teooria peab paika) tulevikus juhtub, milliseid tulevaseid katsetulemusi me peaksime nägema, milliseid valimeid võiksime saada (ja milliseid mitte), mida avastama tulevastel väljakaevamistel vms. Seejärel tehakse vajalik katse või vaatlus ning saadakse teada, kas teooria poolt ennustatud asi juhtus või mitte. Kui kontrollitava teooria ennustus ei pidanud paika, siis on teooria vale. Kui ennustus läks täkkesse, siis võib teooria kehtida (aga ei pruugi, sest paljud erinevad teooriad võivad viia sarnase ennustusele). Seega kontrollitavat teooriat saab vaid kummutada, tavaliselt pole seda aga võimalik tõestada. Tõsi, kui kontrollitav teooria kannatab välja palju erinevaid kontrollimisi, siis tema usaldusväärsus tõuseb.

Hüpoteeside statistiline kontrollimine jälgib sarnast loogikat. Oletatakse, et mingi hüpotees (nimetagem seda nullhüpoteesiks) kehtib. Arutletakse, milliseid hinnangu väärtuseid (näiteks milliseid valimi keskmiseid) me nullhü-

poteesi kehtides tõenäoliselt võiksime näha. Siis minnakse ja võetakse valim. Kui valimi põhjal arvutatud hinnang tuli ootuspärane, selline nagu ta võiks tulla nullhüpoteesi kehtides, siis võib nullhüpotees õige olla. Kui aga meie hinnang polnud selline, nagu ta nullhüpoteesi kehtides oleks tõenäoliselt pidanud olema, siis lükkame nullhüpoteesi ümber. Teisisõnu — me usume siis, et kehtib väide, mis eitab nullhüpoteesi (alternatiivne hüpotees).

Hüpoteeside kontrollimist alustataksegi nullhüpoteesi ja alternatiivse hüpoteesi sõnastamisest. Nullhüpotees ja alternatiivne hüpotees peavad olema teineteist välistavad ja vähemalt üks neist peab kehtima. Traditsiooniliselt sõnastatakse hüpoteesid järgmiselt:

Nullhüpotees (H_0) - ka konservatiivne hüpotees. Väide, et praegu kehtiv traditsiooniline elu/tegu/mõttemis tagab (vähemalt) sama hea tulemuse kui uuendajate/reformijate poolt pakutav lähenemisviis. Nullhüpoteesi kummutatades — esivanemate traditsiooni hülgamist soovitades — võtame harilikult endale üsna suure vastutuse. Teadlasena muutuseid soovitades hakkame vastutama oma hea nimega.

Alternatiivne hüpotees (H_1) - kutsutud ka sisukaks hüpoteesiks. Väide, et uus lähenemisviis tagab parema tulemuse (on õigem) kui traditsiooniline, vaikimisi ja aruteludeta aktsepteeritav lähenemisviis. Üks noor ja ambitsioonikas teadlane arvatahvasti unistab alternatiivse hüpoteesi tõestamisest — mis võiks olla veel ihaldusväärsem, kui näidata, et Sinu idee tagab parema tulemuse kui (teaduslike) esivanemate põlvkondade pikkune kogemus.

Näide 4.1 Nullhüpotees: *Mu vastaspartner mängib ausalt, ka ta täring on aus. Formaalselt kirja panduna $H_0: P(\text{sõber viskab täringuga 6 silma})=1/6$.*

Alternatiivne hüpotees: *Ta kasutab võltstäringuid! Formaalselt $H_1: P(\text{sõber viskab täringuga 6 silma})\neq 1/6$.*

Kuidas statistiliselt (vaid vaatlusandmetele tuginedes) saaks näites esitatud alternatiivset hüpoteesi tõestada? Siin ei ole tegelikult midagi tavamõistusele keerulist.

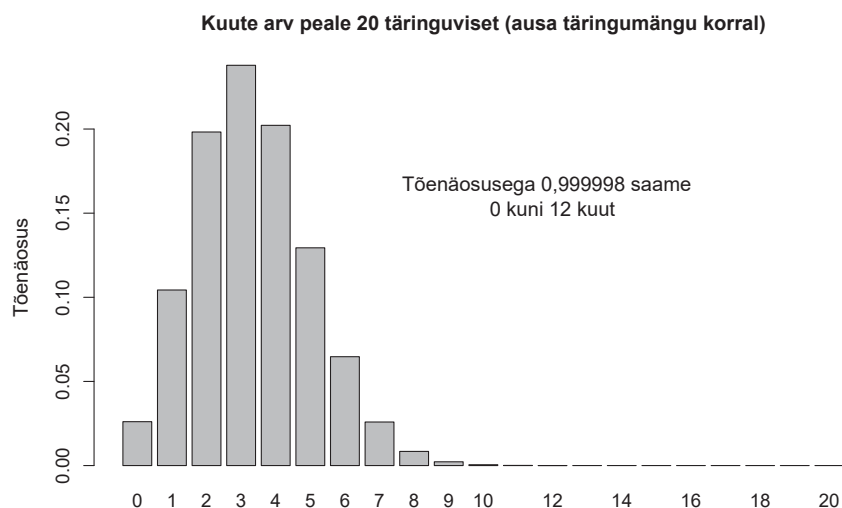
Näite 4.1 jätk...

Te mängite täringumängu. Mängupartner teeb heite ja saab kuus silma, teie kaotate. Halb õnn, mõtlete teie, ja mängite edasi. Vastasmängijal tuleb taas kuus, ja siis veel kord. Teil tekib kahtlus... Aga Te ei saa ju ometi niisama tõusta püsti ja esitada valjul häälel julma süüdistust... Teie nimi saaks määritud, kui kontrollimisel selguks, et täring on täiesti aus ja laua all pole peidus salakavalaid magneteid ega muid vidinaid. Alusetu süüdistus võiks igaveseks rikkuda Teie suhted selle auväärse ja kõrgel positsioonil inimesega kellega koos Te mängite. Te jätkate mängu, ja Teie kaaslane saab jälle kuue, ja jälle...

Peale 20 täringuviset on vastasmängija saanud 19 korda tulemuseks kuus ja vaid ühel korral on tema täring veeretanud kuuest väiksema numbri. Nüüd ei suuda te enam sellist enda pilkamist taluda — ka lauslollus määrab Teie nime.... Te järeldate, et Teie mängupartner teeb sohki ning Te süüdistate oma kaasmängijat pettuses.

Antud juhul töötab meie tunnetus jälgides sama loogikat, mida kasutab ka hüpoteeside statistiline kontrollimine. Juhul, kui täringumäng oleks aus, võiksime 20 viske jooksul näha ehk 9, äärmisel juhul kuni 12 kuut. Aga ausa mängu puhul kahekümne täringuviskega rohkem kuusi saada on peaaegu võimatu. Tõenäosusega 0,999998 jääb 20 ausa täringuviskega saadud kuute arv vahemikku $[0...12]$, vaata ka joonist 4.1 (märkus: 20 ausa täringuviske jooksul saadud kuute arv on binoomjaotusega juhuslik suurus, $X \sim B(20, 1/6)$). Seega, kui me näeme ennekuulmatut, lausa uskumatut tulemust — 19 korda visati kuus — siis me ei suuda enam ausasse mängu uskuda. Loomulikult pole ehk meie sisetunnetuse poolt tehtud tõenäosuse arvutused sedavõrd täpsed, aga küllap järeldus oleks ikkagi seesama — ausa mänguga siin tegemist pole.

Joonis 4.1: Kui mängukaaslane oleks aus...



Kirjeldatud lähenemisviisile on sisse ehitatud üks probleem — lugedes alternatiivse hüpoteesi tõestatuks, võime siiski kogemata teha vea. Ka ausa täringuga on põhimõtteliselt võimalik visata üheksateist korda järjest kuut. Loomulikult tuleks ausas teaduses iseloomustada, soovitavalt kvantitatiiv-

selt, eksimise võimalikkust. See tingib vajaduse paari täiendava termini järgi.

Tehes oma otsust selle kohta, kas õige on nullhüpotees või alternatiivne hüpotees (kas mängupartner teeb sohki või mitte), võime eksida. Juhul, kui loeme alternatiivse hüpoteesi tõestatuks (ütleme, et sõber teeb sohki); aga tegelikult oli õige nullhüpotees (sõber tegelikult ei teinudki sohki) oleme teinud tõsise vea. Seda viga, alternatiivse hüpoteesi ekslikku õigekspidamist, kutsutakse *esimest liiki veaks* (*Type I error*).

Teist liiki viga (*Type II error*) tehakse siis, kui jäädakse ekslikult nullhüpoteesi juurde (ei süüdistata partnerit sohitegemises).

Tabel 4.1: Esimest ja teist liiki viga

	Tegelikult kehtib H_0	Tegelikult kehtib H_1
Jääme H_0 juurde tõestame H_1	õige otsus I liiki viga	II liiki viga õige otsus

Enne hüpoteeside tõestamise juurde asumist tuleks otsustada, kui kindlad soovitakse oma tulemustes olla, kui maailmale minnakse alternatiivse hüpoteesi kehtimisest teatama. Matemaatilises keeles öeldult: fikseeritakse maksimaalne lubatud tõenäosus teha esimest liiki viga. Taolist lubatavat ülempiiri esimest liiki vea tegemise tõenäosusele kutsutakse olulisuse nivooks (*significance level*). Esimest liiki vea tegemise tõenäosus on maksimaalne muidugi siis, kui tegelikult kehtib nullhüpotees (kui alternatiivne hüpotees on õige, siis me ei saagi esimest liiki viga teha). Seega halvimas võimalikus situatsioonis — kui tegelikult on õige nullhüpotees — ei tohiks me alternatiivse hüpoteesi kasuks otsustada suurema tõenäosusega, kui valitud olulisuse nivoo lubab. Põhimõtteliselt võib muidugi kasutada väga erinevaid olulisuse nivoo — perfektsionist võib kasutada näiteks olulisuse nivood 0,001 ja muidulahmija võib näiteks eelistada olulisuse nivood 0,25. Siiski on paljudes teadusvaldkondades / teadusajakirjades esile kerkinud eelistatud olulisuse nivood — kõige sagedamini kasutatakse teaduskirjanduses olulisuse nivood 0,05.

Vähima olulisuse nivoo, mille korral me konkreetse eksperimendi korral oleksime veel saanud alternatiivse hüpoteesi tõestatuks lugeda, on *olulisustõenäosus* (*significance probability; p-value*). Kui olulisustõenäosus on väiksem kui valuläveks valitud olulisuse nivoo, võetakse vastu (loetakse tõestatuks) alternatiivne hüpotees.

Olulisustõenäosusele saab anda ka järgmise interpretatsiooni: olulisustõenäosus näitab, kui suur tõenäosus on näha meie poolt nähtud (või veel

uskumatumat) tulemust siis, kui nullhüpotees kehtib. Pöördume hetkeks tagasi täringumängu näite juurde. Ka aus mängija võib 20 täringuviske käigus saada üheksateist või enam korda tulemuseks kuus silma. See on lihtsalt väga ebatõenäoline — tõenäosus ausa mängu korral visata järjest 19 või enam korda järjest “6” on 0,0000000000000276. Seega võime öelda, et antud juhul on olulisustõenäosus $2,76 \times 10^{-14}$.

Statistiline test on seda parem, mida *tundlikum* ta on - mida väiksem on tõenäosus teha teist liiki viga, kui on fikseeritud maksimaalne lubatud tõenäosus teha esimest liiki viga. Statistilise testi võimsuseks kutsutakse tõenäosust lugeda tõestatuks alternatiivne hüpotees, kui tegelikult ongi õige alternatiivne hüpotees. Seega, mida suurema võimsusega (*power*) on test, seda parem ta on. Testi võimsus võib suuresti sõltuda sellest, milline on tegelikult uurijale huvipakkuv tundmatu protsess/väärtus. Näiteks, kui uuritakse keskkonnapoliitika mõju liigirikkusele, siis on mõistlikult ülesehitatud testide võimsus seda suurem, mida suurem on tegelikult poliitika mõju loodusele.

Näide: Mõõtmiste käigus koguti 50 istiku andmed ($n = 50$), uuritava tunnuse standardhälve olgu 1. Meid huvitavad hüpoteesid on järgmised (H_0 : $\mu = 10$; H_1 : $\mu \neq 10$). Tabelis 4.2 on antud I ja II liiki vea tegemise tõenäosused ja testi võimsus sõltuvalt populatsiooni tegelikust keskväärtusest.

Tabel 4.2: Esimest ja teist liiki vea tegemise tõenäoste sõltuvus uuritud tegelikkusest

Tegelik μ	9.95	10	10.05	10.1	10.2	10.3	10.5	11
I liiki vea tegemise tõenäosus (α)	-	0,05	-	-	-	-	-	-
II liiki vea tegemise tõenäosus	0,937	-	0,937	0,894	0,717	0,453	0,067	> 0,001
testi võimsus	0,063	-	0,063	0,106	0,283	0,547	0,933	0,999

Järgnevalt vaatleme mõningaid väga levinud statistilisi teste.

4.2 T-test hüpoteeside kontrollimiseks keskväärtuse kohta

Võib tulla ette olukordi, kus läheb tarvis kontrollida hüpoteese populatsiooni keskväärtuse kohta. Näiteks võib teaduskorüfee väita midagi aastaste istikute pikkuse keskväärtuse kohta, või on õpikus kirja pandud, milline peaks olema keskmine saagikus kirjeldatud kasvatuse meetodi korral. Kas poleks ahvatlev liig ülbaid autoriteete veidi õpetada ja nende väited kui valed kummutada?

Hüpoteesid keskväärtuse kohta saab kirja panna järgmiselt:

$$\begin{aligned} H_0 : \quad & \mu = \mu_0 \\ H_1 : \quad & \mu \neq \mu_0, \end{aligned}$$

kus μ_0 on mingi konkreetne väidetav number.

Juhul, kui nullhüpotees kehtib (tegelik keskväärtus ongi μ_0), ja vaatlusandmed on normaajaotusega või valim on piisavalt suur ($n > 30$), siis on teisendatud juhuslik suurus (kutsugem teda t -statistikuks) $t := \frac{\bar{x} - \mu_0}{s} \sqrt{n}$ t -jaotusega juhuslik suurus (vaata usaldusintervalli kohta käivat peatükki):

$$t := \frac{\bar{x} - \mu_0}{s} \sqrt{n} \sim t_{n-1}.$$

Märkame, et t -statistiku väärtuse saame oma valimi põhjal kergesti välja rehkendada — teame ju nii oma valimi keskmist $\bar{x}$, valimi standardhälvet s kui ka valimi suurust n . Hüpoteesi sõnastades fikseerisime ka väärtuse μ_0 -le.

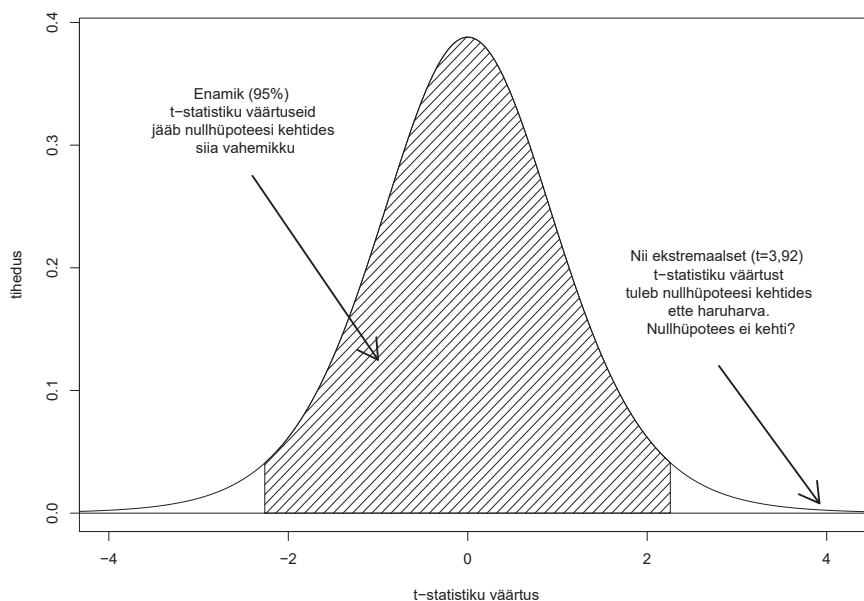
Kui nullhüpotees kehtib, siis iga uurija poolt (iga uue valimi põhjal arvutatud) leitud t -statistiku väärtus tuleb küll erinev, kuid enamikel juhtudel jäävad leitud väärtused üsnagi kitsastesse piiridesse, nulli lähedale. Valimi suuruse $n = 10$ korral on t -statistiku väärtuste jaotus (nullhüpoteesi kehtides) selline, nagu kujutatud joonisel 4.2. Kui meie valimi puhul leitud t -statistiku väärtus ei osutu tüüpiliseks / ootuspäraseks (näiteks $t=3,92$), siis on loomulik hakata kahtlema tehtud eelduse ($\mu = \mu_0$) paikapidavuses.

Mida pidada väga suureks või väga väikeseks, seda saab vaadata t -jaotuse tabelist. Kui nullhüpotees kehtib, siis jääb t -statistik t vahemikku

$$t_{\alpha/2;n-1} < t < t_{1-\alpha/2;n-1}$$

tõenäosusega $1-\alpha$. Väljapoole ülaltoodud vahemikku sattub t -statistik (nullhüpoteesi kehtides) haruharva, kõigest tõenäosusega α . Kui nüüd tõesti juhtub nii, et meie valimi põhjal leitud t on kas väiksem või suurem vaadeldavast

Joonis 4.2: t-statistiku väärtuste jaotus nullhüpoteesi kehtides (df=9)



kvantiilist ($t < t_{\alpha/2;n-1}$ või $t > t_{1-\alpha/2;n-1}$), siis tuleb tunnistada, et taolise situatsiooni tekkimise tõenäosus nullhüpoteesi kehtides on kaduvväike ja järelikult peab õige olema alternatiivne hüpotees. Seega kasutades olulisuse nivood α peaksime vastu võtma alternatiivse hüpoteesi.

Näide 4.2 Sooviti uurida teatud kemikaali mõju närvisüsteemile. Korraldatud katses mõõdeti katseloomade (kasside) reaktsioonikiirust enne ja pärast uuritava kemikaali manustamist. Iga kassi jaoks leiti reaktsioonikiiruse muutus. Kui kemikaal närvisüsteemi ei mõjuta, siis peaks keskmine reaktsioonikiiruse muutus olema null ($\mu_0 = 0$):

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0.$$

Hüpoteeside kontrollimisel valime kasutatavaks olulisuse nivooks $\alpha = 0,05$.

Reaktsioonikiiruse muutused 15 katses osalenud kassi jaoks olid järgmised:

25, 27, -12, 8, 19, 4, -1, -4, 13, 12, 0.05, 2, -21, 24, -1.

Põhistatistikud vaadeldud valimi jaoks:

$$\begin{aligned}\bar{x} &= 6,33\dots \\ s^2 &= 192,05 \\ s &= 13,86\end{aligned}$$

Leiame t -statistiku väärtuse:

$$t = \frac{(6,33\dots - 0)}{13,86} \times \sqrt{15} = 1,770\dots$$

Kas saame tõestatuks lugeda, et kemikaalil oli mõju kasside reaktsiooni-kiirusele? Valitud olulisuse nivoo (0,05) korral peame otsustamiseks tabelist üles otsima t -jaotuse 0,025- ja 0,975-kvantiilid:

$$\begin{aligned}t_{0,025;14} &= -2,14 \\ t_{0,975;14} &= 2,14.\end{aligned}$$

Kuna meie valimi põhjal leitud t -statistiku väärtus (1,77) jääb leitud kvantiilide vahele, siis järelikult ei saa me nullhüpoteesi ümber lükata (selline t -statistiku väärtus on täiesti mõeldav nullhüpoteesi kehtides). Järeldus: antud andmete põhjal ei saa tõestatuks lugeda, nagu mõjutaks vaadeldud kemikaal kasside reaktsioonikiirust.

Tähelepanek: Kui me oleksime valinud kasutatavaks olulisuse nivooks 0,1, siis oleksid võrdluses kasutatavate t -jaotuse kvantiilide väärtusteks tulnud -1,76 ja 1,76 ning me oleksime saanud alternatiivse hüpoteesi vastu võtta. Paraku peab kasutatav olulisuse nivoo olema määratud enne andmetega tutvumist ning olulisuse nivoo tagantjärgi tarkust kasutades on keelatud.

4.3 T-test sõltuvate valimite korral

Juba eelmises näites vaatlesime reaktsioonikiiruse muutuseid. Sarnased situatsioonid tulevad ette üllatavalt sageli — kas oleme teinud mõõtmisi enne ja pärast (millegi tegemist) ning soovime vaadata, kas uuritava tunnuse kesk-väärtus on muutunud; või oleme välja valinud sarnaste katseloomade paarid, ühte paarilist kasvatame ühtedes tingimustes, teist teistes. Soovime taas võrrelda, kas erinevates kasvutingimustes kasvatatud paariliste vahel on uuritava

tunnuse keskmises tasemes toimunud mingit muutust või mitte. Mida sarnasemaid mõõtmistulemusi ootaksime neil kahe paarilise mõõtmisel siis, kui “katsetingimuste muutmisel” mingit “mõju” poleks, seda kindlamalt suudame tõestada ka “mõju” olemasolu.

Sõltuvate valimite korral võib t-testi statistiku kirja panna järgmiselt:

$$t = \frac{\overline{x - y} - 0}{s(\overline{x - y})} = \frac{\overline{x - y} - 0}{s} \times \sqrt{n},$$

kus n on sarnaste paaride arv, x tähistab uuritava tunnuse väärtuseid paarilistel, kes viibisid ühtedes katsetingimustes ja y tähistab uuritava tunnuse väärtuseid paarilistel, kes viibisid teistes katsetingimustes. Tähistus $\overline{x - y}$ märgib paariliste katsetulemuste vahede keskmist. Kontrollitav hüpoteesi-paar — kas uuritava tunnuse keskväärus on sama, sõltumata katsetingimustest — on kirja pandav järgmiselt:

$$H_0 : EX = EY$$

$$H_1 : EX \neq EY.$$

Ülaltoodud valemist arvutatud t-statistik on nullhüpoteesi kehtides (st. kui uuritava tunnuse keskväärus mõlemas katsegrupis on sama) t-jaotusega, vabadusastmete arvuga $n-1$ (vaatluspaaride arv - 1).

4.4 T-test sõltumatute valimite keskväärtuste võrdlemiseks

Alati pole võimalik moodustada sarnast katsealuste paare selliselt, et uuritava “mõju” puudumisel annaksid mõlemad katsealused sama (või peaaegu sama) katsetulemuse. Näiteks võib meid huvitada, kas sordi A saagikuse keskväärtus on samasuur kui sordi B saagikuse keskväärtus. Aga sarnaste põllulappide/põldude valik võib osutuda keeuliseks. Või on katse juba toimunud — ühte sorti kasvatati 6, teist 23-l põllul. Kuidas siis kontrollida hüpoteese keskväärtuste võrdsuse kohta:

$$H_0 : EX = EY$$

$$H_1 : EX \neq EY?$$

Sellisel juhul kasutatakse t-testi sõltumatute vaatluste jaoks. Statistilise testi konstrueerimiseks on kaks võimalust. Ühel juhul eeldatakse, et mõlemas grupis uuritava tunnuse hajuvus on samasuur, teisel juhul taolist täiendavat eeldust ei tehta.

4.4.1 T-test sõltumatute vaatluste jaoks, võrdne hajuvus.

Eeldades, et uuritava tunnuse hajuvus mõlemas vaadeldavas grupis (mõlemas populatsioonis) on samasuur, võib püstitatud hüpoteesipaari kontrollimiseks kasutada järgmist teststatistikut:

$$t = \frac{\bar{x} - \bar{y}}{s(\bar{x} - \bar{y})},$$

kus

$$s(\bar{x} - \bar{y}) = \sqrt{s^2(\bar{x} - \bar{y})} = \sqrt{s^2(\bar{x}) + s^2(\bar{y})} = \sqrt{s^2/n_1 + s^2/n_2} = s/\sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

Viimases valemis tähistab s ühist hinnangut standardhälbele — st hinnang on leitud mõlemat valimit kasutades. Kahe valimi ühine hinnang standardhälbele on leitav järgmise valemi abil:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2}}.$$

Juhul, kui kehtib nullhüpotees, siis on taolisel viisil arvutatud t-statistiku jaotuseks t-jaotus, vabadusastmete arvuga $df = n_1 + n_2 - 2$. Kui arvutatud

t-statistiku väärtus on väga ekstremaalne (ebausutavalt suur/väike võrreldes t-jaotuse poolt lubatud kõrvalekaldega), siis järeldatakse, et tehtud eeldused olid väärad — st. rühmade keskväärtused on erinevad.

Sellisel viisil tehtud t-testi tulemus on usaldusväärne, kui: a) uuritava tunnuse hajuvus mõlemas rühmas on ligikaudu samasuur (mõlema sordi saagikuste varieeruvus on sarnane) b) uuritav tunnus (saagikus) on kas normaaljaotusega või on uuringus kasutatud palju katsepõlde ($n_1 + n_2 > 30$).

4.4.2 T-test sõltumatute vaatluste jaoks, hajuvus erinevates populatsioonides võib olla erinev (Waldi test).

Kahe populatsiooni keskväärtuste kontrollimiseks on võimalik konstrueerida t-testi ka ilma võrdse hajuvuse eeldust tegemata. Kasutatav t-statistik näeks sellisel juhul välja järgmine:

$$t = \frac{\bar{x} - \bar{y}}{s(\bar{x} - \bar{y})} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2(\bar{x}) + s^2(\bar{y})}} = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}}.$$

Asja komplitseerib mõnevõrra tõdemus, et leitud statistik pole nullhüpoteesi kehtides enam täpselt t-jaotusega juhuslik suurus, ta on kõigest peaaegu t-jaotusega. Millist vabadusastmete arvu kasutada? Leidub erinevaid soovitusi (mis võivad anda veidi erinevaid tulemusi), üks levinumaid on nn Satterthwaite'i meetod. See meetod ütleb, et testimisel tuleks kasutada järgmise valemi abil leitud vabadusastmete arvu:

$$df = \frac{(s^2(\bar{x}) + s^2(\bar{y}))^2}{\{s^2(\bar{x})\}^2/(n_1 - 1) + \{s^2(\bar{y})\}^2/(n_2 - 1)}.$$

Enamasti teostab rehkendused arvuti ja sestap pole valemi keerukus eriliseks takistuseks. Kui arvutatud t-statistiku väärtus on väga ekstremaalne (ebausutavalt suur/väike võrreldes t-jaotuse poolt lubatud kõrvalekaldega), siis järeldatakse, et tehtud eeldused olid väärad - st. rühmade keskväärtused on erinevad. Sellisel viisil tehtud t-testi tulemus on usaldusväärne, kui: b) uuritav tunnus (näiteks saagikus) on kas normaaljaotusega või on uuringus osalenud inimesi palju.

4.4.3 Sobiva T-testi valimine

Kas vaatlused on paari pandavad (tehtud samal inimesel, tehtud kaksikutel või sama pesakonna kahel järglasel, tehtud sarnaste põllulappide paaridel)? Kui jah, siis kasuta t-testi sõltuvate valimite jaoks (inglise k. paired t-test). Kui tegemist pole vaatlustega samal isendil või sarnaste isendite paaridel,

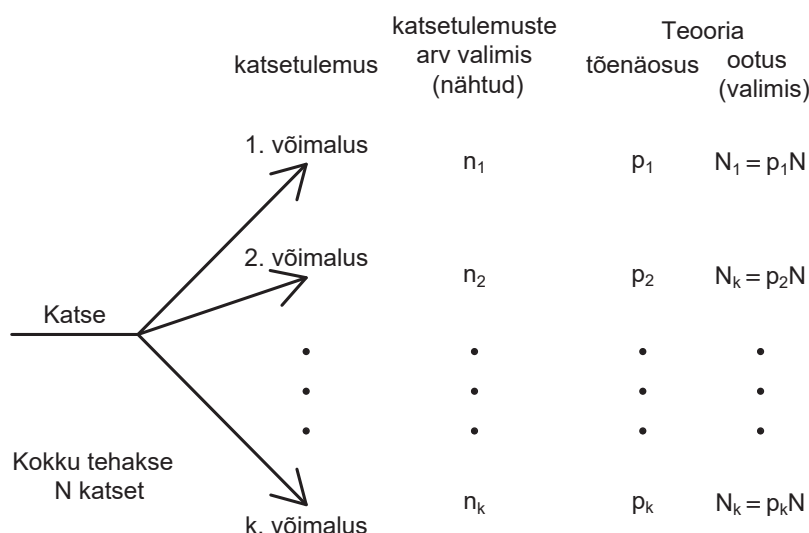
siis otsusta, kas uuritava tunnuse hajuvus võiks mõlemas võrreldavas rühmas (populatsioonis) olla sama? Kui jah, siis kasuta T-testi sõltumatute võrdse hajuvusega vaatluste jaoks. Kui ei, siis kasuta T-testi, mida kirjeldatud alalõigus 4.4.2.

4.5 Hii-ruut test

Üks universaalsemaid ja sagedamini kasutust leidev test on hii-ruut (χ^2 -test, inglise keeles ka *chi-square test*).

Oletame, et sooritataval katsel on k erinevat võimalikku tulemust (lilleseemnest kasvab kas valge, punane või roosa õis — $k = 3$; sündiv linnupoeg on kas isane või emane — $k = 2$ jne). Mõnikord on võimalik teooriat kasutades leida, milline peaks olema ühe- või teise katsetulemuse tulemise tõenäosus. Sellisel juhul on võimalik hii-ruut testi abil kontrollida, kas vaatlustulemused on kooskõlas teooria ennustustega või mitte — kas erinevus teooria ja valimis nähtu vahel võiks olla tingitud valimi juhuslikkusest või on erinevus liiga suur. Vaata ka joonist 4.5.

Joonis 4.3: Tegelikud katsetulemused ja teooria ennustus



Vaatame mõningaid näiteid hüpoteesidest, mida saab testida kasutades hii-ruut testi.

Näide 4.3 Grupi teadlaste arvates on lõvilõua õie värv määratud ühe geeni poolt. Sellel geenil on kaks alleeli, tähistame neid a ja A . Juhul, kui taime genotüüp on AA , peaks tal olema punane õis, genotüübiga Aa lill peaks olema roosa õiega ja aa -genotüübiga lill võiks olla valge. Selle hüpoteesi kontrollimiseks ristasid teadlased roosade õitega (heterosügootseid) taimi. Kui nende

oletus peab paika, peaks järglaste jaotus olema kooskõlas Mendeli seadustega, vt tabel 4.3.

Tabel 4.3: Mendeli seaduste paikapidavuse kontrollimine

Aa x Aa	AA (punane õis)	25% järglastest
	Aa (roosa õis)	50% järglastest
	aa (valge õis)	25% järglastest

Näide 4.4 Soovitakse kontrollida, kas uuritav populatsioon on Hardy-Weinbergi tasakaalus (antud geeni suhtes alleelidega a ja A). Tähistame tõenäosust, et populatsioonist juhuslikult valitud geenialleel on A tähega p . Juhul, kui populatsioon oleks Hardy Weinbergi tasakaalus, peaks genotüüpide esinemistõenäosused olema sellised, nagu antud tabelis 4.4.

Tabel 4.4: Hardy-Weinbergi tasakaalu kontrollimine

genotüüp	esinemistõenäosus
AA	p^2
Aa	$2p(1-p)$
aa	$(1-p)^2$

Märkus: Tõenäosust p saame hinnata oma valimi põhjal — $\hat{p} = \frac{2\#\{AA\} + \#\{Aa\}}{2N}$.

Kuidas siis hii-ruut test kontrollib sedaliiki hüpoteese? Esmalt leiame hii-ruut statistiku väärtuse,

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - N_i)^2}{N_i}$$

Juhul, kui teooria peab paika, siis terooria poolt ennustatud juhtude arv N_i peaks olema ligilähedaselt õige ja vahed $n_i - N_i$ tulevad väikesed ning ka hii-ruut statistiku χ^2 väärtus tuleb väike. Seevastu juhul, kui teooria ei pea paika, siis kipuvad erinevused nähtu ja oodatu vahel olema (absoluutväärtuselt) suured, ning hii-ruut statistiku väärtus tuleb suur.

Soovides kasutada antud statistikut testimaks teooria paikapidavust, peame selgitama, kui suur peab teststatistiku väärtus olema selleks, et me enam ei

tohiks uskuda teooria paikapidavusse. Selle selgitamiseks tuleb esmalt määrata veel üks vajalik parameeter, mida kutsutakse vabadusastmete arvuks. Veidi lihtsustatud (kuid peaaegu alati korrektselt töötav) eeskiri hii-ruut testi vabadusastmete arvu leidmiseks on järgmine:

$$\begin{aligned} df &= \text{Erinevate võimalike tulemuste arv } (k) \\ &\quad - \text{valimi põhjal hinnatud parameetrite arv suuruste } N_i \text{ leidmiseks} \end{aligned}$$

Saamaks näite 4.3 jaoks kontrollitava teooria poolt ennustatavat juhtude arvu, peame oma andmetest piiluma ühte numbrit — valimi suurust. Teades valimi suurust N , saame teooria põhjal öelda, mitu punase õiega, mitu roosa õiega ja mitu valge õiega lille Mendeli seaduste järgi peaks valimis olema. Seega antud näite korral on vabadusastmete arv

$$\begin{aligned} df &= 3 \text{ (kolme värvi õitega järglasi võib esineda)} \\ &\quad - 1 \text{ (valimi suuruse hindamine)} \\ &= 2 \end{aligned}$$

Näite 4.4 korral oleme sunnitud sageduste N_i leidmiseks hindama (kasutades valimit) kahte parameetrit: valimi suurust N ja alleeli A esinemissagedust p :

$$\begin{aligned} df &= 3 \text{ (3 erinevat genotüüpi)} \\ &\quad - 2 \text{ (valimi suurus ja alleeli A esinemissagedus)} \\ &= 1 \end{aligned}$$

Lisaks peame testi läbiviimiseks fikseerima olulisuseniivoo — määratlema, kui kindlad me tulemustes tahame olla, enne kui julgeme nullhüpoteesi kummutada. Teades usaldusniivood, vabadusastmete arvu ja Hii-ruut statistiku väärtust saame otsuse teha kasutades hii-ruut jaotuse tabelit. Kriitilised väärtused on toodud tabelis 4.5.

Näide 4.5 Ühel euroopas esineval liblikaliigil (*Panaxia dominula*) esineb ühe geeni mutatsioon, mis muudab liblika tiibade mustrit. Alleelidega AA liblikail on tiibadel suured valged täpid. Genotüübiga aa isenditel täpid puuduvad ja nende asemel on tiivad ühtlaselt tumedad. Genotüübiga Aa isenditel täpid esinevad, kuigi on väiksemad kui AA tüüpi isenditel. Uurijaid huvitab,

Tabel 4.5: χ^2 - statistiku kriitilised väärtused h

Vab.-astmeid (df)	$P(X > h) = 0,05$	$P(X > h) = 0,01$
1	3,841	6,635
2	5,991	9,210
3	7,815	11,345
4	9,488	13,277
5	11,070	15,068
6	12,592	16,812
7	14,067	18,475
8	15,507	20,090
9	16,919	21,666
10	18,307	23,209
12	21,026	26,217
14	23,685	29,141
16	26,296	32,000
18	28,869	34,805
20	31,410	37,566
25	37,652	45,624
30	43,773	50,892
35	49,802	57,342
40	55,758	63,691
45	61,656	69,957
50	67,505	76,154
60	79,082	88,379
70	90,531	100,425
100	124,32	135,807

kas üks või teine tiivamuster annab liblikale mingit evolutsioonilist eelist või on liblikate populatsioon Hardy-Weinbergi tasakaalus. Selle selgitamiseks läks uurija Ford 1971. aastal aasale ja määras 1612 liblika genotüübid:

Vastamaks meid huvitavale küsimusele tuleb leida geeni A esinemistõenäosus:

$$p := p(A) = (2 * 1469 + 138) / (1612 * 2) = 0,954$$

Nüüd saame leida näites 2 toodud valemeid kasutades, kui palju me ühe või teise genotüübiga isendeid oleksime oodanud olevat oma valimis, kui

Tabel 4.6: h rra Fordi andmed

AA-t��pi isendeid:	1469
Aa-t��pi isendeid:	138
aa-t��pi isendeid:	5

Hardy-Weinbergi tasakaal oleks kehtinud. Genot  biga AA isendite oodatav proportsioon oleks $p_i = p^2 = 0,954^2 \approx 0,9103$, ootusp rane genot  biga AA isendite arv valimis oleks $N_i = Np_i = 0,91031612 \approx 1467,4$, erinevus tegeliku ja oodatava vahel on $n_i - N_i = 1469 - 1467,4 = 1,6$. Samamoodi v ime j tkata arvutusi ka teiste genot  pide tarvis, vaata ka tabel 4.7.

Tabel 4.7: H rra Fordi arvutused

	tegelik n_i	oodatav proportsioon	oodatav arv N_i	erinevus $n_i - N_i$
AA-t��pi isendeid	1469	0,9103	1467,4	1,6
Aa-t��pi isendeid	138	0,0876	141,2	-3,2
aa-t��pi isendeid	5	0,0021	3,4	1,6

N  ud saab v lja arvutada ka h i-ruut statistiku v  rtuse:

$$\chi^2 = \frac{1,6^2}{1467,4} + \frac{(-3,2)^2}{141,2} + \frac{1,6^2}{3,4} = 0,827.$$

Olles valinud usaldusnivooks 0,05 (t  piline teadusartiklites kasutatav usaldusnivoo), v rdleme oma saadud h i-ruut statistikut tabelis antud v  rtustega. Kuna antud juhul on vabadusastmete arv 1 ($3-2=1$), siis peame v rdlema leitud statistiku v  rtust (0,827) kriitilise v  rtusega, mis on antud reas $df=1$ (3,841). Kuna $3,841 > 0,827$, siis j reldame, et erinevus teooria poolt ennustatava ja vaatlustulemuste vahel on v ike. Me pole suutnud nullh pooteesi   ber l kata, Hardy-Weinbergi tasakaal v ib kehtida. Seega ei saa nende vaatlustulemuste p hjal   elda, et   he tiivamustriga j rglastel oleks evolutsiooniline eelis teistsorti tiivamustriga liblikate   le.

*Antud klassikalise n  ite andmed p rinevad raamatust E. B. Ford (1971) *Ecological genetics*. Chapman and Hall, London.*

4.5.1 Hii-ruut testi eeldused

Hii-ruut test kasutab teststatistiku asümptootilist (suurte valimite korral kehtivat) jaotust, st. korrektse tulemuse saamiseks peab valim olema suur. Kui suur peaks olema valim, et praktikas saaks kasutada hii-ruut testi? Harilikult soovitatakse, et igat võimalikku väärtust esineks H_0 kehtides ootuspäraselt enam kui viiel korral ($N_i \geq 5$ iga i korral). Mida teha, kui oletades kontrollitava teooria kehtimist peaks antud valimi suuruse juures mingit väärtust esinema vähem kui 5 korda? Üks võimalus (lisaks valimi suurendamisele) oleks kombineerida kokku mõned harvemad väärtused üheks uueks väärtusklassiks. Seejärel tuleks kontrollida, kas kombineeritud väärtuste esinemissagedus vastab teooria poolt ennustatavale. Loomulikult peame võimalike väärtuste arvu vähendades vähendama ka kasutatavat vabadusastmete arvu. Lisaks mainitud eeldusele peab loomulikult olema tegemist nõuetekohaselt leitud (juhusliku) valimiga.

Järeldus: viimases näites toodud arvutus ei pruugi olla (päris) korrektne, sest genotüübiga aa isendeid ootasime valimis olevat vaid 3,4 tükki (mida on alla 5).

4.5.2 Hii-ruut test seose olemasolu kontrollimiseks

Vahel soovitakse testida, kas kahe (nominaalse, järjestus-) tunnuse vahel eksisteerib statistiline seos või mitte (kas ühe tunnuse väärtuse teadmine aitab öelda midagi selle kohta, milline võiks olla teise tunnuse jaotus — ehk teisisõnu — kas ühe tunnuse väärtuse muutudes muutub teise tunnuse jaotus või mitte). Näiteks võime soovida kontrollida, kas alamliigiti erineb saakloomade eelistus (kas erinevatel alamliikidel on erinev saakloomade jaotus) või võime testida, kas eksisteerib seos inimese töölemineku viisi (jala/rattaga/bussiga/autoga) ja tema tervisliku seisundi (suurepärase; hea; keskmine; halb; väga kehv) vahel.

Kontrollimaks hii-ruut testiga seose olemasolu tunnuste X ja Y vahel (olgu nende tunnuste võimalikud väärtused tähistatud vastavalt sümbolitega $1, 2, \dots, k$ ja $1, 2, \dots, l$) peame leidma, milline on mistahes väärtuste komplekti ($X = i, Y = j$) saamise tõenäosus siis, kui kontrollitav hüpotees (väide: tunnused on sõltumatud) kehtiks. Juhul, kui tunnused X ja Y oleks tõepoolest sõltumatud, siis peaks tunnuse X jaotus olema alati samasugune, ükskõik, milline siis ka tunnuse Y väärtus ka pole. Juhul, kui seost tunnuste vahel pole, on tõenäosus näha tunnuse X väärtust i üks ja seesama sõltumata sellest, milline on tunnuse Y väärtus:

$$P(X = i | Y = 1) = P(X = i | Y = 2) = \dots = P(X = i | Y = l) \quad (= P(X = i)).$$

Ülaltoodud valemis tähistab $P(X = i|Y = j)$ nn tinglikku tõenäosust — juhul kui teame, et tunnuse Y väärtus on j , siis tõenäosus, et tunnuse X väärtus tuleb i on $P(X = i|Y = j)$.

Kui X ja Y on sõltumatud, siis $P(X = i|Y = j) = P(X = i)$ ja tõenäosus, et populatsioonist juhuslikult valitud isendi korral saame sellise looma/taime/inimese, kelle puhul $X = i$ ja $Y = j$ on leitav järgmiselt:

$$P(X = i, Y = j) = P(X = i|Y = j)P(Y = j) \stackrel{\text{sõltumatud}}{=} P(X = i)P(Y = j).$$

Seega saame leida tunnuste mistahes väärtuste puhul, millise tõenäosusega üht- või teistsugune väärtuste komplekt peaks esinema nullhüpoteesi (seost tunnuste vahel pole) paikapidamisel. Ja edasi jätkame juba nii, nagu hii-ruut testi tehakse – leiame oodatud arvud ($N_{X=i,Y=j} = NP(X = i)P(Y = j)$), leiame erinevused teooria poolt ennustatud sageduste ja tegelikult valimis nähtud sageduste vahel ning arvutame hii-ruut statistiku väärtuse. Vaatame vaid üle vabadusastmete arvu leidmise – mitut numbrit või parameetrit me peame enne oma valimi põhjal leidma, enne kui saame arvutada suurused $N_{X=i,Y=j}$? Esiteks peame teadma oma valimi suurust N . Teiseks peame hindama tunnuse X jaotuse ehk tõenäosused $P(X = 1), \dots, P(X = k)$. Paneme aga tähele, et viimast tõenäosust $P(X = k)$ me ei pea piiluma oma valimist — kuna $P(X = 1) + \dots + P(X = k) = 1$ siis $P(X = k) = 1 - P(X = 1) + \dots + P(X = k - 1)$ ja seega peame valimi põhjal hindama kõigest $k - 1$ tõenäosust (ja viimase saame juba eelnevate põhjal leida). Sama juhtub tõenäosuste $P(Y = 1), \dots, P(Y = l)$ leidmisel — viimase tõenäosuse saame leida kasutades teisi. Seega valimi põhjal tuleb leida kokku $1(\text{valimi suurus } N) + k - 1(\text{tunnuse } X \text{ jaotus: } P(X = 1), \dots, P(X = k - 1)) + l - 1(\text{tunnuse } Y \text{ jaotus: } P(Y = 1), \dots, P(Y = l - 1)) = k + l - 1$ parameetrit ja antud juhul tuleb hii-ruut testi vabadusastmete arvuks

$$\begin{aligned} df &= kl - (k + l - 1) \\ &= kl - k - l + 1 \\ &= (k - 1)(l - 1). \end{aligned}$$

Näide 4.6 Vaatame, kas saame tõestada, et tudengite alkoholitarbimise ja nende soo vahel eksisteerib seos. Algandmed on esitatud tabelis 4.8.

Esmalt hindame nii õlletarbimise kui ka soolise jaotuse (tabelid 4.9 ja 4.10).

Nüüd leiame tõenäosused $P(\text{sugu} = i, \text{õlu} = j)$. Leitud tõenäosused on toodud tabelis 4.11.

Nüüd leiame teooria (sõltumatus) poolt ennustatavad tudengite arvud $N_{\text{sugu}=i, \text{õlu}=j}$, vaata tabel 4.12.

Tabel 4.8: Seos tudengite soo ja nädalase õlletarbimise vahel

sugu õlu	ei tarbi	alla pudeli	1-4	5 või enam	kokku
naine	241	222	43	6	512
mees	25	44	49	31	149
kokku	266	266	92	37	661

Tabel 4.9: Õlletarbimise jaotus

x	ei tarbi	alla pudeli	1-4	5 või enam
$P(\text{õlu}=x)$	$0.4024 = 266/661$	0.4024	0.1392	0.0560

Tabel 4.10: Sooline jaotus

x	naine	mees
$P(sugu = x)$	$0.7746 = 512/661$	0.2254

Tabel 4.11: Oodatavad tõenäosused sõltumatuse korral

sugu õlu	ei tarbi	alla pudeli	1-4	5 või enam	kokku
naine	$0.3117 = 0.4024 * 0.7746$	0.3117	0.1078	0.0434	0.7746
mees	$0.0907 = 0.4024 * 0.2254$	0.0907	0.0314	0.0126	0.2254
kokku	0.4024	0.4024	0.1392	0.0560	1

Tabel 4.12: Sõltumatuse korral oodatavad tudengite arvud

sugu õlu	ei tarbi	alla pudeli	1-4	5 või enam	kokku
naine	$0.3117 * 661 = 206.04$	206.04	71.26	28.66	512
mees	59.96	59.96	20.74	8.34	149
kokku	266	266	92	37	661

Ja lõpuks võime leida hii-ruut statistiku väärtuse:

$$\chi^2 = \frac{(241 - 206.04)^2}{206.04} + \frac{(25 - 59.96)^2}{59.96} + \dots + \frac{(31 - 8.34)^2}{8.34} = 161.$$

*Vabadusastmeid oli $df = (2 - 1) * (4 - 1) = 3$, hii-ruut statistiku kriitiline väärtus on 7,815, meie statistiku väärtus on aga palju suurem — palju suurem kui sõltumatute tunnuste puhul oleks võinud tulla. Järeldus: nullhüpotees ei saa kehtida, tunnused ei saa olla sõltumatud. Tunnuste sugu ja õlletarbimine vahel esineb statistiline seos. Seega saame väita, et eri soost tudengite õlletarbimisharjumused on erinevad.*

4.6 Teisi teste

Wilcoxon (Mann-Whitney) test — t-testi analoog, aga ei eelda, et uuritav tunnus oleks normaaljaotusega (aga eeldab, et uuritav tunnus on pidev). Juhul kui Wilcoxon (Mann-Whitney) test otsustab alternatiivse hüpoteesi kasuks — ühes populatsioonis on uuritava tunnuse väärtused suuremad kui teises — võib osutada raskeks kirjeldada, kuidas (ja kui suurelt) need kaks populatsiooni siis ikkagi erinevad teineteisest. Erinevus Wilcoxon testi mõttes ei pruugi tähendada keskväärtuste erinevust, samuti ei saa sellest, et Wilcoxon test lükkas ümber nullhüpoteesi järeldata veel mediaanide erinevust. Seega statistiliselt olulise testitulemuse korrektne interpreteerimine võib osutada keerukaks.

Kolmogorov-Smirnovi test — kontrollib hüpoteesi, kas uuritava tunnuse jaotus kahes populatsioonis on sama või mitte. Seega, kui ühes populatsioonis on näiteks uuritava tunnuse hajuvus suurem kui teises, siis Kolmogorov-Smirnovi test avastab (suure valimi korral) erinevuse. Või kui populatsioonide keskväärtused on erinevad. Või kui leidub mõni muu erinevus uuritava tunnuste jaotustes. Kolmogorov-Smirnovi testi kasutatakse vahel ka kontrollimaks, kas meie uuritava tunnuse jaotus võiks olla mõni kindel, hästituntud teoreetiline jaotus. Testi eeldused: Nõuab, et uuritav tunnus oleks pidev. Paljud statistikaprogrammid võivad anda kahtlaseid tulemusi, kui esineb (palju) kokkulangevaid uuritava tunnuse väärtuseid. Märkus: Kolmogorov-Smirnovi test on enamasti üsna madala võimsusega — tema võime märgata erinevusi on kehvavõitu. Sestap vajame alternatiivse hüpoteesi tõestamiseks suurt vaimu.

F-test — kontrollib hüpoteesi uuritava tunnuse hajuvuse kohta. Näiteks võime F-testi abil kontrollida, kas uuritava tunnuse hajuvus kahes populatsioonis on samasuur või mitte (näiteks kas kahe sordi saagikused — näiteks üle aastate — on sama stabiilsed või on üks vaadeldav sort tundlikum keskkonnamõjudele kui teine). Eeldused — uuritava tunnuse jaotuseks mõlemas populatsioonis peab olema normaaljaotus.

4.7 Mitmese võrdluse probleem

Tõestanud ühe või teise hüpoteesi oleme harjunud tõestatud väidet õigeaks pidama sõltumata sellest, kui mitu korda me oleme varem üritanud meid huvitavat hüpoteesi tõestada. Samuti ei oska me harilikult tähelepanu pöörata

sellele, milliseid ja kui paljusid teisi hüpoteese me oleme proovinud tõestada. Paraku on hüpoteeside statistilise kontrollimise korral olukord keerulisem, sest kasutatav meetodika annab eriti jäärapäisele uurijale “järgi” ja võib mõnigi kord lugeda alternatiivset hüpoteesi tõestatuks ka siis, kui tegelikult kehtib nullhüpotees.

Näide 4.7 *Kujutame ette olukorda, kus kehtib nullhüpotees (imeväetisest pole kasu). Alternatiivse hüpoteesi kehtivuses veendunud inimene korraldab uuringu ja jõuab tulemuseni, et peab jääma nullhüpoteesi juurde. Oma töö õigsusesse uskudes korraldab ta selle peale uue uuringu ja veel ühe uuringu jne. Korraldades 20 korrektselt läbiviidud uuringut, igaüks neist teostatud olulisuse nivool 0,05, on tõenäosus, et ükski neist kahekümnest uuringust ei otsusta alternatiivse hüpoteesi tõestatuks lugeda $(1-0,05)^{20}=0,358$. Kui asjast huvitatud jäärapäine teadlane otsustab avaldada vaid sellise uuringu tulemused, mis alternatiivse hüpoteesi õigeks loevad, siis teeb see hüpoteetiline jäärapäine teadlane esimest liiki vea tõenäosusega $1-0,358=0,642$.*

Näide 4.8 *Otsitakse põhjuseid, mis võiksid soodustada haiguse tekkimist. Viidi läbi uuring, kus haigetel ja tervetel taimedel mõõdeti sadade erinevate tunnuste väärtused (kui toitainerikkas mullas nad kasvavad, kui pikk on neid kasvatav talumees, palju sadas vihma eelmisel kuul jne). Andmed kokku kogutud, hakatakse t-testi abil võrdlema, kas haigeks jäävad taimed kasvasid toitainerikkamas mullas kui terved, kas haigeid taimi kasvas pikem talumees kui terveid jne. Oletame, et andmestikus oli 100 tunnust, mis kirjeldasid taimede kasvutingimusi. Iga kasvutingimuse korral võrreldakse terveid ja haigeid ja võrdlemisel kasutatakse olulisuse nivood 0,05. Oletame, et ükski uuritud kasvutingimus tegelikult ei mõjuta (ei muuda) võimalust haigeks jääda (terveid taimi kasvatanud talunike pikkuse keskväärtus oli sama mis haigeid taimi kasvatanud talunike pikkuse keskväärtus jne). Vaatamata sellele võiksite antud uuringu puhul tõestatuks lugeda umbes 5 haiguse tekkimist soodustavat tegurit (tõenäosus, et saame tõestada vähemalt ühe küsimuse korral keskväärtuste erinevuse ehk teeme esimest liiki vea on 0,994).*

Üks levinud meetod mitmese võrdluse ohte vältida on Bonferroni meetod. Juhul kui on tarvis kontrollida n erinevat hüpoteesi ja soovime, et tõenäosus teha üht või enamat valeotsust ohtlikus suunas (I-liiki vea tegemise tõenäosus) ei ületaks α , siis tuleb Bonferroni meetodi järgi teha iga üksiktest kasutades olulisuse nivood α/n .

Tasub tähele panna, et mida rohkem küsimusi (üksikteste) esitame, seda väiksemaks läheb üksiktesti teostamisel tarvitatav olulisuse nivoo. Seega

läheb ka raskemaks alternatiivse hüpoteesi vastuvõtmine. Seetõttu tuleks tehtavate testide arvu hoida siiski võimalikult väike, kontrollides statistilise hüpoteeside kontrollimise protseduuri abil vaid neid hüpoteese, mis teoreetiliste arutelude põhjal võiksid kõige enam huvi pakkuda. Vältida tuleks sihipäratut katsetamist ja proovimist, sest siis võib statistilise hüpoteeside testimise protseduur anda halbu tulemusi.

Bonferroni meetod kipub paljudes olukordades olema veidi liiga konservatiivne, st. tema kasutamisel on maksimaalne esimest liiki vea tegemise tõenäosus enamasti väiksem kui α (5%). Alternatiivina Bonferroni meetodile võib kasutada näiteks Bonferroni-Holmi meetodit. Bonferroni-Holmi meetod eeldab, et kõigepealt teostatakse kõik testid ja leitakse nende olulisustõenäosused. Saadud olulisustõenäosused järjestatakse kasvavalt: $p_1 \leq p_2 \leq \dots \leq p_k$. Otsused nullhüpoteesi kasuks või kahjuks tehakse siis kasutades olulisuse nivoosid

$$\alpha/k, \alpha/(k-1), \alpha/(k-2), \dots, \alpha/2, \alpha.$$

Summarne esimest liiki vea tegemise tõenäosus on ka Bonferroni-Holmi meetodi puhul maksimaalselt α . Bonferroni-Holmi meetod on aga tsipa vähem konservatiivne kui Bonferroni meetod, st. Bonferroni-Holmi meetod otsustab praktikas sagedamini alternatiivse hüpoteesi kasuks kui Bonferroni meetod. Miinuseks on asjaolu, et Bonferroni-Holmi meetod eeldab, et kõik testid on enne tema kasutamist tehtud. See välistab tema kasutamise olukorras, kus soovime teha osa teste tänaste andmete pealt ja osad testid kahe aasta pärast (siis kui on rohkem andmeid kogunenud).

Kui teostavate testide arv kasvab, väheneb Bonferroni meetodi rakendamisel kiiresti ka kasutatav olulisuse nivoo ja alternatiivse hüpoteesi tõestamine osutub sageli äärmiselt raskeks (nõuab tohutult paljude vaatluste olemasolu). Selle tõttu pole statistilised meetodid mitte eriti sobivad katse/eksituse meetodil teaduse tegemiseks (proovime, kas midagi õnnestub). Ideaaljuhul jõutakse arusaamisele hüpoteeside kehtivuse kohta kasutades mittestatistilisi argumente, lähtudes näiteks teoreetilise bioloogia tõekspidamistest, ja statistilist hüpoteeside kontrolli rakendatakse pigem võimalike eksimuste tuvastamiseks oma arutluskäigus.

Praktikas kipub olukord olema muidugi vastupidine, vähesed viitsivad mõelda ja suhteliselt rohkem on neid, kellele meeldib niisama proovida ja katsetada. Sestap on ka olemas tugev surve arutu proovimise jaoks sobivate statistikameetodite järgi. Mitmese testimise kontekstis on üheks huvitavaks kontseptsiooniks nn “valeavastuste määra” piiramine. (False Discovery Rate). Idee selle termini taga on järgmine: las juhtugu esimest liiki vigasid,

see pole suur probleem, aga teeme nii, et vastuvõetud alternatiivsete hüpoteeside seas poleks ekslikult vastuvõetud alternatiivseid hüpoteese mitte rohkem kui mingi lubatud protsent (5%). Ehk teisisõnu öeldes me kontrollime valepositiivsete ehk ekslike “avastuste” osakaalu kõigi “avastuste” seas. Antud teemaga seotud statistilised meetodid on praegu kiirelt arenev valdkond ja vähesed statistikaprogrammid pakuvad võimalust mitmese testimise probleemi ohjamiseks piirata valepositiivsete osakaalu ehk limiteerida False Discovery Rate-i. Üks lihtne võimalus antud lähenemist ise kasutada oleks järgmine:

1. järjestatult testide poolt raporteeritavad olulisustõenäosused kasvavalt,

$$p_1 \leq p_2 \leq \dots \leq p_k.$$

2. Kirjuta välja kriitilised suurused $a/k, 2*a/k, 3*a/k, \dots, a$, kus a näitab maksimaalset nn False Discovery Rate'i.

3. Võrdle esimest, teist jne järjestatud olulisustõenäosust vastava kriitilise suurusega kuni jõuad paarini, kus olulisustõenäosus on suurem kui vastav kriitiline suurus. Loe kõigi talle eelnenud hüpoteeside jaoks alternatiivne hüpotees tõestatuks ja tema ning järgnevate testide puhul jää nullhüpoteesi juurde.

Näide 4.9 Teostati 10 testi, nende olulisustõenäosused (järjestatult) on järgmised: 0,0002; 0,008; 0,01; 0,04; 0,045; 0,12; 0,22; 0,25; 0,62; 0,83.

Vastavad kriitilised väärtused on ($\alpha = 0,05$ korral): 0,005; 0,01; 0,015; 0,02; 0,025; 0,03; 0,035; 0,04; 0,045; 0,05

Esimene olulisuse nivoo, mis on suurem vastavast kriitilisest väärtusest, on 4. olulisuse nivoo. Seega testide 1-3 korral võtame vastu alternatiivse hüpoteesi ja testide 4-10 jaoks jääme nullhüpoteesi juurde.

Vahel on võimalik kontrollitavad hüpoteesid selliselt ümber sõnastada, et tekiks üks uus kontrollitav hüpotees ja paljude üksiktestide asemel tuleb teha üksainus test. Üks selline erijuht on keskmiste mitmene võrdlus. Oletame, et meil on m gruppi, mille keskmist taset tahame võrrelda. Näiteks õpetavad m õpetajat sama õppeainet m -grupile tudengitele. Soovime teada, kas kõigi õpetajate õpilased saavad peale kursuse läbimist (keskmiselt) võrdselt hästi hakkama, või leidub mõni õpetaja, kes teistest paremini oma ainet seletab. Üks võimalus soovitu väljaselgitamiseks on teha $m(m-1)/2$ t -testi võrdlemaks kõigi õpetajate töö tulemusi kõigi teiste õpetajate töö tulemustega. Sealjuures tuleks kasutada mõnda mitmese võrdluse meetodit (näiteks Bonferroni meetodit). Teine võimalus oleks kasutada t -testi edasiarendust disper-

sioonianalüüsi üheainsa hüpoteesipaari kontrollimiseks. Dispersioonanalüüsi vaatleme lähemalt ühes järgnevas loengus.

Ülesanded

1. Ristati haplotüüpidega Aa, Bb ja Aa, Bb isendeid. Saadud andmed on toodud tabelis 4.13. On teada, et alleelid A ja a päranduvad vastavalt mendeli seadustele, st ristamisel Aa x Aa saadud järglane on tõenäosusega 0,25 genotüübiga AA, tõenäosusega 0,50 genotüübiga Aa ja tõenäosusega 0,25 genotüübiga aa. Sama reegli järgi päranduvad ka alleelid B ja b. Küsimus: kas geeni A alleelid (A,a) ja geeni B alleelid (B, b) päranduvad sõltumatult või on tegemist nn geeniaheldusega (linkage)?

Tabel 4.13: Vaatlusandmed

B\A	AA	Aa	aa	kokku
BB	23	7	2	32
Bb	10	32	7	49
bb	3	10	20	33
kokku	36	49	29	114

Kuidas muutub teststatistik ja vabadusastmete arv, kui me ei eelda üksikute alleelide pärandumist vastavalt mendeli seadustele? Näiteks võib olla võimalik, et genotüüpi AA esineb mingil põhjusel heterosügootsete vanemate lastel “liiga” palju? Ehk mis juhtub siis, kui teeme tavalise hii-ruut testi kahe muutuja sõltumatuse kontrollimiseks? (Vihe — kirjeldatud kahe juhu korral tulevad nii teststatistiku väärtused kui ka vabadusastmete arvud erinevad, samuti võime jõuda erinevate otsusteni!)