Peatükk 7

Hüpoteeside statistiline kontrollimine

Jälgides enda ümber toimuvat, elusloodust vaadeldes või aretusprotsessi juhtides, vahel ka teadusartikleid lugedes, tekib mõtlevatel inimestel paratamatult oletusi või hüpoteese ümbritseva maailma kohta. Oletus või kahtlus pole veel teadmine. Kuidas leida kinnitust oma kahtlustele? Kas kogutud andmed kinnitavad või hoopistükkis kummutavad meie oletuse/hüpoteesi?

7.1 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, millise valimi peaksime saama, 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 ennustuseni). 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 nullhüpoteesi 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/mõtteviis tagab (vähemalt) sama hea tulemuse kui uuendajate/reformijate poolt pakutav lähenemisviis. Nullhüpoteesi kummutades — 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 arvatavasti unistab alternatiivse hüpoteesi tõestamisest — mis võiks olla veel ihaldusväärsem, kui näidata, et Sinu idee tagab parema tulemuse kui (teaduslike) esivanemate põlvkondadepikkune kogemus.

Näide 7.1 *Nullhüpotees: Mu vastaspartner mängib ausalt, ka ta täring on aus. Formaalselt kirja panduna H_0: P(sõber viskab täringuga 6 silma)=1/6.*

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

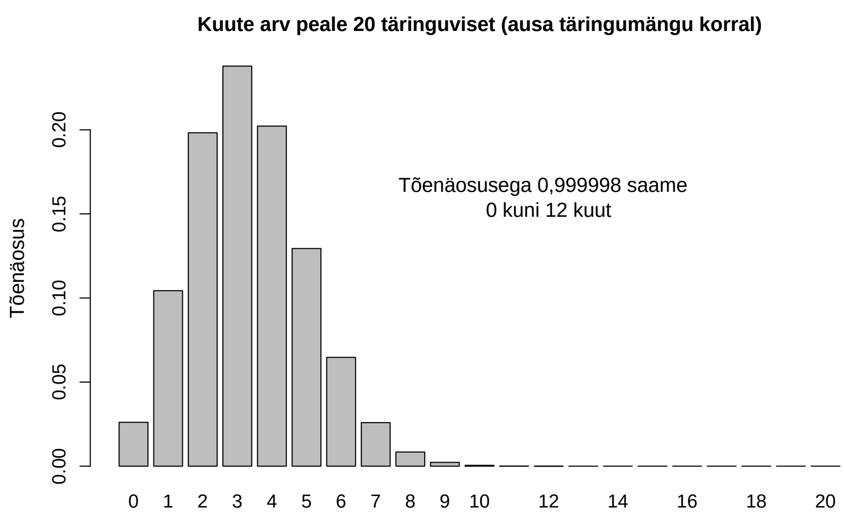
Kuidas statistiliselt (vaid vaatlusandmetele tuginedes) saaks näites esitatud alternatiivset hüpoteesi tõestada? Siin ei ole tegelikult midagi tavamõistusele keerulist.

Näite 7.1 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 auväärt 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 määrib 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 7.1 (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 7.1: 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 kvantitatiivselt, 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 7.1: Esimest ja teist liiki viga

	Tegelikult kehtib H_0	Tegelikult kehtib H_1
Jääme H_0 juurde	õige otsus	II liiki viga
tõestame H_1	I 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 nivoosid — 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 7.2 on antud I ja II liiki vea tegemise tõenäosused ja testi võimsus sõltuvalt populatsiooni tegelikust keskväärtusest.

Tabel 7.2: 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.

7.2 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 kasvatusmeetodi korral. Kas poleks ahvatlev liig ülbeid 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 normaaljaotusega 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.

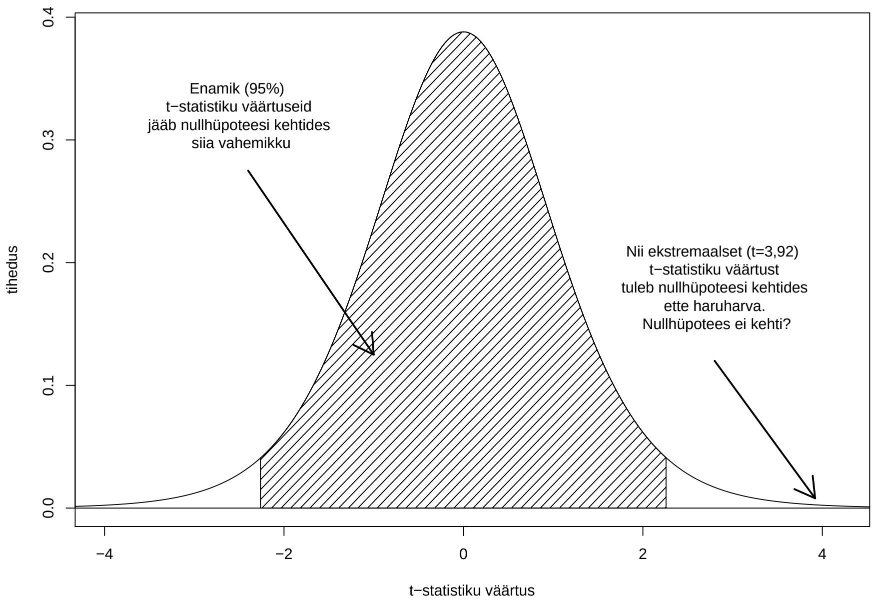
Kui nullhüpotees kehtib, siis iga uurija poolt (iga uue valimi põhjal arvutatud) leitud t-statistiku väärtus tuleb küll erinev, kuid enamikel juhtudel jäävad leitud väärtused üsnagi kitsastesse piiridesse, nulli lähedale. Valimi suuruse $n = 10$ korral on t-statistiku väärtuste jaotus (nullhüpoteesi kehtides) selline, nagu kujutatud joonisel 7.2. 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 7.2: 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 7.2 *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 ($\mu_0 = 0$):*

$$H_0: \quad \mu = 0$$
$$H_1: \quad \mu \neq 0.$$

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:

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

Leiame t-statistiku väärtuse:

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

Kas saame tõestatuks lugeda, et kemikaalil oli mõju kasside reaktsioonikiirusele? 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.

7.3 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 keskväärtus 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 hüpoteesipaar — kas uuritava tunnuse keskväärtus on sama, sõltumata katsetingimustest — on kirja pandav järgmiselt:

$$\begin{aligned} H_0: &\quad EX = EY \\ H_1: &\quad EX \neq EY. \end{aligned}$$

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

7.4 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:

$$\begin{aligned} H_0 :&\quad EX = EY \\ H_1 :&\quad EX \neq EY? \end{aligned}$$

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.

7.4.1 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 ($n_1 + n_2 > 30$).

7.4.2 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.

7.4.3 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 7.4.2.

7.5 Teisi teste

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

Kolmogorov-Smirnovi test — kontrollib hüpoteese, 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 valimit.

F-test — kontrollib hüpoteese 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 normaaljaotus.