Hüpoteeside testimine
t-test

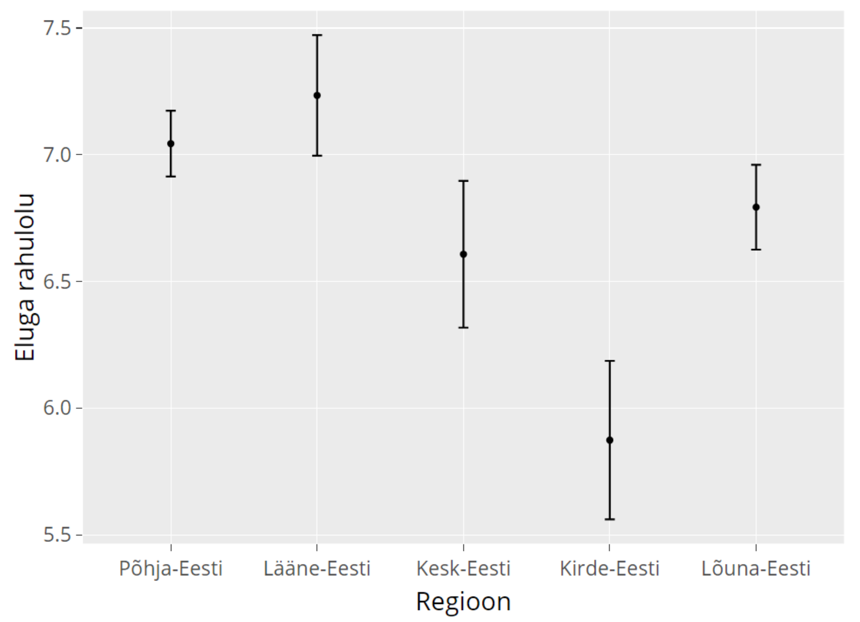
Sissejuhatus kvantitatiivsetesse meetoditesse

Indrek Soidla

Eelmine kord

- Kas keskmine eluga rahulolu erineb Eestis regiooniti?
- Andmed valimi kohta, tahtsime järeldusi teha populatsiooni kohta (rahvastiku tasandil)
- Arvutasime keskmistele usaldusvahemikud
- Usaldusnivool 95% saime öelda, et
  - eluga rahulolu alusel eristuvad kolm regioonide rühma
  - Põhja- ja Lääne-Eestis oli eluga rahulolu keskmine üle 7 või 7 piirimail
  - Kesk- ja Lõuna-Eestis jäi eluga rahulolu keskmine alla 7 palli
  - Kirde-Eestis oli eluga rahulolu teistest regioonidest madalam ja keskmine jäi alla 6,2 palli
- Kas keskmine eluga rahulolu Põhja-Eestis ja Lõuna-Eestis erineb?
- Usaldusnivool 95% usaldusvahemikud kattuvad, kuigi vähesel määral
- Antud usaldusvahemike võrdlemisel seda väita ei saa
- Kui usaldusvahemikud mingil määral kattuvad, siis keskmiste erinevuse osas saaks täpsema järelduse statistilise testi põhjal, mis testib keskmiste erinevust
- Teeme läbi hüpoteeside testimise t-testi põhjal

Regioon	Keskmine	Standardviga	Alumine usalduspiir	Ülemine usalduspiir
Põhja-Eesti	7,04	0,07	6,91	7,17
Lääne-Eesti	7,23	0,12	7,00	7,47
Kesk-Eesti	6,61	0,15	6,32	6,90
Kirde-Eesti	5,87	0,16	5,56	6,19
Lõuna-Eesti	6,79	0,09	6,63	6,96



Hüpoteeside testimine

- Statistiline hüpotees
  - oletus üldkogumi (nt selle jaotuse või parameetri) kohta,
  - mida testitakse valimiandmete alusel
- Kontrollime / testime hüpoteesi, et Põhja- ja Lõuna-Eestis keskmine eluga rahulolu erineb
- Peame sõnastama alternatiivhüpoteesi ja nullhüpoteesi

Hüpoteeside testimine

- Alternatiivhüpotees (H1) – reeglina sõnastatud nii, et väidab erinevuse esinemist / seose kehtimist
- Nullhüpotees (H0) – vastupidine sellele
- Praegusel juhul:
  - H1: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu erineb ($m_{PE} \neq m_{LE}$)
  - H0: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu ei erine ($m_{PE} = m_{LE}$)
- Oluline, et hüpoteesipaari hüpoteesid
  - välistavad teineteist
    - nt ei saa olla, et H1: $m_{PE} \geq m_{LE}$ ja H0: $m_{PE} \leq m_{LE}$
  - katavad kõik võimalikud stsenaariumid
    - nt ei saa olla, et H1: $m_{PE} \geq m_{LE}$ ja H0: $m_{PE} = m_{LE}$
- NB! Hüpoteesides ei pruugi olla öeldud, et need käivad populatsiooni kohta, aga meeles tuleb pidada, et tegu on statistiliste hüpoteesidega

Hüpoteeside testimine

- Praegusel juhul:
  - H1: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu erineb ($m_{PE} \neq m_{LE}$)
  - H0: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu ei erine ($m_{PE} = m_{LE}$)
- Tegu on kahepoolse hüpoteesipaariga – H1 seondub arvtelje kahe lõiguga
  - H1 kohaselt võib olla, et $m_{PE} < m_{LE}$, aga võib olla ka, et $m_{PE} > m_{LE}$
- Kui tegu ühepoolse hüpoteesipaariga, H1 seondub arvtelje ühe lõiguga
- Võimalik ühepoolne hüpoteesipaar
  - H1: Põhja-Eestis on keskmine eluga rahulolu kõrgem, kui Lõuna-Eestis ($m_{PE} > m_{LE}$)
  - H0: Põhja-Eesti keskmine eluga rahulolu ei ole kõrgem kui Lõuna-Eestis ($m_{PE} \leq m_{LE}$)

Esimest ja teist liiki viga

- Hüpoteesipaari testitakse statistilise testi põhjal, valimiandmete alusel
- Kaks võimalikku otsust: kas
  - võtame vastu alternatiivhüpoteesi või
  - jääme nullhüpoteesi juurde
- Tulenevalt sellest, et juhuvalim võib juhuslikkuse tõttu mõnevõrra populatsioonist erineda, on meil võimalik eksida
- Esimest liiki viga – populatsioonis kehtib H0 (erinevust ei esine), aga jõuame valimiandmete põhjal järeldusele, et kehtib H1 (keskmiste erinevus)
  - Millisel juhul teeksime 1. liiki vea oma hüpoteesipaari puhul?
- Teist liiki viga – populatsioonis kehtib H1 (erinevus esineb), aga jõuame valimiandmete põhjal järeldusele, et kehtib H0 (erinevust ei esine)
  - Millisel juhul teeksime 2. liiki vea oma hüpoteesipaari puhul?

Esimest ja teist liiki viga

Populatsioonis kehtib	Otsus valimi alusel, et kehtib	
	nullhüpotees H0	alternatiivhüpotees H1
nullhüpotees H0	viga ei teki	tekib esimest liiki viga
alternatiivhüpotees H1	tekib teist liiki viga	viga ei teki

- Hüpoteesipaar sõnastatakse nii, et esimest liiki veast tulenev probleem suurem (tahame eelkõige vältida selle vea tegemist) =>
  - nullhüpotees väidab soovitule (sellele, mis meid huvitab) vastupidist
- Nt mõõdame silla tugevust, viime läbi hulga katseid juhuslikes kohtades silla kandekonstruktsioonis
  - H0: sild ei ole piisavalt vastupidav
  - H1: sild on piisavalt vastupidav

Esimest ja teist liiki viga

	Otsus valimi alusel, et kehtib	
Populatsioonis kehtib	nullhüpotees H0	alternatiivhüpotees H1
nullhüpotees H0	viga ei teki	tekib esimest liiki viga
alternatiivhüpotees H1	tekib teist liiki viga	viga ei teki

- Hüpoteeside testimise loogika: me ei tea, kas populatsioonis kehtib H0 või H1
- Püüame jõuda tõele jälile, keskendudes mitte sellele, kas kehtib H1, vaid sellele, kas kehtib H0
- Seejuures prioriteet on vältida esimest liiki viga (ka teist liiki viga oleks halb, aga esimest liiki vea vältimine olulisem)
- Analoogia: kohtuprotsess nullhüpoteesi üle
- Meie oleme süüdistajad, kes püüavad näidata, et H0 on süüdi (et populatsioonis H0 ei kehti)
- Tõendusmaterjal: valimiandmed, statistiline test
- Kehtib süütuse presumptsioon: nullhüpotees on süütu kuniks meil ei õnnestu mõistliku kahtluseta tõendada vastupidist
- William Blackstone: *On parem, et kümme süüdlast pääsevad, kui et üks süütu kannatab.*
- Oluline vältida süütu süüdimõistmist (esimest liiki viga)

Esimest ja teist liiki viga

	Otsus valimi alusel, et kehtib	
Populatsioonis kehtib	nullhüpotees H0	alternatiivhüpotees H1
nullhüpotees H0	viga ei teki	tekib esimest liiki viga
alternatiivhüpotees H1	tekib teist liiki viga	viga ei teki

- Oluline vältida süütu süüdimõistmist (esimest liiki viga)
- Esimest liiki vea suurust mõõdetakse tõenäosusega seda viga teha
- Hüpoteeside testimisel määratakse eelnevalt kindlaks esimest liiki vea lubatav ülempiir α
    - Seda nimetatakse olulisuse nivooks
    - Tavaliselt 0,05, rangem 0,01 või 0,001, leebem 0,1
- H0 õnnestub kummutada, kui esimest liiki vea tõenäosus on väiksem olulisuse nivoost
- St nullhüpotees mõistetakse süüdi, kui võimalus eksida (süütu süüdi mõista) on piisavalt väike

Hakkame testima – kuidas?

- Niisiis, hüpoteesipaari testitakse valimiandmete alusel, otsus tehakse populatsiooni kohta
- Keskmised ja nende erinevuse arvutame valimi alusel, aga mille alusel teha otsus populatsiooni kohta?
- Selleks ongi statistiline test – võimaldab hinnata erinevuse / seose esinemist populatsioonis

Valime statistilise testi

- Aritmeetiliste keskmiste võrdlemiseks: t-test
- Sõltumatute kogumite t-test: võrdleme arvulise tunnuse keskmisi kategoriseeriva tunnuse lõikes

$$t = \frac{m_1 - m_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

- $m_1,\ m_2$ – arvulise tunnuse keskmised kogumites
- $s_1^2,\ s_1^2$– arvulise tunnuse dispersioonid kogumites
- s – kahe kogumi ühise standardhälbe hinnang
- n_1, n_2 – indiviidide arvud kogumites

Sõltumatute kogumite t-test

- Aritmeetiliste keskmiste võrdlemiseks: t-test
- Sõltumatute kogumite t-test: võrdleme arvulise tunnuse keskmisi kategoriseeriva tunnuse lõikes

$$t = \frac{m_1 - m_2}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

- Eeldused:
  - Arvulise tunnuse jaotus gruppide lõikes normaaljaotuse lähedane
  - Arvulise tunnuse hajuvus gruppide lõikes sarnane
    - kui ei ole => Welchi t-test

$$t = \frac{(m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Arvutame teststatistiku

- Kui kasutame t-testi, nimetatakse teststatistikut t-statistikuks (t väärtus valemis)

$$t = \frac{m_1 - m_2}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{0,25}{1,94} * 18,09 = 2,33$$

- Saime teada t väärtuse, aga kuidas saame selle alusel midagi öelda populatsiooni kohta või teha järelduse hüpoteesipaari kohta?

Teeme järelduse populatsiooni osas

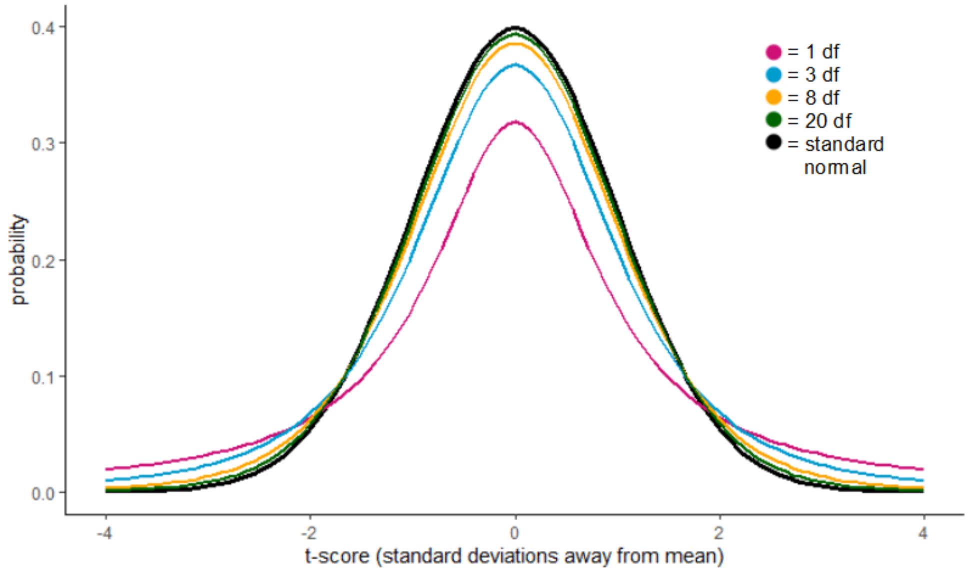
- Võtame appi t-jaotuse
- Analoogne usaldusvahemiku mõtteharjutusega
  - Olgu meil populatsioon, kus kehtib H0 (keskmiste erinevust pole)
  - Kui võtaksime sellest populatsioonist palju valimeid,
  - mõõdaksime samu tunnuseid ja arvutaksime iga valimi andmete alusel t-statistikud, siis
  - jaotuksid saadud t-statistikud t-jaotuse kohaselt
  - indiviidide suure arvu korral ühtib t-jaotus normaaljaotusega, seega
  - sajast valimist
    - 10 valimis $t > 1{,}64$ või $t < -1{,}64$
    - 5 valimis $t > 1{,}96$ või $t < -1{,}96$
    - 1 valimis $t > 2{,}58$ või $t < -2{,}58$
- Eelnevaid piire nimetatakse t-statistiku kriitilisteks väärtusteks
- t-statistiku kriitiliseks piirkonnaks nimetatakse
  - t positiivsest kriitilisest väärtusest suuremate väärtuste piirkonda ja
  - t negatiivsest kriitilisest väärtusest väiksemate väärtuste piirkonda
- Reaalne olukord: me ei tea, kas populatsioonis kehtib H0 või H1; meil on ainult üks valim
- Kui valimi alusel saadud t-statistik asub kriitilises piirkonnas, saame H0 kummutada ja kinnitada H1
- Miks?

Teeme järelduse populatsiooni osas

- Tuletame meelde: olulisuse nivoo – esimest liiki vea lubatav ülempiir
  - Olulisuse nivool 0,05 on t-statistiku kriitilised väärtused $t = 1{,}96$ ja $t = -1{,}96$
  - See tähendab, et kui võtaksime populatsioonist, kus kehtib H0 (keskmiste erinevust pole) 100 valimit, siis
  - viies neist saaksime t väärtuse, mille absoluutväärtus on suurem 1,96-st
- Reaalsuses saame võtta ühe valimi
- Me ei tea, kas populatsioonis kehtib H0 või H1
- Kui populatsioonis kehtib H0, siis tõenäosus, et saaksime valimi, kus t väärtuse absoluutväärtus on üle 1,96, on väga väike (5%)
  - Kui t väärtus on piisavalt suur, st üle olulisuse nivoo,
  - siis võime eeldada, et on kogunenud piisavalt tõendusmaterjali, et
  - n-ö nullhüpoteesi süüdimõistmisel (väites, et H0 ei kehti), mitte eksida (mitte teha esimest liiki viga) –
  - saame kinnitada nullhüpoteesi
- Oluline siiski mõista, et otsus on tingimuslik
  - põhimõtteliselt on võimalik ka populatsioonist, kus kehtib H0, saada valim, kus $|t| > 1{,}96$ (kuigi see võimalus on väike)

Kriitilised väärtused sõltuvad vabadusastmete arvust

- Eelnevad kriitilised piirid täpsed eeldusel, et indiviide on palju
- Kui indiviide on vähe, ei lange t-jaotus täpselt kokku normaaljaotusega
- Täpsemalt: t-statistikud jaotuvad t-jaotuse kohaselt vabadusastmete arvu $n - 1$ korral
- Vabadusastmete arv – väljendab tõsiasja, et kui arvutame t-statistikut, siis meil on andmed n indiviidi kohta, kuid ka teatud piirangud
- Sisuline tähtsus on selles, et sõltuvalt vabadusastmete arvust => indiviidide arvust on t-jaotus mõnevõrra erineva kujuga =>
    - t kriitilised väärtused mõnevõrra erinevad
- Nt olulisuse nivool 0,05
    - kui $df = n - 1 = 80$, siis $t = 1{,}99$
    - kui $df = n - 1 = 30$, siis $t = 2{,}04$
    - kui $df = n - 1 = 8$, siis $t = 2{,}31$



Olulisuse tõenäosus

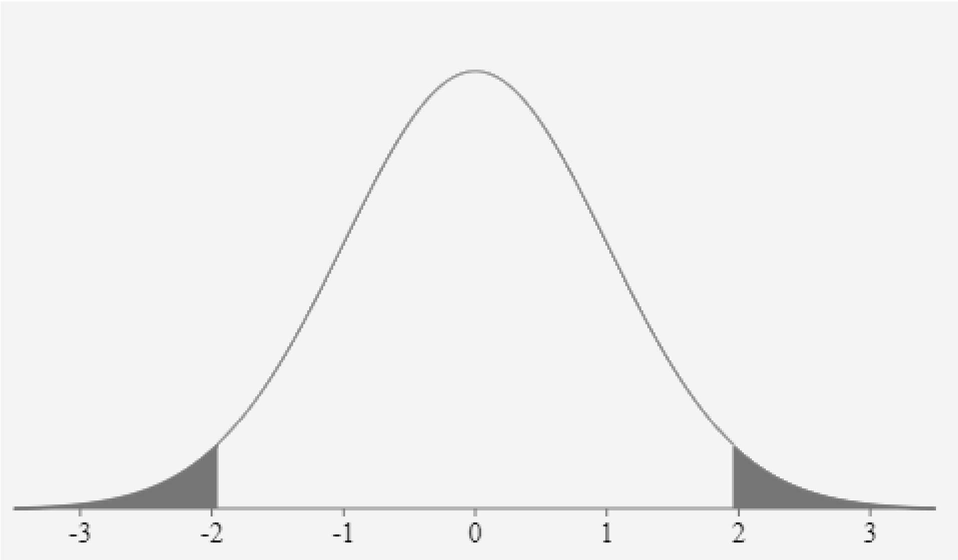
- Kas peame erinevad t väärtused meeles pidama?
- Ei pea, ei pea ka kuskilt üles otsima
- Kui andmeanalüüsiprogramm arvutab t väärtuse, annab ka olulisuse tõenäosuse (*p*-väärtus) vastava vabadusastmete arvu juures
- Olulisuse tõenäosus
  - tõenäosus teha esimest liiki viga – laias laastus pole otseselt vale, aga pole päris täpne
  - tõenäosus saada nullhüpoteesi kehtimise korral vastav või sellest veel suurem teststatistiku väärtus
  - ehk tõenäosus saada populatsioonist, kus kehtib H0, valimiandmed, mille põhjal saaksime vähemalt nii suure teststatistiku väärtus
- Meie näites $t = 2{,}33, df = 1384$, siis $p = 0{,}02$
- Kui olulisuse tõenäosus < olulisuse nivoo, tuleb kummutada / tagasi lükata nullhüpotees ja vastu võtta / kinnitada alternatiivhüpotees
  - Saame öelda, et keskmiste erinevus on *statistiliselt oluline* vastaval olulisuse nivool
- Kui olulisuse tõenäosus ≥ olulisuse nivoo, ei saa kinnitada alternatiivhüpoteesi, tuleb jääda nullhüpoteesi juurde
- NB! Nullhüpoteesi ei saa kinnitada / tõestada – kui ei õnnestu kinnitada alternatiivhüpoteesi, tuleb jääda nullhüpoteesi juurde
  - Me ei saa öelda, et nullhüpotees tingimata kehtib, meil lihtsalt ei õnnestunud leida tõendust vastupidisele

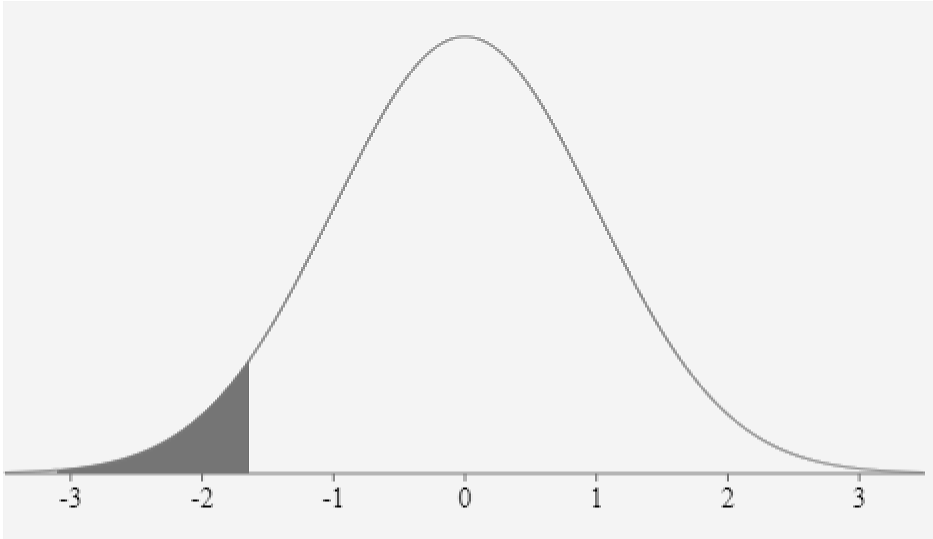
Tulemuste esitamine

- Võib tunduda, et järelduste tegemiseks polegi vaja muud, kui vaadata olulisuse tõenäosust ja raporteerida see
- Päris nii see pole
- Olulisuse tõenäosus (nagu ka usaldusvahemik) käib alati konkreetse näitaja kohta
- Seega, t-testi tulemuste põhjal saame olulisuse nivool 0,05 väita, et 2016. aastal keskmine eluga rahulolu Põhja- ja Lõuna-Eestis erines ($t = 2{,}33, df = 1384$, $p = 0{,}02$)
- Või: …erines (t-statistiku väärtus 2,33, vabadusastmete arv 1384, olulisuse tõenäosus 0,02)

Aga kui meil on ühepoolne hüpoteesipaar?

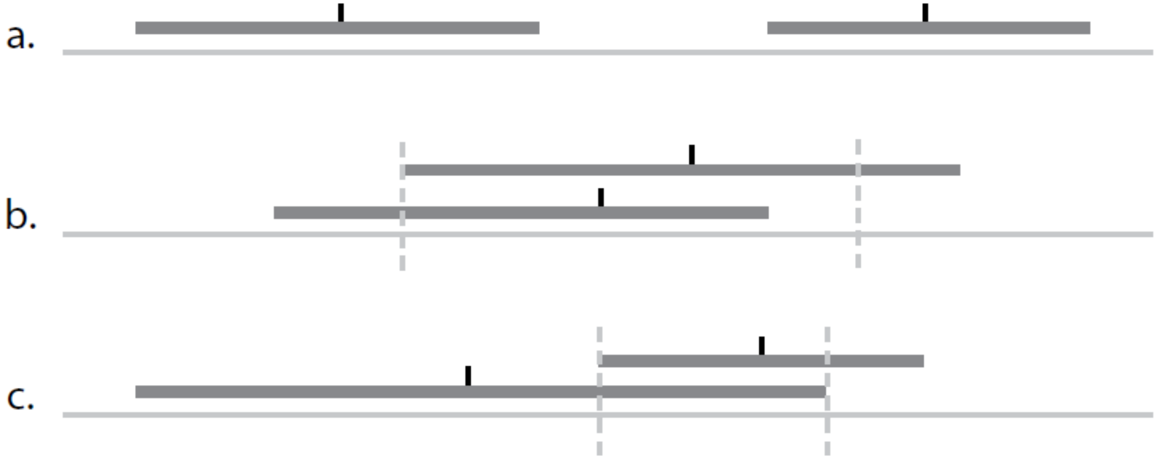
- Eelnev arutelu ja näide põhinesid kahepoolse hüpoteesipaari testimisel
  - H1: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu erineb ($m_{PE} \neq m_{LE}$)
  - H0: Põhja- ja Lõuna-Eestis keskmine eluga rahulolu ei erine ($m_{PE} = m_{LE}$)
- Kui hüpoteesipaar on ühepoolne, tuleb
  - teststatistiku kriitilised väärtused arvutada natuke teisel põhimõttel
- Kahepoolne hüpoteesipaar: kriitilised piirkonnad asuvad t-jaotuse mõlema „saba“ all
- Ühepoolne hüpoteesipaar: kriitiline piirkond asub t-jaotuse selle „saba“ all, mida väidab H1





Keskmiste usaldusvahemikud vs t-test

- Keskmiste usaldusvahemike osalisel kattumisel võivad keskmised siiski statistiliselt oluliselt erineda



a. Saab erinevust väita
b. Erinevust ei saa väita
c. Kontrolli keskmiste erinevust t-testi alusel

Tooding (2015: 165)

Ühe kogumi t-test

- Põhimõte sarnane sõltumatute kogumite (kahe kogumi) t-testile
- Ühe valimi alusel mõõdetud kogumi keskmise asemel on mingi kindel väärtus
- Testime, kas valimi alusel arvutatud keskmine sellest erineb
- Nt kui tahame teada, kas valimi alusel saab väita, et Eesti elanike keskmine eluga rahulolu on üle skaala keskpunkti
  - H1: $m > 5$
  - H0: $m \leq 5$

Paariskogumite t-test

- Kasutatakse, kui vaatlused on omavahel seotud
- St me ei võrdle kahe erineva grupi keskmisi, vaid samade indiviidide keskmisi, mis nt
  - on mõõdetud erinevatel ajahetkedel
    - nt samade indiviidide rahulolu eluga enne ja majanduskriisi ajal
  - mõõdavad erinevaid nähtusi
    - nt samade indiviidide rahulolu majanduse olukorraga ja valitsuse tegevusega