

# Hüpoteeside testimine & t-test

Indrek Soidla

28 | 10 | 2022

Hüpoteeside testimise mõte

# Võtame ühe näite

- Kas usaldus Riigikogu vastu on sisseerändajate ja Eestis sündinute võrdluses keskmiselt erinev?
  - Teisiti sõnastades: kas sisseerändajad ja Eestis sündinud usaldavad Riigikogu keskmiselt erineval määral?
  - Analüüsime seda ESS 2018. aasta andmete alusel
- Näeme, et keskmised erinevad 0,4 palli võrra, mis pole just suur erinevus, aga võib siiski sisulisel tasandil oluline olla
- See erinevus käib selle valimi kohta, kas saame keskmiste erinevust väita ka populatsioonis (rahvastikus)?

```
library(haven)
library(tidyverse)
library(TAM)

ee9 <- read_spss("data/ee9.sav")

ee9 <- ee9 %>%
  mutate(brncntr2 = recode(as.factor(brncntr),
                           "1" = "Jah",
                           "2" = "Ei"))

ee9 %>%
  group_by(brncntr2) %>%
  summarise(keskmine = weighted_mean(trstpr1, w = pspwght))
```

```
## # A tibble: 2 x 2
##   brncntr2 keskmine
##   <fct>      <dbl>
## 1 Jah       4.84
## 2 Ei        5.24
```

# Arvutame keskmiste usaldusvahemikud

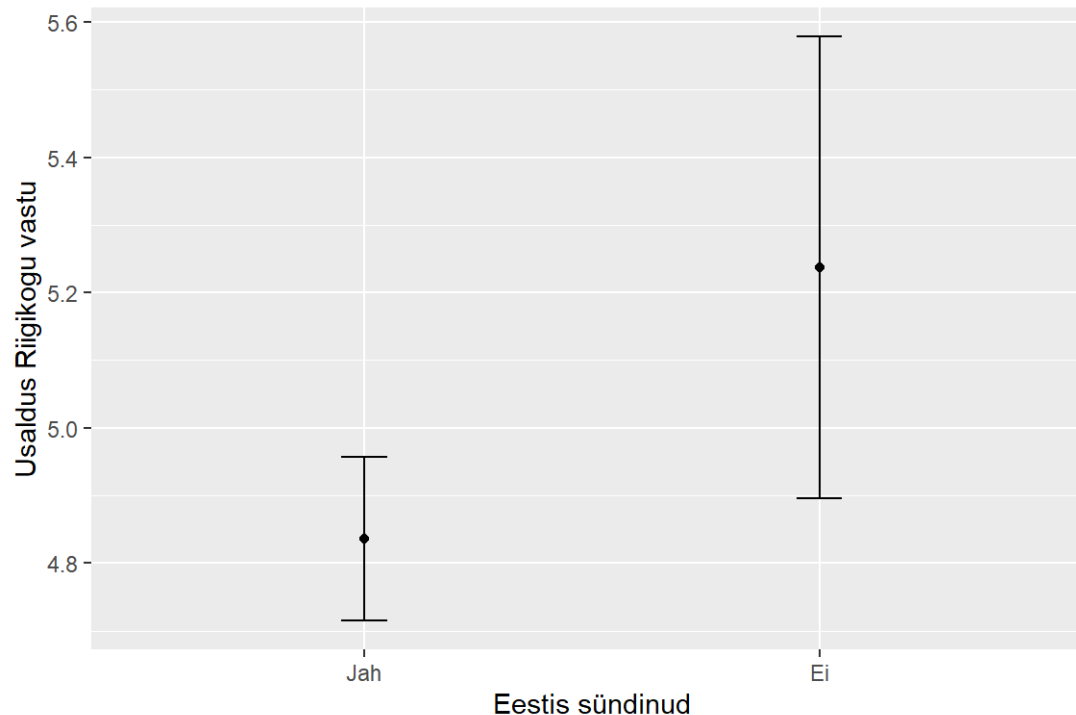
- Arvutame keskmiste usaldusvahemikud usaldusnivool 95% (vaikeseadena seadistatud, ei pea koodis eraldi välja kirjutama)

```
library(survey)

ee9w <- svydesign(id = ~1, data = ee9, weights = ~pspwght)

svyby(~trstprl, ~brncntr2, design = ee9w, FUN = svymean, na.rm = TRUE, vartype = c("se", "ci")) %>%
  ggplot(aes(x = brncntr2, y = trstprl)) +
  geom_point(stat = "identity") +
  geom_errorbar(aes(ymin = ci_l, ymax = ci_u), width = 0.1, color = "black") +
  xlab("Eestis sündinud") +
  ylab("Usaldus Riigikogu vastu")
```

# Arvutame keskmiste usaldusvahemikud



Joonis 1. Usaldus Riigikogu vastu (möödetud skaalal 0-10) Eestis sündinute ja sisseännanute seas: aritmeetilised keskmised koos usaldusvahemikega usaldusnivool 95%. Andmed: Euroopa Sotsiaaluuring 2018.

- Usaldusnivool 95% saame öelda, et
  - Eestis sündinud inimeste seas on parlamendi usaldamise tegelik keskmine vahemikus 4,71 kuni 4,96 palli
  - mujal riigis sündinute seas on tegelik keskmine vahemikus 4,9 kuni 5,58 palli
- Usaldusnivool 95% usaldusvahemikud kattuvad, kuigi vähesel määral
- Keskmiste usaldusvahemike võrdlemise põhjal keskmiste erinevust 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

# Hüpoteeside testimine

- Statistiline hüpotees
  - oletus üldkogumi (nt selle jaotuse või parameetri) kohta,
  - mida testitakse valimiandmete alusel
- Kontrollime / testime hüpoteesi, et Eestis sündinute ja sisse­rännanute usaldus Riigikogu vastu keskmiselt erineb
- Milles on hüpoteeside testimise iva?
  - Meil on olemas ainult valimiandmed, valimis näeme, et Eestis sündinute ja sisse­rännanute usaldus parlamendi vastu on keskmiselt mõnevõrra erinev
  - Kas me saame öelda, et keskmiste erinevus esineb ka populatsioonis?
  - Me ei tea seda, sest valim ei pruugi olla täiesti täpne koopia populatsioonist
  - Eeldusel, et meie valim on tõenäosuslik (st juhuvalim), saame hinnata, kas võib väita keskmiste erinevust ka populatsioonis
- Selleks peame esmalt sõnastama alternatiivhüpoteesi ja nullhüpoteesi

# Hüpoteeside testimise loogika

# Hüpoteeside sõnastamine

- Alternatiivhüpotees ( $H_1$ ) – reeglina sõnastatud nii, et väidab erinevuse esinemist / seose kehtimist
- Nullhüpotees ( $H_0$ ) – vastupidine sellele
- Praegusel juhul:
  - $H_0$ : Eestis sündinute ja sisse­rännanute keskmine usaldus Riigikogu vastu ei erine ( $m_e = m_s$ )
  - $H_1$ : Eestis sündinute ja sisse­rännanute keskmine usaldus Riigikogu vastu erineb ( $m_e \neq m_s$ )
- Oluline: hüpoteesipaari hüpoteesid peavad
  - teineteist välistama
    - ei saa olla  $H_0: m_e \geq m_s$  ja  $H_1: m_e \leq m_s$
  - katma kõikvõimalikud olukorrad
    - ei saa olla  $H_0: m_e = m_s$  ja  $H_1: m_e \leq m_s$
- NB! Hüpoteesides ei pruugi olla öeldud, et need käivad populatsiooni kohta, aga meeles tuleb pidada, et tegu on statistiliste hüpoteesidega



# Esimest ja teist liiki viga

- Hüpoteesipaari testitakse statistilise testiga valimiandmete alusel
- Kaks võimalikku otsust: kas
  - võtame vastu alternatiivhüpoteesi või
  - jääme nullhüpoteesi juurde
- Tuletame meelde usaldusvahemiku teemat:
  - aritmeetilise keskmise usaldusvahemiku arvutasime *valimiandmete alusel* ja püüdsime selle varal midagi öelda *populatsiooni kohta*
- Nii ka hüpoteeside testimisel püüame *valimi alusel hinnata*, kas/kuivõrd tõepärane on väita keskmiste erinevust *populatsioonis*
- Juhuvalim võib juhuslikkuse tõttu mõnevõrra populatsioonist erineda
- Seetõttu eksisteerib võimalus, et teeme valimi alusel populatsiooni kohta eksliku otsuse
- Esimest liiki viga – populatsioonis kehtib  $H_0$  (erinevust ei esine), aga jõuame valimiandmete põhjal järeldusele, et kehtib  $H_1$  (keskmised erinevad)
- Kuigi juhuvalik võimaldab saada valimi, mis on populatsioonile võimalikult sarnane, võib ikkagi mõningatel juhtudel valim juhuse tõttu populatsioonist niivõrd palju erineda, et jõuame eksliku järelduseni
  - Millisel juhul teeksime 1. liiki vea oma hüpoteesipaari puhul?
- Teist liiki viga – populatsioonis kehtib  $H_1$  (erinevus esineb), aga jõuame valimiandmete põhjal järeldusele, et kehtib  $H_0$  (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

- Hüpoteeside testimise loogika: me ei tea, kas populatsioonis kehtib H0 või H1
- Püüame jõuda tõe jälele, keskendudes mitte sellele, kas kehtib H1, vaid sellele, kas kehtib H0
- Kui valimiandmete põhjal leiame, et H0 kehtivus ei ole tõepärane, saame kinnitada H1, sest H0 ja H1 välistavad teineteise
- Seejuures prioriteet on vältida esimest liiki viga (ka teist liiki viga oleks halb, aga esimest liiki viga vältimine olulisem)
  - Reeglina väidetakse nullhüpoteesiga võrdumist või seose puudumist, alternatiivhüpoteesiga erinevust, seose esinemist
  - Seega prioriteet on vältida olukorda, kus väidame millegi (erinevuse, seose) esinemist, kui seda tegelikult ei eksisteeri
  - Ehk kinnitame H1 ainult juhul, kui tõendusmaterjal on kaalukalt nullhüpoteesi paikapidavuse vastu
  - Lähemalt on seda loogikat kirjeldanud nt [Tooding \(2015: 143-144\)](#) ja [Navarro \(2018: ptk 11.1\)](#)

t-test

# Hakkame testima – millega?

- Niisiis, hüpoteesipaari testitakse valimiandmete alusel, otsus tehakse populatsiooni kohta
- Keskmised ja nende erinevuse arvutame valimi alusel, aga kuidas nende alusel hinnata, kas saame keskmiste erinevust väita 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
- $n_1, n_2$  – indiviidide arvud kogumites
- $s$  – kahe kogumi ühise standardhälbe hinnang

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- $s_1^2, s_2^2$  – arvulise tunnuse dispersioonid kogumites

# Valime statistilise testi

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

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

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

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

# Kontrollime testi eelduste täidetust

- Statistilise testi tulemuseks on teststatistik, mis võimaldab hüpoteesipaari kohta otsuse teha
- Kui kasutame  $t$ -testi, nimetatakse teststatistikut  $t$ -statistikuks ( $t$  väärtus valemis)
- Kui meil poleks vaja andmeid kaaluda, saaksime selle arvutada järgmiselt
- Kontrollime  $t$ -testi eelduste täidetust

```
library(summarytools)

ee9 %>%
  group_by(bmcntr2) %>%
  descr(trstpr1)
```

```
## Descriptive Statistics
## trstpr1 by bmcntr2
## Data Frame: ee9
## Label: Trust in country's parliament
## N: 1642
##
##           Jah      Ei
## -----
##           Mean    4.80  5.22
##           Std.Dev  2.41  2.69
##           Min     0.00  0.00
##           Q1      3.00  3.00
##           Median   5.00  5.00
##           Q3      7.00  7.00
##           Max     10.00 10.00
##           MAD      2.97  2.97
##           IQR      4.00  4.00
##           CV       0.50  0.52
##           Skewness -0.26 -0.06
##           SE.Skewness 0.06  0.15
##           Kurtosis -0.48 -0.63
##           N.Valid 1635.00 255.00
##           Pct.Valid 99.57 97.33
```

- Parlamendi usaldushinnangute jaotused on normaaljaotuselähedased, hajuvused on sarnased

# Arvutame teststatistiku

- Andmeid mitte kaaludes saame kasutada R-i baasversioonis olevat funktsiooni `t.test`

```
t.test(trstpr1 ~ brncntr, data = ee9, var.equal = TRUE)
```

```
##  
## Two Sample t-test  
##  
## data: trstpr1 by brncntr  
## t = -2.5294, df = 1888, p-value = 0.01151  
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0  
## 95 percent confidence interval:  
## -0.74062186 -0.09370086  
## sample estimates:  
## mean in group 1 mean in group 2  
##      4.802446      5.219608
```

- Andmeid kaalumata on kahes grupis Riigikogu usaldushinnangute keskmised 4,8 ja 5,22 palli
- $t$ -statistiku väärtus on antud juhul -2,53



# Arvutame teststatistiku

- Kui eeldus arvilise tunnuse hajuvuste sarnasuse kohta gruppides pole täidetud, saab kasutada Welchi *t*-testi, mis ongi funktsiooni `t.test` vaikeseadistus

```
t.test(trstprl ~ brncntr, data = ee9)
```

```
##  
## Welch Two Sample t-test  
##  
## data: trstprl by brncntr  
## t = -2.3314, df = 320.54, p-value = 0.02035  
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0  
## 95 percent confidence interval:  
## -0.76918588 -0.06513684  
## sample estimates:  
## mean in group 1 mean in group 2  
## 4.802446 5.219608
```

# Kontrollime testi eelduste täidetust andmeid kaaludes

- Küsitlusandmete puhul on tavaliselt vaja esinduslikkuse parandamiseks andmeid ka kaaluda
- Seega eeldatavasti saaksime antud juhul täpsema tulemuse, arvutades  $t$ -statistiku kaale arvesse võttes
- Kontrollime  $t$ -testi eelduste täidetust

```
ee9 %>%
  group_by(brncntr2) %>%
  summarise(keskmine = weighted_mean(trstpr1, w = pspwght),
            sd = weighted_sd(trstpr1, w = pspwght),
            asymm = weighted_skewness(trstpr1, w = pspwght),
            kurt = weighted_kurtosis(trstpr1, w = pspwght))
```

```
## # A tibble: 2 x 5
##   brncntr2 keskmine    sd  asymm  kurt
##   <fct>      <dbl> <dbl>   <dbl> <dbl>
## 1 Jah        4.84  2.42 -0.278 -0.498
## 2 Ei         5.24  2.69 -0.0983 -0.625
```

- Andmeid kaaludes on jaotusparameetrid võrreldes eelnevaga väga sarnased,  $t$ -testi eeldused on täidetud ka siin

# Arvutame teststatistiku andmeid kaaludes

- Kasutame andmete kaalumise  $t$ -testi tegemiseks funktsiooni `wtd.t.test` paketest `weights` (teeb Welch  $t$ -testi)

```
brn_y <- ee9 %>% filter(brncntr2 == "Jah")
brn_n <- ee9 %>% filter(brncntr2 == "Ei")

library(weights)
wtd.t.test(brn_y$strstprl, brn_n$strstprl, brn_y$pspwght, brn_n$pspwght)

## Warning in wtd.t.test(brn_y$strstprl, brn_n$strstprl, brn_y$pspwght,
## brn_n$pspwght): Treating data for x and y separately because they are of
## different lengths

## $test
## [1] "Two Sample Weighted T-Test (Welch)"
##
## $coefficients
##      t.value      df    p.value
## -2.2458486 323.0041991  0.0253894
##
## $additional
## Difference    Mean.x    Mean.y Std. Err
## -0.4011964  4.8361036  5.2372999  0.1786391
```

- Antud juhul  $t$ -statistiku väärtus kaale arvestades eelnevast praktiliselt ei erine
- Alati see siiski nii ei pruugi olla, andmete kaalumise võib mõnede teiste tunnuste puhul tulemusi mõjutada küll
- Saime teada  $t$  väärtuse, aga kuidas saame selle alusel midagi öelda populatsiooni kohta või teha järelduse hüpoteesipaari kohta?

Järelduste tegemine

# Teeme järelduse populatsiooni kohta

- Nullilähedane teststatistiku väärtus viitab sellele, et ilmselt on keskmiste erinevus nullilähedane ka populatsioonis
- Alternatiivhüpotees võetakse vastu, kui teststatistik on suurem mingist kriitilisest väärtusest
  - Täpsemalt, kriitilistest väärtustest, sest nii suur negatiivne kui positiivne teststatistiku väärtus viitab keskmiste erinevusele populatsioonis
- Millel need kriitilised väärtused põhinevad?
- Võtame appi  $t$ -jaotuse
- Analooogne usaldusvahemiku mõtteharjutusega
  - Olgu meil populatsioon, kus kehtib  $H_0$  (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 maksimaalselt
    - 10 valimis  $t > 1.64$  või  $t < -1.64$
    - 5 valimis  $t > 1.96$  või  $t < -1.96$
    - 1 valimis  $t > 2.58$  või  $t < -2.58$
- Eelnevad piirid ongi  $t$ -statistiku võimalikud kriitilised väärtused
- Kasutatakse ka kriitilise piirkonna mõistet
- $t$ -statistiku kriitiliseks piirkonnaks nimetatakse
  - $t$  positiivsest kriitilisest väärtusest suuremate väärtuste piirkonda ja
  - $t$  negatiivsest kriitilisest väärtusest väiksemate väärtuste piirkonda
- Kui  $t$ -testiga saadav  $t$ -statistiku väärtus on kriitilises piirkonnas, võetakse vastu alternatiivhüpotees

# Teeme järelduse populatsiooni kohta

- Reaalne olukord: me ei tea, kas populatsioonis kehtib  $H_0$  või  $H_1$  ja meil on ainult üks valim
- Aga teame, et *kui* populatsioonis kehtib  $H_0$ , siis sealt valimit võttes on tõenäosus, et saaksime valimi, kus  $t$  absoluutväärtus on üle 1,96, väga väike (5%)
- Seega, kui oleme saanud nii suure  $t$ -statistiku väärtuse, siis on tõepärasem eeldada, et see ei juhtunud mitte seetõttu, et meil on väga erandlik valim, vaid seetõttu, et populatsioonis keskmised ei võrdu
- Seega kinnitame alternatiivhüpoteesi
- Me võime muidugi ka eksida – võib-olla meie valim *on* juhuse tahtel nende 5% valimite seas – sellisel juhul teeksime I liiki vea
- Mis on piisavalt range protsent, et mitte teha I liiki viga?
- See on subjektiivne, ka rangemate kriitiliste väärtuste puhul on võimalik I liiki viga teha, tõsi küll, oht selleks on väiksem
- Traditsiooniliselt on selleks kujunenud 5% ehk 0,05, kuigi võidakse kasutada ka 0,1, 0,01 ja 0,001
- NB! Seda suhtarvu nimetatakse *olulisuse nivooks* (tähistatakse  $\alpha$ )
- Meie Riigikogu usaldamise näites on  $t$ -statistiku väärtus -2,25, seega
  - võttes otsuse tegemise aluseks olulisuse nivoo 0,05,
  - saame kinnitada alternatiivhüpoteesi,
  - sest  $t$  väärtus asub valitud olulisuse nivool kriitilises piirkonnas ( $t < -1.96$  |  $t > 1.96$ )
  - Ääremärkus:  $t$  väärtus on antud juhul negatiivne lihtsalt seetõttu, et esimese grupi (Eestis sündinud) keskmine on teise grupi keskmisest madalam (vt  $t$ -statistiku arvutamise valemit)
- Kui võtaksime otsuse tegemise aluseks rangema olulisuse nivoo 0,01, siis antud juhul alternatiivhüpoteesi kinnitada ei saaks (ei kehti tingimus  $t < -2.58$  |  $t > 2.58$ ) ja tuleks jääda nullhüpoteesi juurde
- Nagu näha, on otsus alati tinglik – milline olukord eksisteerib populatsioonis, me lõpuni teada ei saagi; otsus sõltub olulisuse nivoo valikust ja võimalus eksimiseks jääb
- Milline olulisuse nivoo valida, sõltub sellest, kuivõrd oluline on vältida I liiki viga
  - Ääremärkus: jõudsimme antud näites olulisuse nivoo mõisteni pärast  $t$ -statistiku arvutamist, andmeid analüüsides oleks aus olulisuse nivoo valida (ehk kehtestada) enne arvutamist

# Järelduse tegemise võimalused, olulisuse tõenäosus

- Niisiis, hüpoteesipaari kohta otsuse tegemise loogika:
  - Kui valimi alusel saadud  $t$ -statistik asub valitud olulisuse nivoole vastavas kriitilises piirkonnas, saame  $H_0$  kummutada ja kinnitada  $H_1$
- Kas peame erinevad  $t$  kriitilised väärtused meeles pidama?
- Ei pea, ei pea ka kuskilt üles otsima; on ka lihtsam võimalus
- Kui andmeanalüüsiprogramm arvutab  $t$  väärtuse, annab ka *olulisuse tõenäosuse* ( $p$ -väärtus) vastava vabadusastmete arvu juures
- Mis see olulisuse tõenäosus õigupoolest on?
- “Tõenäosus teha esimest liiki viga” – sage tõlgendus, kuid ebatäpne, pole õige
- Korrektne: tõenäosus saada nullhüpoteesi kehtimise korral vastav või sellest veel suurem teststatistiku väärtus
- Täpsemalt: tõenäosus saada populatsioonist, kus kehtib  $H_0$ , valimiandmed, mille põhjal saaksime vähemalt nii suure teststatistiku väärtuse
- Meie näites kaalutud andmetega  $t = -2.25$ ,  $df = 323$ , siis  $p = 0.025$
- Kui populatsioonis kehtib  $H_0$  ehk Eesti sündinute ja sisseännanute keskmine usaldus Riigikogu vastu ei erine,
  - siis tõenäosus saada populatsioonist juhuvalikuga antud suurusega valim,
  - kus nende kahe kogumi keskmiste alusel arvutatav  $t$ -statistik on  $-2,25$  või sellest suurem,
  - on  $0,025$  ehk  $2,5\%$  ehk  $2,5$  sajast
- Alternatiivne võimalus hüpoteesipaari kohta otsuse tegemiseks:
  - kui olulisuse tõenäosus  $<$  valitud olulisuse nivoo, tuleb kummutada / tagasi lükata nullhüpotees ja vastu võtta / kinnitada alternatiivhüpotees
  - Teisisõnu saame öelda, et keskmiste erinevus on *statistiliselt oluline* valitud olulisuse nivool
- Kui olulisuse tõenäosus  $\geq$  olulisuse nivoo, ei saa kinnitada alternatiivhüpoteesi, vaid tuleb jääda nullhüpoteesi juurde

# Järelduse tegemise võimalused

- NB! Nullhüpoteesi ei saa kinnitada / tõestada – kui ei õnnestu kinnitada alternatiivhüpoteesi, tuleb jääda nullhüpoteesi juurde
  - Kui ei ole alust  $H_1$  kinnitamiseks, ei saa öelda, et nullhüpotees tingimata kehtib, meil lihtsalt ei õnnestunud leida tõendust vastupidisele
  - Meil ei ole piisavalt tõendeid, et nullhüpoteesi n-ö süüdi mõista, aga see ei tähenda, et saame kinnitada nullhüpoteesi süütust
- Olulisuse tõenäosuse põhjal on otsust hüpoteesipaari kohta teha lihtsam kui teststatistiku väärtuse põhjal
- Samas on oluline aru saada mõlemast loogikast, sest olulisuse tõenäosuse väärtus põhineb teststatistiku väärtusel



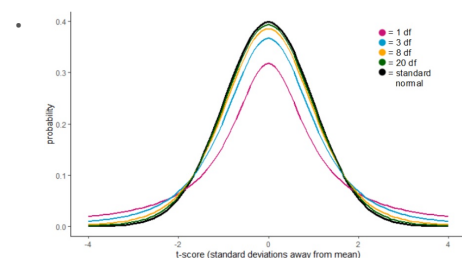
Mõned täpsustused

# Olulisuse nivoo ja usaldusnivoo

- Mille poolest sarnanevad / erinevad?
- Sama mündi kaks poolt
- Hüpoteeside testimisel võtame aluseks olulisuse nivoo (0,05 või 0,1 või 0,01 (või 0,001))
- Usaldusvahemiku arvutamisel usaldusnivoo (95% või 90% või 99%)
- Nii olulisuse nivoo 0,05 (ehk 5%) kui usaldusnivoo 95% puhul võetakse arvutuste aluseks väärtused -1,96 ja 1,96
- Miks siis kaks erinevat mõistet?
- Usaldusvahemiku puhul huvitab meid see, mis jääb nende väärtuste alusel arvutatavate piiride (usalduspiiride) *vahele*
  - Vahemik, mis *katab* tegeliku väärtuse populatsioonis
- Hüpoteeside testimisel huvitab meid see, kas teststatistik jääb nende väärtuste alusel arvutatavatest piiridest (kriitilistest väärtustest) *väljapoole*
  - Kui jääb, saame kinnitada alternatiivhüpoteesi
- Oluline tehniline erinevus:
  - usalduspiiride väärtused sõltuvad valitud usaldusnivoost
    - arvutamisel on vaja usaldusnivoo väärtus arvesse võtta
    - programmides vaikimisi usaldusnivoo väärtus 95%, arvutuse tulemus sõltub sellest
  - olulisuse tõenäosus ja muud statistilise testi tulemused ei sõltu valitud olulisuse nivoo
    - arvutamisel ei ole vaja (ei saagi) olulisuse nivood kuskil seadistada
    - järelduse tegemisel on siiski olulisuse nivood vaja

# Kriitilised väärtused sõltuvad vabadusastmete arvust

- Eelnevalt  $t$ -statistiku puhul välja toodud kriitilised piirid on 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 (*degrees of freedom*) – 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  $df = n - 2 = 80$ , siis  $t_{1-\frac{\alpha}{2}} = 1.99$  ja  $t_{\frac{\alpha}{2}} = -1.99$
  - kui  $df = n - 2 = 30$ , siis  $t_{1-\frac{\alpha}{2}} = 2.04$  ja  $t_{\frac{\alpha}{2}} = -2.04$
  - kui  $df = n - 2 = 8$ , siis  $t_{1-\frac{\alpha}{2}} = 2.31$  ja  $t_{\frac{\alpha}{2}} = -2.31$
- Kui teeme otsuse hüpoteesipaari kohta mitte  $t$ -statistiku väärtuse, vaid olulisuse tõenäosuse alusel, siis selle arvutusel on vabadusastmete arv juba arvesse võetud

# Tulemuste esitamine

- Võib tunduda, et järelduste tegemiseks polegi vaja muud, kui vaadata olulisuse tõenäosust ja raporteerida see
- Päril nii see pole
- Olulisuse tõenäosus (nagu ka usaldusvahemik) käib alati konkreetse näitaja kohta
- Raporteerida tuleks ka see näitaja (teststatistik) ja selle tõlgendamiseks vajalikud näitajad
- Seega,  $t$ -testi tulemuste põhjal saame olulisuse nivool 0,05 väita, et
  - aastal 2018 oli Eestis sündinud inimeste ja sisseärrannute usaldus Riigikogu vastu keskmiselt erinev ( $t = -2.25, df = 323, p = 0.025$ )
  - Või: ...erinev ( $t$ -statistiku väärtus 2,33, vabadusastmete arv 1384, olulisuse tõenäosus 0,02)

# Kui suur on keskmiste erinevus?

- Meie näites valimikeskmised kahes grupis olid 4,84 ja 5,24, keskmiste erinevus 0,4 palli
- Olulisuse nivool 0,05 saime öelda, et keskmiste erinevus esineb ka populatsioonis
- Aga kui suur see keskmiste erinevus populatsioonis on?
- Kas H1 kinnitamisel saame öelda, et erinevus 0,4 palli esineb ka populatsioonis?
- Ei saa, selle hindamiseks peaks arvutama keskmiste erinevuse usaldusvahemiku
- Kui otsuse hüpoteesipaari kohta tegime olulisuse nivool 0,05, peaks keskmiste erinevuse usaldusvahemiku arvutama usaldusnivool 95%
- Funktsioon `t.test` (kaalumata andmetega) annab selle väljundis, funktsiooni `wtd.t.test` saame selle ise standardvea põhjal arvutada

```
wtd.t.test(brn_y$trstprl, brn_n$trstprl, brn_y$pspwght, brn_n$pspwght)      -0.4 - 1.96 * 0.179 # Alumine usalduspiir
```

```
## $test  
## [1] "Two Sample Weighted T-Test (Welch)"  
## [1] -0.75084
```

```
## $coefficients  
##      t.value      df      p.value  
## -2.2458486 323.0041991 0.0253894  
## -0.4 + 1.96 * 0.179 # Ülemine usalduspiir
```

```
## $additional  
## Difference      Mean.x      Mean.y      Std. Err  
## -0.4011964 4.8361036 5.2372999 0.1786391  
## [1] -0.04916
```

# Kui suur on keskmiste erinevus?

- Kui soovime arvutuse teha otse väljundinäitajate põhjal

```
ttest <- wtd.t.test(brn_y$trstprl, brn_n$trstprl, brn_y$pspwght, brn_n$pspwght)
str(ttest)
```

```
## List of 3
## $ test      : chr "Two Sample Weighted T-Test (Welch)"
## $ coefficients: Named num [1:3] -2.2458 323.0042 0.0254
## $ .. attr(*, "names")= chr [1:3] "t.value" "df" "p.value"
## $ additional : Named num [1:4] -0.401 4.836 5.237 0.179
## $ .. attr(*, "names")= chr [1:4] "Difference" "Mean.x" "Mean.y" "Std. Err"
```

```
ttest$additional[1] - 1.96 * ttest$additional[4] # Alumine usalduspiir
```

```
## Difference
## -0.751329
```

```
ttest$additional[1] + 1.96 * ttest$additional[4] # Ülemine usalduspiir
```

```
## Difference
## -0.05106374
```

- Võttes aluseks usaldusnivoo 95%, saab väita, et keskmiste erinevus võib populatsioonis olla nii väike kui 0,05 palli, aga võib ka olla nii suur kui 0,75 palli
- *t*-test võimaldab lihtsalt öelda, kas keskmised erinevad, keskmiste erinevuse usaldusvahemik võimaldab erinevust sisulisemalt hinnata
- Ka kolmas võimalus keskmiste erinevuse kohta populatsioonis otsuse tegemiseks: kas keskmiste erinevuse usaldusvahemik katab nullpunkti või mitte
  - Antud viisil tehtav otsus ei saa eelnevatel viisidel tehtavate otsustega vastuollu minna, kui usaldusvahemik arvutatakse eelnevate otsuste olulisuse nivoole vastaval usaldusnivool (olulisuse nivoole 0,05 vastab usaldusnivoo 95%)

# Kahepoolne ja ühepoolne hüpoteesipaar

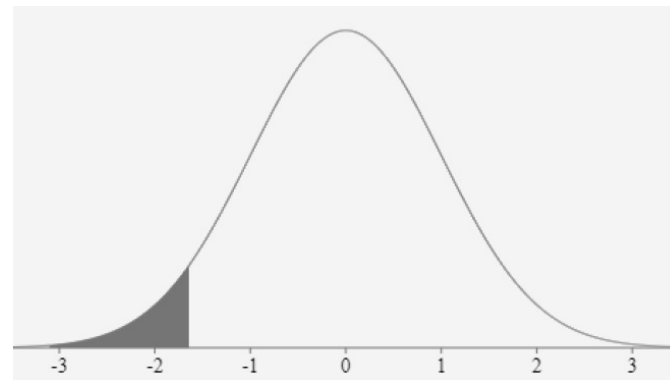
# Kahepoolne ja ühepoolne hüpoteesipaar

- Praegusel juhul:
  - $H_0$ : Eestis sündinute ja sisseännanute keskmine usaldus Riigikogu vastu ei erine ( $m_e = m_s$ )
  - $H_1$ : Eestis sündinute ja sisseännanute keskmine usaldus Riigikogu vastu erineb ( $m_e \neq m_s$ )
- Tegu on kahepoolse hüpoteesipaariga –  $H_1$  seondub arvtelje kahe lõiguga
- Hüpoteesipaar (täpsemalt, alternatiivhüpotees) võib olla ka ühepoolne
- Sel juhul seondub  $H_1$  arvtelje ühe lõiguga, hüpoteesipaar võib olla
  - $H_0: m_e \geq m_s$  ja  $H_1: m_e < m_s$  või
  - $H_0: m_e \leq m_s$  ja  $H_1: m_e > m_s$
- Võimalik ühepoolne hüpoteesipaar lahti sõnastatult meie näite puhul
  - $H_0$ : Eestis sündinute seas on keskmine usaldus Riigikogu vastu kõrgem või võrdne usaldusega sisseännanute hulgas ( $m_e \geq m_s$ )
  - $H_1$ : Eestis sündinute seas on keskmine usaldus Riigikogu vastu madalam kui sisseännanute hulgas ( $m_e < m_s$ )



# Kahepoolne ja ühepoolne hüpoteesipaar

- Eelnev arutelu ja näide põhinesid kahepoolse hüpoteesipaari testimisel
  - $H_0$ : Eestis sündinute ja sisseännanute keskmine usaldus Riigikogu vastu ei erine ( $m_e = m_s$ )
  - $H_1$ : Eestis sündinute ja sisseännanute keskmine usaldus Riigikogu vastu erineb ( $m_e \neq m_s$ )
- Kui hüpoteesipaar on ühepoolne, tuleb
  - teststatistiku kriitilised väärtused arvutada natuke teisel põhimõttel
- Kahepoolne hüpoteesipaar: kriitilised piirkonnad asuvad  $t$ -jaotuse mõlema „saba“ all
- Olulisuse nivool 0,05 on kriitilised piirid -1,96 ja 1,96, sest
  - nullhüpoteesi kehtides jäävad nendest kaugemale 5% võimalike valimite  $t$ -statistikud
  - hallid alad joonisel moodustavad kokku 5% joonealusest pindalast
- Ühepoolne hüpoteesipaar: kriitiline piirkond asub  $t$ -jaotuse selle „saba“ all, mida väidab  $H_1$
- Olulisuse nivool 0,05 peab 5% joonealusest alast mahtuma vaid ühe saba alla =>
- Kriitiline väärtus (neid on nüüd ainult üks!) on madalam kui kahepoolse hüpoteesipaari korral
- Nt kui hüpoteesipaar on  $H_0: m_e \geq m_s$  ja  $H_1: m_e < m_s$ , siis kriitiline väärtus on -1,64:



# Ühepoolse hüpoteesipaari testimise näide

- Testime ühepoolset hüpoteesipaari
  - $H_0$ : Eestis sündinute seas on keskmine usaldus Riigikogu vastu kõrgem või võrdne usaldusega sisserännanute hulgas ( $m_e \geq m_s$ )
  - $H_1$ : Eestis sündinute seas on keskmine usaldus Riigikogu vastu madalam kui sisserännanute hulgas ( $m_e < m_s$ )

```
wtd.t.test(brn_y$strstprl, brn_n$strstprl, brn_y$pspwght, brn_n$pspwght, alternative = "less")
```

```
## Warning in wtd.t.test(brn_y$strstprl, brn_n$strstprl, brn_y$pspwght,  
## brn_n$pspwght, : Treating data for x and y separately because they are of  
## different lengths
```

```
## $test  
## [1] "Two Sample Weighted T-Test (Welch)"  
##  
## $coefficients  
##      t.value      df      p.value  
## -2.2458486 323.0041991 0.0126947  
##  
## $additional  
## Difference      Mean.x      Mean.y      Std. Err  
## -0.4011964 4.8361036 5.2372999 0.1786391
```

- Kõik testitulemused on samad, mis kahepoolse hüpoteesipaari puhul, välja arvatud olulisuse tõenäosus, mis on poole madalam (0,013 vs 0,025)

- Nagu eelnevalt märgitud, ka  $t$ -statistiku kriitiline väärtus on vaadeldava saba all ühepoolse hüpoteesipaari puhul madalam (olulisuse nivool 0,05 antud juhul -1,64 vs -1,96)
- Paradoksaalsel kombel on spetsiifilisema tingimuse täidetust (esimese grupi keskmine on madalam) n-ö kergem tõestada kui leebema tingimuse puhul (esimese grupi keskmine on madalam või kõrgem)
- Võib kallutada uurijat valima testimise viisi, mis sobib eeldatavate tulemustega paremini:
  - kui kahepoolse hüpoteesipaari puhul on olulisuse tõenäosus ainult natuke üle olulisuse nivoo, võimaldab ühepoolse hüpoteesipaari testimine saada statistiliselt olulise tulemuse
  - seda nimetatakse *p-hacking*
  - selle vältimiseks tuleks olulisuse nivoo ja hüpoteesipaar tuleks sõnastada enne testimist

# Valik ühe- ja kahepoolse hüpoteesipaari vahel

- Hüpoteesipaari võib sõnastada ühepoolsena, aga sel juhul peame arvestama, et teistpidine erinevus võib jääda meil tuvastamata
- Nt antud juhul saaksime alternatiivhüpoteesiga Eestis sündinute kõrgemat usalduse taset väites väga kõrge olulisuse tõenäosuse ( $p$ -väärtus on nüüd 1 - eelnevalt saadud olulisuse tõenäosuse väärtus):

```
wtd.t.test(brn_y$trstprl, brn_n$trstprl, brn_y$pspwght, brn_n$pspwght, alternative = "greater")
```

```
## Warning in wtd.t.test(brn_y$trstprl, brn_n$trstprl, brn_y$pspwght,  
## brn_n$pspwght, : Treating data for x and y separately because they are of  
## different lengths
```

```
## $test  
## [1] "Two Sample Weighted T-Test (Welch)"  
##  
## $coefficients  
##      t.value      df      p.value  
## -2.2458486 323.0041991  0.9873053  
##  
## $additional  
## Difference      Mean.x      Mean.y      Std. Err  
## -0.4011964  4.8361036  5.2372999  0.1786391
```

- Antud näide sobib selle ohu illustreerimiseks hästi

# Valik ühe- ja kahepoolse hüpoteesipaari vahel

- Pigem võiks intuiitiivselt eeldada, et Eestis sündinud inimesed usaldavad Riigikogu rohkem
- Vastavat ühepoolset alternatiivhüpoteesi testides ( $t$ -test eelneval slaidil) saaksime teada, et see pole üldse tõepärane
- Vastupidise alternatiivhüpoteesi kehtivuse kohta me midagi teada ei saaks, selle kehtivus jääks tuvastamata
  - Mõlemat ühepoolset hüpoteesipaari (kumbki olulisuse nivool 0,05) testida ei oleks korrektne, sest sisuliselt teeksime siis kogujärelduse olulisuse nivool 0,1.
- Seega, praktikas võib ja saab ühepoolse hüpoteesipaari testimist kasutada, aga
  - piiripealsete tulemuste puhul, kus valikust ühe- ja kahepoolse hüpoteesipaari vahel sõltuvad sisulised järeldused,
  - tuleks seda valikut kindlasti põhjendada
- Muidu riskime oma analüüsi tõsiseltvõetavusega
- Praktikas kasutatakse hüpoteeside testimisel reeglina kahepoolset hüpoteesipaari (tihti ka juhul, kui sisulises mõttes ollakse huvitatud ühepoolsetest järeldustest)

t-testi variatsioonid

# Ühe kogumi $t$ -test

- Põhimõtte 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 usaldus parlamendi vastu on alla skaala keskpunkti
- Saab alternatiivhüpoteesi ühepoolsena sõnastada, kui oluline on just teada saada, kas usaldus on *alla* skaala keskpunkti
- Kas või kui palju on keskmine usaldus üle skaala keskpunkti, see meid ei huvita, sest nt tarvis on midagi Riigikogu usaldamise suhtes ette võtta ainult siis, kui sellega on nii kehvad lood, et jääb alla skaala keskpunkti
  - $H_0: m \geq 5$
  - $H_1: m < 5$

# Ühe kogumi *t*-test

- Andmeid kaalumata
  - kontrollime testi eeldusi
- Asümmeetria- ja järskuskordaja on nullile küllaltki lähedal, seega normaaljaotuse läheduse eeldus on täidetud. Kuna antud *t*-testi puhul võrreldakse tunnuse keskmist etteantud väärtusega, ei ole hajuvuste sarnasuse eelduseks alust.

```
descr(ee9$trstprl)
```

```
## Descriptive Statistics
## ee9$trstprl
## Label: Trust in country's parliament
## N: 1904
##
##          trstprl
## -----
##          Mean    4.86
##          Std.Dev  2.45
##          Min     0.00
##          Q1      3.00
##          Median   5.00
##          Q3      7.00
##          Max     10.00
##          MAD     2.97
##          IQR     4.00
##          CV      0.50
##          Skewness -0.21
##          SE.Skewness 0.06
##          Kurtosis -0.47
##          N.Valid  1890.00
##          Pct.Valid 99.26
```

```
t.test(trstprl ~ 1, mu = 5, data = ee9, alternative = "less")
```

```
##
## One Sample t-test
##
## data: trstprl
## t = -2.5036, df = 1889, p-value = 0.006188
## alternative hypothesis: true mean is less than 5
## 95 percent confidence interval:
##      -Inf 4.951589
## sample estimates:
## mean of x
##  4.85873
```

- *t*-statistik on -2,5, vabadusastmete arvu 1889 juures on olulisuse tõenäosus 0,006, alternatiivhüpoteesi saab olulisuse nivool 0,05 kinnitada, saaksi segi olulisuse nivool 0,01.

# Ühe kogumi $t$ -test

- Andmeid kaaludes

```
wtd.t.test(ee9$trstprl, 5, weight = ee9$pspwght, alternative = "less")
```

```
weighted_skewness(ee9$trstprl, w = ee9$pspwght)
```

```
## [1] -0.2363142
```

```
weighted_kurtosis(ee9$trstprl, w = ee9$pspwght)
```

```
## [1] -0.4877105
```

- Kaalutud kujuparameetrid ei erine kuigivõrd kaalumata kordajatest

```
## $test
## [1] "One Sample Weighted T-Test"
##
## $coefficients
##      t.value      df      p.value
## -2.02513458 1891.06800254  0.02149689
##
## $additional
## Difference      Mean Alternative      Std. Err
## -0.11452442  4.88547558  5.00000000  0.05655151
```

- $t$ -statistiku väärtus on -2,03, vabadusastmete arvu 1891 juures on olulisuse tõenäosus 0,021. Võttes aluseks olulisuse nivoo 0,05, saab öelda, et usalduse keskmine on alla skaala keskpunkti



# Statistilise ja sisulise olulisuse erinevus

- Statistiliselt on eelnev järeldus õige, küll aga tuleb siin hästi esile, et on vaja eristada statistilist olulisust ja sisulist olulisust
- Tulemus on statistiliselt oluline, aga keskmise ja skaala keskpunkti erinevus on usaldusnivool 95% vahemikus  
 $-0.1145 - 1.96 \cdot 0.0566 = -0.225$  ja  $-0.1145 + 1.96 \cdot 0.0566 = -0.004$
- See tähendab, et erinevus on maksimaalselt alla veerand palli ja ilmselt veel väiksemgi - liiga väike, et mingeid märkimisväärseid sisulisi järeldusi teha
- Statistiliselt oluline tulemus ei tähenda veel sisuliselt olulist tulemust
  - Aga statistiliselt olulise tulemuse puudumisel pole mõtet ka sisulise tõlgendamisega vaeva näha
  - Sest puudub kindlus, et erinevus/seos populatsioonis üldse esineks
- Statistiline olulisus konkreetne (kuigi tinglik!), sisulise olulisuse aluseks (vähemalt mingis ulatuses) subjektiivne hinnang

# 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 õnnelikkuse ja eluga rahulolu hinnangud
    - tegu on sarnaste kontseptsioonidega, aga sisuliselt on nad siiski erinevad
    - kas erinevus esineb ka hinnangutes?
- Kontrollime testi eelduste täidetust, näite mõttes esmalt kaalumata andmetega

# Paariskogumite *t*-test

```
descr(ee9$happy)
```

```
descr(ee9$stflife)
```

```
## Descriptive Statistics
```

```
## ee9$happy
```

```
## Label: How happy are you
```

```
## N: 1904
```

```
##
```

```
##
```

```
## -----
```

```
## Mean 7.33
```

```
## Std.Dev 1.81
```

```
## Min 0.00
```

```
## Q1 6.00
```

```
## Median 8.00
```

```
## Q3 9.00
```

```
## Max 10.00
```

```
## MAD 1.48
```

```
## IQR 3.00
```

```
## CV 0.25
```

```
## Skewness -0.80
```

```
## SE.Skewness 0.06
```

```
## Kurtosis 0.80
```

```
## N.Valid 1904.00
```

```
## Pct.Valid 100.00
```

```
## Descriptive Statistics
```

```
## ee9$stflife
```

```
## Label: How satisfied with life as a whole
```

```
## N: 1904
```

```
##
```

```
##
```

```
## -----
```

```
## Mean 7.02
```

```
## Std.Dev 1.97
```

```
## Min 0.00
```

```
## Q1 6.00
```

```
## Median 7.00
```

```
## Q3 8.00
```

```
## Max 10.00
```

```
## MAD 1.48
```

```
## IQR 2.00
```

```
## CV 0.28
```

```
## Skewness -0.87
```

```
## SE.Skewness 0.06
```

```
## Kurtosis 1.03
```

```
## N.Valid 1903.00
```

```
## Pct.Valid 99.95
```

- Asümmeetria- ja järskuskordajad näitavad märkimisväärselt erinevust normaaljaotusest, kuid erinevus ei ole nii suur, et *t*-testi ei võiks teha. Piiripealsete tulemuste tõlgendamisel peaks siiski ettevaatlik olema.

# Paariskogumite *t*-test

```
t.test(Pair(happy, stflife) ~ 1, data = ee9)
```

```
##  
## Paired t-test  
##  
## data: Pair(happy, stflife)  
## t = 8.3597, df = 1902, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##  0.2360947 0.3808260  
## sample estimates:  
## mean of the differences  
##           0.3084603
```

- Tõlgendame testitulemusi kaalutud andmete näite juures, sest tulemused sisuliselt ei erine.

# Paariskogumite $t$ -test

- Kaalutud andmetega

```
weighted_skewness(ee9$happy, w = ee9$pspwght)
```

```
## [1] -0.8120742
```

```
weighted_kurtosis(ee9$happy, w = ee9$pspwght)
```

```
## [1] 0.8097254
```

```
weighted_skewness(ee9$stflife, w = ee9$pspwght)
```

```
## [1] -0.9021268
```

```
weighted_kurtosis(ee9$stflife, w = ee9$pspwght)
```

```
## [1] 1.086036
```

- Jaotusparameetrid on enam-vähem samad, mis kaalumata andmetega
- Paariskogumite  $t$ -testi tegemiseks arvutame esmalt iga indiviidi kohta tema õnnelikkuse ja eluga rahulolu hinnangu vahe. Seejärel kontrollime ühe kogumi  $t$ -testiga vahede keskmise erinevust nullist (mis ongi paariskogumite  $t$ -testi iva).

# Paariskogumite *t*-test

```
ee9$happy_stflife <- ee9$happy - ee9$stflife  
wtd.t.test(ee9$happy_stflife, 0, weight = ee9$pspwght)
```

```
## $test  
## [1] "One Sample Weighted T-Test"  
##  
## $coefficients  
##      t.value      df      p.value  
## 8.157890e+00 1.902252e+03 6.661338e-16  
##  
## $additional  
##      Difference      Mean Alternative      Std. Err  
## 0.29508951 0.29508951 0.00000000 0.03617228
```

- Tulemus on selgelt statistiliselt oluline (ka kõige rangemal olulisuse nivool), sest *t*-statistiku väärtus on suur ja olulisuse tõenäosus on  $6.66 \times 10^{-16}$  ehk  $p = 0.0000000000000000666$  (16 nulli enne esimest 6)
- Arvu väiksuse tõttu kasutataksegi siin arvu standardkuju (kümne astmete näol), tekstis võiks tulemuse ka esitada rangeima olulisuse nivoo abil:  $p < 0.001$
- Kuigi olulisuse tõenäosus on väga madal, ei ole sisulises plaanis tegu suure erinevusega, populatsioonis jääb keskmiste erinevus 0,22 ja 0,37 vahele (usaldusnivool 95%)