


# Hypoteesien muodostaminen ja testaus

Tekijä: VS - älä  
levitä

- 
- Mikä on kategoristen ja jatkuvien muuttujien välinen ero?
  - Mikä ero on t-testillä ja khii toiseen (khiin neliö) testillä?

- 
- <https://youtu.be/I10q6fjPxJ0>

# T-testi

- T-testiä voidaan käyttää vain kahden ryhmän keskiarvojen vertailussa (numeeriset jatkuvat muuttujat) --> **erojen tutkiminen**
- Pitäisikö ryhmien keskiarvojen välillä olla ero?
- Esimerkki: Mitataan kahden eri merkkisen energiapatukan proteiinigrammat. Kaksi ryhmäämme ovat kaksi tuotemerkkiä. Muuttujamme on proteiinigrammat jokaista energiapatukkaa kohden. Ajatus on, että näiden kahden tuotemerkin taustalla olevien populaatioiden keskimääräiset proteiinigrammat voivat olla erilaisia. Haluamme tietää, onko meillä todisteita siitä, että näiden kahden energiapatukkamerkin proteiinin keskimääräiset grammat ovat erilaisia.

# Khiin neliö -testi

- Khii toiseen testiä käytetään kahden kategorisen muuttujan ryhmien välisten erojen välisen **suhteen** tutkimiseen
- esimerkki: Haluamme selvittää, onko paidan värin ja kuolemien välillä tilastollisesti merkitsevää yhteyttä. Meidän on käytettävä tätä testiä, koska nämä muuttujat ovat molemmat kategorisia muuttujia. Paidan väri voi olla vain sininen, kulta tai punainen. Kuolemantapaukset voivat olla vain kuolleita tai eläviä.

# P-arvo

- Sopimuksenvarainen raja, joka kertoo, kuinka todennäköistä on, että havaittu tulos aineistossa esiintyy vain sattumalta
- Jos suurempi kuin 0.05, on yli 5% todennäköisyys, että havaintosi on vain sattumaa, jos pienempi kuin 0.05, sattumanvaraisen tuloksen löytämisen todennäköisyys on siis pienempi kuin 5% prosenttia
- Voi itse määrittää, kuinka suuren sattumanvaraisuuden haluaa hyväksyä, yleensä 0.05 (5%), joskus myös 0.01 (1%)
- Täysin sopimuksenvarainen tilastotieteen historiasta kumpuava käytäntö - se nyt vain täytyy hyväksyä (yleensä)

# Tehtävä

- Löydätkö visualisoinneista joitain havaintoja, joiden paikkansapitävyyttä (tilastollista merkittävyyttä) voisi testata?
- Tehkää hypoteesit ja nollahypoteesit (nollahypoteesi hylkää varsinaisen hypoteesin väitteen) -->jos p-arvo suurempi kuin sovittu arvo (esim.  $>0.05$ ), nollahypoteesi hyväksytään ja varsinainen hypoteesi hylätään, koska havaittu tulos voi olla liian todennäköisen sattuman vuoksi löydetty

# Apumateriaaleja

- Kvantimot (esim. Summamuuuttuja)
- Youtube, todella paljon selostavia hyviä videoita eri testien välisistä eroista ja siitä, miten analyysijä voi käytännössä toteuttaa Excelissä tai R:llä