

Biomeetria

Märt Möls
martm@ut.ee

Korralduslik informatsioon

Kursuse kodulehekülg:

<http://www.ms.ut.ee/mart/biomeetria2015/>

projekt 30% hindest

eksam 70% hindest

Loengukursusel osalejad peavad iseseisvalt teostama ühe andmestiku analüüsi. Soovitavalt peaks andmestik olema seotud tudengi enda uurimistöö või huvidega. Need tudengid, kes millestki ei huvitu ega midagi ei uuri, võiksid abipalvega pöörduda sõprade poole. Kui teil pole aga ka sõpru, siis küsige andmestikku õppejõult.

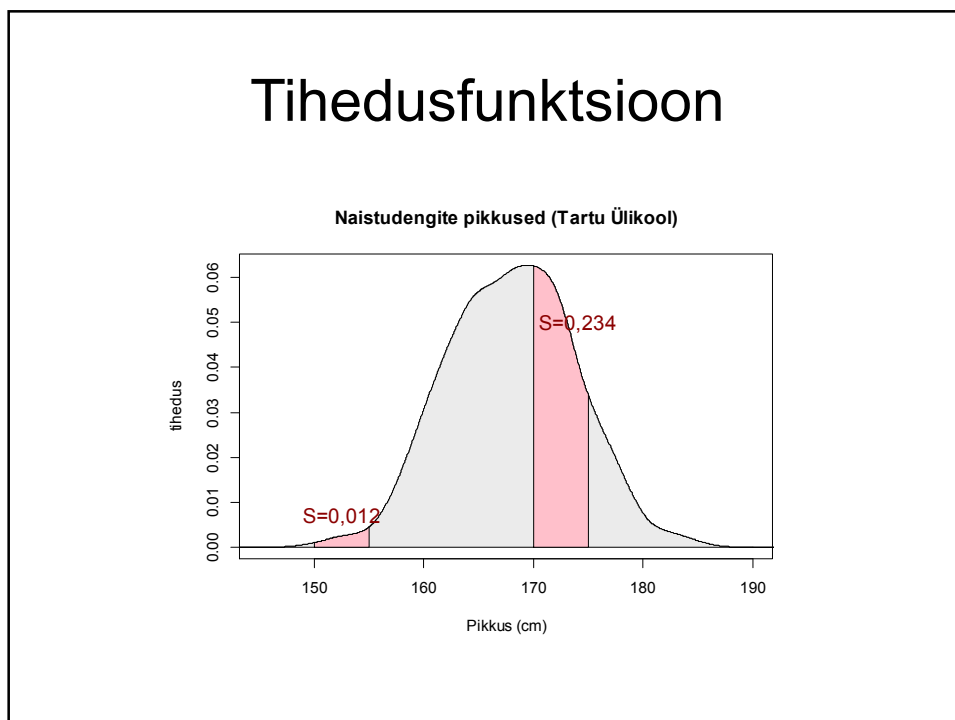
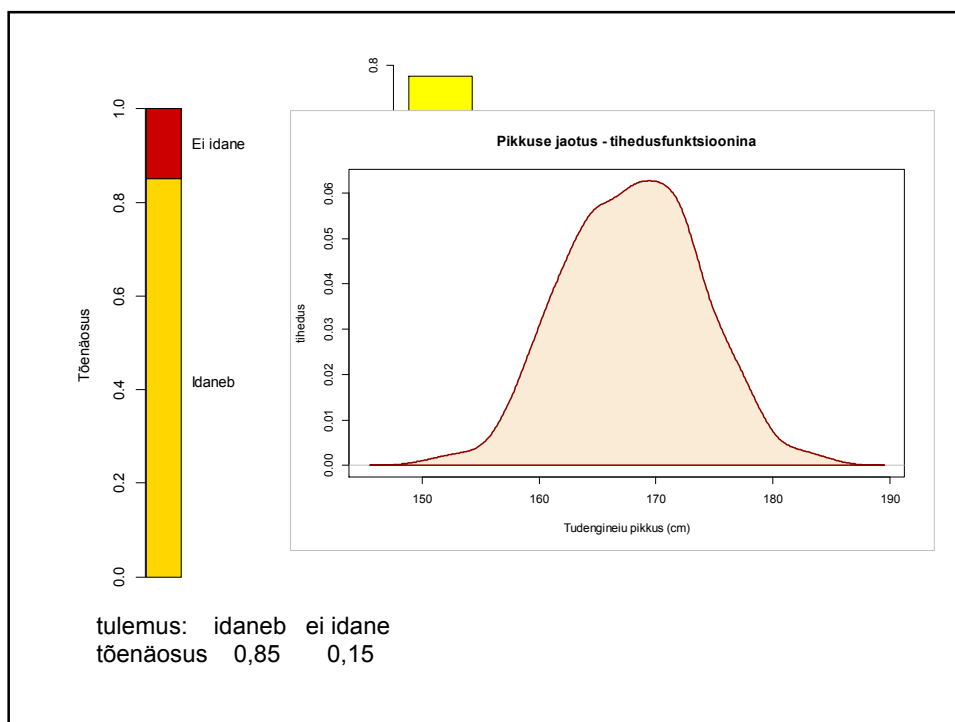
Mis on biomeetria?

Katkend Rahvusvahelise Biomeetriasetsi koduleheküljelt (www.tibs.org):

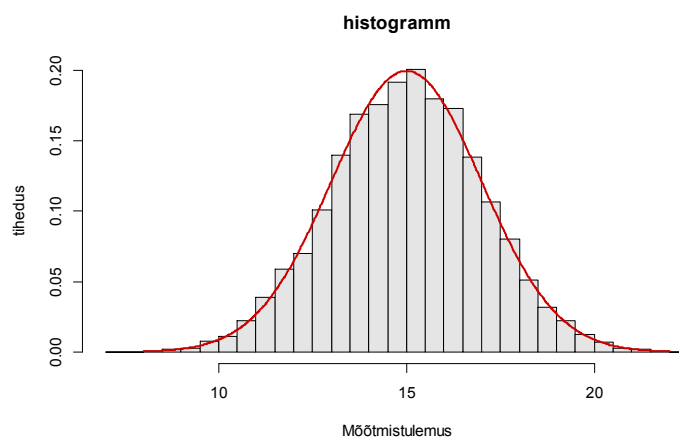
The terms "Biometrics" and "Biometry" have been used since early in the 20th century to refer to the field of development of statistical and mathematical methods applicable to data analysis problems in the biological sciences. Statistical methods for the analysis of data from agricultural field experiments to compare the yields of different varieties of wheat, for the analysis of data from human clinical trials evaluating the relative effectiveness of competing therapies for disease, or for the analysis of data from environmental studies on the effects of air or water pollution on the appearance of human disease in a region or country are all examples of problems that would fall under the umbrella of "Biometrics".

Mudelid juhuslikkusele

Kuidas kirjeldada seda, mis on juhuslik?

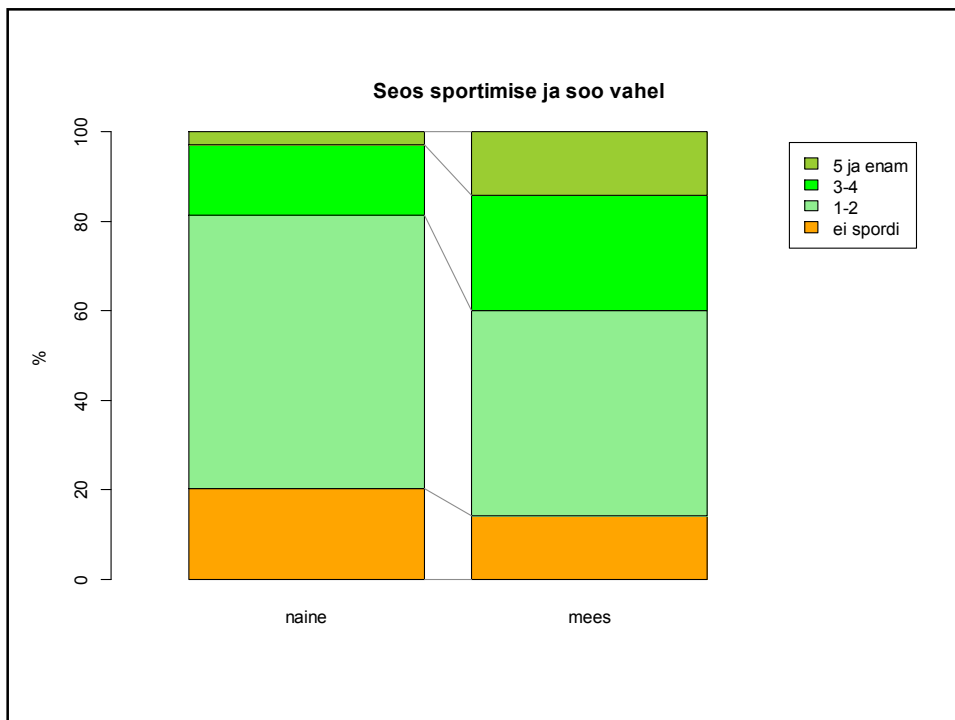
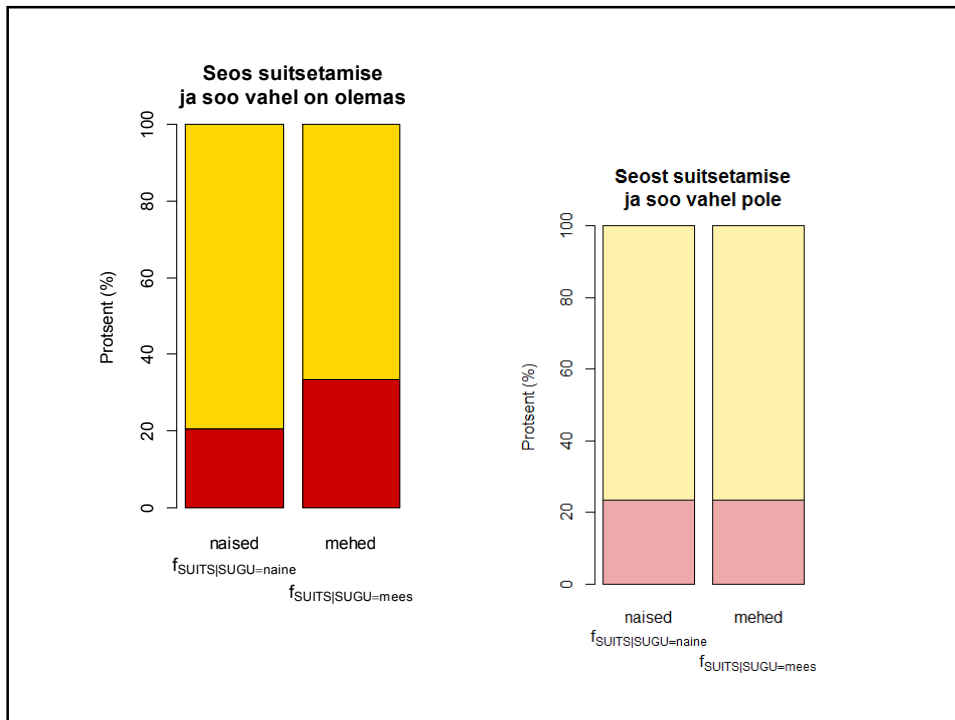


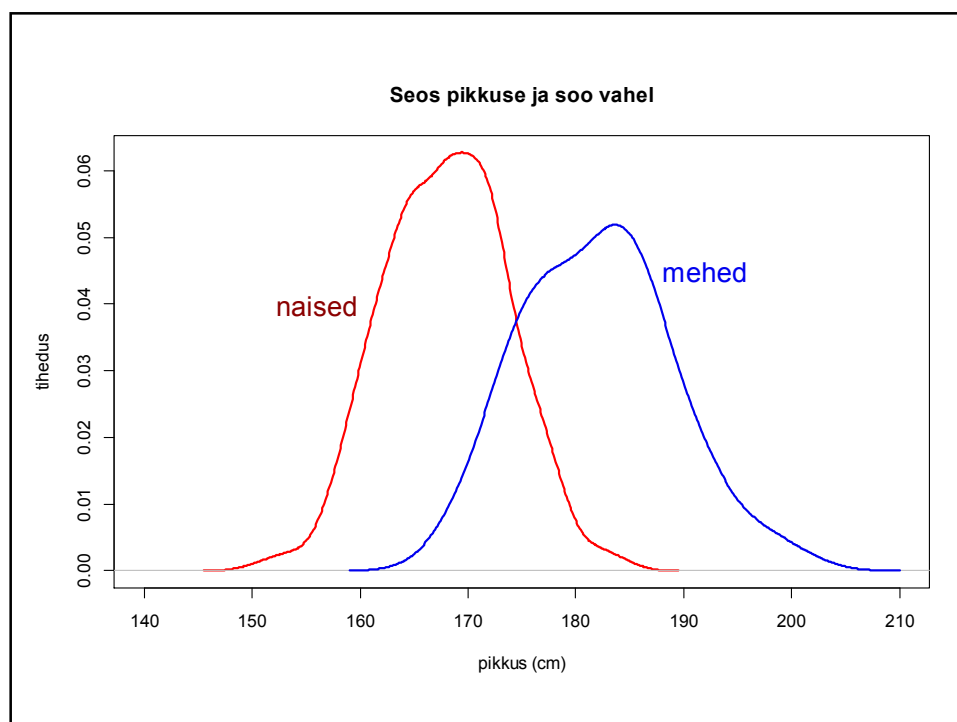
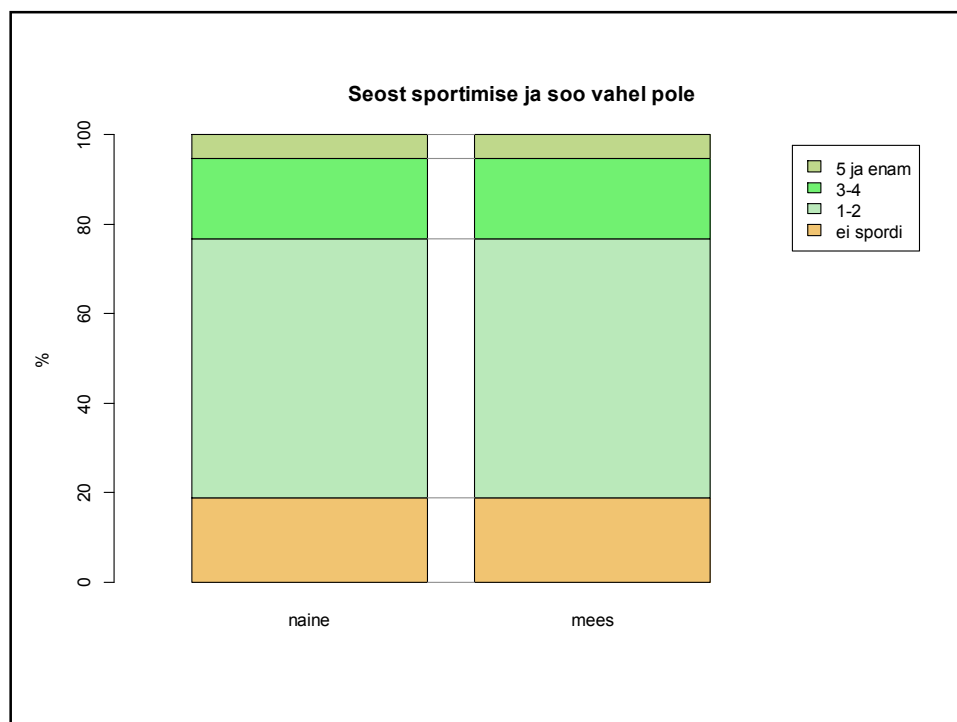
Tihedusfunktsioon ja histogramm

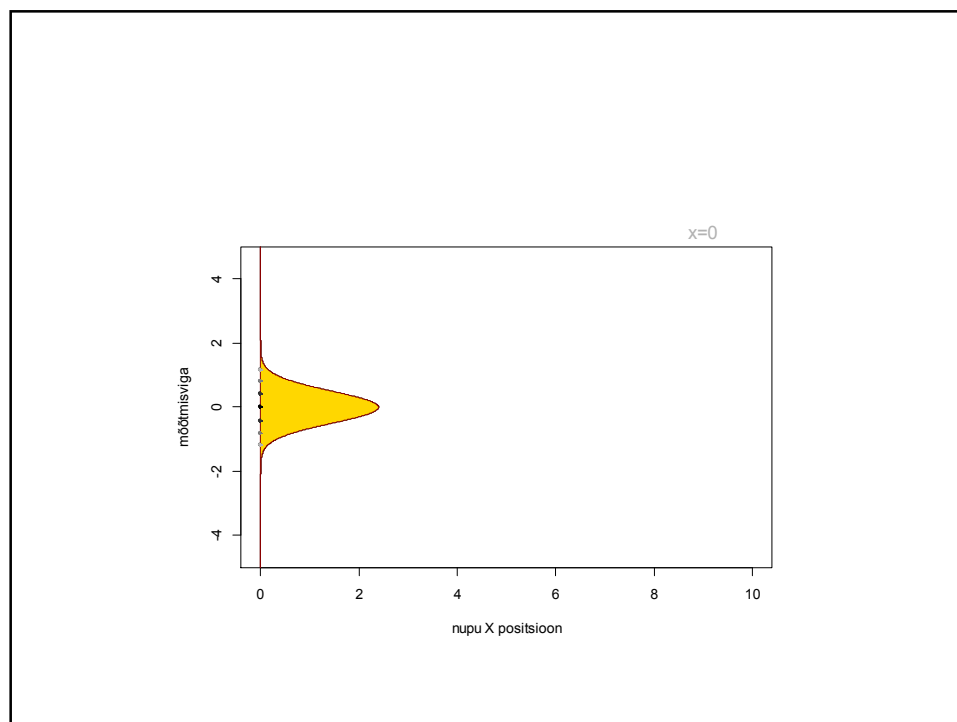
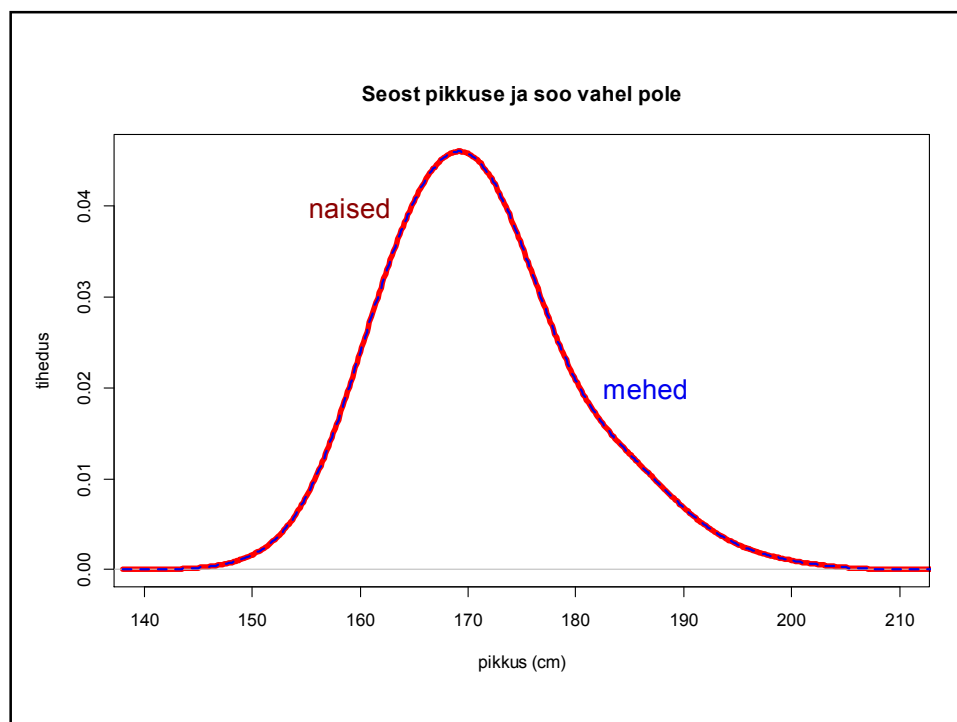


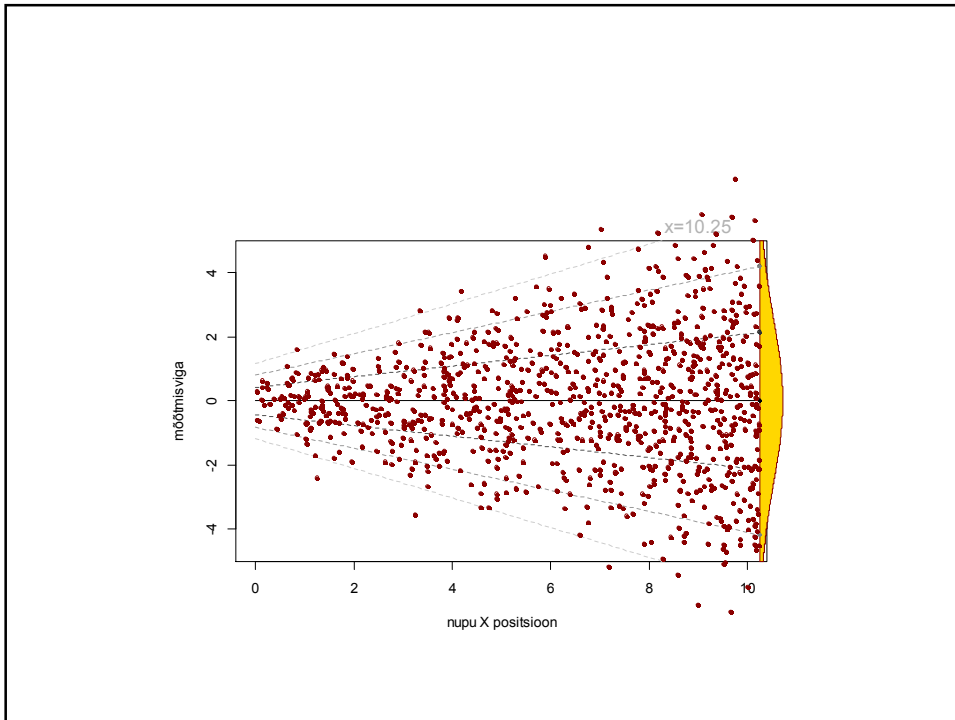
Statistiline seos (statiline sõltuvus)

statistical dependence; statistical association

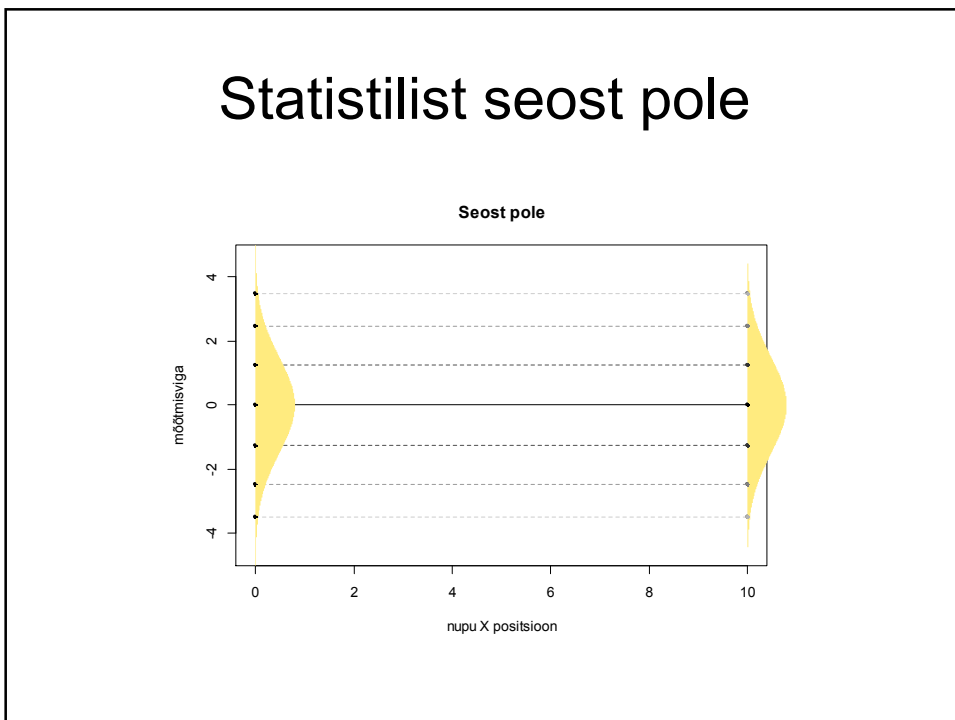








Statistilist seost pole



Mis on statistiline seos?

Matemaatiline definitsioon (pideva tunnuse jaoks):

Tunnused Y ja X on sõltumatud, kui

$$f_{Y|X=c} = f_Y \quad \forall c \text{ (mida } Y \text{ võib omandada)}$$

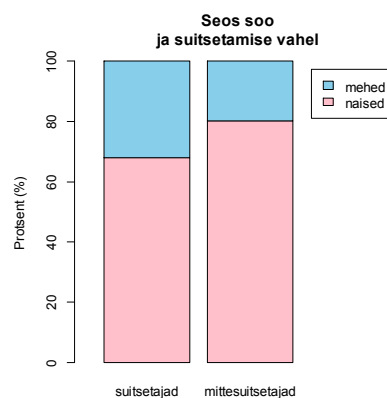
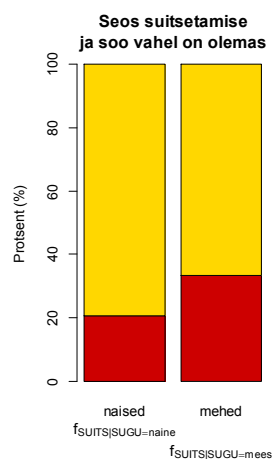
tunnuse Y tinglik jaotus ei sõltu valitud X -tunnuse väärtusest

Tunnused Y ja X on sõltuvad, kui

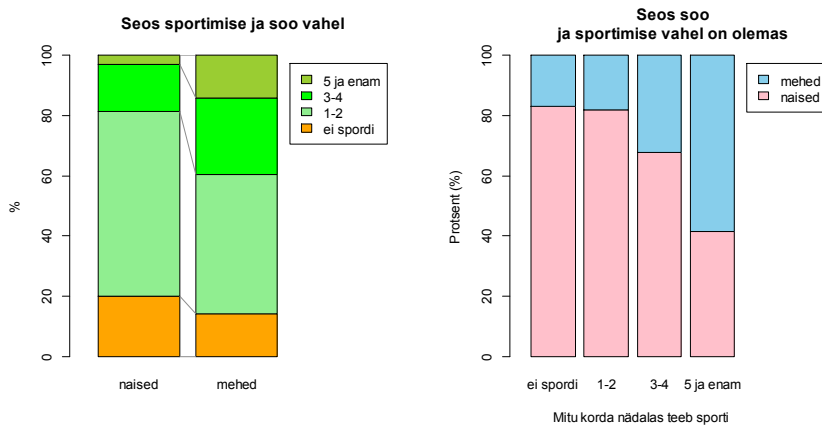
$$\exists c, f_{Y|X=c} \neq f_Y$$

tunnuse Y tinglik jaotus sõltub valitud X -tunnuse väärtusest

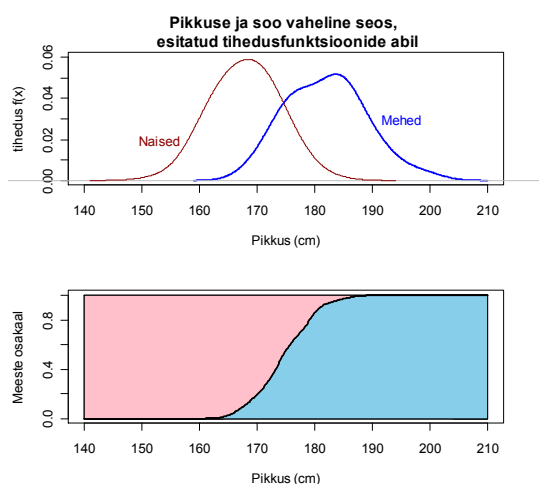
Statistiline seos on sümmeetriline – kui on seos tunnuste X ja Y vahel, siis eksisteerib ka seos tunnuste Y ja X vahel!



Statistiline seos on sümmeetriline – kui on seos tunnuste X ja Y vahel, siis eksisteerib ka seos tunnuste Y ja X vahel!



Statistiline seos on sümmeetriline – kui on seos tunnuste X ja Y vahel, siis eksisteerib ka seos tunnuste Y ja X vahel!

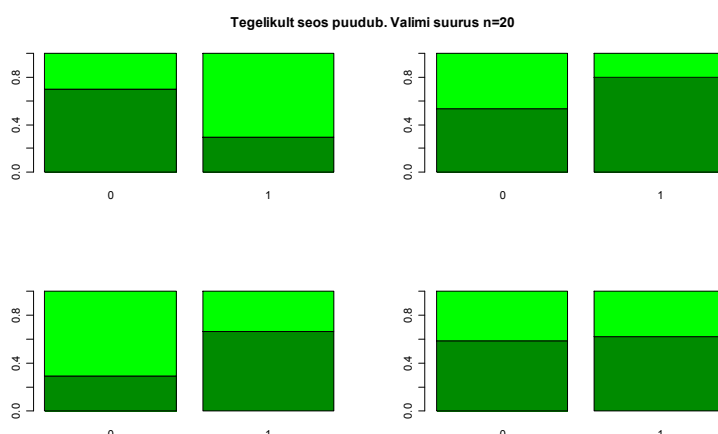


Järeldus

Ainuüksi statistilise seose olemasolust (ravimit tarvitanud inimesed on haigemad kui ravimit mittetarvitanud inimesed) ei saa teha põhjuslikke järeldusi (ravimi tarvitamine teeb inimesed haigemaks).

Meeldetuletuseks: statistilise seose sümmeetrilisuse tõttu on seos ka haiguste ja ravimi tarvitamise vahel: haigemad inimesed tarvitavad rohkem ravimeid, seega võib nähtud seose taga olla hoopis see, et haigus meelitab inimesi rohkem ravimeid tarvitama...

Kas seos on tegelik või näiline
(valimi juhuslikkus petab meid)?



Mida sooviksin antud aines teile õpetada

- **Kuidas kirjeldada juhuslikkuse muutlikust**
 - kui uuritav tunnus on pidev (pikkus, kaal, toimeaine kontsentratsioon...)
 - kui uuritav tunnus on diskreetne (millegi loendamisel saadud: linnukeste arv põllul, jalutuskäigul kohatud liblikate arv, ühe positsiooni katvus sekveneerimisel, ...)
 - kui uuritav tunnus on binaarne (suri/ei surnud; idanes/ei idanenud; katse tuli välja/ei õnnestunud...)
- **Kuidas kontrollida statistilise seose olemasolu tunnuste vahel** (kas seos võib olla tingitud valimi juhuslikkusest või on ikka tegelikult olemas?)
- **Kuidas leida üles ja kirjeldada põhjuslikku mõju?** (Üks asi võib teist põhjustada, aga Mitte iga statistilise seose taga ei ole peidus põhjuslik mõju...)
- **Õpetada kasutama statistikatarkvara R**