

Hüpoteeside statistiline testimine Võimsusfunktsioon

Märt Möls

1

Hüpoteesid

Nullhüpotees (H_0) – hüpotees mida oleme meie (või ühiskond või uuringu sihtauditoorium – näiteks teadlaskond) nõus vaikimisi õigeks pidama.

Võimalikke stiilinäiteid:

- üldrelatiivsusteooria kehtib;
- Märk Möls ei ole mõrvar;
- Inimese silmade värvi määrav pärilik informatsioon on peidus inimese DNAs

Hüpoteesid

Nullhüpotees (H_0) – hüpotees mida oleme meie (või ühiskond või uuringu sihtauditoorium – näiteks teadlaskond) nõus vaikimisi õigeks pidama.

Alternatiivne hüpotees (H_1 või H_A) – nullhüpoteesi eitus.

- Üldrelatiivsusteooria ei kehti;
- Märk Möls on mõrvar;
- Inimese silmade värvi määrav pärilik informatsioon pole peidus inimese DNAs vaid on kirjas kuskil mujal
-

Hüpoteesid

Nullhüpotees (H_0) – hüpotees mida oleme meie (või ühiskond või uuringu sihtauditoorium – näiteks teadlaskond) nõus vaikimisi õigeks pidama.

Alternatiivne hüpotees (H_1 või H_A) – nullhüpoteesi eitus.

NB! Üks kahest – alternatiivne hüpotees või nullhüpotees – peab igal juhul õige olema (kolmandat võimalust ei tohi eksisteerida)!

Kuidas otsustame?

(üldine põhimõte)

- Sõnastame hüpoteesid;
- valime teststatistiku – näitaja, mille põhjal teeme otsuse;
- Nuputame välja milliseid teststatistiku väärtuseid võiksime näha H_0 kehtides;
- Võta valim.
- Arvuta teststatistiku väärtus. Kui tulemus oli ootuspärane (H_0 vaatevinklist) siis otsusta H_0 kasuks.
- Kui teststatistik polnud ootuspärane, siis on nullhüpoteesiga midagi valesti – otsusta H_1 kasuks

Uinunud teadlase probleem.

Vaene väsinud teadlane uinus ja andmed jäid kogumata.

Kas ta peaks oma hüpoteese statistiliselt kontrollides otsustama H_0 või H_1 kasuks?

Otsus: H_0

Kui nullhüpotees kehtiks, siis uinunud teadlasele poleks vaatluseid – seega on tema vaatlustulemused kooskõlas H_0 –ga.

Kui nullhüpoteesi kahjuks rääkivaid andmeid ei ole siis jääme H_0 juurde.



H_0 : Märt Möls ei ole mõrvar
 H_1 : Märt Möls on mõrvar

Põhjendus: süütõendite puudumise tõttu õigeks mõistetud

Klassikaline sõnastus... ... ja tänapäevane kasutus...

Tänapäeval ei ole nullhüpotees sageli endastmõistetav. Sageli kontrollitakse ka selliselt sõnastatud nullhüpoteese, millesse meie või ühiskond pigem ei usu (sest loodame tõestada, et nullhüpotees on vale). See on toonud kaasa ka vajaduse tulemuste tõlgendamist muuta.

Ajalooliselt (klassikaline arusaam)

Loe kehtivaks H_0 (võta vastu H_0 , otsusta H_0 kasuks, jää H_0 juurde – *accept the H_0*)

kui andmeid pole, siis

Tänapäev / tavakasutus

Nullhüpoteesi ei õnnestunud kummutada (nullhüpotees ei ole vastuolus kogutud andmetega, *do not reject H_0 , failing to reject H_0 , ...*)

Näide 2

Hüpoteesid:

H_0 : Mängupartner on aus

H_1 : Ta ei mängi ausat mängu!

Mille põhjal teeme oma otsuse (teststatistik)?

Visatud 6-te arv.

Kavandatav valim:

10 täringuviske tulemused

Näide 2

Hüpoteesid:

$$H_0: P(6) = 1/6$$

$$H_1: P(6) \neq 1/6$$

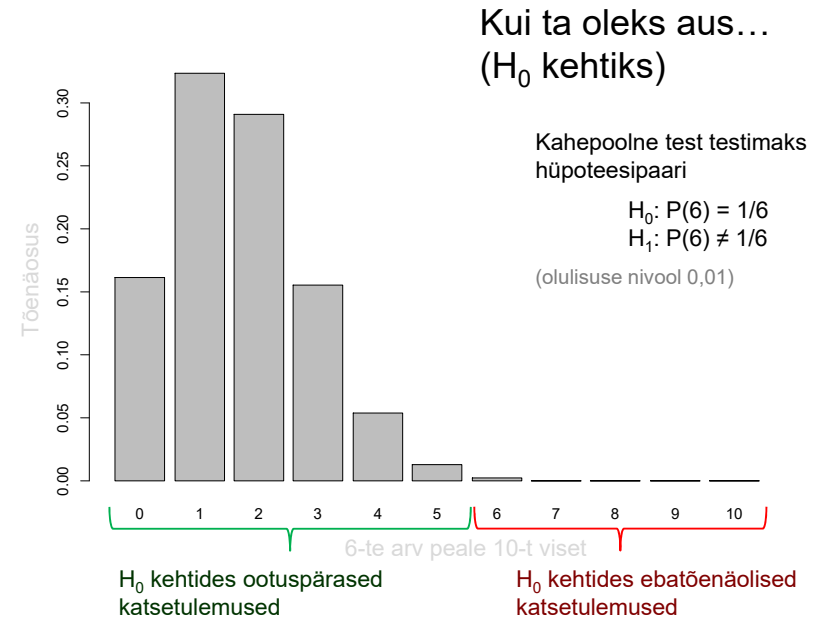
Mille põhjal teeme oma otsuse (teststatistik)?

T = visatud 6-te arv.

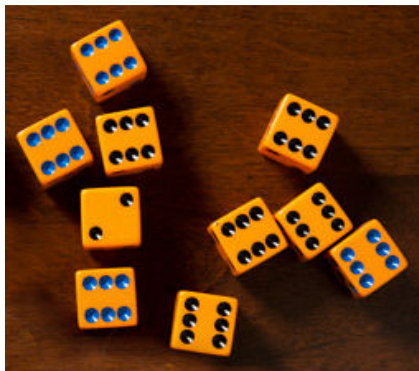
Milline on
teststatistiku jaotus
 H_0 kehtides?

Kavandatav valim:

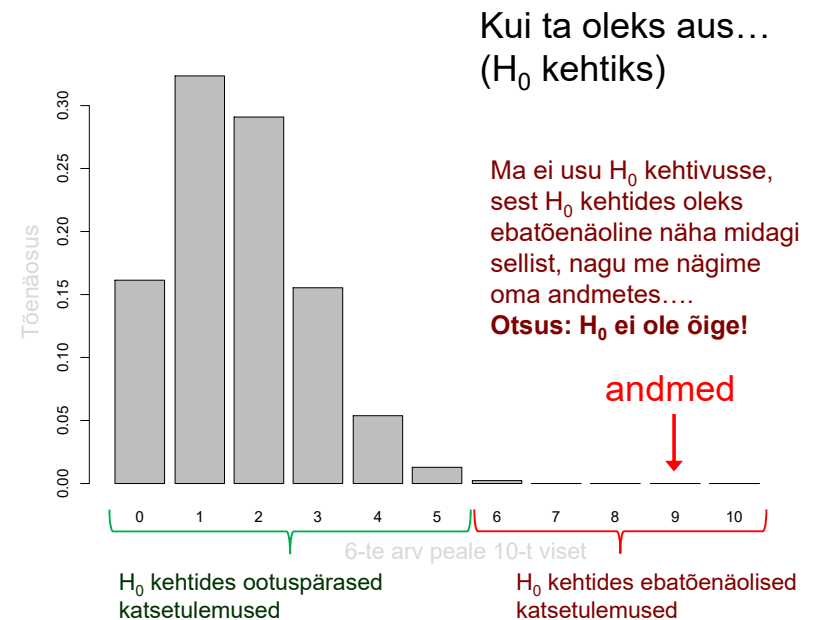
10 täringuviske tulemused



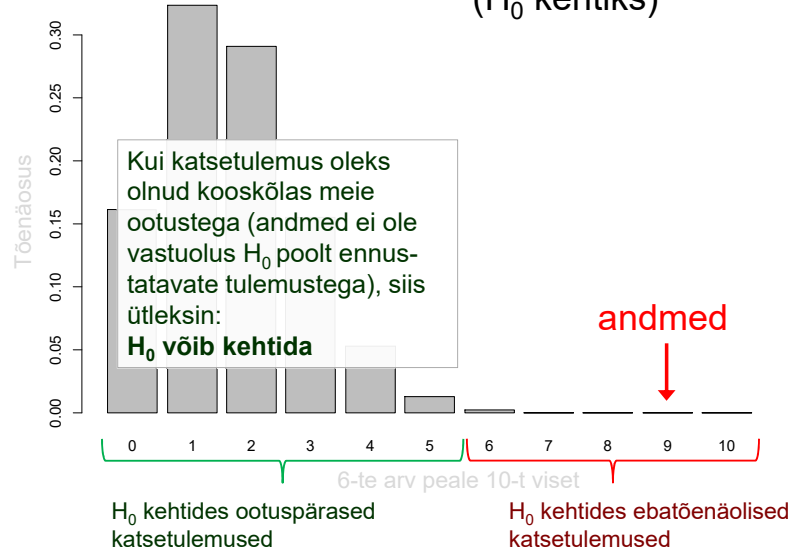
Kogume andmed (10 täringuviske tulemused):



Teststatistiku väärtus: 9



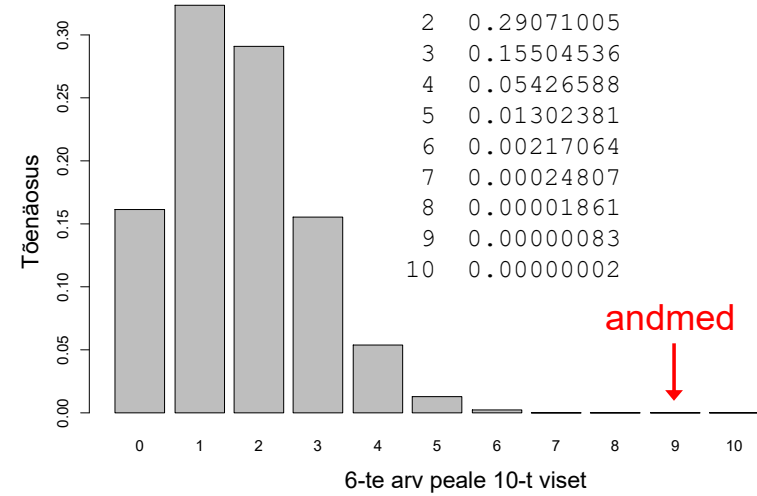
Kui ta oleks aus...
(H_0 kehtiks)



Kuusi tõenäosus

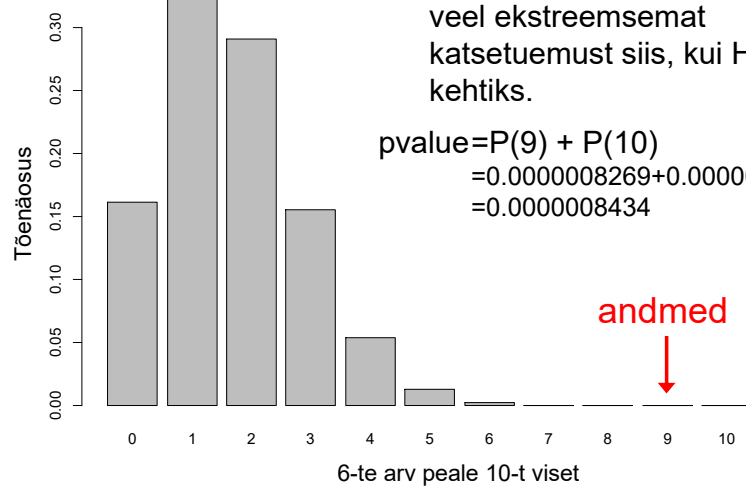
Kuusi	tõenäosus
0	0.16150558
1	0.32301117
2	0.29071005
3	0.15504536
4	0.05426588
5	0.01302381
6	0.00217064
7	0.00024807
8	0.00001861
9	0.00000083
10	0.00000002

p-väärtus?



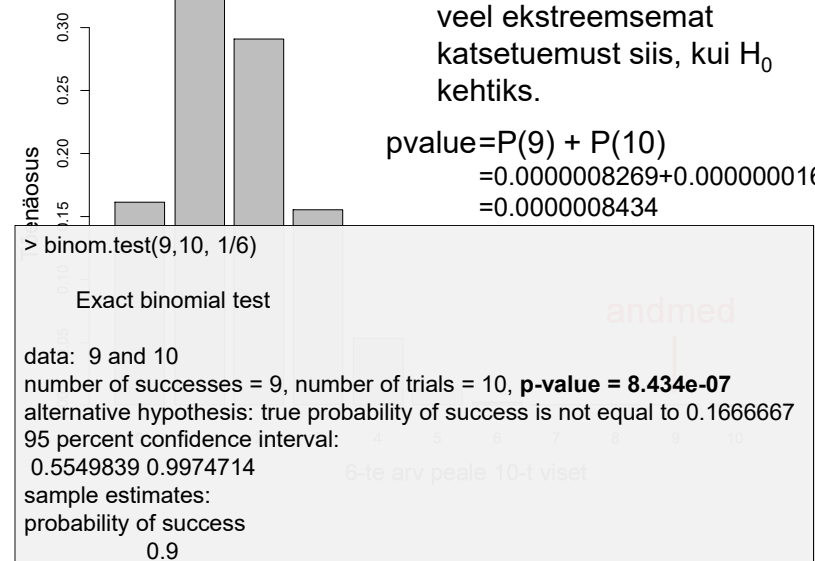
p-väärtus – tõenäosus näha sedavõrd ekstreemset või veel ekstreemsemat katsetulemust siis, kui H_0 kehtiks.

$$\begin{aligned} \text{pvalue} &= P(9) + P(10) \\ &= 0.0000008269 + 0.0000000165 \\ &= 0.0000008434 \end{aligned}$$



p-väärtus – tõenäosus näha sedavõrd ekstreemset või veel ekstreemsemat katsetulemust siis, kui H_0 kehtiks.

$$\begin{aligned} \text{pvalue} &= P(9) + P(10) \\ &= 0.0000008269 + 0.0000000165 \\ &= 0.0000008434 \end{aligned}$$



```
> binom.test(2, 10, 1/6)
```

Exact binomial test

data: 2 and 10

number of successes = 2, number of trials = 10, p-value = 0.677

alternative hypothesis: true probability of success is not equal to 0.1666667

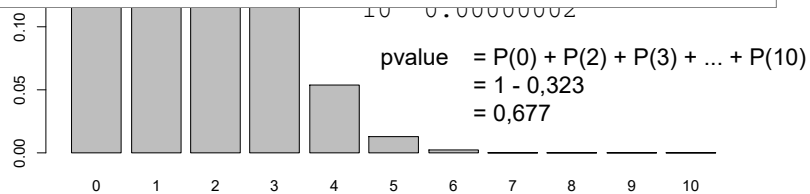
95 percent confidence interval:

0.02521073 0.55609546

sample estimates:

probability of success

0.2



6-te arv peale 10-t viset

andmed

Teadmiste kontroll

H_0 : "99% tudengitest on lollid"

Valim: 1 tudeng (osutus mittelolliks)

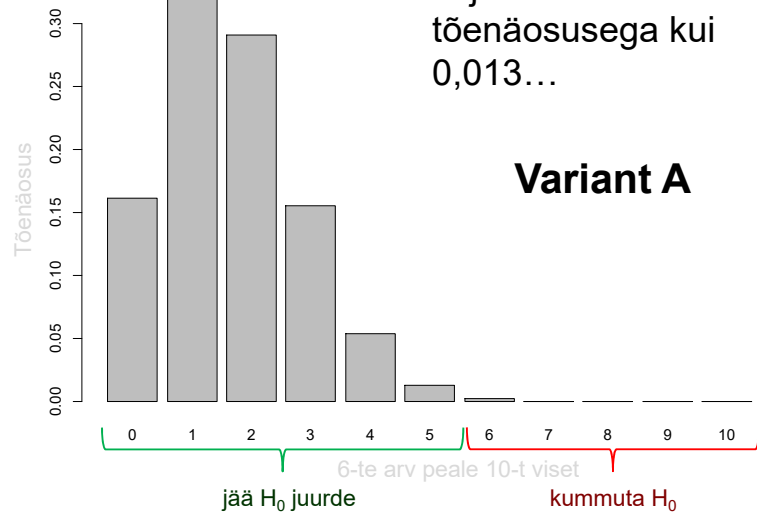
p-väärtus:0,01.....

Otsus (olulisuse nivoo 0,05):

..... H_1

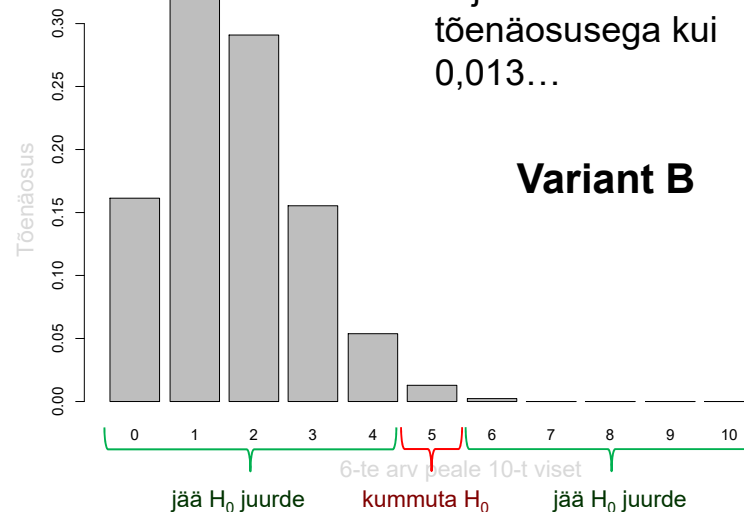
Kui tahame, et I-liiki viga ei juhtuks suurema tõenäosusega kui 0,013...

Variant A

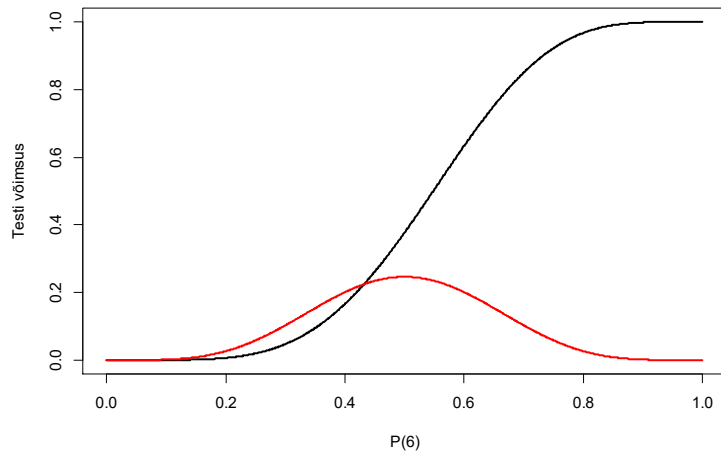


Kui tahame, et I-liiki viga ei juhtuks suurema tõenäosusega kui 0,013...

Variant B



Miks eelistame variant A-d variandile B?



I ja II liiki vead

Suure vea tegemise tõenäosus on piiratud olulisuse nivooga (α , significance level). Enamasti 0,05.

		Tõde	
		H_1	H_0
Sinu otsus	H_1	Tubli!	suur viga (I-liiki viga)
	H_0	viga (II-liiki viga)	Tubli!

I ja II liiki vead

Tõenäosus (β) tõestada alternatiivne hüpotees (kui alternatiivne hüpotees kehtib).

		Tõde	
		H_1	H_0
Sinu otsus	H_1	Tubli!	suur viga (I-liiki viga)
	H_0	viga (II-liiki viga)	Tubli!

Teadmiste kontroll

H_0 : "90% tudengitest on lollid"

Olulisuse nivoo: $\alpha=0,05$
 $n=1$

Leia testi võimsus!

$\beta=.....$

Testi võimsusfunktsioon

Hea testi korral on nii I-liiki kui ka II-liiki vea tegemise tõenäosus väike. Testide omavaheliseks võrdluseks kasutatakse võimsusfunktsiooni.

Võimsusfunktsioon $h(\theta)$ on nullhüpoteesi kummutamise tõenäosus vaadatuna parameetri θ funktsioonina:

$$h(\theta) = P(\text{kummutame } H_0 \mid \theta)$$

Kuna testi võimsus sõltub ka valimi suurusest siis võime võimsusfunktsiooni vaadelda lisaks parameetrile ka valimi suuruse funktsioonina.

Sample Size in Cardioprotection Trials Engblom et al

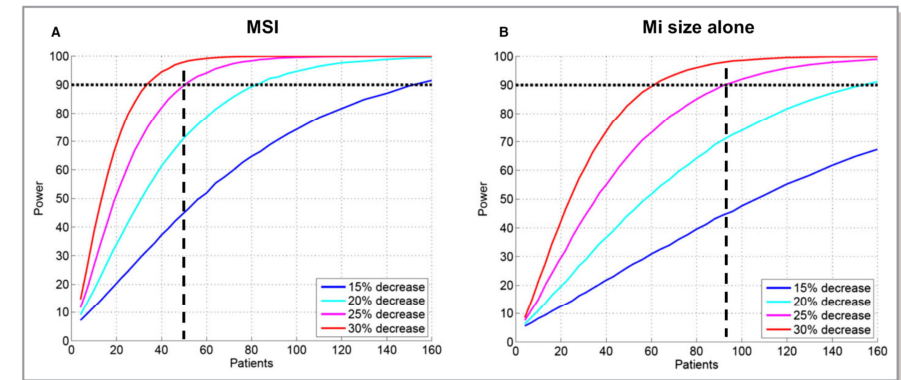


Figure 1. Difference in number of patient needed per treatment arm for different expected treatment effects when using (A) MSI and (B) MI size alone in order to reach sufficient statistical power. Two-sided probability $\alpha=0.05$ of type 1 error was assumed. Dashed lines indicate the number of patients needed in each treatment arm to detect a decrease of 25% in outcome variables. MaR indicates myocardium at risk; MI, myocardial infarction; MSI, myocardial salvage index.