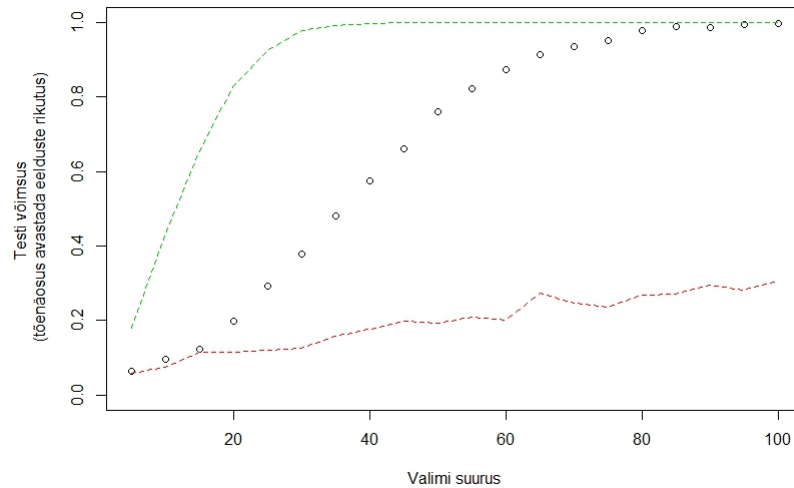


# 1 Ülesannete lahendused. Peatükk 1

1. Kuna normaaljaotuse korral langevad mediaan ja keskvärtus kokku ( $med(X) = \mu$ ), siis saame hüpoteesi  $H_0 : med(X) = 3$  testimiseks kasutada t-testi.
2. Nullhüpoteesi ei saa statistilise testi abil tõestada. Seega ka Shapiro-Wilk'i testi suur olulisustõenäosus ei tõesta veel väidet, et andmed on normaaljaotusega.



Joonis 1: Tõenäosus avastada, et uuritav tunnus pole normaaljaotusega

Nullhüpoteesi väite usutavuse tõestamiseks peaks kasutama teoreetilist tõestust, sellist, mis ei kasuta vaatlusandmeid. Võib näiteks arutleda, et antud uuritavat tunnust (oletame, et uuriti inimese pikkust) mõjutavad tuhanded erinevad geenid ja tuhanded erinevad keskkonnamõjud (lisa viide kirjandusele). See lubab tsentraalset piirteoreemi kasutades eeldada, et uuritava tunnuse jaotus võiks olla normaaljaotus (sadade ja tuhandete juhuslike suuruste summa jaotus on tsentraalse piirteoreemi järgi normaaljaotus). Lisa veel viide sajakonnale autoriteedile, kes kõik on leidnud, et inimeste pikkus on normaaljaotusega. Sellise põhjendusega oled jõudnud lodetavasti punkti, kus normaaljaotuse eelduse juurde jäämine tundub isenesestmõistetav. Igaks juhuks kontrollid täiendavalt Shapiro-Wilk'i testiga väidet täiendavalt. Tehtud test ei lükka normaaljaotuse eeldust ümber, seega võid sina (ja teadustöö

lugejad) tõepoolest eeldada, et andmed on normaaljaotusega — sest seda väidet oleks usutud ka ilma algandmeid nägemata.

Tuleb meeles pidada, et statistikas nullhüpoteesi juurde jäämine tähendab umbes sedasama, mis kohtus süütõendite puudusel õigeksmõistmine — nullhüpotees kas kehtib (kohtualune on süütu) või statistik/uuringu teostaja on teinud oma töö lohakalt ja pole kogunud piisavalt tõendusmaterjali (politsei ei teinud oma tööd piisavalt hästi).

Vaevalt, et endast lugupidav statistik oleks rõõmus lugeja kommentaari üle: “Kuna ta tegi oma tööd lohakalt ja ei viitsinud tõendusmaterjali koguda, siis võis ta normaaljaotuse eelduse juurde jääda”.

Vaata ka graafikut 1, kus on näidatud Shapiro-Wilk’i testi võimsus (tõenäosus tuvastada normaaljaotuse eelduse rikutust). Vaadeldud on kolme erinevat juhtu: uuritav tunnus on tegelikult kas ühtlase jaotusega, t-jaotusega ( $df=8$ ) või binoomjaotusega  $B(10, 0.4)$ .

3. Statistiku  $S(X; \Theta_0)$  jaotus nullhüpoteesi kehtides näeks välja järgmine:

$x$	$P(S(X, \Theta_0) = x)$	$\Theta_0$ väärtuste vahemik
-8	0.0039	102 ... $\infty$
-6	0.0313	55 ... 102
-4	0.1094	30 ... 55
-2	0.2187	20 ... 30
0	0.2734	15 ... 20
2	0.2187	12 ... 15
4	0.1094	11 ... 12
6	0.0313	10 ... 11
8	0.0039	$-\infty$ ... 10

Kus tulbas  $\Theta_0$  on toodud testitavate väärtuste vahemik, mille puhul saame statistiku  $S$  väärtuseks  $x$ -i.

Meie valimi korral  $S(X; 10, 5) = 6$ . Märgitesti olulisustõenäosus on seega

$$\begin{aligned}
 pvalue &= P(S = -8) + P(S = -6) + P(S = 6) + P(S = 8) \\
 &= 2 * (0.0039 + 0.0313) \\
 &= 0,0704.
 \end{aligned}$$

Seega olulisuse nivool 0,05 peaksime jääma nullhüpoteesi juurde (mediaan võib olla 10,5). Olulisuse nivool 0,1 oleksime aga saanud nullhüpoteesi kummutada.

Usaldusintervalli leidmiseks märkame, et  $P(-4 \leq S(X; \Theta) \leq 4) = 0,93$ . Seega moodustavad usaldusintervalli sellised  $\Theta$  väärtused, mille puhul me nullhüpoteesi kummutada ei suuda (sellised  $\Theta_0$  väärtused, mille puhul  $-4 \leq S(X; \Theta_0) \leq 4$ ). Sellisteks väärtusteks on väärtused vahemikus (11...55).

4. Kui alumine kvartiil on  $\Theta$ , siis  $\Theta$ -st suuremate mõõtmistulemuste arv valimis ( $n = 8$  korral) on binoomjaotusega  $B(8; 0,25)$ . Seda teadmist kasutades saame välja kirjutada ka teststatistiku  $S(X; \Theta_0)$  jaotuse nullhüpoteesi kehtides ( $H_0$ : alumine kvartiil on  $\Theta_0$ ):

$x$	$P(S(X, \Theta_0) = x)$	$\Theta_0$ väärtuste vahemik
-8	<0.0001	102...∞
-6	0.0004	55...102
-4	0.0038	30...55
-2	0.0231	20...30
0	0.0865	15...20
2	0.2076	12...15
4	0.3115	11...12
6	0.2670	10...11
8	0.1001	-∞...10

Näeme, et olulisuse nivool 0,214 (= 1 - 0,786) jääme nullhüpoteesi juurde siis, kui statistiku  $S(X, \Theta_0)$  väärtus on kas 6, 4 või 2. Sellise statistiku väärtuse saame, kui  $\Theta_0$  väärtus on vahemikus 10...15. Olemegi saanud 78,6%-usaldusintervalli alumisele kvartiilile: 10...15.