Johdatus tilastolliseen päättelyyn, 3. harjoitus (Viikko 15: 9.–12.4.2013)

1. V. 2012 pidetyn JTP-kurssin ensimmäisellä luennolla tehtiin kysely, jonka mukaan miesopiskelijoiden painon (kg) otoskeskiarvo ja otosvarianssi ovat

$$\bar{y} = 77.07, \qquad s^2 = 184.56,$$

kun otoskoko oli $n = 37$. Analysoimme aineiston ajattelemalla, että vastaavat satunnaismuuttujat ovat satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, jossa molemmat parametrit ovat tuntemattomia.

Mitkä ovat populaation odotusarvon μ ja sen varianssin σ^2 piste-estimaatit? Mikä on populaation keskihajonnan σ piste-estimaatti? Mikä on populaation odotusarvon estimaatin keskivirhe?

2. Tarkastellaan havaintoja $y_1, \ldots, y_n$, jossa jokainen $y_i > 0$. Mallimme mukaan havaintoja vastaavat satunnaismuuttujat ovat satunnaisotos gammajakaumasta $\mathrm{Gamma}(\alpha, \lambda)$. Gammajakauma on jatkuva jakauma, jonka tiheysfunktio on

$$g(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \, \mathrm{e}^{-\lambda x}, \quad x > 0.$$

jossa parametrit $\alpha > 0$ ja $\lambda > 0$. Jos $X \sim \mathrm{Gamma}(\alpha, \lambda)$, niin sen odotusarvo ja varianssi ovat

$$EX = \frac{\alpha}{\lambda}, \qquad \operatorname{var} X = \frac{\alpha}{\lambda^2}$$

Parametrien $\alpha > 0$ ja $\lambda > 0$ SU-estimaatteja ei saada ratkaistua suljetussa muodossa, mutta momenttimenetelmällä saadaan kätevät kaavat estimaateille. Johda ne.

Opastus: toinen populaatiomomentti saadaan kaavalla $EX^2 = \operatorname{var} X + (EX)^2$.

3. Laske tehtävän yksi aineiston perusteella miesopiskelijan painon odotusarvolle 95 %:n t-luottamusväli käyttämällä kaavaa (5.17). Tarvittavan t-jakauman kvantiilin voi lukea esim. oheisesta taulukosta (tai voit laskea sen tietokoneella).

Laske lisäksi odotusarvolle z-luottamusväli kaavalla (5.10), jossa populaatiovarianssin σ^2 kohdalle sijoitat otosestimaatin s^2.

4. Tarkastelemme normaalijakauman odotusarvon luottamusväliä, kun normaalijakauman molemmat parametrit ovat tuntemattomia. Joissakin lähteissä kerrotaan, että suuren otoskoon tapauksessa (esim. $n \geq 30$) luottamusväli voidaan laskea t-luottamusvälin kaavan sijasta z-luottamusvälin kaavalla, jossa sijoitetaan $\sigma^2 = s^2$. Tällä tavalla saatava z-luottamusväli on aina lyhyempi kuin t-luottamusväli, mutta riittävän suurella otoskoolla tällä erolla ei enää ole käytännön merkitystä.

a) Kuinka monta prosenttia leveämpi 95 %:n t-luottamusväli on kuin 95 %:n z-luottamusväli, jos $n = 30$?

b) Miten suuri otoskoon pitää olla, jotta 95 %:n t-luottamusväli olisi enintään yhden prosentin leveämpi kuin 95 %:n z-luottamusväli? Voit approksimoida t-jakauman yläkvantiilia standardinormaalijakauman yläkvantiilin z_u avulla (eräästä asymptoottisesta kehitelmästä saatavalla) kaavalla

$$t_\nu(u) \approx z_u + \frac{z_u^3 + z_u}{4} \, \frac{1}{\nu}.$$

5. Tahtoisimme saada miesopiskelijoiden keskimääräisen painon selville niin tarkasti, että kaksipuolisen 95 %:n t-luottamusvälin leveys olisi vain noin yksi kg. Miten suuri otos tähän (osapuilleen) tarvittaisiin?

Opastus: Tässä laskussa joudutaan tekemään approksimaatioita. Kaksipuolisen t-luottamusvälin leveys on $2\, t_{n-1}(0.05/2)\, s/\sqrt{n}$, jossa s on uudesta otoksesta laskettu otoskeskihajonta ja n on sen otoskoko. Tee approksimaatio $t_{n-1}(0.05/2) \approx 2$ ja arvioi, että uudesta otoksesta saadaan suunnileen sama otoskeskihajonta kuin tehtävän yksi otoksesta. (Tässä on järkeä sen takia, että sekä tehtävän yksi otoskeskihajonta että uuden otoksen otoskeskihajonta estimoivat samaa populaatioparametria.)

6. Naispuolisten yliopisto-opiskelijoiden paino noudattaa likimain normaalijakaumaa. Otantatutkimuksen mukaan naisopiskelijoiden populaation keskimääräisen painon μ (kg) 95 %:n luottamusväliksi saatiin $[59.4, 65.8]$. Mitkä seuraavista tulkinnoista ovat vääriä? Perustele väärien väiteiden kohdalla, miksi ne ovat väärin.

a) Parametrin μ:n todennäköisyysmassasta 95 % sijaitsee ko. välillä.

b) 95 % naisista tässä populaatiossa kuuluu painoltaan ko. välille.

c) Mikäli samanlainen otantatutkimus toistettaisiin useita kertoja, ja joka kerta laskettaisiin vastaava luottamusväli, niin μ kuuluisi noin 95 %:iin lasketuista väleistä.

Taulukko 1: t-jakauman u-yläkvantiileja $t_\nu(u)$ eri vapausasteluvun ν arvoilla. Tässä $u = P(Y \geq t_\nu(u))$, kun $Y \sim t_\nu$.

$\nu \backslash u$	0.05	0.025	0.01	0.005	0.001
1	6.314	12.706	31.821	63.657	318.309
2	2.920	4.303	6.965	9.925	22.327
3	2.353	3.182	4.541	5.841	10.215
4	2.132	2.776	3.747	4.604	7.173
5	2.015	2.571	3.365	4.032	5.893
6	1.943	2.447	3.143	3.707	5.208
7	1.895	2.365	2.998	3.499	4.785
8	1.860	2.306	2.896	3.355	4.501
9	1.833	2.262	2.821	3.250	4.297
10	1.812	2.228	2.764	3.169	4.144
11	1.796	2.201	2.718	3.106	4.025
12	1.782	2.179	2.681	3.055	3.930
13	1.771	2.160	2.650	3.012	3.852
14	1.761	2.145	2.624	2.977	3.787
15	1.753	2.131	2.602	2.947	3.733
16	1.746	2.120	2.583	2.921	3.686
17	1.740	2.110	2.567	2.898	3.646
18	1.734	2.101	2.552	2.878	3.610
19	1.729	2.093	2.539	2.861	3.579
20	1.725	2.086	2.528	2.845	3.552
21	1.721	2.080	2.518	2.831	3.527
22	1.717	2.074	2.508	2.819	3.505
23	1.714	2.069	2.500	2.807	3.485
24	1.711	2.064	2.492	2.797	3.467
25	1.708	2.060	2.485	2.787	3.450
26	1.706	2.056	2.479	2.779	3.435
27	1.703	2.052	2.473	2.771	3.421
28	1.701	2.048	2.467	2.763	3.408
29	1.699	2.045	2.462	2.756	3.396
30	1.697	2.042	2.457	2.750	3.385
31	1.696	2.040	2.453	2.744	3.375
32	1.694	2.037	2.449	2.738	3.365
33	1.692	2.035	2.445	2.733	3.356
34	1.691	2.032	2.441	2.728	3.348
35	1.690	2.030	2.438	2.724	3.340
36	1.688	2.028	2.434	2.719	3.333
37	1.687	2.026	2.431	2.715	3.326
38	1.686	2.024	2.429	2.712	3.319
39	1.685	2.023	2.426	2.708	3.313
40	1.684	2.021	2.423	2.704	3.307
45	1.679	2.014	2.412	2.690	3.281
50	1.676	2.009	2.403	2.678	3.261