Johdatus tilastolliseen päättelyyn, kesä 2015
Harjoitus 3 (2.6.)

Tehtävät 1–4 liittyvät luottamusväleihin. Niitä käsitellään monisteen jaksoissa 5.1–5.6, joista 5.6 on keskeisin. Tehtävä 4 ei kuulu kurssin ydinainekseen, ja sen käsittely voidaan jättää vähemmälle, jos aika ei riitä. Tehtävissä 1 ja 2 tarvittava taulukko on ohessa. Vaihtoehtoisesti voit käyttää netistä löytyviä online-laskimia; esimerkiksi osoitteessa http://surfstat.anu.edu.au/surfstat-home/tables/t.php (valitse tarvittaessa "Standard normal").

1. Tilastollinen mallimme on satunnaisotos kokoa n normaalijakaumasta $N(\mu, \sigma^2)$, jossa $\sigma^2 > 0$ ajatellaan tunnetuksi luvuksi (halutessasi voit valita $\sigma^2 = 1$). Havaintoja vastaaville satunnaismuuttujille pätee siis $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$. Palauta mieleen, miten parametrille μ muodostettiin luottamusväli annetulla luottamustasolla $1 - \alpha$ (ns. z-luottamusväli, ks. monisteen kaava (5.10)).

a) Miten otoskoon n kasvu vaikuttaa luottamusvälin pituuteen? Jos luottamusvälin pituus halutaan kymmenenteen osaan aikaisemmasta eli halutaan "kymmenkertainen tarkkuus", mitä n:lle on tehtävä?
b) Tutki, miten luottamustason muuttaminen vaikuttaa luottamusvälin pituuteen. Jos luottamustaso halutaan nostaa 90 %:sta 95 %:een, kuinka moninkertaiseksi luottamusväli tulee pituudeltaan? Entä jos siirrytään vielä 95 %:n luottamustasosta 99 %:n luottamustasoon? Kuinka pitkä olisi 100 %:n luottamusväli?
c) Oletetaan, että $\sigma^2 = 1$. Kuinka suuri on otoskoon n likimain oltava, jotta 95 %:n luottamusväli μ:lle olisi pituudeltaan noin 0.1 eli muotoa $[\bar{y} - 0.05,\ \bar{y} + 0.05]$?

2. Erääseen ammattiryhmään kuuluvien miesten päivittäistä energiansaantia selvitettiin pienellä otantatutkimuksella, jossa seitsemän miehen päivittäiset energiansaannit (seurantajakson aikana) olivat 9730, 8240, 8060, 9100, 8880, 10070 ja 9810 kilojoulea päivässä.

a) Laske energiansaantimittausten keskiarvo $\bar{y}$ ja keskihajonta s.
b) Muodosta aineiston perusteella 95 %:n luottamusväli ammattiryhmään kuuluvien miesten keskimääräiselle päivittäiselle energiansaannille μ. Oletamme, että energiansaanneissa esiintyvä (satunnais)vaihtelu on normaalisti jakautunutta.

Vihje: Jos lasket keskihajonnan "käsin" laskimella, kaava $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ voi olla avuksi. Se seuraa harjoituksen 2 tehtävästä 5a.

3. Merkitään I:llä edellisessä tehtävässä muodostettua luottamusväliä. Mitkä seuraavista tulkinnoista ovat oikein ja mitkä väärin?

a) μ:n todennäköisyysjakaumasta 95 % sijaitsee välillä I.
b) Jos samanlaisia energiansaantimittauksia kuin edellä tehtäisiin suuri määrä, niistä noin 95 % sijoittuisi välille I.
c) Jos samanlainen seitsemän mittauksen tutkimus kuin edellä toistettaisiin useita kertoja ja joka kerta muodostettaisiin vastaava luottamusväli I, kuuluisi μ:n arvo noin 95 %:iin näistä väleistä.

Perustele väärien väitteiden kohdalla, miksi ne ovat väärin. (Monisteessa luottamusvälin tulkintaa käsitellään erityisesti jaksossa 5.6.2.)

4. Mallimme on yksi havainto Y_1 jakaumasta $\mathrm{Tas}(0, \theta)$, jossa $\theta > 0$ on tuntematon parametri. Käytännön esimerkkinä voisi olla harjoituksen 1 tehtävän 4a asetelma.

a) Päättele, että satunnaismuuttujan Y_1/θ jakauma ei riipu θ:n arvosta. Kyseessä on siis *saranasuure* parametrille θ (ks. monisteen jaksot 5.3 ja 5.5). Totea, että itse asiassa pätee

$$P_\theta(Y_1/\theta \le q) = q, \quad \text{kun } 0 < q < 1,$$

eli Y_1/θ noudattaa tasajakaumaa $\mathrm{Tas}(0, 1)$. Tässä P:n alaindeksi korostaa sitä, että ajattelemme Y_1:n noudattavan jakaumaa $\mathrm{Tas}(0, \theta)$ parametriarvolla θ ja samaa lukua θ käytetään osamäärän Y_1/θ nimittäjässä.
b) Totea, että edellisen perusteella pätee $P_\theta(0.05 \le Y_1/\theta \le 1) = 0.95$ ja päättele tästä 95 %:n luottamusvälin lauseke θ:lle, kun on havaittu $Y_1 = y_1$. Minkä luottamusvälin saat harjoituksen 1 tehtävän 4a tilanteessa, jossa $y_1 = 4.8$?

Lisätehtävä innokkaille: Yleistä tarkastelu malliin, joka koostuu n riippumattomasta havainnosta jakaumasta $\mathrm{Tas}(0, \theta)$ (harjoituksen 1 tehtävän 6a malli).

Tehtävät 5 ja 6 liittyvät testaukseen, jota käsitellään monisteen luvussa 6. Testien voiman käsitettä (jakso 6.3 ja osittain 6.5), testien ja luottamusjoukkojen duaalisuutta (jakso 6.6) ja binomijakauman testausta (jakso 6.8) emme kuitenkaan kurssilla käsittele, vaikkakin erityisesti jakso 6.6 on hyödyllistä luettavaksi itse.

5. Eräässä väestössä on otantatutkimuksen avulla tutkittu tuloeroja yhtäältä miesten ja naisten välillä ja toisaalta mustien ja valkoisten välillä. Tutkimuksen tulokset raportoidaan tällaiseen hyvin tyypilliseen tapaan: "Miesten ja naisten keskituloissa havaittiin merkitsevä ero ($p = 0.009$), miesten ansaitessa keskimäärin enemmän. Mustien ja valkoisten keskituloissa sen sijaan ei havaittu eroa ($p = 0.11$)."

a) Ystäväsi ei ole opiskellut lainkaan tilastotiedettä. Miten selität tutkimuksen tulokset ja siinä esiintyvät lukuarvot hänelle?

b) Ystäväsi kysyy: "Onko nyt siis osoitettu, että ko. väestössä miehet ansaitsevat keskimäärin enemmän kuin naiset ja että mustilla ja valkoisilla on olennaisesti samat keskitulot?" Miten vastaat? Perustele.

6. Professorilla on tavoitteena mitoittaa kurssinsa vaatima viikottainen työmäärä siten, että opiskelijat keskimäärin joutuisivat käyttämään kurssiin liittyvään opiskeluun enintään 10 tuntia. Eräällä viikolla tehdyssä otoksessa (kokoa 13) opiskelijat käyttivät aikaa opiskeluun keskimäärin 11.1 tuntia ja aikojen keskihajonta oli 1.8 tuntia. Testaa t-testillä ja 5 %:n merkitsevyystasolla, onko professorin syytä katsoa epäonnistuneen tavoitteessaan. Oletamme (paremman puutteessa), että vaihtelu opiskeluajoissa opiskelijoiden kesken on likimain normaalisti jakautunutta. Taulukko on ohessa.

Muistutus: Harjoitustehtävien tekemiseen saa neuvoja ohjaustilaisuuksissa; ks. tarkemmin kurssin kotisivulta.

Taulukko: t_ν-jakauman u-yläkvantiileja $t_\nu(u)$, joille $u = P(X > t_\nu(u))$, kun $X \sim t_\nu$. Tässä ν on jakauman vapausasteluku ja t_∞ tarkoittaa standardinormaalijakaumaa $N(0,1)$, jolloin on tapana merkitä $t_\infty(u) = z_u$.

$\nu \backslash u$	0.05	0.025	0.01	0.005	0.001
1	6.314	12.706	31.821	63.657	318.31
2	2.920	4.303	6.965	9.925	22.327
3	2.353	3.182	4.541	5.841	10.214
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
∞	1.645	1.960	2.326	2.576	3.090