Johdatus tilastolliseen päättelyyn, kevät 2015
Harjoitus 3 (31. 3. – 10. 4.)

Opastus: Nämä harjoitukset liittyvät monisteen jaksoihin 5.1–5.6, joista 5.6 on keskeisin. Tärkeimmät tehtävät ovat 1–4. Tehtävien 5 ja 6 käsittely voidaan harjoitusryhmissä jättää vähemmälle, jos aika ei riitä. Tehtävissä 1, 2, 4 ja 5 tarvittavat taulukot ovat ohessa. Vaihtoehtoisesti voit käyttää netistä löytyviä online-laskimia; esimerkiksi osoitteessa http://surfstat.anu.edu.au/surfstat-home/tables/normal.php (valitse tarvittaessa ”Student's t” tai ”Chi-squared”).

Pääsiäisloma on to 2.4. – ke 8.4. Harjoituksia käsitellään tiistain ja keskiviikon ryhmissä ennen pääsiäistä sekä torstain ja perjantain ryhmissä sen jälkeen.

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?

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. Erään auton polttoaineen keskikulutusta todellisessa kaupunkiajossa haluttiin tutkia. Tätä varten suoritettiin kuusi 100 km:n testiajoa sattumanvaraisesti valittuina ajankohtina ja eri kuljettajien toimesta. Mitatut kulutukset litroina olivat 6.1, 5.7, 5.9, 6.8, 6.7 ja 6.0.

a) Laske mittausten keskiarvo $\bar{y}$ ja keskihajonta s.

b) Muodosta aineiston perusteella 95 %:n luottamusväli auton keskikulutukselle μ sadan kilometrin matkalla kaupunkiajossa. Oletamme, että kulutusmittauksissa esiintyvä (satunnais)vaihtelu on normaalisti jakautunutta. Millaisista tekijöistä se voisi johtua?

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 kulutusmittauksia kuin edellä tehtäisiin suuri määrä, niistä noin 95 % sijoittuisi välille I.

c) Jos samanlainen kuuden 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. Tarkastellaan normaalijakauman $N(\mu, \sigma^2)$ odotusarvon luottamusväliä, kun sekä μ että σ^2 ovat tuntemattomia. Joissakin lähteissä kerrotaan, että suuren otoskoon tapauksessa (esim. $n \geq 30$) luottamusväli voidaan laskea t-luottamusvälin kaavan sijasta z-luottamusvälin kaavalla, johon sijoitetaan $\sigma = s$. Tällä tavalla saatava väli on aina lyhyempi kuin ”oikea” t-luottamusväli, mutta riittävän suurella otoskoolla tällä erolla ei ole käytännön merkitystä.

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

b) Kuinka suuri otoskoon n 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

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

5. Opiskele, miten monisteen jaksossa 5.6.4 johdetaan mallissa $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$ parametrille σ^2 luottamusväli muotoa $[c_1(\alpha, n)s^2,\ c_2(\alpha, n)s^2]$ (ks. kaava (5.18)).

a) Muokkaa tätä tarkastelua ja esitä, miten muodostetaan σ^2:lle luottamusväli muotoa $(0,\ b(n, \alpha)s^2]$, kun luottamustaso on $1 - \alpha$. Miten kerroin $b(n, \alpha)$ on valittava?

Lukua $b(n, \alpha)s^2$ kutsutaan σ^2:n *ylemmäksi luottamusrajaksi*. Se on monissa sovelluksissa luontevampi tieto varianssista σ^2 kuin ”kaksisuuntainen” luottamusväli: Kuvaahan σ^2 nimittäin havaintoihin liittyvän satunnaisvaihtelun suuruutta ja usein on tarpeen esittää arvioita siitä, kuinka suurta tämä satunnaisvaihtelu ”pahimmillaan” voi olla. Sen pienuus sen sijaan harvoin on ongelma.

Vastaavalla tavalla voitaisiin määritellä *alempi luottamusraja*.

b) Oletetaan, että $n = 30$ ja $s^2 = 9.0$, ja toimitaan luottamustasolla 95 %. Mikä on kaavan (5.18) antama luottamusväli σ^2:lle? Mikä on σ^2:n ylempi luottamusraja?

6. Tilastollinen 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 pätee itse asiassa

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

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) Valitse yllä $q = 0.95$ ja päättele siitä 95 %:n luottamusvälin lauseke θ:lle, kun on havaittu $Y_1 = y_1$. Minkä luottamusvälin saat harjoituksen 1 tehtävän 4a tilanteessa, jossa $y_1 = 4.5$?

Taulukko 1: 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

Taulukko 2: χ^2_ν-jakauman u-alakvantiileja $q_\nu(u)$, joille $u = P(X < q_\nu(u))$, kun $X \sim \chi^2_\nu$. Tässä ν on jakauman vapausasteluku.

$\nu\backslash u$	0.001	0.005	0.01	0.025	0.05	0.95	0.975	0.99	0.995	0.999
1	0.000	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	59.703
31	12.196	14.458	15.655	17.539	19.281	44.985	48.232	52.191	55.003	61.098
32	12.811	15.134	16.362	18.291	20.072	46.194	49.480	53.486	56.328	62.487
33	13.431	15.815	17.074	19.047	20.867	47.400	50.725	54.776	57.648	63.870
34	14.057	16.501	17.789	19.806	21.664	48.602	51.966	56.061	58.964	65.247
35	14.688	17.192	18.509	20.569	22.465	49.802	53.203	57.342	60.275	66.619
36	15.324	17.887	19.233	21.336	23.269	50.998	54.437	58.619	61.581	67.985
37	15.965	18.586	19.960	22.106	24.075	52.192	55.668	59.893	62.883	69.346
38	16.611	19.289	20.691	22.878	24.884	53.384	56.896	61.162	64.181	70.703
39	17.262	19.996	21.426	23.654	25.695	54.572	58.120	62.428	65.476	72.055
40	17.916	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766	73.402