**Johdatus tilastolliseen päättelyyn, kevät 2014**
**Harjoitus 3 (31. 3.–4. 4.)**

1. Olkoon $Y_1, \ldots, Y_n$ satunnaisotos jakaumasta $N(\mu, 1)$. Kuinka suuri otoskoon $n$ on oltava, jotta havaintojen $y_1, \ldots, y_n$ perusteella muodostettava 95 %:n luottamusväli $\mu$:lle olisi pituudeltaan (noin) 1.0 eli muotoa $[\bar{y} - 0.5,\ \bar{y} + 0.5]$? Entä kuinka suureksi $n$ on valittava, jos halutaan kymmenkertainen tarkkuus eli luottamusvälin pituudeksi vain (noin) 0.1?

   [Muista, että jos $Z \sim N(0, 1)$, niin $P(Z > 1.96) \approx 0.025$.]

2. JTP-kurssin 2012 ensimmäisellä luennolla tehdyn kyselyn mukaan miesopiskelijoiden painon (kg) otoskeskiarvo ja otosvarianssi olivat $\bar{y} = 77.1$ ja $s^2 = 184.6$, kun otoskoko oli $n = 37$. Mallinnamme aineiston ajattelemalla, että vastaavat satunnaismuuttujat ovat satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, jossa molemmat parametrit ovat tuntemattomia.

   a) Mitkä ovat miesopiskelijan painon odotusarvon $\mu$ ja varianssin $\sigma^2$ tavalliset piste-estimaatit? Mikä on keskihajonnan $\sigma$ piste-estimaatti?
   b) Laske aineiston perusteella 95 %:n luottamusväli miesopiskelijan painon odotusarvolle. Tarvittavan $t$-jakauman kvantiilin voit lukea oheisesta taulukosta tai etsiä jollakin netistä löytyvällä laskimella, esim. http://surfstat.anu.edu.au/surfstat-home/tables/t.php.

3. Oletetaan, että vastasyntyneen tyttölapsen päänympärys noudattaa likimain normaalijakaumaa. Eräässä väestössä on otantatutkimuksen perusteella johdettu keskimääräiselle päänympärykselle $\mu$ (senttimetriä) 95 %:n luottamusväli $[34.3, 34.9]$. Mitkä seuraavista tulkinnoista ovat oikein ja mitkä väärin?

   a) $\mu$:n todennäköisyysjakaumasta 95 % sijaitsee kyseisellä välillä.
   b) 95 %:lla kaikista vastasyntyneistä tyttölapsista päänympärys on kyseisellä välillä.
   c) Mikäli samanlainen otantatutkimus toistettaisiin useita kertoja ja joka kerta muodostettaisiin vastaava luottamusväli, kuuluisi $\mu$:n arvo noin 95 %:iin saaduista väleistä.

   Perustele erityisesti väärien väitteiden kohdalla, miksi ne ovat väärin.

4. Tarkastellaan normaalijakauman $N(\mu, \sigma^2)$ odotusarvon luottamusväliä, kun sekä $\mu$ että $\sigma^2$ 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

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

KÄÄNNÄ!

5. a) Olkoot $y_1, \ldots, y_n$ reaalilukuja ja $\bar{y} = \sum_{i=1}^n y_i/n$ niiden aritmeettinen keskiarvo. Näytä, että jos $\mu$ on reaaliluku, niin

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2.$$

Luennoilla tätä hajotelmaa käytettiin normaalijakauman su-estimoinnin yhteydessä (ks. muistiinpanojen sivu 29). [*Ehdotus.* Kirjoita aluksi $y_i - \mu = (y_i - \bar{y}) + (\bar{y} - \mu)$.]

b) Olkoot $Y_1, \ldots, Y_n$ riippumattomia satunnaismuuttujia, joille $E(Y_i) = \mu$ ja $\mathrm{var}(Y_i) = \sigma^2$ (esimerkiksi $Y_i \sim N(\mu, \sigma^2)$). Osoita a-kohdan avulla, että tavanomaiselle otosvarianssille

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y})^2$$

pätee $E(S^2) = \sigma^2$. Tämä kertoo, että $S^2$ on *harhaton estimaattori* $\sigma^2$:lle.

**Taulukko:** $t$-jakauman $u$-yläkvantiileja $t_\nu(u)$, joille $u = P(X > t_\nu(u))$, kun $X \sim t_\nu$. Tässä $\nu$ on jakauman vapausasteluku ja $t_\infty$ tarkoittaa standardinormaalijakaumaa $N(0,1)$.

| $\nu \backslash u$ | 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 |
| $\infty$ | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |