TILASTOTIETEEN JATKOKURSSI, 5+5 OP, 19.1.–25.4.2012. Kirjallisuus: Ilkka Mellinin Johdatus tilastotieteeseen, 2. kirja. Luennoi: yliopistonlehtori Pekka Pere.

1. välikokeen 29.2.2012 kysymysten ratkaisuehdotukset

1.

a) Todennäköisyys, että kukaan ei hälytä poliisia, on

$$(1-0,95)^{38} \approx 0,00000.$$

(Todennäköisyys, että edes viisi naapuria ei olisi hälyttänyt eli $(1-0,95)^5$ olisi viiden desimaalin tarkkuudella 0! Todennäköisyys, että edes kolme naapuria ei olisi hälyttänyt eli $(1-0,95)^3$, olisi kolmen desimaalin tarkkuudella 0!)

b) Todennäköisyys, että vähintään yksi ihminen hälyttää poliisin on $1-\mathsf{P}$(kukaan ei hälytä poliisia):

$$1-(1-0,95)^{38} \approx 1,00000.$$

c) Lasketut todennäköisyydet ovat niin pieniä, että oletusten pätevyys on syytä kyseenalaistaa: Ihmiset eivät mahdollisesti hälytä poliisia toisistaan riippumattomasti samalla todennäköisyydellä, kun monta ihmistä havaitsee kriisitilanteen. Toisin sanoen tehtävässä oletettu poliisille soittamisen todennäköisyyden riippumattomuus ihmisten välillä ei päde todellisuudessa. (Tällainen ajattelu johtaa kurssin osassa II tilastollisen testin käsitteeseen.)

2. Binomijakauman normaalijakauma-approksimaatio on kirjan (s:n 110) peukalosäännön mukaan käyttökelpoinen, kun $np > 5$ ja $nq > 5$, jossa n on havaintojen lukumäärä ja p ja q ovat vaihtoehtoisten tapahtumien todennäköisyydet. Tehtävän tilanteessa $n = 96$ ja oletuksen mukaan $p = q = 0,5$, joten

$$np = nq = 96 \times 0,5 = 48 > 5.$$

Peukalosäännön mukaan normaalijakauma-approksimaatio on käytettävissä. Merkitään X:llä satunnaismuuttujaa "lukumäärä valituille naiskirkolliskokousedustajille". Tehtävän tilanteessa odotusarvo $\mathsf{E}(X) = np = 48$ ja varianssi $\mathsf{D}^2(X) = npq = 96\times 0,5\times 0,5 = 24$. Näin ollen normaalijakauma-approksimaation mukaan todennäköisyys, että 96:n havainnon aineistossa nainen valittaisiin kirkolliskokousedustajaksi 34 tai pienemmässä lukumäärässä on

$$\begin{aligned}
\mathsf{P}(X \le 34) &= \mathsf{P}\left(\frac{X-48}{\sqrt{24}} \le \frac{34-48}{\sqrt{24}}\right) \\
&\approx \mathsf{P}\left(Z \le \frac{-14}{\sqrt{24}}\right) \\
&\approx \Phi(-2,858) \\
&\approx 0,002.
\end{aligned}$$

Yllä Z on standardinormaalijakautunut satunnaismuuttuja. Standardinormaalijakauman kertymäfunktion arvo $0,002$ on saatu kokeessa jaetusta normaalijakauman kertymäfunktiotaulukosta (kirjan s. 306). Todennäköisyys on noin kaksi tuhannesosaa.

Ylimääräistä pohdintaa:

Mikäli käytetään jatkuvuuskorjausta (kirjan s. 109 ja kurssin kotisivun lisämateriaali), tulos on oleellisesti sama eli noin kolme tuhannesosaa:

$$\Phi\left(\frac{34+\frac{1}{2}-48}{\sqrt{24}}\right) \approx \Phi(-2,756) \approx 0,003.$$

Kurssin kotisivulle linkitetty binomitodennäköisyyslaskuri antaa oleellisesti saman vastauksen:

$$\mathsf{P}(X \leq 34) = \sum_{k=0}^{34} \binom{96}{k} 0,50^k \times 0,50^{96-k} = 0,0028.$$

Tätä kaavaa saisi toki käyttää kokeessakin, mutta se vaatisi 35 termin suuruuden selvittämisen ja yhteenlaskemisen.

Tehtävän voi ratkaista yhtäpitävästi myös vertaamalla havaitun suhteellisen osuuden ($\hat{p} = 34/96 \approx 0,354$) ja tehtävässä oletetun arvon ($p = 0,5$) erotusta edellisen keskihajontaan ($\sqrt{pq/n} = \sqrt{0,5 \times 0,5/96} \approx \sqrt{0,0026}$):

$$\begin{aligned}
\mathsf{P}(\hat{p} \leq 0,354) &= \mathsf{P}\left(\frac{\hat{p}-0,5}{\sqrt{0,0026}} \leq \frac{0,354-0,5}{\sqrt{0,0026}}\right) \\
&\approx \mathsf{P}\left(Z \leq \frac{-0,146}{\sqrt{0,0026}}\right) \\
&\approx \Phi(-2,861) \\
&\approx 0,002.
\end{aligned}$$

Yllä ja ensin lasketut osamäärät eivät ole aivan samat, mikä johtuu konkreettisten laskujen numeerisesta epätarkkuudesta (ei ole laskettu riittävällä määrällä desimaaleja tms.). Laskut ovat kuitenkin yhtäpitäviä, sillä viimeisessä laskussa alkuperäisestä epäyhtälöstä

$$\frac{X-48}{\sqrt{24}} \leq \frac{34-48}{\sqrt{24}}$$

on saatu jälkimmäinen epäyhtälö

$$\frac{\hat{p}-0,5}{\sqrt{0,0026}} \leq \frac{0,354-0,5}{\sqrt{0,0026}}$$

jakamalla osamäärien osoittajat ja nimittäjät n:llä eli 96:lla ($\sqrt{24}/96 = \sqrt{24/96^2} \approx \sqrt{0,0026}$).

Mikäli kirkolliskokousedustajaksi valikoituisi naisia ja miehiä yhtä todennäköisesti, olisi aineisto erittäin poikkeuksellinen. Sen vuoksi on syytä epäillä oletuksen $p = 0,5$ paikkansapitävyyttä.c

3.

a) Kyse on otannasta ilman takaisinpanoa, joten käytetään hypergeometrisesti jakautuneen satunnaismuuttujan pistetodennäköisyyden kaavaa

$$\mathsf{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

(aputulos ja kirjan s. 102). Tehtävän tilanteessa merkinnät yllä tulkitaan seuraavasti: $N = 17$ on valiokunnan koko, $K = 4$ on Suomen Keskustan jäsenten lukumäärä valiokunnassa, k on Suomen Keskustan jäsenten lukumäärä poimitussa ryhmässä ja $n = 4$ on poimitun ryhmän koko.

Lasketaan todennäköisyys, että poimittuun ryhmään tulee 4 Suomen Keskustan jäsentä:

$$\begin{aligned}
\mathsf{P}(X = 4) &= \frac{\binom{4}{4}\binom{17-3}{4-4}}{\binom{17}{4}} \\
&= \frac{1 \times \binom{14}{0}}{\frac{17!}{4!13!}} \\
&= \frac{1 \times 1}{\frac{17 \times 16 \times 15 \times 14}{4!}} \\
&= \frac{1}{2380} \\
&\approx 0,0004.
\end{aligned}$$

Todennäköisyys on noin neljä kymmenestuhannesosaa.

Vaihtoehtoinen tapa laskea todennäköisyys on

$$\frac{4}{17} \times \frac{3}{16} \times \frac{2}{15} \times \frac{1}{14} = \frac{1}{2380} \approx 0,0004.$$

(On todennäköisyys ainolle mahdolliselle jonolle, jossa on neljä Suomen Keskustan jäsentä. Vrt. HT 4.2.)

Ylimääräistä kommentointia: Tapahtuma on hyvin epätodennäköinen, kun ryhmän jäsenet on poimittu sattumanvaraisesti. Ilmeisesti esitutkintaa tarpeettomana pitäneeseen ryhmään jäsenet eivät ole valikoituneet sattumanvaraisesti.

4. Merkintä-takaisinpyyntimenetelmällä arvioidaan populaation kokoa. Menetelmän ensimmäisessä vaiheessa hankitaan (yksinkertaisella satunnaisotannalla) K:n suuruinen otos N:n suuruisesta populaatiosta. Otoksen alkiot merkitään ja palautetaan populaatioon. Menetelmän toisessa vaiheessa otetaan uusi n:n suuruinen otos populaatiosta ja lasketaan merkittyjen alkioiden määrä (k) siinä. Intuitiivisesti merkittyjen alkioiden suhde populaatiossa tulisi olla sama kuin otoksessa:

$$\frac{K}{N} = \frac{k}{n}.$$

Yhtälöstä saadaan estimaatti populaation koolle:

$$N = \frac{nK}{k}.$$

(Kirjan s. 100.)