Johdatus tilastolliseen päättelyyn, 6. harjoitus (2.–4.5.2012)

1. Mendelin sääntöjen mukaan eräässä herneiden risteytyskokeessa saadaan tietyillä todennäköisyyksillä kasveja, joiden herneillä on tiettyjä ominaisuuksia. Kun tämä koe tehtiin, saatiin 200 kasvia, joiden herneiden ominaisuudet olivat seuraavat.

- Keltainen ja pyöreä: teorian mukainen todennäköisyys 9/16; havaittu lukumäärä 120.
- Keltainen ja ryppyinen: teorian mukainen todennäköisyys 3/16, havaittu lukumäärä 40.
- Vihreä ja pyöreä: teorian mukainen todennäköisyys 3/16; havaittu lukumäärä 35.
- Vihreä ja ryppyinen: teorian mukainen todennäköisyys 1/16; havaittu lukumäärä 5.

Testaa merkitsevyystasolla 0.05 pitävätkö teorian mukaiset todennäköisyydet paikkaansa. Suorita testi toisaalta laskemalla Pearsonin χ^2-testisuure ja toisaalta laskemalla uskottavuusosamäärän testisuure.

2. Britanniassa tutkittiin 1940-luvulla erilaisia tuberkuloosin hoitomuotoja. Eräässä kokeessa verrattiin kolmea erilaista lääkitystä: a) para-aminosalyyli (PAS), b) streptomysiini, c) yhdistelmähoitoa, jossa käytettiin sekä PAS-lääkettä että streptomysiiniä. Tuberkuloosin diagnosointiin käytettiin ensisijaisesti irtosolunäytettä. Jos irtosolunäyte oli negatiivinen, tehtiin vielä soluviljelmä. (Lääketieteen terminologiassa positiivinen testitulos tarkoittaa viitettä siitä, että potilaalla on tauti ja negatiivinen testitulos viitettä sitä, että potilaalla ei ole tautia.) *British Medical Research Council* julkisti v. 1950 seuraavat tulokset, joissa testin tulos + tarkoittaa sitä, että irtosolunäytteen tulos oli positiivinen; $(-,+)$ tarkoittaa sitä, että irtosolunäytteen tulos oli negatiivinen, ja soluviljelmän tulos positiivinen ja viimein $(-,-)$ sitä, että molempien testien tulos oli negatiivinen.

	Testin tulos			Yhteensä
Hoito	+	(-, +)	(-, -)	
PAS	56	30	13	99
Strepotomysiini	46	18	20	84
PAS ja streptomysiini	37	18	35	90
Yhteensä	139	66	68	273

Testaa merkitsevyystasolla 0.05 nollahypoteesia, jonka mukaan jokaisella hoitomuodolla näiden kolmen eri testituloksen jakauma on sama.

3. Perinteinen tapa muodostaa luottamusvälejä (tai suorittaa testejä) kahden riippumattoman binomijakautuneen populaation tapauksessa perustuu suuren otoskoon asymptoottisiin tuloksiin ja niistä saataviin z-luottamusväleihin (tai z-testeihin). Jos satunnaismuuttujat X_1 ja X_2 ovat riippumattomia, ja niillä on binomijakaumat

$$X_1 \sim \mathrm{Bin}(n_1, p_1), \qquad X_2 \sim \mathrm{Bin}(n_2, p_2),$$

jossa otoskoot n_1 ja n_2 ovat tunnettuja ja onnistumistodennäköisyydet p_1 ja p_2 tuntemattomia, niin tällöin onnistumistodennäköisyyksien erotusta $\tau = p_1 - p_2$ arvoidaan vastaavilla suhteellisten osuuksien erotuksella

$$\hat{\tau}(X_1, X_2) = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

Tämä estimaattori on harhaton, ja sen (otantajakauman) varianssi on helppo selvittää. Tämän jälkeen voimme käyttää jakson 4.7 kaavan (4.19) mukaista z-luottamusväliä. Anna konkreettiset kaavat sille, mitä tällöin pitää laskea, kun havaitut onnistumislukumäärät populaatiosta yksi ja kaksi ovat x_1 ja x_2.

4. Kaupasta hankittiin 4 samanlaista valkoista palloa, ja ne käytiin läpi yksitellen siten, että kukin pallon kohdalla heittiin harhatonta lanttia, ja pallo värjättiin mustaksi mikäli lantinheiton tulos oli kruuna. (Eri lantinheitot oletetaan toisistaan riippumattomiksi, ja kruunan todennäköisyys on puoli.) Lopuksi pallot asetettiin kulhoon. Tiedät, että kulhossa on neljä palloa ja tiedät kuinka valmistelut tehtiin. Sen sijaan et tiedä lantinheittojen tuloksia, joten et myöskään tiedä kulhossa olevien valkoisten pallojen lukumäärää θ, joka on mallin parametri.

Nostat kulhosta pallon kolme kertaa silmät sidottuna. Kunkin noston jälkeen nostettu pallo palautetaan kulhoon. Ennen kutakin nostoa kulhoa ravistellaan perusteellisesti. Nostettujen pallojen värit ovat musta, musta ja valkoinen.

a) Esitä valmisteluja vastaavat prioritodennäköisyydet (joko kaavalla tai luettelemalla arvot).

b) Esitä uskottavuusfunktio (joko kaavalla tai luettelemalla arvot).

c) Laske posteriorijakauma (luettele arvot).

5. Nastaa ravistettiin purkissa 100 kertaa, ja se päätyi 42 ravistuksen jälkeen selälleen ja muissa tapauksissa kyljelleen. Tahdomme estimoida parametria θ, joka on todennäköisyys että nasta päätyy selälleen yhden ravistuksen jälkeen. Määritä posteriorijakauma, kun priorijakauma on tasajakauma. Määritä myös posteriorijakauman odotusarvo ja varianssi.