Johdatus tilastolliseen päättelyyn, 6. harjoitus (Viikko 18: 30.4.–3.5.2013)

1. Herneiden risteytyskokeessa saadaan kasveja, joiden herneillä on tiettyjä ominaisuuksia. Kyseessä olevien luokkien todennäköisyydet voidaan laskea Mendelin säännöillä. 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 Mendelin teorian mukaiset todennäköisyydet paikkaansa.

2. Britanniassa tutkittiin 1940-luvulla erilaisia tuberkuloosin hoitomuotoja. Eräässä kokeessa verrattiin kolmea erilaista lääkitystä, jotka olivat a) para-aminosalyyli (PAS), b) streptomysiini, c) yhdistelmähoito, jossa käytettiin sekä PAS-lääkettä että streptomysiiniä. Tuberkuloosin diagnosointiin käytettiin ensisijaisesti irtosolunäytettä. Jos irtosolunäyte oli negatiivinen, tehtiin vielä soluviljelmä. (Lääketieteessä positiivinen testitulos tarkoittaa viitettä siitä, että potilaalla on tauti ja negatiivinen testitulos viitettä siitä, 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 5.7 kaavan (5.19) mukaista z-luottamusväliä. Olkoot havaitut onnistumislukumäärät populaatiossa yksi x_1 ja populaatiossa kaksi x_2. Anna kaavat, joilla saadaan laskettua 95 %:n luottamusväli onnistumistodennäköisyyksien erotukselle τ.

4. V. 2012 JTP-kurssin kyselytutkimuksesta saadaan miespuolisten opiskelijoiden pituudelle x (cm) ja painolle y (kg) seuraavat yhteenvedot (otosvariansseissa ja otoskovarianssissa jakajana on $n-1$).

$$n = 37, \quad \bar{x} = 180.1, \quad \bar{y} = 77.07$$
$$s(\mathbf{x}, \mathbf{x}) = 56.52, \quad s(\mathbf{x}, \mathbf{y}) = 46.37, \quad s(\mathbf{y}, \mathbf{y}) = 184.56$$

Mallinnamme aineiston siten, että se vastaa satunnaisotosta kaksiulotteisesta normaalijakaumasta, joten lineaarisen regression kaavoja voidaan soveltaa.

a) Laske lineaarisen mallin kertoimien SU-estimaatit.

b) Laske virhevarianssin estimaatti s^2.

c) Laske kulmakertoimelle 95 %:n luottamusväli.

5. Oletaan, että lineaarisen mallin oletukset (9.9) pitävät paikkansa. Johda kulmakertoimen estimaattorille $b(\mathbf{Y})$ sen

a) odotusarvo,

b) varianssi.

Käytä hyväksesi jaksossa 3.3 kerrattuja todennäköisyyslaskennan tietoja.

6. 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).