Frekvenssiaineistojen analyysi
Harjoitukset 4.10.2001

3.1. Oletetaan, että X ja Y ovat toisistaan riippumattomia Poissonin jakaumaa parametrein μ_1 [$X \sim \mathrm{Pois}(\mu_1)$] ja μ_2 [$Y \sim \mathrm{Pois}(\mu_2)$] noudattavia satunnaismuuttujia. Osoita, että X:n ehdollinen jakauma ehdolla $X + Y = n$ noudattaa binomijakaumaa (n on annettu positiiviluku).

3.2. Doll ja Hill (1952) saivat Taulukon 3.16 mukaiset tulokset tutkiessaan keuhkosyöpäpotilaiden tupakoimista useiden eri kaupunkien sairaaloissa Englannissa. Taulukossa vertaillaan keuhkosyöpäpotilaita kontrollipotilaisiin, joilla ei ole keuhkosyöpää. Tupakointikäyttäytymistä vertailtiin yli 10 vuoden ajanjaksolta ennen taudin puhkeamista.

Taulukko 3.16 Tupakointi ja keuhkosyöpä

	Ryhmä	
Päivittäin poltettujen savukkeiden lukumäärä (keskimäärin)	Keuhkosyöpäpotilaat	Kontrollipotilaat
– 4	7	61
5–14	55	129
15–24	489	570
25–49	475	431
50–	38	12

(a) Käsittele tupakointia järjestysasteikollisena muuttujana kahdella tavalla (vrt. luentoesimerkki). Testaa tupakoinnin ja sairastamisen välinen lineaarinen riippuvuus.

(b) Käsittele tupakointia välimatka-asteikollisena muuttujana. Käytä mittalukuina luokkakeskipisteitä. Testaa tupakoinnin ja sairastamisen välinen lineaarinen riippuvuus.

3.3. Taulukossa 3.17 on Ladislaus von Bortkiewiczin aineisto, jossa on kirjattu hevosenpotkuun kuolleiden lukumäärät/armeijakunta Preussin armeijassa. Aineisto koskee 10:ntä armeijakuntaa, joista jokaista on havainnoitu 20 vuotta. Kaikkiaan altistumisvuosia on siis 200. Taulukon mukaan yhtään kuolemantapausta ei sattunut 109 vuotena ja yli neljää tapausta ei ole sattunut minään vuonna.

Taulukko 3.17 Hevosenpotkuun kuolleiden lukumäärät/armeijakunta Preussin armeijassa.

Kuolleitten lkm	Vuosien lkm
0	109
1	65
2	22
3	3
4	1
ainakin 5	0

(a) Estimoi keskimääräinen hevosen potkuun kuolleiden lkm/vuosi.

(b) Testaa, noudattaako lukumäärä Poissonin jakaumaa.

3.4. Tarkastellaan taulukkoa

n_{11}	n_{12}	n_{13}
n_{21}	n_{22}	n_{23}

ja sen osataulukkoja

		Yhteensä
n_{11}	n_{12}	$n_{1+}^{(1)}$
n_{21}	n_{22}	$n_{2+}^{(1)}$

ja

$n_{1+}^{(1)}$	n_{13}
$n_{2+}^{(1)}$	n_{23}

Osoita, että 2×3–taulukosta laskettu Wilksin testisuure G^2 on yllä esitetyistä 2×2–osataulukoista laskettujen vastaavien Wilksin testisuureiden summa.

3.5. Analysoi ja tulkitse Taulukon 3.18 opiskelijoiden tupakointikäyttäytymista koskeva aineisto.

Taulukko 3.18 Opiskelijoiden tupakointi vanhempien tupakoinnin mukaan.

	Opiskelija polttaa	Opiskelja ei-polta
Molemmat vanhemmat polttavat	400	1380
Toinen vanhemmista polttaa	416	1823
Kumpikaan vanhemmista ei polta	188	1168

3.6. Taulukko 3.19 on saatu Yhdysvaltain vuoden 1991 yleisen väestötiedustelun tuloksista (Agresti 1996, s. 31).

Taulukko 3.19 Puoluekanta ja rotu. Otos Yhdysvaltain vuoden 1991 yleisen väestötiedustelun tuloksista.

	Puoluekanta		
Rotu	Demokraatti	Riippumaton	Republikaani
Valkoinen	341	105	405
Musta	103	15	11

(a) Testaa puoluekannan ja rodun riippumattomuus. Tulkitse.

(b) Käytä suhteutettuja residuaaleja tuloksen kuvailuun.

(c) Osita χ^2–testisuure kahteen komponenttiin ja käytä hajotelmaa tuloksen kuvailuun.