

# TILA311 Harjoitustyö 2

Alexi Pekkala (alvianpe@student.jyu.fi)

26.10.2017

## 1 Aineiston kuvaus

Tarkasteltava aineisto käsittelee beetakaroteenin ja muiden taustavaikuttajien suhdetta ihosyövän uusiutumistodennäköisyyteen. Raportissa tarkastelomme laajemmasta tutkimusaineistosta (Greenberg ym., 1990) koostettua poikkileikkausaineistoa, joka sisältää  $n = 1683$  riskiryhmään kuuluvan henkilön seuraavat tiedot:

**id** tutkimushenkilön id

**age** tutkittavan ikä

**skin** 0 = normaali iho, 1 = helposti palava iho

**gender** 0 = nainen, 1 = mies

**exposure** aikaisempien ihosyöpien lukumäärä

**cancer** 1 = jos potilaalla havaitaan uusi ihosyöpä, 0 muuten

**trt** 0 = plasebo, 1 = 50mg beetakaroteenia / päivä

Aineiston 1683 henkilöstä naisten osuus on 31%, havaittujen ihosyöpien osuus 18% ja helposti palavan ihon osuus 46%. 51%:lle henkilöistä on annosteltu plasebo ja lopuille 50mg beetakaroteenia. Aineiston keski-ikä on 63 vuotta ja aikaisempien ihosyöpähavaintojen keskiarvo on 2,9.

## 2 Menetelmät

Tavoitteena on muodostaa ihosyövän esiintymistä selittävä malli. Koska vastemuuttujia *cancer* on dikotominen, sovelletaan logistista regressiota. Aluksi poistettiin aineistosta *id*-muuttujan ja korvasin *age*- sekä *exposure*-muuttujat keskitetyillä *cage*- ja *ceposure*-muuttujilla. Formuloin alustavan mallin käyttämällä kaikkia muuttujia selittäjinä. Selittäjistä ei-merkitseviksi osoittautuivat *cage* ( $p = 0,16$ ) sekä *trt* ( $p = 0,11$ ). Lisäksi tarkastelin eri interaktioita, joista ainoana merkitsevä oli *cage* · *ceposure*. Merkitsevyydeltään vähäisten *cage*- ja *trt*-selittäjien jättäminen malliin alentaa devianssia 4,7 yksiköllä mutta ero ei  $\chi^2$ -testin mukaan ole merkitsevä ( $p = 0,095$ ). Lisäksi *trt*-selittäjän vaikutus on vasten odotuksia positiivinen (sairastavuutta lisäävä), joten jätin selittäjät pois mallista. Jäljelle jäävistä muuttujista on muodostettu kaava 1, missä  $\pi_C$  merkitsee uuden ihosyövän havaitsemisen todennäköisyyttä.

$$\text{logit}(\pi_C) = \beta_0 + \beta_1 \text{ceposure} + \beta_2 \text{skin} + \beta_3 \text{gender} + \beta_4 \text{cage} \cdot \text{ceposure} \quad (1)$$

## 3 Tulokset

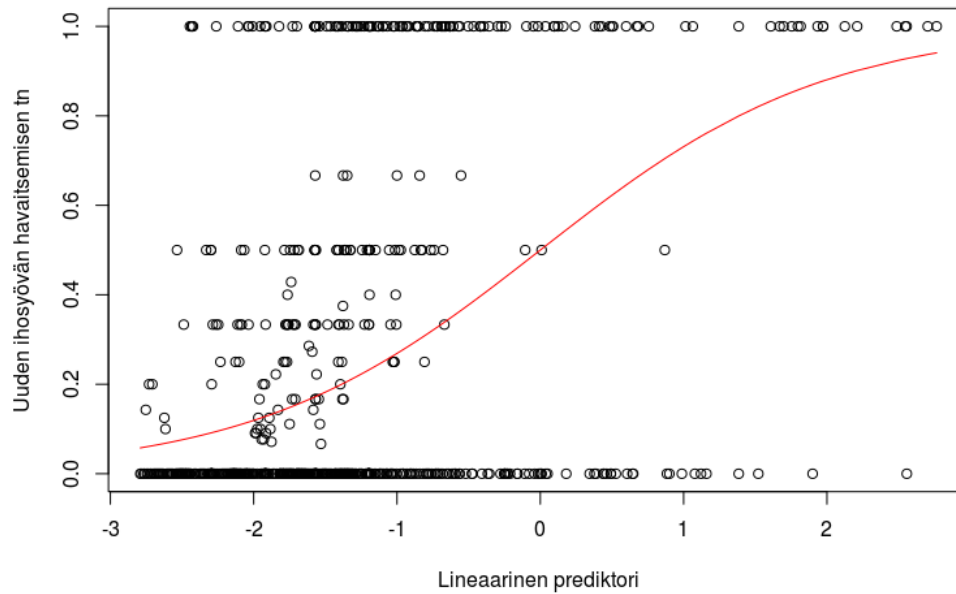
Mallin kertoimien estimaatit on koottu taulukkoon 1. Mallin mukaan suurin sairastavuutta lisäävä vaikutus on sukupuolella mies. Kun ikä ja aikaisempien ihosyöpähavaintojen lukumäärä ovat aineiston keskiarvon mukaisia, on miesten ja naisten välinen vedonlyöntisuhde  $e^{0,72}$ , eli miesten todennäköisyys ihosyövän ilmentymään on 2,1-kertainen naisiin verrattuna. Samoin ennakkoletuksin naisten vedonlyöntisuhde on  $e^{-2,31}$ , eli naisilla todennäköisyys ihosyövän ilmentymään on 10% suhteessa siihen, että syöpää ei ilmene. Helposti palava iho lisää odotetusti ihosyövän todennäköisyyttä: normaaliin ihoon verrattuna syövän havaitsemisen todennäköisyys on 1,46-kertainen kun muut selittäjät ovat samanarvoisia. Lisäksi verrattaessa ryhmiä joissa aikaisempien ihosyöpien määrä eroaa yhdellä, on korkeamman ryhmän todennäköisyys 1,19-kertainen kun muut selittäjät ovat samanarvoisia ja ikä aineiston keski-ikä mukainen. Henkilön ikä hieman vahvistaa edellistä vaikutusta, mikä ilmenee selittäjän *cage* · *ceposure* kertoimesta.

Kuvassa 1 on piirretty sovitetun mallin estimaatit sekä ennusteet lineaarista prediktorista vasten. Kuva 2 esittää estimaatit sekä ennusteet *exposure*-

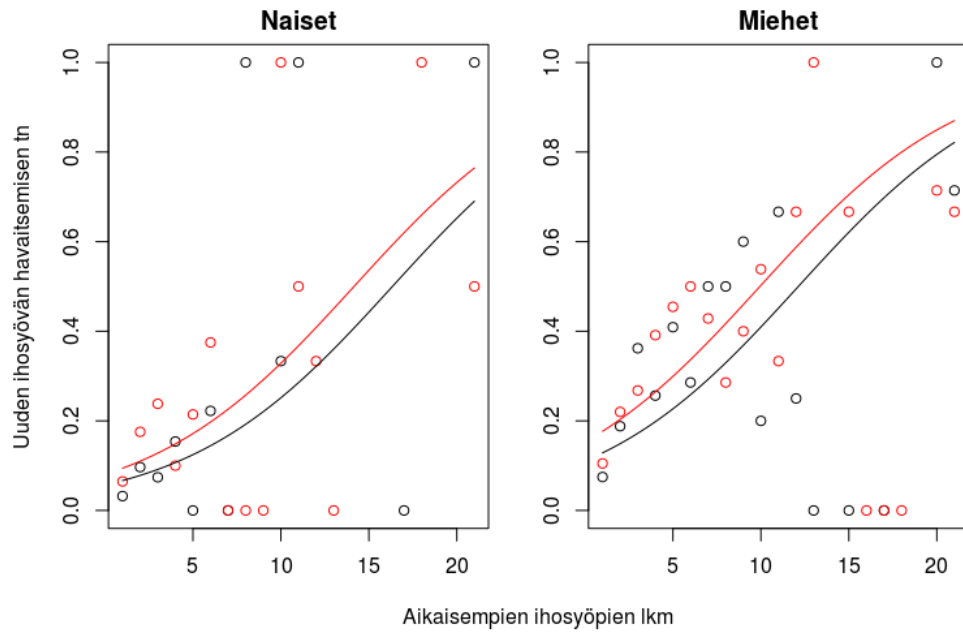
	estimaatti	keskivirhe	z-arvo	p-arvo
vakiotermin	-2.31	0.16	-14.2	$< 0,001$
cexposure	0.17	0.03	9.54	$< 0,001$
skin	0.38	0.14	2.78	$< 0,01$
gender	0.72	0.16	4.42	$< 0,001$
cage:cexposure	0.004	0.002	1.82	$< 0,1$

Taulukko 1: Ennustemallin muuttujat

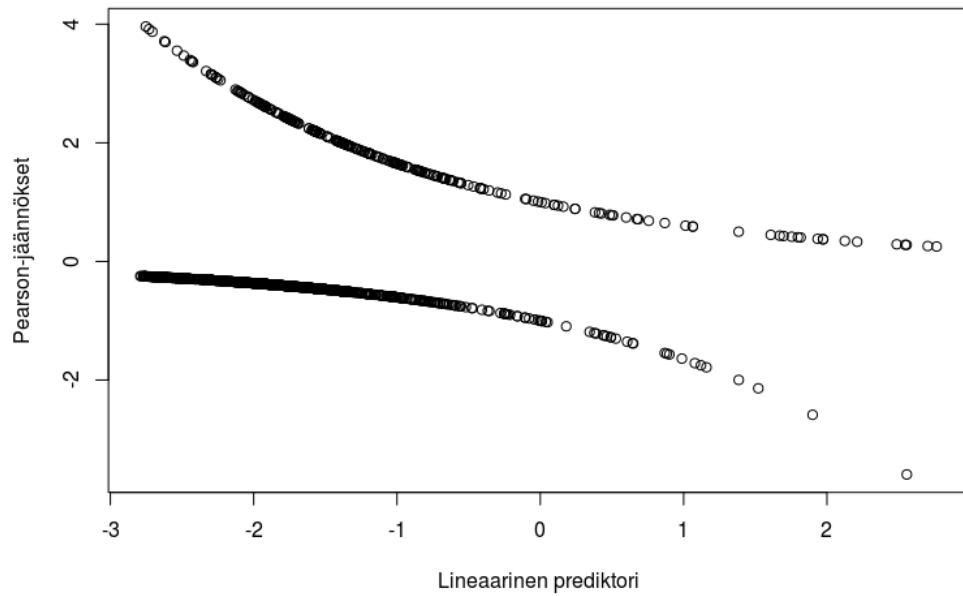
muuttujaa vasten sukupuolen sekä ihotyyppin mukaan ryhmitellen, olettaen että  $age$  = aineiston keski-ikä. Kuva 3 esittää mallin Pearson-jäännökset sovitteen suhteen.



Kuva 1: Mallin suhteelliset frekvenssit sekä ennustetut todennäköisyydet lineaarisen prediktorin suhteen



Kuva 2: Mallin suhteelliset frekvenssit sekä ennustetut todennäköisyydet *exposure*-muuttujan suhteen ryhmittäin. Musta vastaa normaalia ihoa ja punainen helposti palavaa ihoa.



Kuva 3: Mallin Pearson-jäännökset sovitteen suhteen

## 4 Johtopäätökset

Aineiston perusteella päivittäisen beetakaroteeniannoksen ei voida sanoa laskevan ihosyövän uusiutumisriskiä. Vaikuttavia tekijöitä ovat sen sijaan sukupuoli, ihon palamisherkyys sekä aikaisempien ihosyöpien lukumäärä. Näistä jälkimmäisellä havaittiin lisäksi yhteisvaikutus henkilön iän kanssa, mutta ikä itsessään ei näytä vaikuttavan sairastuvuuteen.

Toinen mahdollinen tulkinta on se, että beetakaroteenin vaikutus perustuu syöpäriskin ehkäisemiseen vain niiden henkilöiden keskuudessa, joilla ei ole aikaisempia tartuntoja. Koska käsiteltävä aineisto koostuu yksinomaan korkean riskiryhmän henkilöistä, tällainen vaikutus ei pääse ilmenemään muodostetussa mallissa. Aineisto ei täten täysin kiistä beetakaroteenin vaikuttavuutta ihosyöpäriskin suhteen.