



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# Poistuman tutkiminen logistisella regressiolla (aihe-esittely)

*Lauri Suoknuuti*

*12.06.2020*

Ohjaaja: FM *Markus Linnakaari*

Valvoja: Apul.prof. *Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

# Tausta

## Asiakkaiden poistuma engl. *customer churn*

- Asiakkaan toimesta tapahtuva asioinnin lopettaminen yrityksen kanssa
- Uusien asiakkaiden hankkiminen tyypillisesti kalliimpaa kuin vanhojen asiakkaiden pitäminen
- Asiakaspoistuman ennustaminen laajalti tutkittu ongelma
- Voidaan määritellä binäärisenä luokitteluongelmana

# Tavoitteet

- Tutustutaan asiakaspoistumaan ilmiönä
- Tutustutaan logistiseen regressioon menetelmänä
- Rakennetaan logistiseen regressioon pohjautuva ennustemalli
- Tutkitaan rakennetun mallin ominaisuuksia
- Käytetään mallia asiakaspoistuman ennustamiseen

# Aineisto

- Suomalaisen vakuutusyhtiön autoliikkeille tarjottavien kampanjatarjousten asiakasdataa
- Datassa tietoja vakuutetusta ajoneuvosta ja vakuutuksenottajasta
- Datassa suhteellisen suuri poistuman todennäköisyys
  - Helpottaa mallintamista

# Menetelmät

## Logistinen regressio

- Regressioanalyysin tavoitteena on selvittää vasteen ja selittävien muuttujien välinen riippuvuus
- Logistinen regressio kuuluu yleistettyihin lineaarisiin malleihin
- Logistisessa regressiomallissa vastemuuttuja on tyypillisesti binäärinen
- Malli antaa vastemuuttujan odotusarvon, joka tulkitaan todennäköisyytenä sille, että vastemuuttuja saa arvon 1

# Menetelmät

## Logistinen regressio

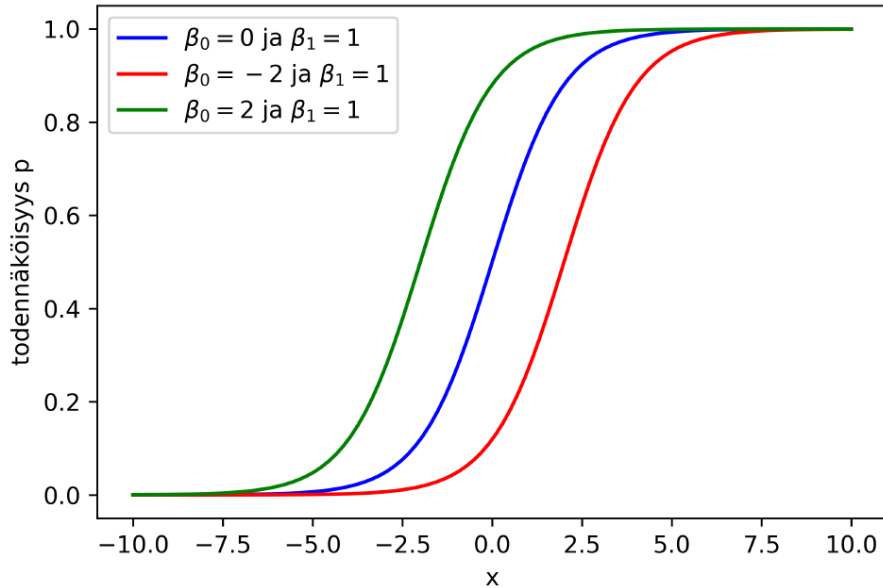
$$\pi(\mathbf{x}_i) = \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ij})} = \frac{1}{1 + \exp(-(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))} \quad 1)$$

$$\pi(\mathbf{x}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}} \quad 2)$$

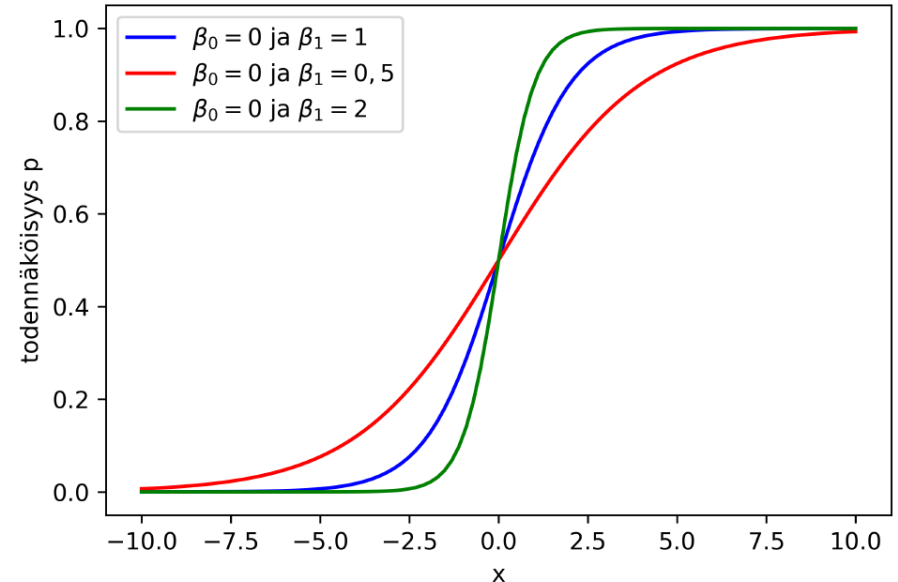
- 1) Malli, jossa on p selittävää muuttujaa
- 2) Malli, jossa on yksi selittävä muuttuja

# Menetelmät

## Logistinen regressio



Kuva 1: Logistinen funktio kuvattuna parametrin  $\beta_0$  eri arvoilla.



Kuva 2: Logistinen funktio kuvattuna parametrin  $\beta_1$  eri arvoilla.

# Aikataulu

Toukokuu

Aiheeseen  
perehtyminen

Kirjoituksen  
aloittaminen

Kesäkuu

Kirjoituksen  
jatkaminen

Aihe-esittely

Mallin  
rakentaminen

Heinäkuu

Mallin  
rakentaminen

Tulosten  
valmistuminen

Kirjoituksen  
viimeistely

Elokuun alku

Kirjoituksen  
viimeistely

Valmiin työn esittely