Evaluating face-to-face fundraisers

Albert Wendsjö

Bachelor Thesis in Statistics (15 ECTS)

Supervisor: Jonas Wallin

Lund University
School of Economics and Management
Department of Statistics
Autumn 2019

Abstract

Face-to-face fundraising is a method used for raising money by having fundraisers working on public spaces. Fundraisers either ask for one time donations, or sign up people to give on a monthly basis, in this thesis the focus is on the latter. For this method to be useful all fundraisers must meet their target goal, meaning that they must get a certain amount of sign ups when they work. By using data on fundraiser results for Sweden for UNHCR we study how fundraisers develop. Using a Poisson regression model with mixed effects we examine whether fundraisers become better over time, whether the results differ between cities and if helpful predictions can be made. On a population level it is found that fundraisers become 0.9% better for each week worked. This differs for all fundraisers, and at the extremes some fundraisers become $\mp$7% better/worse over time. It is also found that the results in most cities are rather equal, with the exception that the results are 30% higher in Stockholm. Lastly we used the model for predicting some results using the data of the first 5 weeks of two fundraisers. In both cases the prediction intervals are too broad to determine whether the fundraisers will reach the goal or not.

Keywords: fundraising, poisson regression, mixed effects, GLMM, bootstrap.

Acknowledgments

I wish to thank Sweden for UNHCR for giving me the oppertunity to write this thesis. I'm especially grateful for Samuel taking the time to help me out with the data and answering my questions regarding Sweden for UNHCR.

Contents

1 Introduction

Face-to-face fundraising is a method for raising money that is often used by charitable organizations. It is the act of asking the public for contributions by stopping people on public spaces such as busy streets or shopping malls. Many programs are run so that people sign up to contribute on a monthly basis, making this an effective strategy to secure funding the long run. Most often this work is performed by paid employees or contractors, and some estimates say that this method of raising money gives a 1:3 turnout (Nguyen 2019).

To what degree this method is successful depends fully on the performance of each individual fundraiser. Because of this it is important that each hired fundraiser get the signups needed. When hiring new fundraisers there is always a risk that the new fundraiser won't reach the target goal, there is no certainty until they started working. But even when the fundraisers started working practically interpreting the results of fundraisers turn out to be a difficult task. Most often the results are very volatile, maning that the results vary from week to week so that it is difficult to distinguish the trend from the deviations. It might also be difficult to evaluate new fundraisers since it is believed to be a learning curve in the job. So even if the results initially are low it might be a good investment to keep the fundraiser who then gets a chance to evolve.

From a statistical perspective the central problem lies in the relationship between the time a fundraiser have worked and the how well they perform in terms of signups. Modelling this relationship is complicated mainly due to two reasons. The first reason is that the relationship is on an individual level, not on a group level. The second reason is that the random variable signups is a count data. These two issues will be further discussed in section 3 where a Poisson regression model with mixed effects will be proposed to approach this problem. This is a model for count data that can capture individual variation, it falls under the broader term Generalized Linear Mixed Models (GLMM).

This thesis is written in collaboration with Sweden for UNHCR[1]. They are the organisation responsible for collecting money in Sweden for UNHCR and have been running a face-to-face program since 2011. The program is active in various different cities, but the most consistent ones are located in Stockholm, Gothenburg and Malmö.

Face-to-face fundraising on the streets is one of many methods that Sweden for UNHCR use for raising money, other methods include the usage of ads and recruiting over the phone. Last year they raised 284 million SEK, in which 83% of that money were sent directly to UNHCR.

The current goal of Sweden for UNHCR is that each fundraiser should get an weekly average of 0.33 signups per hour fundraised. If the lasting trend is that

[1] UNHCR (United Nations High Comissioner for Refugees) is a United Nations program working to assist refugees. They have the international mandate to aid displaced people all over the world, and they work to secure everyone's right to seek asylum and safe refuge. When an acute crisis emerges they provide assistance in form of clean water, sanitation, health care and shelter. They also provide support for people who wishes to return home or resettle. (UNHCR 2019)

the fundraiser doesn't reach this goal they don't get extended contracts. But bearing in mind that the results are volatile and that fundraisers may evolve, the results are interpreted with caution. Currently all coordinators make an qualitative assessment of each fundraiser trying to figure out who has the potential to reach the goal or not. They especially look for individual circumstances such as temporary stress or personal issues that might explain deviance from the target goal.

Objectives. *The objective of this study is to explore how fundraisers evolve, in particular the difficulties that come with evaluating new fundraisers.*

(i) Is there a learning curve for fundraisers?

(ii) What variables affect the results of fundraisers?

(iii) To what extent are predictions possible?

Answering these questions can hopefully provide coordinators with a new perspective in how fundraiser results can be interpreted. Whether or not fundraiser results increase substantially over time is crucial when it comes to deciding if newly hired fundraisers should be kept. Whether or not circumstances such as city or season have an effect on the results could also improve the decision making.

2 Data

In this section the data material is described. The data is provided by Sweden for UNHCR, and contains all results since 2015. Firstly the variables available are presented in Table 1.

Table 1: Variables in the data.

Variable	*Description*	*Values*
Signups	Amount of signups	$1, 2, ...$
Recruitment Hours	Amount of hours spent fundraising that week	$[0, 40]$
City	The city that the fundraiser worked in that week	1 = Gothenburg 2 = Stockholm 3 = Malmö 4 = Uppsala 5 = Linköping-Norrköping 6 = Borås 7 = Jönköping 8 = Västerås 9 = Karlstad 10 = Trollhättan 11 = Helsingborg 12 = Falun-Borlänge 13 = Halmstad 14 = Eskilstuna
Week	Week of the year	$1, 2, ..., 52$
Weeks worked	Amount of weeks worked	$1, 2, ...$
ID	Identification number for each fundraiser	

There are mainly two issues with the data, whereas one can be dealt with. The first issue is that the data set does not cover all years the program has been active, because of this it is not possible to follow the fundraisers that started before 2015 from their first working week. Since the initial progress is in focus in this thesis, and we do not really know how long these fundraisers have worked it is decided to drop those fundraisers from the data. This is relatively easily done since it is known that all fundraisers with a ID number higher than 511 started working during or after 2015. The second issue with the data is that there might be a potential bias due to fundraisers not getting extended contracts when results are low. This has the implication that it is not possible to follow how the fundraisers that start of the lowest develop over longer periods. This can not be dealt with, but should be kept in mind when interpreting the results later on.

2.1 Descriptive statistics

In this section some descriptive statistics regarding the data is presented. Note that this is without the complete data set, still with the fundraisers that started before 2015.

Firstly in Table 2 the data available is described divided by city and year.

Table 2: Amount of observations divided by city and year.

Year / *City*	*2015*	*2016*	*2017*	*2018*	*2019*	*Total*
Gothenburg	418	266	244	213	199	1340
Stockholm	400	384	651	864	562	2861
Malmö	119	176	324	280	321	1220
Uppsala	54	91	184	27		356
Linköping-Norrköping	47			82		129
Borås						0
Jönköping				34		34
Västerås	67	37				104
Karlstad	79		106	23		208
Trollhättan		7	2			9
Helsingborg		38	39			77
Falun-Borlänge			95			95
Halmstad		41				41
Eskilstuna		27				27
Total	1184	1067	1645	1523	1082	6501

As seen in Table 2 there are totally 6501 observations, one observation is the result of one fundraiser on a particular week. Totally there are 796 different fundraisers in the data, implying that the fundraisers work 8.2 weeks in average.

To further study the duration that fundraisers work in the organisation we can take the maximum value of the variable weeks worked for each each individual. This lets us see how long the fundraisers stay. In Figure 1 the duration is visualized in a histogram.

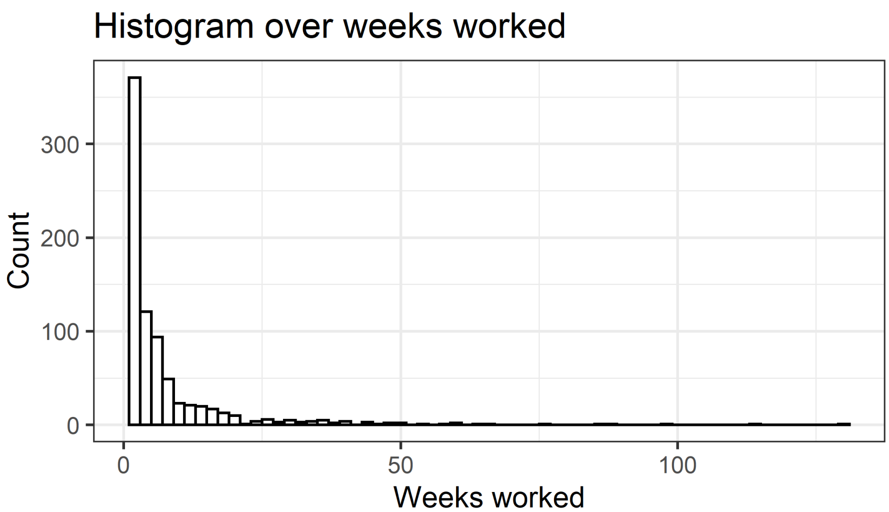


Figure 1: Histogram over the amount of weeks that fundraisers work.

We can here see that most fundraisers work less than 12 weeks. Further description is given below.

Table 3: Summary statistics for the duration that fundraisers work.

Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
1	3	4	8.2	8	130

From the summary statistics we can conclude that 75% of the fundraisers work less than 8 weeks, the mean of 8.2 is due to some fundraisers working really long in comparison to the others.

To study how the results differ among cities a boxplot is presented in Figure 2 where signups per hour is presented for each city. For every city 50% of the observations lies within the blue box, and the black line in the box is the median. The two lines going out from the boxes are called whiskers, and their length are equal to 1.5 times the value at the top of the box minus the value at the bottom of the box. Values that are lower or higher than the whiskers are called outliers, and those values are represented with a dot. Each dot therefore represent the result of a fundraiser on a certain week, when the results were unusually high (or low). The dashed line represent the 0.33 threshold which fundraisers must meet.

Since there are not that many observations for some cities it does not really make sense analyzing them in to detail. Instead it is decided to analyze the five cities with the most signups, treating the other cities as a group.

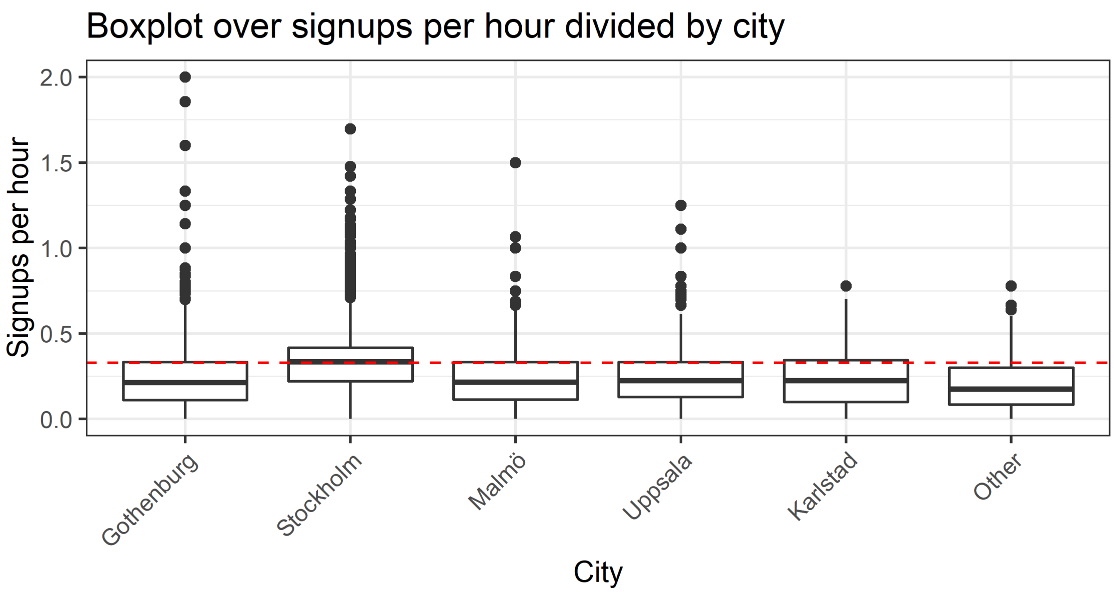


Figure 2: On the Y axis we see signups per hour, and each boxplot is calculated using all values of signups per hour for that particular city.

Worth to be noting here is that Stockholm seems to perform better than the other cities and that the other cities are rather equal.

In Figure 3 the mean hours worked throughout the year is shown. It is here clarified that the fundraisers work more hours over the summer period, approximately from week 20 to 40.

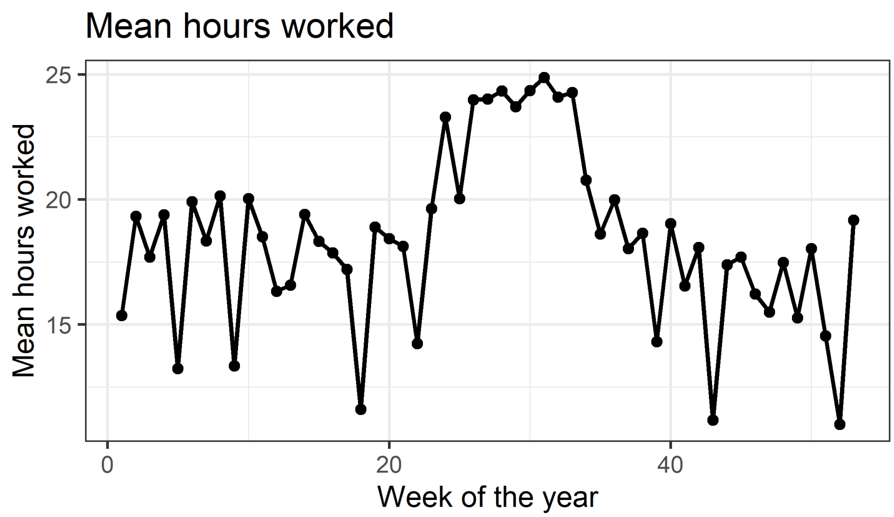


Figure 3: On the X axis is the mean of the hours worked and on the Y axis is each week of the year.

In Figure 4 the mean signups per hour over the year is presented. This plot indicates a positive trend from week 20 up to week 40, and some other patterns that might indicate a seasonal relationship. However in Figure 5 a similar plot is presented but divided by year. Now the results seem mostly constant throughout the year, ranging from 0.2 to 0.4 signups per hour - with some random deviation from year to year. The exception is around week 33-40 where the result seem to be more volatile, and for most years the results are unusually high.

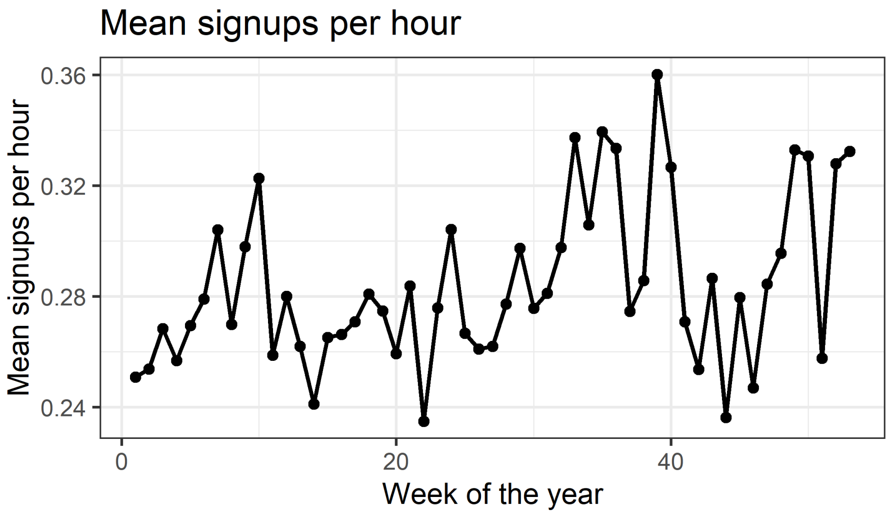


Figure 4: Signups per hour is represented on the Y axis, and each week of the year on the X axis. Each dot on the plot represent the mean of all results on that particular week.

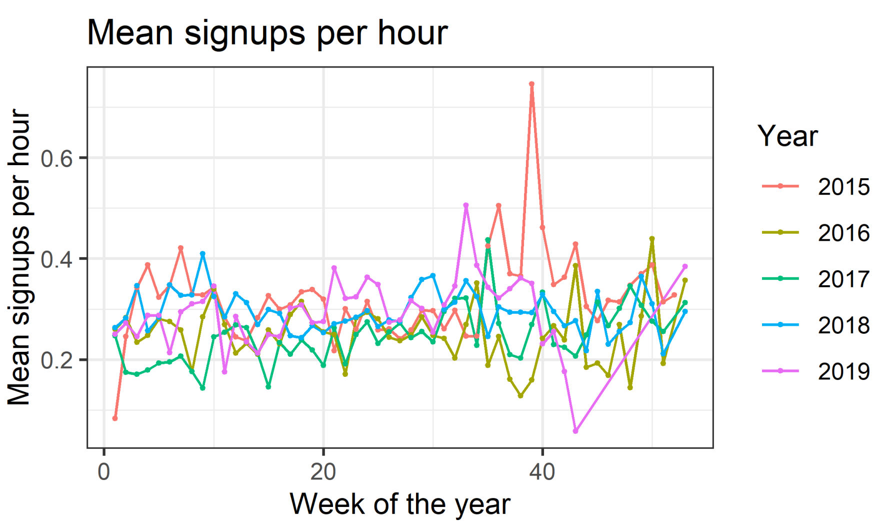


Figure 5: This plot is similar to 4, but now we calculate the mean of the results for each year, and plot each year with a separate line.

3 Methodology

In this section the methods used in this thesis will be described. First the reader is provided with some theoretical background about the underlying concepts before the model used is presented. Following this the model assumptions and some tools for examining these are discussed.

The calculations are done in the statistical software R version 3.4.4 (R Core Team 2019). Main packages used for analysis will be presented throughout the text.

3.1 Theoretical background

3.1.1 Linear Mixed Models

When examining the linear relationship between a dependent and independent variable, regression analysis a popular method. It can provide insight in how strong a relationship between two variables are, and can be used for predictions. The model can be written as (Sheather 2009)

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (1)$$
$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2),$$

here β_0 and β_1 is the intercept and the slope respectively, and ε is normally distributed random deviation from the expected value. A regression analysis approach for analyzing fundraisers could explain signups per hour using weeks worked as an explanatory variable.

Figure 6: Signups per hour for each individual is plotted on the Y axis, and weeks worked for each result on the X axis. A regression model is adapted to the data, with weeks worked explaining signups per hour.

This model however is not sufficient. It does not recognize the dependency in all observations, namely that they are the result of individual developments. It instead treats all observations as a big group. The parameter estimations of this model will likely be misleading, and all predictions for individual fundraisers will be the same no matter how the individual fundraiser perform - and this will likely be misleading.

Another solution is to perform a regression analysis for each individual separately, but then all other results will be ignored. Since there aren't that many observations per individual these models would likely be insignificant.

The solution to these problems is to introduce the concept of random effects. Models with random effects are widely used in psychological studies where the goal is to draw conclusion on individuals, or the group as a whole with individual variation taken into consideration. A model with only random effects can be written as (Bauer 2011)

$$Y_{ij} = v_{0i} + v_{1i}x_{ij} + \varepsilon_{ij}. \tag{2}$$

In the case of fundraisers Y_{ij} would represent signups per hour for individual i after working j weeks and x_{ij} are the amount of weeks worked for individual i, after j weeks. v_0 and v_{1i} are the individual intercepts and slopes. The random effect model basically uses a categorical variable with a level for each individual, allowing for different estimations on each level (Bates 2010).

This model can highlight how the relationship between weeks worked and signups per hour for all fundraisers on an individual level. But often it is of interest to also estimate what the common characteristics of the population is, then the model above can be expanded to include a fixed effect

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + v_{0i} + v_{1i}x_{ij} + \varepsilon_{ij}. \tag{3}$$

This model as both the the fixed effects that are common population, β_0 and β_1, as well as the individual random effects v_{0i} and v_{1i}. These types of models are often referred to as mixed effects models since they have both fixed and random effects.

In general terms one could define random and fixed effects this way. Fixed effects are levels which we chose before the study. This has the implication that we only can make statements about those levels. Random effects on the other hands have levels which we don't choose on beforehand. They are rather seen as sample of all the possible levels; from this sample we wish to draw conclusions about the population of levels. Montgomery (2013)[2]

For fundraisers this could mean that for example cities could be a fixed effect, we can only draw conclusions about the cities we have data on and results can't

[2]But as the Swedish saying goes "Kärt barn har många namn", it should be noted that many of these concepts go under different names in different disciplines and contexts. For example mixed effects models might sometimes be referred to as multilevel models, hierarchical linear models or random coefficient models. The same goes for the term random effect, which may have different meaning or denominations in different literature. (McElreath 2016)

be generalized to other cities. But the effect for each fundraiser is regarded as random.

As an initial analysis a model based on this method is applied to the data. Using the lme4 package by Bates et al. (2015) in R a linear mixed effects model is applied, using signups per hour as dependent variable and worked weeks as independent variable, with a fixed intercept and slope aswell as random intercepts and slopes for each individual. In Figure 7 a plot is provided with a sample of 10 individuals, it shows the intuition of the model.

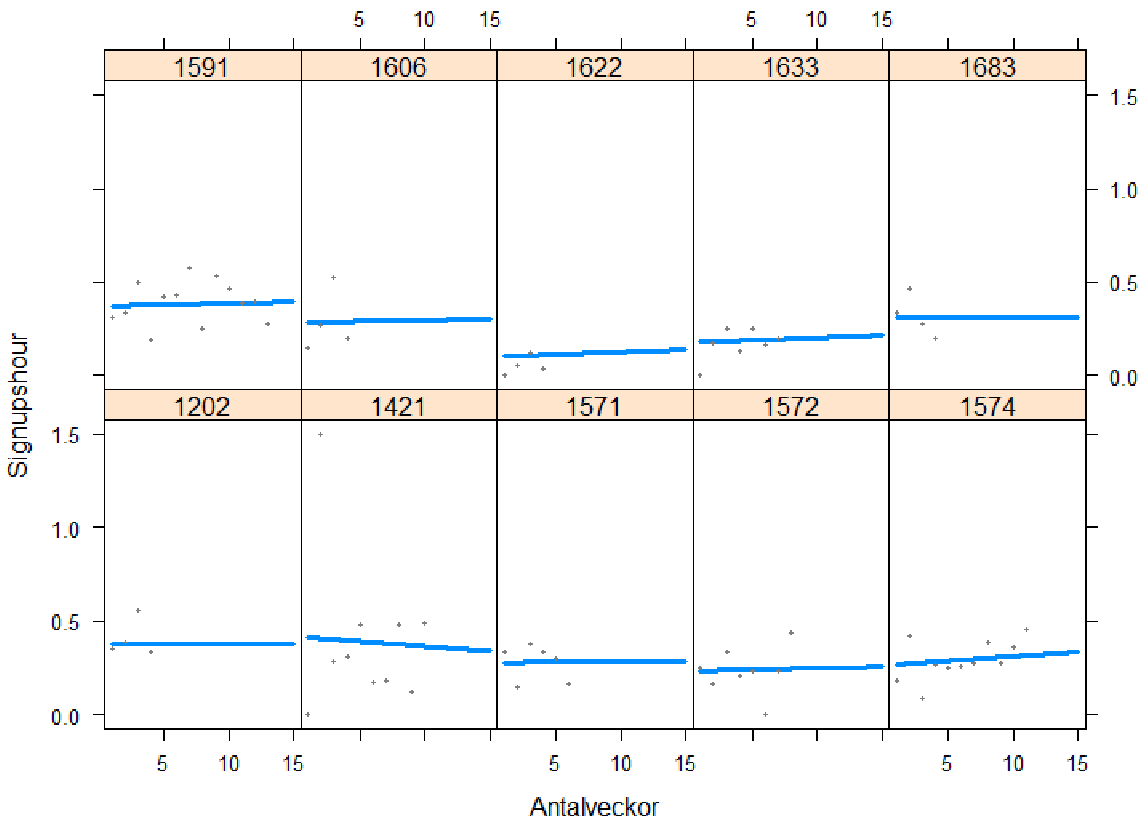


Figure 7: Each graph shows the results for one fundraiser, with weeks worked on the X axis and signups per hour on the Y axis. Above each graph is the individual ID number. The regression line in each graph is the sum of the fixed effects coefficients and the random effects coefficients.

As the graph shows the intercepts as well as the slopes differ for all individuals, capturing the individual characteristics. Mixed effects models bases its predictions both on the population parameters and the individual deviations.

It assumes that there are characteristics common to the population so that instead of basing predictions solely on the individual result, it uses the population information as well. This has the implication that the predictions are *pulled* towards the population mean. (Bates 2010)

A linear mixed effect model could be sufficient investigating how fundraisers develop, but it is a bit misleading. In this model the explained variable is be signups per hour, which is a quota of the two variables signups and hour. Agresti (2007) suggests using a count data model when explaining a rate such as this one. Intuitively this can be motivated by the fact that signups is the random variable of interest, not the amount of hours worked. A count data model lets us explain the amount of signups a fundraiser gets given that they have worked for some t hours. One of the most common models for count data is Poisson regression. (Agresti 2007)

Poisson regression is a model which falls under the the broader term Generalized Linear Models. In the next section the components of Generalized Linear Models (GLMs) is described, and just as before this model will be expanded to include random effects. To make the definitions more compact and general matrix notation will be used further on.

3.1.2 Generalized Linear Mixed Models

The regression model presented earlier eq. (1) require that the dependent variable Y follows the normal distribution, but as just shown this may not always be the case in real world situations. To escape the assumption of normality, one can use Generalized Linear Models (GLMs). As the name suggests, they are a generalization of the normal linear models, and can be used to model data that follows several different distribution that are a part of the exponential distribution family (ex. Binomial, Poisson, Negative-Binomial). It can be shown that the regression model presented earlier eq. (1) is a special case of the GLM. The concept of GLMs where first brought up by Nelder and Wedderburn (1972).

All GLMs consists of three components, the *random* component, the *systematic* component and the *link function* (Agresti 2007; Nelder and Wedderburn 1972)

(i) The random component specifies the underlying distribution of the dependent variable. Hence the observations $Y_1, Y_2, ..., Y_n = \{Y_i\} = \mathbf{Y}$, are independent observations from a distribution connected to the exponential family.

(ii) The systematic component describes the linear predictors in terms of a set of p explanatory variables x, and p coefficients β. The linear predictor is denoted η.

Let

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,p-1} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,p-1} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

Then η can be described as a vector in the following way

$$\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\eta} \tag{4}$$

(iii) The link function, $g(\cdot)$, is the connection between the linear predictor η and μ which is the expected value of Y.[3]

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} \tag{5}$$

Correspondingly $\boldsymbol{\mu}$ can be described as:

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) \tag{6}$$

If we let $\mathbf{x}_i'$ be row i in the model matrix, then the fitted value for each Y_i can be described as

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}) \tag{7}$$

This is the basic components of all Generalized Linear Models. This is essentially the same for Generalized Linear Mixed Models, the only difference is that the systematic component is expanded to include random effects as well. To do this we first go through some definitions.

We denote i fundraisers, and j the the number of weeks that the fundraiser have worked. We let M denote the number of fundraisers, and n_i the number of weeks that each fundraiser have worked. Thus $\sum_{i=1}^{M} n_i = N$ is the total number of observations in the data.

In Table 4 all other symbols are defined.

[3]If no transformation of μ is neded, meaning that $\mu = \eta$, it is said that we use the *identity* link function.

Table 4: Matrix notation

Symbol	*Size*	*Description*
$\mathbf{Y}_i$	$n_i \times 1$	Dependent variable vector for individual i
$\mathbf{X}_i$	$n_i \times p$	Fixed effects model matrix for individual i
$\boldsymbol{\beta}$	$p \times 1$	Fixed effects coefficients
$\mathbf{Z}_i$	$n_i \times q$	Random effects model matrix for individual i
$\mathbf{v}_i$	$q \times 1$	Random effects coefficients for individual i
$\boldsymbol{\Psi}$	$q \times q$	Covariance-variance matrix for the random effects.

The model can be described as

$$g(E[\mathbf{Y}_i|\mathbf{v}_i]) = g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i \tag{8}$$

$$\mathbf{v}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}) \tag{9}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma^2_{v_0} & \sigma^2_{v_0,v_1} & \dots & \sigma^2_{v_0,v_{q-1}} \\ \sigma^2_{v_1,v_0} & \sigma^2_{v_1} & \dots & \sigma^2_{v_1,v_{q-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{v_{q-1},v_0} & \sigma^2_{v_{q-1},v_1} & \dots & \sigma^2_{v_{q-1}} \end{bmatrix}, \tag{10}$$

in this model $\boldsymbol{\beta}$ is a parameter describing the characteristics of the population as a whole, and $\mathbf{v}_i$ are the random effects demonstrating how individuals vary from the population characteristics. $\mathbf{v}_i$ is assumed to be an unobserved random variable, with an underlying multivariate normal distribution. Since we assume that $\mathbf{v}_i$ is random variable, it cannot be a parameter. Instead we describe $\mathbf{v}_i$ with the parameter $\boldsymbol{\Psi}$ which can be understood as a measurement of how much the individual random effects vary between each other. (Agresti 2007)

The diagonal elements of $\boldsymbol{\Psi}$ are the variance components for each random effect, and the other terms are the covariance between the random effects. Sometimes it can be valuable to skip the covariance terms, this reduces model complexity (Bates et al. 2015).

The linear predictor for a certain individual a certain week is (Hedeker 2005):

$$g(E[Y_{ij}|\mathbf{v}_i]) = g(\mu_{ij}) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v_i}. \tag{11}$$

The model will then explain the results of a single fundraiser i after working a certain amount of weeks j.

The estimation of $\boldsymbol{\beta}$ allows us to draw conclusions of the population characteristics with all the variation between individual eliminated. The condition on $\mathbf{v}_i$ means that we wish to draw conclusions on certain individuals with a certain value of $\mathbf{v}_i$. If we do not condition on $\mathbf{v}_i$ all point estimations on different individuals will be the same since the expected value of $\mathbf{v}_i$ is 0.

Estimating the parameters of the model is often done the maximum likelihood estimation. It is a common method for estimating parameters in a model, and it can intuitively be understood as a method for choosing the parameters so that the observed sample is the most likely. In likelihood estimation in the discrete case one sets up the joint probability mass function of the observed sample, and sets this as a function to the parameter(s) of interest. Generally the parameter(s) of interest is denoted θ. (Hogg, Tanis, and Zimmerman 2015)

The joint pmf of the sample with θ as the argument is what is referred to as the likelihood function

$$L(\theta|Y), \tag{12}$$

when the function is defined, parameter estimation is done by choosing the value of θ that maximises the function. This is equivalent with maximising the likelihood[4] of the observed sample. One of the pros with using Likelihood estimation is that the estimates are asymptotically normally distributed. (Hogg, Tanis, and Zimmerman 2015)

The conditional pmf for GLMMs is defined as (Hedeker 2005; Casals et al. 2015)

$$f(\mathbf{Y}_i|\mathbf{v}_i) = \prod_{j=1}^{n_i} f(Y_{ij}|\mathbf{v}_i). \tag{13}$$

The assumption that we can multiply the individual observations together for the joint probability of each individual is called the conditional independence assumption. It means that the n_i observations for each individual, conditional on $\mathbf{v}_i$, are independent each other. (Hedeker 2005)

To obtain the marginal pmf one has to get rid of the condition of $\mathbf{v}_i$. Following Hedeker (2005) we can multiply the expression with the pmf of the random effects to obtain the joint distribution of $\mathbf{Y}_i$ and $\mathbf{v}_i$, and then integrate $\mathbf{v}_i$ out

$$\int f(\mathbf{Y}_i|\mathbf{v}_i) f(\mathbf{v}_i) d\mathbf{v}_i = \int f(\mathbf{Y}_i, \mathbf{v}_i) d\mathbf{v}_i = f(\mathbf{Y}_i|\boldsymbol{\beta}, \boldsymbol{\Psi}). \tag{14}$$

This is an integral evaluated over $\mathbb{R}^q$, solving it require us to integrate over the multivariate distribution of the random effects. The lme4 package used in this thesis does an analytical approximation using Laplace approximation, for a further discussion of Laplace approximation and other methods available for solving this integral see Bolker et al. (2009).

The marginal pmf tells us the probability of the observed data for individual i, when we don't know their random effect. We view their random effect as a

[4]Generally probability is described as a function of the data given the parameter θ, but in the likelihood function it is the other way around. Because of this one can not state that we maximises the probability, instead we say that we maximize the likelihood.

random variable with a variance we can estimate. To obtain the joint pmf for the whole sample we can multiply their probabilities with each other, given that their results are independent each other. (Hedeker 2005)

$$L(\boldsymbol{\beta}, \boldsymbol{\Psi}|\mathbf{Y}_i) = \prod_i^M f(\mathbf{Y}_i|\boldsymbol{\beta}, \boldsymbol{\Psi}). \tag{15}$$

The parameter estimations is done by selecting the values of $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ that maximizes eq. (15).

When it comes to estimate the random effects $\mathbf{v}_i$ a philosophical discussion arises. What are random effects? In some cases they are just model errors which one wishes to eliminate (Bolker et al. 2009). In other situations however they have a natural part of the model, such as in this thesis where all predictions will be done on an individual level, conditioned on each random effect. The difference between conditional and unconditional inference is a central concept in GLMMs and will be brought up throughout the the thesis.

In this thesis the random effects play a vital role, and it is therefore desirable to estimate them. But since they are not parameters they can't be estimated in the same manner as $\boldsymbol{\beta}$. Instead it is said that the random effects are predicted. (Bates 2010)

The prediction of $\mathbf{v}_i$ is done by maximizing the conditional density of the random effects and $\mathbf{Y}_i$ given the estimated paremeters, (Bates 2009)

$$\widehat{\mathbf{v}_i} = \arg\max_{\mathbf{v}_i} f(\mathbf{Y}_i|\mathbf{v}_i)f(\mathbf{v}_i), \tag{16}$$

these predictions of the random effects can be called mode values, since they are the maximum of a probability density function.

With the description of mixed effect, GLMs and the connection between them we can now go describe the Poisson regression model which is used in this thesis.

3.2 Poisson Regression with Mixed Effects

Poisson regression models are a category of GLMs that are often used when modelling count data. In its simplest form they explain a dependent variable Y, with a set of independent variables X, and assumes that the dependent variable Y follow a Poisson process.

The Poisson distribution expresses the probability of having Y events occurring in a certain time frame, with λ being the mean amount of events happening in that certain time frame.

The Poisson distribution has the following properties

$$E[Y] = V[Y] = \lambda, \tag{17}$$

meaning that the variance and the expected value are both equal to λ. Also, the Poisson process can only take positive integers as values, an attribute that makes it especially useful to model random events such as signups. Since the amount of signups a fundraiser gets can also only be positive integers. (Hogg, Tanis, and Zimmerman 2015)

Most often the Poisson regression model uses a log transformation as the link function (Agresti 2007). One reason for this is that this guarantees that the estimates is positive, something which wouldn't be guaranteed if we used the identity link function (Rodríguez 2007). Given the log link function, the relationship between λ and X can then in accordance with eq. (5) be described as

$$\log E[Y_i] = \log \lambda_i = \eta_i = \mathbf{x}_i' \boldsymbol{\beta}.$$

Assuming a Poisson distribution on Y gives (Hogg, Tanis, and Zimmerman 2015)

$$Y_i \sim Po(\lambda_i),$$
$$P(Y_i|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}.$$

To conclude, the Poisson regression estimate of Y_i is the mean of a Poisson process with a certain parameter λ_i that is decided by $\exp(-\mathbf{x}_i' \boldsymbol{\beta})$

If we assume that signups per hour follows a Poisson distribution, we can use Poisson regression to study this situation. But since the amount of hours worked for each fundraiser each week differs, the model needs to be adjusted. As mentioned earlier, λ represent the mean amount of signups occurring in a certain time frame, if the time frame differs we can define the situation as (McElreath 2016)

$$\lambda = \frac{\mu}{t}, \tag{18}$$

here μ is the amount of events occurring, and t is the duration of the time frame.

The model can then be written as

$$\log(\mu_i/t_i) = \eta_i = \mathbf{x}_i' \boldsymbol{\beta}.$$

By the logarithmic rules this term can be moved over to the right side. That term is then what is often called an offset term. It will have an multiplicative effect on our estimates of Y, the logarithmic transformation makes sure that it is on the same scale as Y so that (Agresti 2007).

$$\log(\mu_i) = \log(t_i) + \mathbf{x}_i' \boldsymbol{\beta}, \tag{19}$$
$$\mu_i = t_i \exp(\mathbf{x}_i' \boldsymbol{\beta}). \tag{20}$$

Note that t_i is not an explanatory variable, it is only a different way of viewing the Poisson process that lets us use Poisson regression. The interpretation of $\boldsymbol{\beta}$ is still signups per hour.

To add random effects to this model, we follow eq. (8)

$$\log(\mu_{ij}) = \eta_{ij} = \log(t_{ij}) + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v_i}. \tag{21}$$

This is the complete Poisson GLMM that will be used in this thesis. To apply this model to the data the lme4 package by Bates et al. (2015) will once again be used.

3.3 Diagnostics

In this section some techniques for model building and model evaluation will be discussed. Mainly R^2 and AIC will be used for comparing different models. When it comes to checking the model assumptions some diagnostic plots will be used for informal validation. Testing whether the fixed effect parameters are significantly different from zero Wald Z-test is used.

In addition to this it will be discussed how the model can be used for predictions, and how to measure the uncertainty in these predictions. Bootstrapping will be proposed to compute approximate prediction intervals.

3.3.1 Akaike's Information Criterion (AIC)

A common difficulty in model building is the trade-off between model simplicity and goodness-of-fit. If the model is to simple it might miss important information, but if it is to complex it might become overfitted. What this means is that the model might explain the observed data really well, but fails in explaining unobserved data. Some of the characteristics of the sample might just be because of randomness, and not because of the underlying characteristics common to the population.

A popular tool for assisting in this trade of is Akaike's Information Criterion (AIC) (Sheather 2009). It rewards goodness-of-fit and punishes model complexity, and one should always strive after a model with as low AIC as possible. The general AIC is defined as (Sheather 2009; Lian 2012)

$$AIC = -2\left[\log L(\hat{\theta}|Y) - K\right] \tag{22}$$

Here K is the amount estimated parameters in the model, and $L(\hat{\theta}|Y)$ is the maximum value of Likelihood function with respect to θ (see eq. 12). This result has it roots in the Kullback–Leibler divergence, and generally it can be said that K is a bias correction due to the same data being used for estimating θ and maximizing the likelihood function. Furthermore AIC is a measurement of a models capability to explain new data, compared to other models. (Sheather 2009; Lian 2012).

In a GLMM context this is referred to as *marginal* AIC, which can be written as (Lian 2012)

$$mAIC = -2\left[\log L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}}|\mathbf{Y}_i) - K\right]. \tag{23}$$

Here $L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}}|\mathbf{Y}_i)$ is the maximum of the marginal likelihood eq. (15), and similarly to the previous equation K is once again the amount of estimated parameters, which is the coefficients of $\boldsymbol{\beta}$ and variance components in $\boldsymbol{\Psi}$. (Lian 2012)

mAIC should be used in GLMM when the fixed effects are the primary focus. In the case of fundraisers it would be useful if the main interest was to predict results for entirely new fundraisers with a random effect that is unknown. In this thesis this is not entirely the case, since we wish to draw conclusions on fundraisers after that they have worked for some weeks, letting us predict their individual random effects. This means that the conditional inference is central in this case, and this calls for the use of the *conditional* AIC which was first derived by Lian (2012)

$$cAIC = -2\left[\log L(\hat{\boldsymbol{\beta}}|\mathbf{Y}_i, \hat{\mathbf{v}}_i) - K\right]. \tag{24}$$

In this expression $L(\hat{\boldsymbol{\beta}}|\mathbf{Y}_i, \hat{\mathbf{v}}_i)$ the value of the conditional likelihood, with any estimator of $\mathbf{v}_i$ and $\boldsymbol{\beta}$. The conditional likelihood describes the probability of the sample, conditioned on the random effects. It can thus be written as

$$L(\boldsymbol{\beta}|\mathbf{Y}_i, \mathbf{v}_i) = \prod_{i}^{M} f(\mathbf{Y}_i|\mathbf{v}_i; \boldsymbol{\beta}), \tag{25}$$

here we do not need to integrate out the random effects as in the marginal likelihood.

Once again K is also a bias corrector but it is a bit more complex, see Lian (2012) for details.

The cAIC4 package by Saefken et al. (2018) will be used to calculate the cAIC for all models. It can do it both approximate using bootstrap or analytically by Chen-Stein formula, it was decided to go by the latter.

3.3.2 $\mathbf{R^2}$

In linear regression R^2 is a good tool for understanding how well a model explains a dependent variable (Sheather 2009). It can be understood as a proportion of how much the model explains the observed variation, i.e.

$$\frac{\text{Explained variation}}{\text{Total variation}}$$

The coefficient of determination R^2 can take any value between 0 and 1, and when building models one should strive for as high R^2 as possible. An estimated R^2 of 1 would mean that *all* variation observed can be explained by the model. R^2 is an useful measurement since it is an absolute value of the goodness-of-fit for the model, has an intuitive meaning and it can be compared with other models for other data materials (Nakagawa and Schielzeth 2013).

Estimating an R^2 value for Mixed Effects models is not as straight forward as in the linear regression case. Mainly due to the fact that we assume variation due to the data originating from different individuals - and it is not clear how a definition of R^2 should relate to this variability and whether the concept of R^2 is helpful in evaluating mixed effects models. There are also some dilemmas when defining the unexplained variance, residual variance, in GLM models. (Nakagawa and Schielzeth 2013)

One recent proposal of R^2 for GLMMs is made by Nakagawa and Schielzeth (2013), and they define two different R^2 values that can be calculated for GLMMs. They refer to marginal R^2 as the variability explained by the fixed effects only, and conditional R^2 as the variability explained by the fixed and random effects in the model. With some extension done by Johnson (2014), this value can be calculated for GLMM models with random intercept and random slope using the MuMin package by Bartoń (2013).

Let σ_f^2 be the fixed effect variance attribute, σ_v^2 the random effect variance and σ_ε^2 the residual variance (Nakagawa and Schielzeth 2013), then the two R^2 measurements are defined as

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sum_{v=1}^{q} \sigma_v^2 + \sigma_\varepsilon^2} \tag{26}$$

$$R_c^2 = \frac{\sigma_f^2 + \sum_{v=1}^{q} \sigma_v^2}{\sigma_f^2 + \sum_{v=1}^{q} \sigma_v^2 + \sigma_\varepsilon^2} \tag{27}$$

This does not have the same finesse as an R^2 for linear regression model. In general R_m^2 will be low since it ignores variability between individuals, whilst R_c^2 will be high if we have great variability among individuals. R_c^2 is the closest we can come to an regular R^2 since it is the *total* variation explained by the model, but it should be interpreted with caution.

Following Nakagawa and Schielzeth (2013) their R^2 values will be used for evaluating different models alongside the AIC criteria. They view their R^2 as a appropriate compliment to AIC since AIC does not tell anything about the variation explained. Furthermore, R^2 will solely be used to compare models, and will not be interpreted more in detail.

3.3.3 Wald Z-test

In most models it is of interest to see whether a parameter is significantly different from zero. If this is proved, one could argue that the parameter most definitely is relevant. To test whether the fixed effects are signifantly different from zero the lme4 package by default uses Wald Z test, making use of the asymptotic normality assumption of the likelihood estimation. The Wald Z test is defined as (Agresti 2007)

$$Z_{obs} = \frac{\theta - \theta_0}{SE(\widehat{\theta})}, \tag{28}$$

it then compares this value with the standardized normal distribution to calculate P-values. These are rough estimates, popular due to being easy to calculate.

It is safer to use bootstap to calculate confidence intervals, and/or P-values, for the fixed effects. But the computation time for this is rather unforgiving, forcing us to stick to the Wald Z test. (Bates et al. 2015)

3.3.4 Diagnostic plots

As mentioned earlier in eq. (9), it is assumed that the random effects are multivariate normally distributed. One efficient way of testing normality is using a QQ-plot (Quantiles quantiles plot). What it does is that it takes the observed data and plots it along the Y axis, whilst the theoretical quantiles of the choosen distribution is plotted along the X axis. If the observed data follows the theoretical distribution a straight line should be formed. In this case the theoretical distribution is the normal distribution. (Sheather 2009)

To explore how the fitted values differ from the observed data two plots will be used. They will give us an understanding of the model and can give hints of the model adequacy. The first plot is to simply plot the fitted values against the observed data, in a linear regression model, this should roughly be a straight line if the model is correct. But since we assume a Poisson distribution of Y_{ij} conditional on $\mathbf{v_i}$, the variance will increase for higher values - creating a funnel resemblance. To solve this we will divide both values with the theoretical standard deviation, which is the square root of the fitted value given that the Poisson assumption holds.

$$Y_{ij}/\sqrt{\widehat{\mu}_{ij}} \quad \text{vs} \quad \widehat{\mu}_{ij}/\sqrt{\widehat{\mu}_{ij}}$$

We can also plot the standardized the residuals, also called Pearson residuals, against the fitted values. The pearson residuals for the GLM Poisson case is defined as (Agresti 2007)

$$_p\varepsilon_{ij} = \frac{Y_{ij} - \widehat{\mu}_{ij}}{\sqrt{\widehat{\mu}_{ij}}}.$$

The same principle will be followed when plotting

$$_p\varepsilon_{ij} \quad \text{vs} \quad \widehat{\mu}_{ij}.$$

These diagnostic tools are often used in linear regression where the residuals are assumed independently normally distributed, this is not entirely the case in the GLM nor the GLMM case. But they will give us a hint of the misspecification of the model and can tell us if there are extreme outliers affecting the results, or other potential misspecifications.

3.3.5 Prediction

Lastly the model will be used for predictions, and this is done with a two-sided purpose. Firstly prediction can be used as a tool for evaluating how well adapted the model is to the data. This is a robust way for finding errors and misspecifications in the model (McElreath 2016). Secondly it is of great interest to explore the predictive capabilities of the model since it is one of the underlying purposes in the thesis.

The model can be used for two types of predictions, conditioned and unconditioned. Unconditional predictions are only made based on the population parameters $\boldsymbol{\beta}$. This type of prediction can be used to predict how a newly hired fundraiser will perform, before they've even started working. Conditional predictions on the other hand can be made when we can predict $\mathbf{v}_i$, and wish to see how an individual will progress.

It is of interest in this thesis to do conditional predictions. To recap what has earlier been mentioned the point estimation of the model in eq.(21) is the mean of a Poisson process which is decided by the linear predictioner η.

Given that we have observed our data $\mathbf{Y}_i$, and wish to make a prediction on individual i using new data, and their predicted random effects $\widehat{\mathbf{v}}_i$, we can see this prediction as a result of

$$t_{ik} \exp\left(\underbrace{\mathbf{x}'_{ik}}_{1\times p} \underbrace{\widehat{\boldsymbol{\beta}}}_{p\times 1} + \underbrace{\mathbf{z}'_{ik}}_{1\times q} \underbrace{\widehat{\mathbf{v}}_i}_{q\times 1} \right) \tag{29}$$

where we say that index k represent some unobserved week.

Noteworthy is that the expression within the parentheses represents signups per hour, which is what we originally strived to explain. To standardize the results later we can divide all values with t_{ij}

The point estimation does not tell the whole picture though - it leaves out the uncertainty. To capture the uncertainty of these predictions will be complemented with a prediction interval. A prediction interval tells us the range of which we can expect new values to fall under, with a certain degree of confidence. The ciTools package in R by Haman and Avery (2019) lets us do this using parametric bootstrapping. A simplified explanation of what the package does is that it refits the model again and again, and for each fitted value it calculates the span of which 95% of the observations fall under. A bootstrap sample of 1000 seems sufficient, increasing the sample does not change the intervals noticeably.

This is a method which dates back to Efron (1979). The method can be used to estimate for example variance, confidence intervals, prediction intervals in an observed sample by simply resampling new data based of the observed data (or from a fitted model, this is then called parametric bootstrapping). The method is rather counterintuitive at first glance, just like pulling yourself upwards by pulling your bootstraps.[5]

[5] In orginal paper from 1979 the author finishes off with the following: "I also wish to thank the many friends who suggested names more colorful than Bootstrap, including Swiss Army Knife, Meat Axe,Swan-Dive, Jack-Rabbit, and my personal favorite, the Shotgun,which to paraphrase Tukey, 'can blow the head of any problem if the statistician can stand the resulting mess.'"(Efron 1979)

4 Results

In this section the results of some Poisson models will be presented. The goal is to create a model with as low AIC as possible, and with a high R^2. Diagnostic and prediction plots can be found in the Appendix.

When fitting the models with the lme4 package with all data available the models failed to converge, one problem of this may lie in the fact that it is difficult to estimate a parameter common to all fundraiser when some work more than 100 weeks, and some only 2. Only using the data of the first 15 weeks fundraisers work solved the convergence problem. Another issue came up when using the city variable, since all cities except Stockholm performed pretty similar they where highly correlated - which made model fitting difficult. Therefore it is decided to only use Stockholm as one variable, contrasting all other cities against it.

4.1 Parameter estimation

One popular approach for selecting variables in a model is forward selection. The method starts with a simple model, then adding explanatory variables until they do not improve the model (Sheather 2009).

Following this method three models are presented in Table 5. Model one has only a fixed and random intercept, in model two we add the variable weeks worked as a fixed and random effect, this improves the model. Lastly we add the city variable as a fixed effect. This last model seems to be the best in terms of AIC and R_m^2, model 2 does have a better R_c^2, but the difference is minimal. To conclude: model 3 seem to be the best and will be further explored down below.[6]

[6] Since it is assumed that the fundraisers have a learning curve, implying that the development diminish over time a quadratic term was added. This however seemed to be a to complex model for the package to handle, and there were many reported errors when fitting the model. Since we only look at the first 15 weeks of the development it might be reasonable to assume a linear development.

Table 5: Summary of parameter estimation and diagnostic values.

	Dependent variable:		
	Signups		
	Model 1	Model 2	Model 3
Weeks worked		0.008^{*}	0.009^{**}
		(0.004)	(0.004)
Other cities			-0.363^{***}
			(0.037)
Intercept	-1.613^{***}	-1.637^{***}	-1.401^{***}
	(0.022)	(0.026)	(0.034)
Random effects			
Weeks worked		0.0010	0.0010
Intercept	0.255	0.2950	0.2667
Observations	4461	4461	4461
No. fundraisers	766	766	766
Log Likelihood	−10276.0	−10232.6	−10186.1
Marginal AIC	20555.9	20475.1	**20384.2**
Conditional AIC	19743.95	19622.90	**19598.07**
R^2m	0	0.0021	**0.0692**
R^2c	0.5155	**0.5219**	0.5202

Note: For fixed effects the estimated coefficient is reported, with the estimated standard error in parentheses. For random effects the estimated variance is reported. The city variable uses Stockholm as the intercept and the other cities in contrast to Stockholm. All results are presented in the scale of the linear predictor.

The stargazer package in R by Hlavac (2018) have been helpful in creating this and other Tables in this thesis.

*p<0.05; **p<0.01; ***p<0.001

4.2 Diagnostics and prediction

All diagnostic plots can be found in Appendix A. In summary the random effects seem normally distributed, and the residuals look fine. They should however be interpreted with caution as they do not share the same properties as in the linear regression case.

Next we wish to examine whether the model works well in predicting new data. We do this by visualising the model predictions on two individuals. For each individual their results after their fifth week are removed, so that we predict unobserved results. As seen in the prediction plots in Appendix B the prediction intervals a rather wide, which is expected since the results mostly are volatile. Therefore it is difficult to rule out whether or not a fundraiser will reach the goal, even if they initially are below it.

5 Discussion

5.1 Interpretation of the results

The objectives presented earlier where the following:

(i) Is there a learning curve for fundraisers?

(ii) What variables affect the the results of fundraisers?

(iii) To what extent are predictions possible?

To answer these questions we first need to put the estimated parameters from Table 5 in the correct scale. This is done below

Table 6: Interpretation of the parameter estimations

Parameter	*Estimation*	*Meaning*
Intercept	$e^{-1.401} = 0.2464$	Fundraisers in Stockholm generally start at 0.25 signups per hour.
Other cities	$e^{-0.363} = 0.6956$	Results outside of Stockholm are generally 30% lower.
Weeks worked	$e^{0.009} = 1.0090$	The results improve 0.9% for each week worked.

Regarding the first question we can prove a significant learning curve for fundraisers generally, but it is not very big. We estimated that if someone starts at 0.2 signups per hour and work 15 weeks, their results will improve to 0.223 approximately. We can not really state what will happens after 15 weeks since we only use the initial data in this thesis. Since many fundraisers quit working early on when the results are low there simply is not enough data to draw valid conclusions on the long term development.

When taking the random effects into consideration we can see a more nuanced picture. The random effects vary from 0.06 to - 0.08, which tells us that fundraisers at best become 7% better, while some become 6-7% worse over time. These predictions of the random effects are of course also unsure, and has some variability in them. But they tell us the range of how we can expect the best and worst developments.

Regarding the second question it seems like whether or not fundraisers are in Stockholm or not plays an important role in predicting the results. This is the result of all models fitted where the city variables is used. It can not however be stated if that is because of fundraisers in Stockholm being better, or because it is easier to fundrais in Stockholm - but it should be looked in to.

If it is the latter, coordinators could perhaps but higher demands on Stockholm fundraisers than in other cities.

Lastly we can conclude that predictions based on the model tend to be unsure since the predictions intervals are rather wide. Because of this it is difficult to tell whether or not a fundraiser will reach the target goal in most cases. In one sense this is expected since the results are so volatile, and to expect otherwise would be misleading. We can still use the model however to get an understanding of what direction the development is going, and how strong it is.

5.2 Future studies

Something that has been overseen in this thesis is the affect of potential autocorrelation. Autocorrelation is basically a dependency structure where the observed values are dependent on the previous values. Many times it can be observed when examining a dependent variable over time, and it should be dealt with.

Ben Bolker (2016) discusses possible approaches to solve this problem. One solution is to incorporate a model term in to the random effects with a autocorrelated dependency structure. The model is the same as before but with a redfined $\mathbf{v}_i$

$$\mathbf{v}_i = \mathbf{v}_{0,i} + \epsilon_i$$
$$\mathbf{v}_{0,i} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi})$$

Here $\mathbf{v}_{0,i}$ is the regular random effect, earlier defined as $\mathbf{v}_i$, while ϵ_i is the autoregressive component for each individual i. Applying this approach for the problem of fundraisers could perhaps improve the model. But this is left out for further studies, partly due to the fact that the focus in this study is the first 15 weeks of fundraisers, and this leaves us with little data for doing reasonable estimations of the individual autocorrelation structure. In appendix C the autocorrelation for the residuals of 4 individuals are presented.

Another factor that has been overseen in this thesis is that some fundraisers work less hours for some weeks, this could potentially impact how the fundraisers develop. It would be interesting to see if a model which takes this in to consideration could perform better. This can be achieved for example by having total hours worked as a explanatory variable, this would however be difficult to put in to practical use. Another solution could instead be to put weights on the weeks worked based upon how many hours they worked each week.

References

Agresti, Alan (2007). *An Introduction to Categorical Data Analysis*. 2nd ed. New York: Wiley. ISBN: 978-0-471-22618-5.

Bartoń, Kamil (2013). *MuMIn: Multi-model inference*. R package version 1.10.0. URL: https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf.

Bates, Douglas (2009). *Linear mixed model implementation in lme4*.

— (2010). *lme4: Mixed-effects modeling with R*. Springer.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Bauer, Daniel J (2011). "Evaluating Individual Differences in Psychological Processes". In: *Current Directions in Psychological Science* 20(2), pp. 115–118. DOI: 10.1177/0963721411402670.

Bolker, Ben (2016). *fitting mixed models with (temporal) correlations in R*. URL: https://bbolker.github.io/mixedmodels-misc/notes/corr_braindump.html (visited on 12/09/2019).

Bolker, Benjamin, Mollie Brooks, Connie Clark, Shane Geange, John Poulsen, Hank Stevens, and Jada-Simone White (2009). "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution". In: *Trends in ecology & evolution* 24, pp. 127–35. DOI: 10.1016/j.tree.2008.10.008.

Casals, Martí, Klaus Langohr, Josep Carrasco, and Lars Rönnegård (2015). "Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a simulation study". In: *SORT-Statistics and Operations Research Transactions* 39, pp. 281–308.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". In: *The Annals of Statistics* 7.1, pp. 1–26.

Haman, John and Matthew Avery (2019). *ciTools: Confidence or Prediction Intervals, Quantiles, and Probabilities for Statistical Models*. R package version 0.5.1. URL: https://CRAN.R-project.org/package=ciTools.

Hedeker, Donald (2005). "Generalized Linear Mixed Models". In: *Encyclopedia of Statistics in Behavioral Science*. Ed. by B. S. Everitt and D. C. Howell. DOI: 10.1002/0470013192.bsa251.

Hlavac, Marek (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.2. Central European Labour Studies Institute (CELSI). Bratislava, Slovakia. URL: https://CRAN.R-project.org/package=stargazer.

Hogg, Robert V., Elliot A. Tanis, and Dale L. Zimmerman (2015). *Probability and Statistical Inference*. 9th ed. Harlow: Pearson. ISBN: 1-292-06235-5.

Johnson, Paul C.D. (2014). "Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models". In: *Methods in ecology and evolution* 5 (9), pp. 944–946. DOI: 10.1111/2041-210X.12225.

Lian, Heng (2012). "A note on conditional Akaike information for Poisson regression with random effects". In: *Electron. J. Statist.* 6, pp. 1–9. DOI: 10.1214/12-EJS665. URL: https://doi.org/10.1214/12-EJS665.

McElreath, Richard (2016). *Statistical Rethinking : A Bayesian Course with Examples in R and Stan.* Chapman & Hall/CRC Texts in Statistical Science Series. Chapman and Hall/CRC. ISBN: 9781482253443. URL: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1338291&site=eds-live&scope=site.

Montgomery, Douglas C. (2013). *Design and Analysis of Experiments*. 8th ed. Singapore: John Wiley & Sons. ISBN: 978-1-118-09793-9.

Nakagawa, Shinichi and Holger Schielzeth (2013). "A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models". In: *Methods in Ecology and Evolution* 4, pp. 133–142. DOI: 10.1111/J.2041-210x.2012.00261.X.

Nelder, J. A. and R. W. M. Wedderburn (1972). "Generalized Linear Models". In: *Journal of the Royal Statistical Society. Series A (General)* 135.3, pp. 370–384. ISSN: 00359238. URL: http://www.jstor.org/stable/2344614.

Nguyen, Kevin (2019). "Street fundraisers raised more than $120 million for Australian charities in 2018, report reveals". In: *ABC*. URL: https://www.abc.net.au/news/2019-01-04/charity-street-fundraisers-raise-plenty-of-money-report-reveals/10614810 (visited on 12/09/2019).

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rodríguez, Germán (2007). *Lecture Notes on Generalized Linear Models*. URL: http://data.princeton.edu/wws509/notes/ (visited on 12/09/2019).

Saefken, Benjamin, David Ruegamer, with contributions from Sonja Greven, and Thomas Kneib (2018). *cAIC4: Conditional Akaike information criterion for lme4*.

Sheather, Simon (2009). *A Modern Approach to Regression with R*. New York: Springer. DOI: 10.1007/978-0-387-09608-7.

UNHCR (2019). *What We Do*. URL: https://www.unhcr.org/what-we-do.html (visited on 12/09/2019).

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

A Diagnostic plots

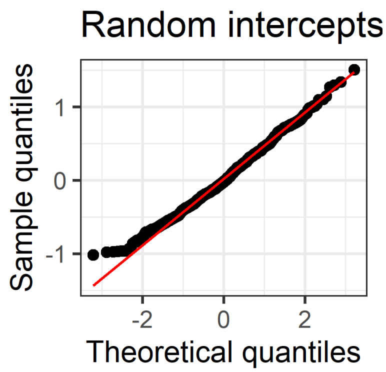


Figure 8: The X axis represent the observed quantiles for the predictions of the random effect intercept, and the Y axis the theoretical quantiles of the standardized normal distribution.

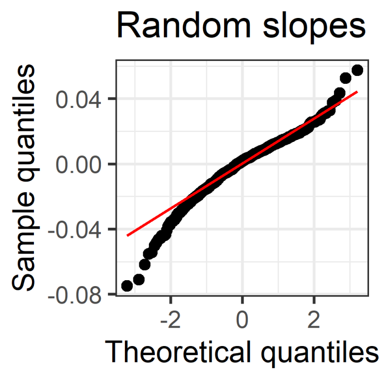


Figure 9: The X axis represent the observed quantiles for the predictions of the random effect weeks worked, and the Y axis the theoretical quantiles of the standardized normal distribution.

In Figure 8 we see that the predictions of the random intercepts are almost perfectly normally distributed. In Figure 9 the predictions of the random slopes (weeks worked), they are not equally well normally distributed, the tails deviate quite a lot from the theoretical distribution, but overall it is rather well.

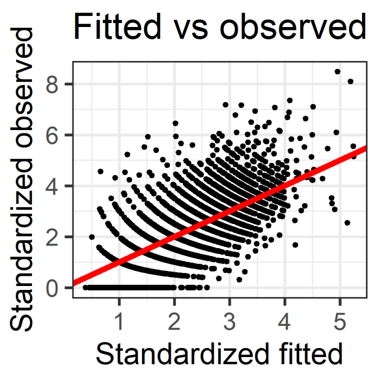


Figure 10: The X axis represent the standardized observed values, and the Y axis the standardized fitted values.

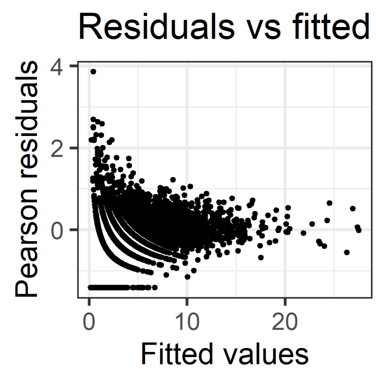


Figure 11: The X axis represent the fitted values and the Y axis the residuals.

In Figure 10 and 11 we study errors of the model in two different ways. Generally they look fine. There are larger residuals around zero, but that may be due to individual variation or due to the fact that we have more observations around for the lower values.

B Prediction plots

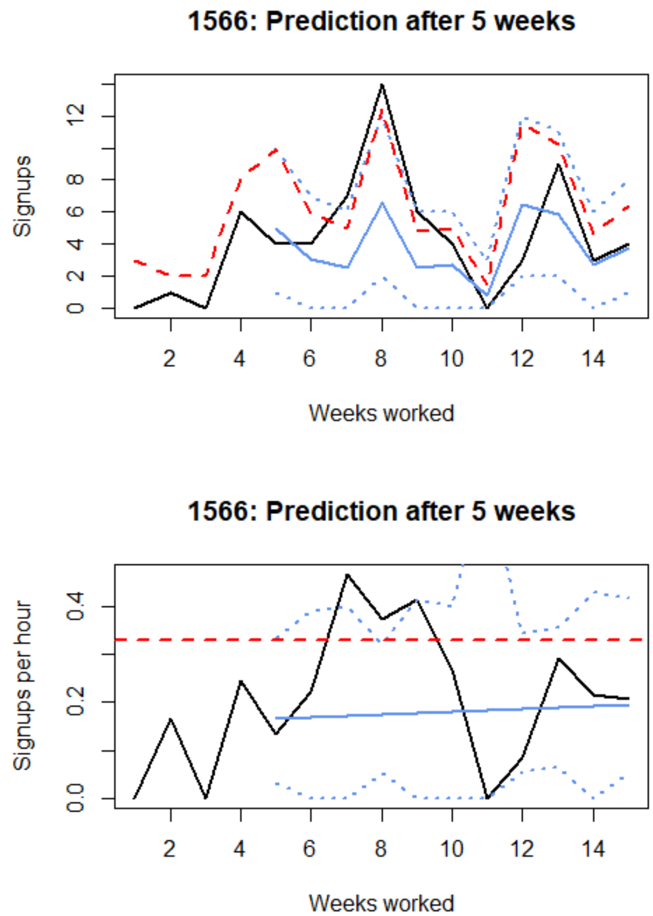
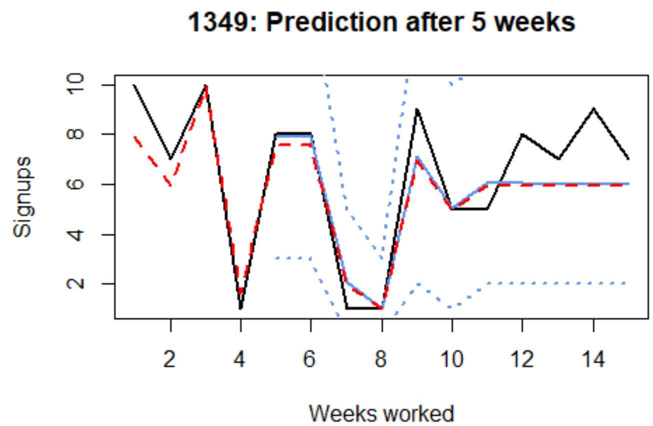


Figure 12: Prediction and prediction intervals based on the first 5 weeks of fundraiser 1566. The blue dotted lines are 95% prediction intervals, the red dashed line is the 0.33 threshold fundraisers must meet. To get the results in a standardized form all values are divided by the amount of hours worked t_{ij}.



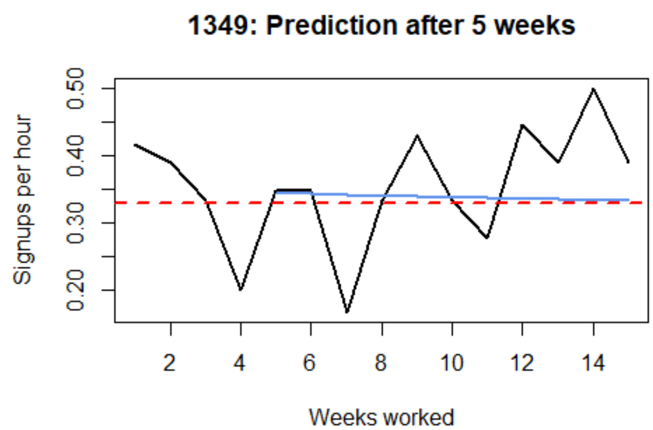


Figure 13: Prediction and prediction intervals based on the first 5 weeks of fundraiser 1349. The blue dotted lines are 95% prediction intervals, the red dashed line is the 0.33 threshold fundraisers must meet. To get the results in a standardized form all values are divided by the amount of hours worked t_{ij}.

C Autocorrelation plots

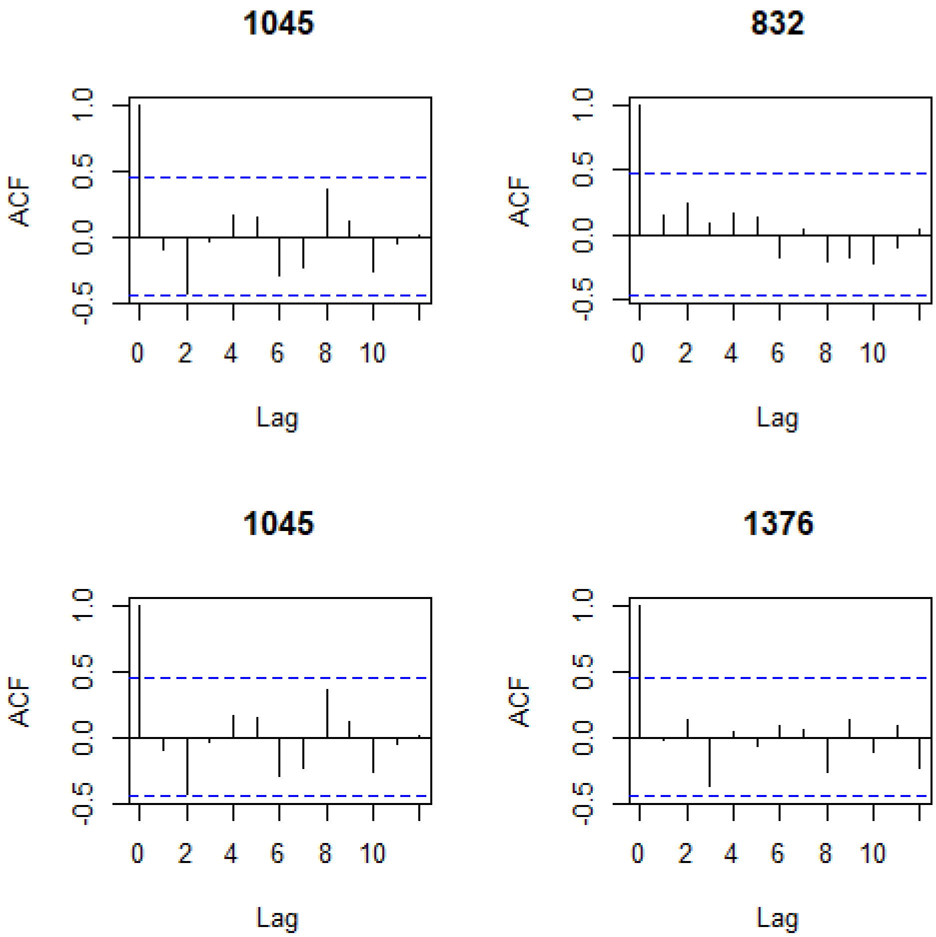


Figure 14: In these plots the observed autocorrelation in the residuals for four individuals are presented. The Y axis represent the observed autocorrelation, and the X axis the different lags.