Assignment 3

02424 Advanced Dataanalysis and Statistical Modelling

CLARA BRIMNES GARDNER (s153542)
TOBIAS ENGELHARDT RASMUSSEN (s153057)
CHRISTIAN DANDANELL GLISSOV (s146996)

May 26, 2021

Part 1: Strength of Ready Mixed Concrete

Problem A

Presentation of Data

Figure 1 shows the strength of the concrete plotted against time. The plot shows that the concrete is stronger after 28 days compared to 7 days. It also shows that there is more variation in the strength after 28 days than in the strength after 7 days. The strength seems to depend on the batch with batch 2 having a high strength, and batch 5 having a low strength.
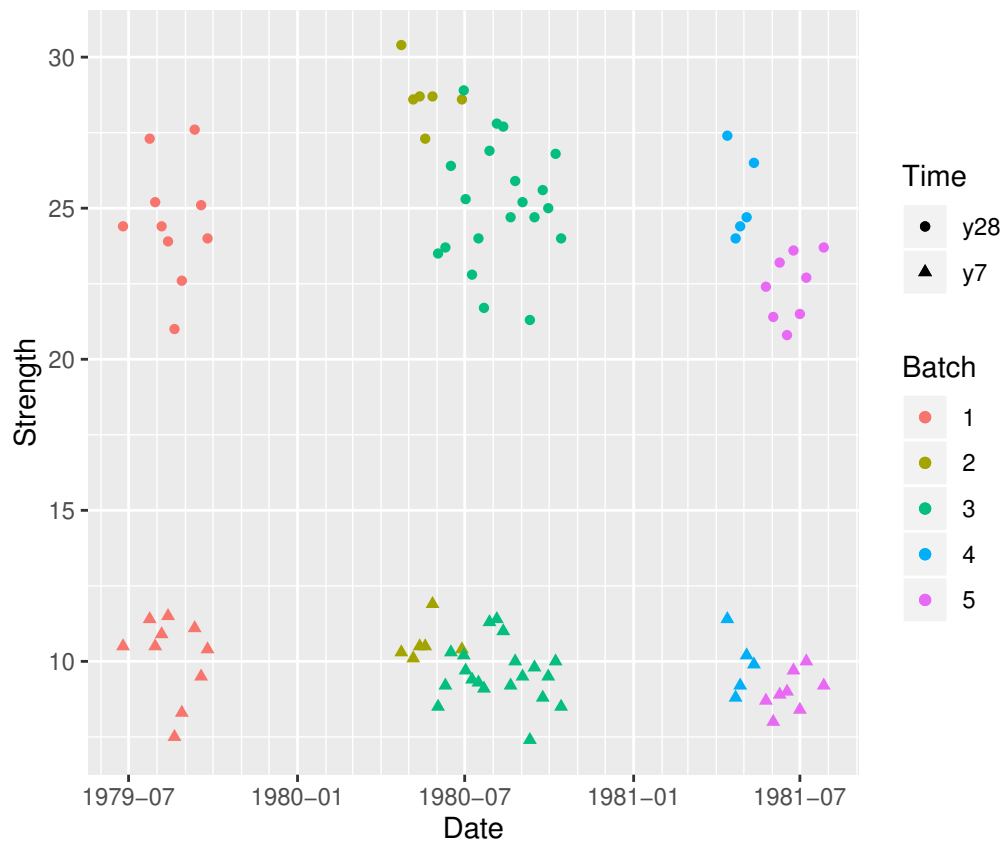

Figure 1: *The strength of the concrete against time. The shape of the amrker shows the shape of the measurement, and the color shows the batch.*

Batch Means

Table 1 shows the mean strength after 7 days and 28 days of each batch. Both after 7 and 28 days the largest mean is found for batch 2. The lowest mean are in both cases found for batch 5.

Batch	y7	y28
1	10.16	24.55
2	10.62	28.72
3	9.61	25.10
4	9.90	25.40
5	8.99	22.41

Table 1: *The mean strength of each batch after 7 days (y7) and after 28 days (y28)*

Mixed Effects Model

We now formulate a mixed effects model for the 28 day strength. We include the 7 day strength and the air-temperature as fixed effects and the batch number as a mixed effect.

$$\texttt{y28}_i = \mu + \alpha \cdot \texttt{y7}_i + \beta \cdot \texttt{air.temp}_i + a(\texttt{batch}_i) + \epsilon_i, \tag{1}$$

where we assume $\epsilon_i \sim N(0, \sigma^2$ and $a(\texttt{batch}_i) \sim N(0, \sigma_B^2)$. The inclusion of the 7-day strength as a fixed effect, means that the model can first be used after 7-days.

Testing for air-temperature dependency

It is investigated whether $\texttt{y28}$ depends on $\texttt{air.temp}$ - that is, it is tested if $\texttt{air.temp}$ is a significant variable. As $\texttt{air.temp}$ is a fixed effect, this can be investigated with an F-test. Table 2 shows the anova-table for the model given in (1). The table shows that $\texttt{air.temp}$ is in-significant, and it should therefore be removed from the model.

	numDF	denDF	F-value	p-value
Intercept	1	42	41.51312	<.0001
y7	1	42	59.46544	<.0001
air.temp	1	42	3.13301	0.084

Table 2: *Type III anova-table for the model given in* (1)

.

For completion it is also tested if $\texttt{y7}$ can be removed from the model. Table 3 shows the anova-table for the model without $\texttt{air.temp}$. It is seen that $\texttt{y7}$ cannot be removed from the model

	numDF	denDF	F-value	p-value
Intercept	1	43	40.51081	<.0001
y7	1	43	54.30682	<.0001

Table 3: *Type III anova-table for the model without* $\textit{air.temp}$

.

The final model thus becomes

$$\texttt{y28}_i = \mu + \alpha \cdot \texttt{y7} + a(\texttt{batch}_i) + \epsilon_i,$$

where we again assume $\epsilon_i \sim N(0, \sigma^2)$ and $a \cdot \texttt{batch}_i \sim N(0, \sigma_B^2)$ The estimates for the parameters are

$$\mu = 12.15$$
$$\alpha = 1.33$$
$$\sigma_B^2 = 2.31$$
$$\sigma = 1.42.$$

Furthermore the random effects are estimated to be

$$a(\texttt{batch}_1) = -1.015$$
$$a(\texttt{batch}_2) = 2.253$$
$$a(\texttt{batch}_3) = 0.198$$
$$a(\texttt{batch}_4) = 0.105$$
$$a(\texttt{batch}_5) = -1.541$$

.

As expected the random effect is highest for batch 2 and lowest for batch 5.

Problem B

We now consider the vector $Y = (Y_7, Y_{28})^T$, which contains both the 7 day and the 28 day measurements.

Multivariate mixed effect model

We formulate a very simple multivariate model, with a general mean, μ, and the batch number, which is included as a random effect.

$$Y_{ij} = \mu + a(\texttt{batch}_i) + \epsilon_{ij},$$

$i = 1, 2, \ldots, 5$, $j = 1, 2, \ldots, n_i$. Here the subscript i is associated with the batch number, and j is associated with the repetition number within the batch. n_i is the number of repetitions within batch i. We assume $a(\texttt{batch}_i) \sim N_2(0, \Sigma_0)$ (i.i.d) and $\epsilon_{i,j} \sim N_2(0, \Sigma)$ (i.i.d).

Estimate parameters

The parameters of the model can be estimated using both maximum likelihood (ML) and restricted maximum likelihood (REML). The estimate of μ is either way the grand means. That is

$$\mu = (9.77, 25.02)^T.$$

To find the estimates of the variance parameters, Σ_0 and Σ, the total variation (SST) should be decomposed into the between batches variation (SSB) and the within groups variation (SSE)

$$SST = SSB + SSE. \tag{2}$$

Following equation (5.85-5.89) in the book, SSB, SSE and SST can be found as

$$SSE = \sum_{i=1}^{5} \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{i+})(Y_{i,j} - \bar{Y}_{i+})^T$$

$$SST = \sum_{i=1}^{5} \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{++})(Y_{i,j} - \bar{Y}_{++})^T$$

$$SSB = \sum_{i=1}^{5} n_i (\bar{Y}_{i+} - \bar{Y}_{++})(\bar{Y}_{i+} - \bar{Y}_{++})^T,$$

where $\bar{Y}_{i+}$ is the group-means, and $\bar{Y}_{++}$ is the grand mean. Inserting the concrets data, we obtain

$$SSE = \begin{pmatrix} 43.07 & 55.46 \\ 55.46 & 132.33 \end{pmatrix}$$

$$SST = \begin{pmatrix} 54.42 & 88.73 \\ 88.73 & 271.76 \end{pmatrix}$$

$$SSB = \begin{pmatrix} 11.35 & 33.27 \\ 33.27 & 139.43, \end{pmatrix},$$

which indeed fits the decomposition formula given in (2). The REML estimates are called the moment estimates, and are given as the following

$$\tilde{\mu} = \bar{Y}_{++}$$

$$\tilde{\Sigma} = \frac{1}{N-k} SSE$$

$$\tilde{\Sigma}_0 = \frac{1}{n_0} \left(\frac{SSB}{k-1} - \tilde{\Sigma} \right),$$

where N is the number of observations, k is the number of groups and $n_0 = \dfrac{N - \sum_{i=1}^{k} n_i^2 / N}{k-1}$. Plugging in the already known numbers, we obtain the following moment estimates

$$\tilde{\mu} = \begin{pmatrix} 9.77 \\ 25.02 \end{pmatrix}$$

$$\tilde{\Sigma} = \begin{pmatrix} 0.979 & 1.260 \\ 1.260 & 3.008 \end{pmatrix} \tag{3}$$

$$\tilde{\Sigma}_0 = \begin{pmatrix} 0.205 & 0.779 \\ 0.779 & 3.515 \end{pmatrix}$$

Rewriting the Σ and Σ_0 as correlation matrices gives the correlations

$$\rho = 0.735$$
$$\rho_0 = 0.917.$$

The parameters can also be estimated by maximum likelihood. For the maximum likelihood estimation, there is no closed form solution so the estimates have to be found using numerical optimization. The log-likelihood function is given in theorem 5.12 in the book as

$$l(\mu, \Sigma, \Sigma_0; \bar{Y}_{1+}, \dots \bar{Y}_{k+}) = -\frac{N-k}{2}\log(\det(\Sigma)) - \frac{1}{2}\text{tr}((SSE)\Sigma^{-1})$$
$$- \sum_{i=1}^{k}\left[\log\left(\det\left(\frac{\Sigma}{n_i} + \Sigma_0\right)\right) + \frac{1}{2}(\bar{Y}_{i+} - \mu)^T\left(\frac{\Sigma}{n_i} + \Sigma_0\right)^{-1}(\bar{Y}_{i+} - \mu)\right] \tag{4}$$

Optimizing this function results in the following maximum-likelihood estimates

$$\tilde{\Sigma} = \begin{pmatrix} 0.968 & 1.254 \\ 1.254 & 3.007 \end{pmatrix}$$
$$\tilde{\Sigma}_0 = \begin{pmatrix} 0.197 & 0.800 \\ 0.800 & 3.619 \end{pmatrix}. \tag{5}$$

Rewriting the matrices as correlation matrices given the parameters

$$\rho = 0.735$$
$$\rho_0 = 0.947.$$

Comparing the results in (3) and (5), it is seen that the results are close to each other, but that the REML estimates for the variances are smaller - just as expected. The correlation estimate for ρ_0 is however smaller for the ML estimate.

Confidence Intervals for the Correlation

The correlation between the levels of 7-day strength and 28-day strength is given by the parameter ρ_0. Note that this is the correlation between 7-day strength and 28-day strength for observations from the same batch. The parameter has already been estimated to be 0.917 (REML) or 0.947 (ML).

The ML confidence intervals can be found by computing the profile-likelihood. This is done by considering the likelihood in (4). By fixing ρ_0 and optimizing for all other parameters the profile-likelihood of each value of ρ_0 can be found. A confidence interval is found by computing the profile-likelihood for a number of ρ_0-values, and normalizing such that the maximum log-likelihood is 0. The $1 - \alpha\%$ confidence interval is then found as the values of ρ_0, which are larger than the $1 - \alpha\%$ quantile in a $\chi^2(p)$, where p is the number of parameters which are profiled. As we only profile one parameter, the quantile is divided by 2, to account for boundary behaviour. Figure 2 shows the profile-likelihood function. From the figure it is seen that the confidence interval is quite large, namely [0.18;1]
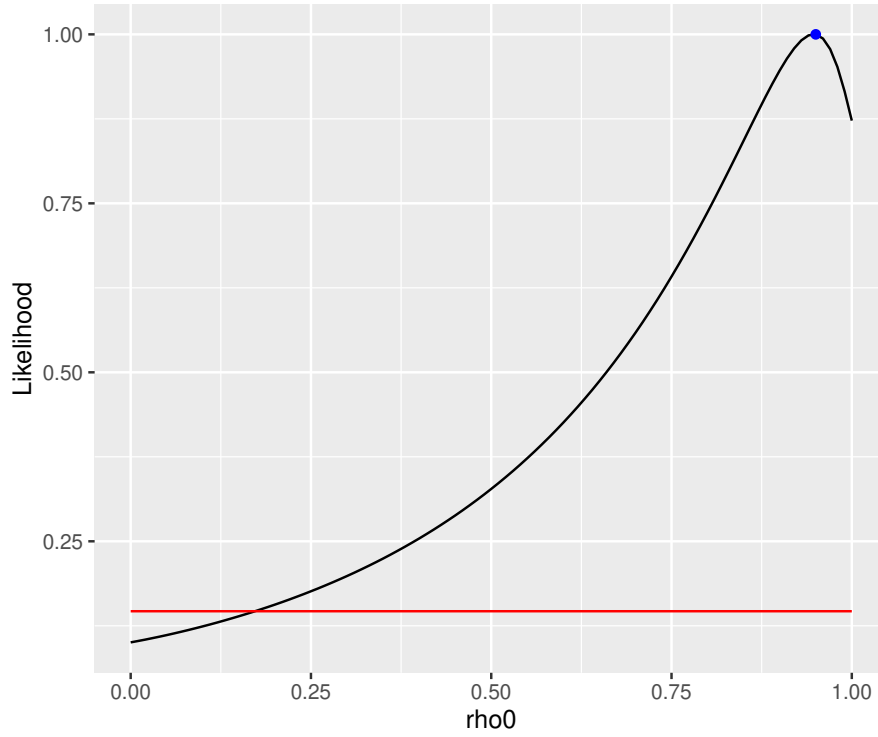
Figure 2: *Profile-likelihood for ρ_0. The red line is the relevant quantile line, and the blue dot is the ML-estimate.*

The confidence interval can also be based on REML, which requires a method to calculate the restricted likelihood of a set of parameters. To do this a one-dimensional model is considered, as this makes it easier to formulate the design-matrix. To formulate a one-dimensional model, we include an extra factor, which indicates if the measurement represent the 7-day strength or the 28-day strength. The model then becomes

$$Y_i = \mu + \alpha(\texttt{day}_i) + a(\texttt{batch}_i) + \epsilon_i,$$

where $\epsilon_i \sim N(0, \Sigma)$ and $a(\texttt{batch}_i) \sim N(0, \Sigma_0)$. By ordering the observations in the vector Y in the order $Y = [Y_{1,1,7}, Y_{1,1,28}, Y_{1,2,7}, Y_{1,2,28}, \ldots Y_{5,8,7}, Y_{5,8,28}]$, where $Y_{i,j,k}$ indicates the k-day strength of observation j in batch i, the variance-covariance matrix between the observations becomes

$$\Sigma_Y = \begin{pmatrix} \mathbf{S}_1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{S}_2 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{S}_3 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{S}_4 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{S}_5 \end{pmatrix},$$

where $\mathbf{S}_i$ is the variance-covariance matrix for observations within batch i. The structure of $\mathbf{S}_i$ is

$$\mathbf{S}_i = \begin{pmatrix} \Sigma + \Sigma_0 & \Sigma_0 & \cdots & \Sigma_0 \\ \Sigma_0 & \Sigma + \Sigma_0 & \cdots & \Sigma_0 \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_0 & \Sigma_0 & \cdots & \Sigma + \Sigma_0, \end{pmatrix}$$

with a total of n_i block-matrices. This makes the dimension of S_i $2n_i \times 2n_i$. The restricted log-likelihood-function is then given as

$$l(Y; \Sigma, \mu) = -0.5 \cdot \log(\det(\Sigma_Y)) - 0.5 \cdot (Y - \mu)^T \Sigma_y^{-1} (Y - \mu) - 0.5 \cdot \log(\det(X^T \Sigma^{-1} X)).$$

The profile-likelihood plot is seen in figure 3 The confidence interval is [0.1,1] which is slightly broader than the ML-estimate. This could however be due to REML-estimate of ρ_0 being smaller than the MLE.
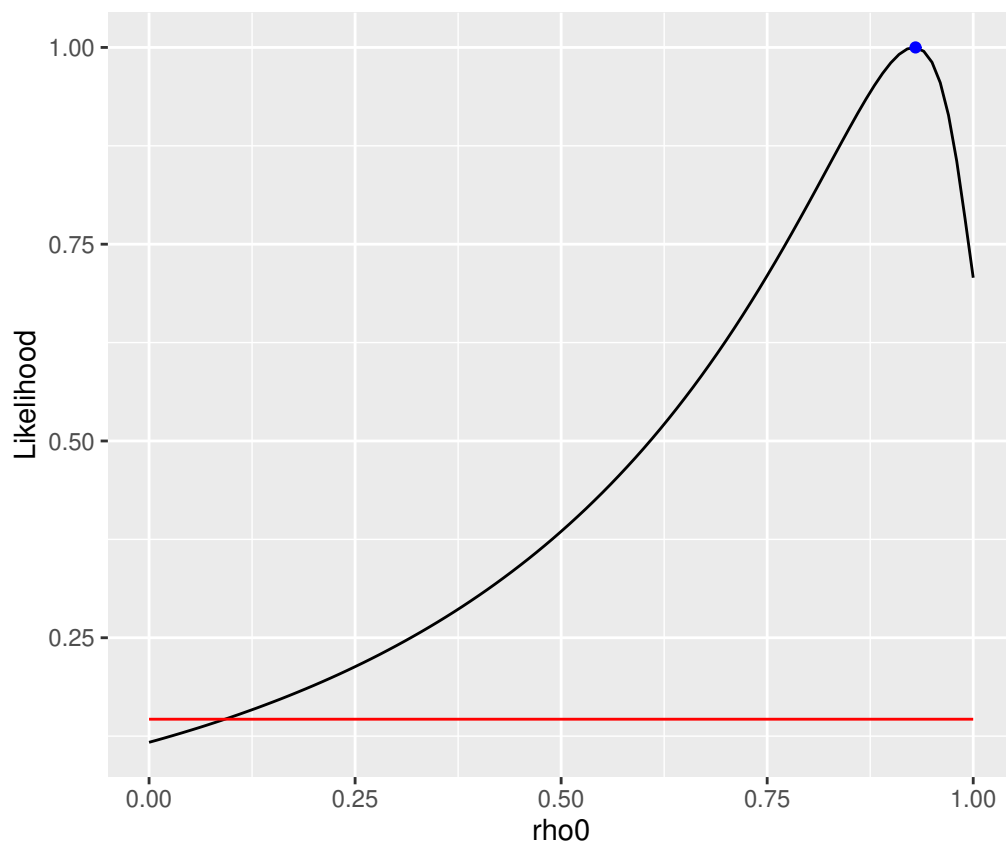
Figure 3: *Profile-likelihood for ρ_0 estimated with REML. The red-line is the relevant quantile line, and the blue dot is REML-estimate.*

The issue with the REML-estimated profile-likelihood confidence interval, is that REML-estimates of the other parameters does not correspond to the moment-estimates - which they should. The REML-estimates found through optimization are given below. It is seen that the estimates for Σ_0 does not correspond. The differences in the values of ρ and ρ_0 are however small.

$$\Sigma = \begin{pmatrix} 0.969 & 1.254 \\ 1.254 & 3.005 \end{pmatrix}$$
$$\Sigma_0 = \begin{pmatrix} 0.277 & 1.056 \\ 1.056 & 4.668 \end{pmatrix}$$
$$\rho = 0.734$$
$$\rho_0 = 0.928$$

Part 2: Clothing insulation count data

This part of the assignment concerns the clothing data. Table 4 gives a short description.

Variable	Type	Description
clo	Continuous	Number of times the subject changes clothes
nobs	Factor	Number of observations during the day
tOut	Continuous	Outdoor temperature
tInOp	Continuous	Indoor operating temperature
sex	Factor	Sex of the subject
time	Continuous	Total time of observation
subjId	Factor	Identifier for subject
day	Factor	Day (within the subject)

Table 4: *List of included variables in the clothing level data set.*

Problem A

Presentation of data

Figure 4 shows histograms of `clo` for each level of the variable `subjId`. It is seen that there is a great difference between each `subjId`.
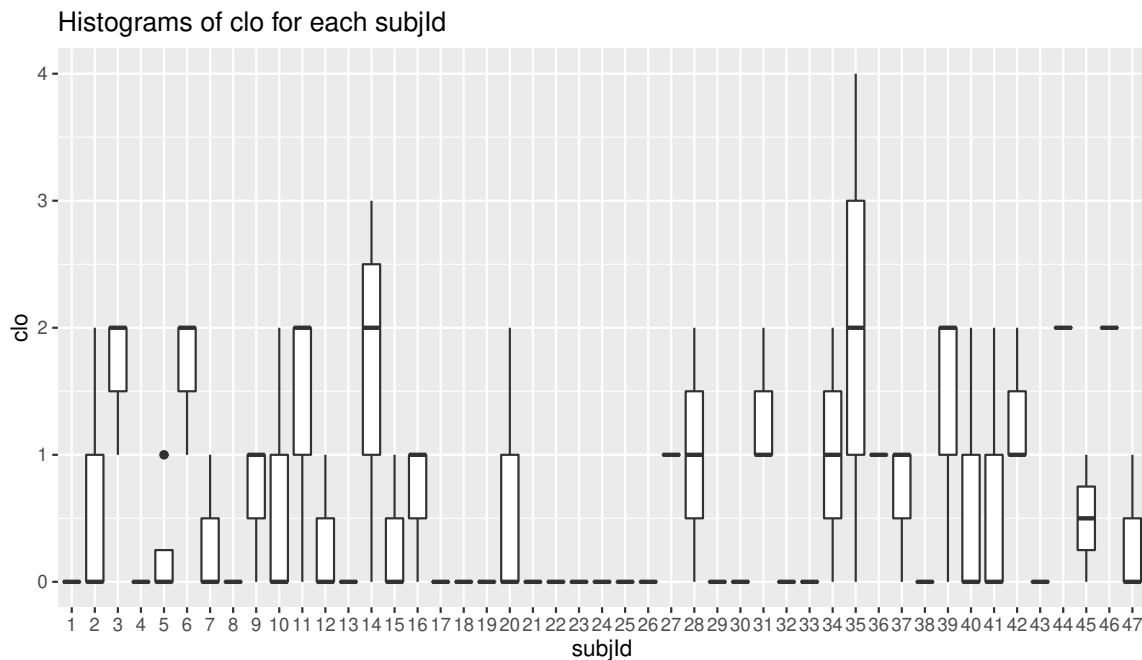
Figure 4: *Histogram of clo for each subjId*

We are not interested in a specific subject, but rather want to model a representation of the population. `subjId` can be seen as a representative of the population selected at random under varying experimental conditions, as to why `subjId` will be the random variable when modelling.

A General Mixed Model with glmmTMB

The variable `clo` is now modelled with a generalized mixed linear model with `subjId` as a random effect. The likelihood of such a model can be written in integral form by the marginal likelihood of the joint likelihood.

$$\mathcal{L}_M(\theta, y) = \int_{\mathbb{R}^q} L(\theta, u, y) \ du, \tag{6}$$

where u is the random effect, y is the observation and q is the dimension of the random effect space. A hierarchical setup is used. The first stage model, $f_{y|u}(y; u, \beta)$, specifies how we would model the observations,

for a known random effect (`subjId`). For the `clo` data-set both a binomial-model and a poisson-model can be used, as discussed in assignment 2. The second stage model, $f_U(u; \psi)$ specifies how the random effects are modelled. In both cases a Gaussian second stage model is used. This hierarchical set-up means that the likelihood described in (6) can be written as a product

$$L(\beta, \psi, u, y) = f_{y|u}(y; u, \beta) \cdot f_U(u; \psi), \tag{7}$$

and the parameters to estimate are β and ψ. The R-package glmmTMB can be used to estimating parameters on this form.

The generalized linear mixed model assumes a linear mean-value structure in some domain specified by a link-function, we can write.

$$\eta = X\beta + ZU$$
$$U \sim N(0, \psi)$$
$$\mathbb{E}[Y \mid U] = g^{-1}(\eta),$$

where $f_{y|u}(y; u, \beta)$ should belong to the exponential family. In this exercise the families used are the Poisson distribution, $y_i \sim \text{Pois}(\mu_i)$, and binomial distribution, $y_i \sim \text{Bin}(n_i, p_i)$ (first stage models). Several link functions have been tested, but for simplicity and because of unstable or bad results, these have been omitted. We found best results (lowest negative log-likelihood and convergence stability) when the Poisson model uses a log-link function, while the binomial distribution uses a logit link function. That is for the Poisson-model it holds that

$$Y \mid U \sim Pois(\mu_i)$$
$$\mu_i = g^{-1}(\eta) = \exp(\eta)$$

and for the binomial model it holds that

$$Y \mid U \sim Bin(n_i, p_i)$$
$$p_i = g^{-1}(\eta) = \frac{1}{1 + \exp(-\eta)}$$

For the Poisson-model the following structure of η is used as a start-guess. Note that `subjId` is used as the random effect. As a poisson-model estimates a rate, that is time-dependent, `time` is included as an offset

$$\eta = \mu + \alpha_1 (\text{nobs}_i) + \alpha_2 (\text{sex}_i) + \alpha_3 (\text{tOut}_i) + \alpha_3 (\text{tInOp}_i) + U_1 (\text{subjId}_i) + \log(\text{time}_i) \tag{8}$$

with $y_i \sim \text{Pois}(\mu_i)$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, $U_1 \sim N(0, \sigma_u^2)$. By formulating smaller models we find that the model with the lowest AIC, is a model including only `sex`, the off-set and the random effect

The linear structure of the binomial model follows the same structure as the Poisson model, except that the time is included as a variable and not an offset, and `nobs` is removed, as it is part of the Y-value. The linear structure thus becomes

$$\eta = \mu + \alpha_1 (\text{sex}_i) + \alpha_2 (\text{tOut}_i) + \alpha_3 (\text{tInOp}_i) + \alpha_4 (\text{time}_i) + U_1 (\text{subjId}_i)$$

After reducing this model, only `sex` and the random effect is left. Table 5 shows the AIC, BIC and log-likelihood of the best Poisson model and the best binomial model.

	AIC	BIC	logLik	deviance	df.resid
Poisson	265.5	274.2	-129.8	259.5	133
Binomial	263.4	272.1	-128.7	257.4	133

Table 5: *Comparison of the best binomial model and the best Poisson model*

Looking at the AIC and BIC of both models it can be seen that the binomial model has a slightly lower AIC and BIC. Using the restricted maximum likelihood (REML) we can get unbiased estimates Table 6.

	μ	$\alpha_1(sex)$	$g^{-1}(\alpha_1(female))$	$g^{-1}(\alpha_1(male))$
Poisson	-2.138	-1.083	0.1179	0.0399
Binomial	-1.686	-1.248	0.1563	0.0505

Table 6: *Estimates of the glmmTMB. The exponential and inverse logit of the estimates is used to show them in the original domain for each of the sexes*

From the estimates it is seen that female are more likely than men to change insulation level.
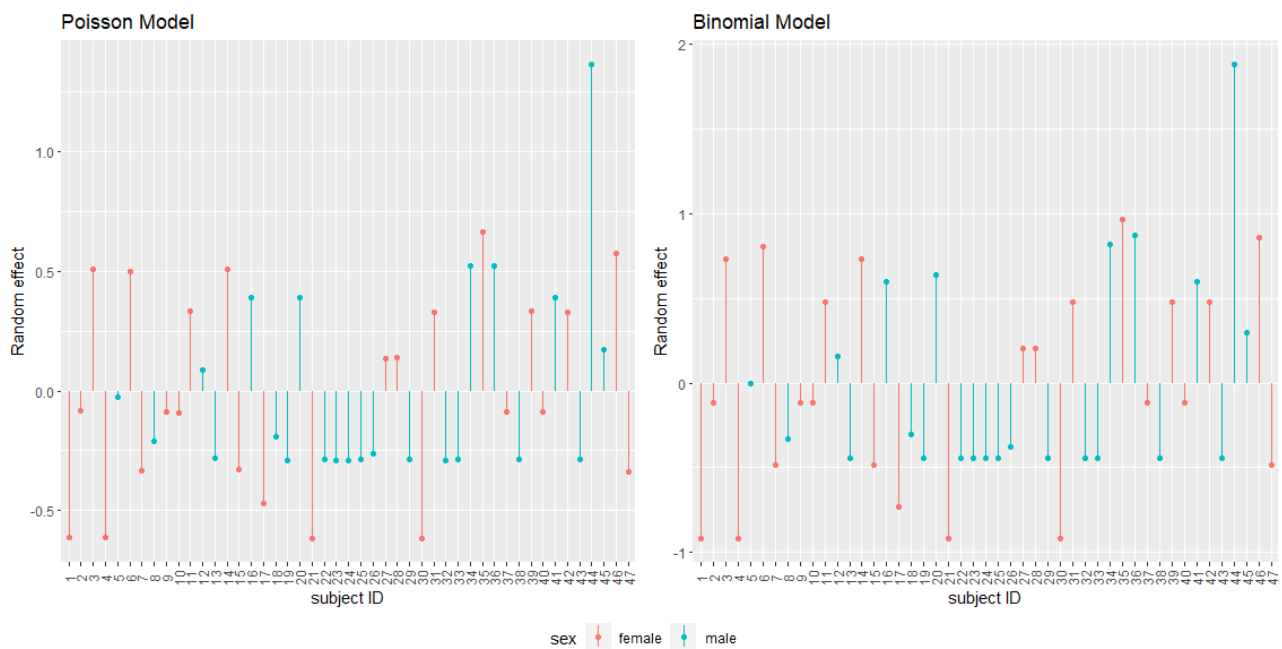

Figure 5: *To the left is the random effects of the Poisson model and to the right, the binomial model.*

One can see that the intercepts will change by quite a lot depending on subjects, as expected. Checking the assumptions of normality of the residuals for the random effects and fixed effects can be seen below,
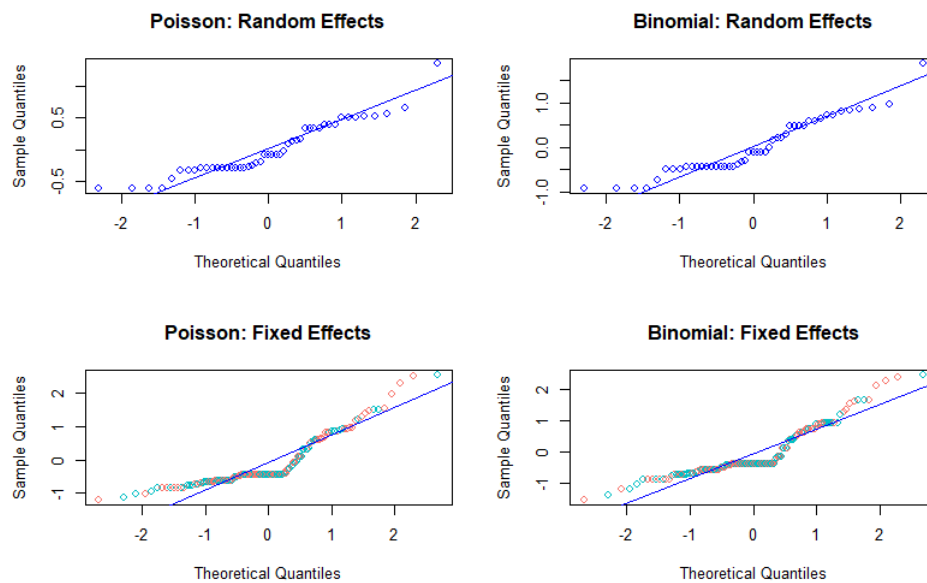
Figure 6: *QQ-plots of the random effects and the Pearson residuals of the models.*

Some deviations away from the normality assumptions is observed, this is expected based on the low counts of each subject. The same reason is why the Pearson residuals will be omitted, they do not make much sense to show or evaluate for this problem.

Looking at a goodness of fit test, it becomes clear that the models does not fit very well.

$$\text{Binomial Model: } P(\chi^2(133) \geq 257.4) = 5.76 \cdot 10^{-10},$$
$$\text{Poisson Model: } P(\chi^2(133) \geq 259.5) = 3.41 \cdot 10^{-10},$$

As the p-value is below 0.05, the models are not sufficient, indicating over-dispersion. One of the reasons might be that the random effects should not follow a Gaussian distribution assumption, a hierarchical model should therefore be taken in to consideration to model the over-dispersion.

At last we compare the predictions of the Poisson-model and the Binomial model. Figure Figure 7 shows boxplots of the predictions for the data-set for the two data-sets. The figure shows that the predictions in general are a bit smaller for the Poisson model.
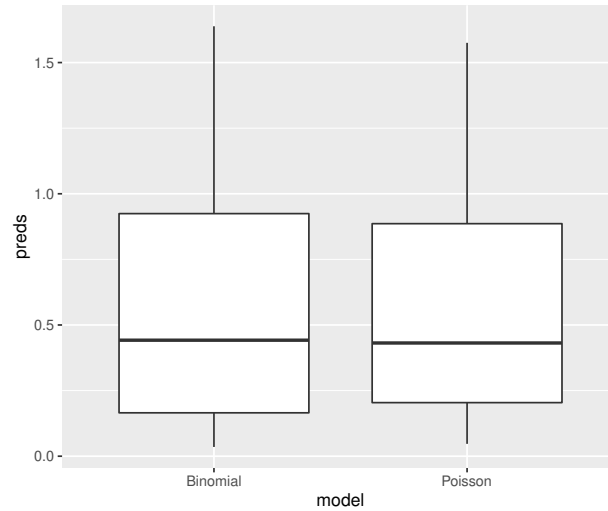
Figure 7: *Predictions of the glmmTMB model in the response-domain.*

Problem B

Implementing the Laplace Approximation

Now in an attempt to increase the fit of the data, we will consider adding non-Gaussian random effects in a hierarchical model. In practice we try fitting a Poisson model to the data, in which we include gamma-distributed random effects for each of the subjects. That means that we will fit a model of the form:

$$
\begin{aligned}
Y_{ij}|U_i &\sim \text{Pois}\left(\mu_{ij}U_i\right) \\
U_i &\sim G(\alpha, \beta) \\
\mu_{ij} &= e^{x_{ij}^T\theta}
\end{aligned}
\tag{9}
$$

Where Y_{ij} is the number of clothing changes in the jth observation for subject i and U_i is the gamma-distributed random effect for the ith subject. The intensity in the Poisson-distribution is given by the product of the random effect and μ_{ij}, which is given by the exponential of the explanatory variables with the coefficients. For this hierarchical model it is possible to calculate the exact marginal distribution by integrating out the random effects, which we will do in the next section.

To estimate the model given in Equation 9, we will first use Laplace approximation. The Laplace approximation is the second order Taylor approximation of the joint log-likelihood taken in the optimum ($\tilde{u}$) with respect to the random effects, i.e.

$$
\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) \approx \ell(\boldsymbol{\theta}, \tilde{\boldsymbol{u}}, \boldsymbol{y}) - \frac{1}{2}(\boldsymbol{u} - \tilde{\boldsymbol{u}})^T \boldsymbol{H}(\tilde{\boldsymbol{u}})(\boldsymbol{u} - \tilde{\boldsymbol{u}})
\tag{10}
$$

This means that in order to estimate we need to find the optimal random effects every time we try a new set of coefficients. Due to the numerous random effects, this is high dimensional and often time consuming. However, since the random effects are assumed to be independent, it is possible to optimise each of them individually. Using Equation 7 and the pdfs of the Poisson- and gamma distributions, the joint log-likelihood of the model becomes:

$$
l(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{x}, \alpha, \beta) = \log f_{Y|u}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{u}) + \log f_U(\boldsymbol{u}, \alpha, \beta)
\tag{11}
$$

But since we assume the random effects to have mean of 1, i.e $E[U_i] = 1$, and we know that the mean of the gamma-distribution is given as $\alpha \cdot \beta$, we need to have $\beta = \frac{1}{\alpha}$, which means that we do not have to consider the β parameter. This means that in the joint log-likelihood we have:

$$
\log f_{Y|u}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}) = \sum_{j,i} \log \frac{\lambda_{ij}^{y_{ij}} e^{\lambda_{ij}}}{y_{ij}!} \qquad \text{where} \quad \lambda_{ij} = \mu_{ij} \cdot u_i
\tag{12}
$$

$$\log f_U(\boldsymbol{u}, \alpha) = \sum_i \log \frac{\alpha^{\alpha}}{\Gamma(\alpha)} u_i^{\alpha-1} \cdot e^{-u_i \alpha} \tag{13}$$

Which means that the joint log-likelihood can also be written as:

$$l(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{x}, \alpha) = \sum_i \left[\log \left(\frac{\alpha^{\alpha}}{\Gamma(\alpha)} u_i^{\alpha-1} \cdot e^{-u_i \alpha} \right) + \sum_j \log \frac{\lambda_{ij}^{y_{ij}} e^{\lambda_{ij}}}{y_{ij}!} \right], \qquad \text{where} \quad \lambda_{ij} = \mu_{ij} \cdot u_i \tag{14}$$

In this expression $\mu_{ij} = \exp(\eta) = \exp\left(\theta_0 + \theta_1(\texttt{sex}_i = \texttt{Male}) + \log(time_{ij})\right)$, and this is the joint log-likelihood that will be approximated using the 2nd order Taylor approximation. Using Equation 10, we get that the Laplace Approximation of the log-likelihood of the model is given by:

$$l_{LA}(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x}) \approx \sum_i \left[\log \left(\frac{\alpha^{\alpha}}{\Gamma(\alpha)} u_i^{\alpha-1} \cdot e^{-u_i \alpha} \right) + \sum_j \log \frac{\lambda_{ij}^{y_{ij}} e^{\lambda_{ij}}}{y_{ij}!} \right] - \frac{1}{2} |\boldsymbol{H}(\tilde{\boldsymbol{u}})| \tag{15}$$

In this expression we have:

$$\boldsymbol{H}(\tilde{\boldsymbol{u}}) = - \left. l_{uu}''(\mathbf{u}, \boldsymbol{\theta}, \mathbf{y}, \boldsymbol{x}) \right|_{\mathbf{u}=\tilde{\mathbf{u}}} \tag{16}$$

Which is the second derivative in the optimal point of the random effects. Due to the complexity of the expression, the hessian $\boldsymbol{H}$ will be approximated numerically using the "numderiv" library in R. The optimal us at which the Taylor approximation is to be taken, is estimated using "nlminb", and keeping all the other us equal to 1. This will be done for every u which becomes the diagonal elements of the hessian for the whole model. Again due to the assumption on individual random effects, the hessian for the full model will become a diagonal matrix. The optimal coefficients of the model is estimated to be:

$$\theta_0 = -2.055853 \quad \theta_1 = -1.057773 \quad \alpha = 3.6980561 \tag{17}$$

Which means that in the model with the structure in Equation 9 we have that:

$$\mu_{ij} = \exp(-2.055853 - 1.057773(\texttt{sex}_i = \texttt{Male}) + \log(time_{ij})) \tag{18}$$

From these coefficients we see that the rate with which the males do clothing changes is $\exp(-1.05773) = 0.3472$ that of women in general. The random effects stems from a gamma-distribution with parameters:

$$\alpha = 3.6980561 \qquad \beta = \alpha^{-1} = 0.2704123$$

A histogram of the random effects for each of the 47 subjects is given in Figure 10. On the histogram the theoretical distribution with the estimated coefficients is drawn. It seems that low values of the random effects are too well represented.

Now in order to compare the estimated random effects to those found for the Poisson and binomial mixed effects models, we plot the random effects against each other. The result is shown in figure Figure 9. In this plot also the line with slope 1 is visible. If the random effects were to be equal for the different models, the points would follow the line perfectly which is not the case. For instance the Poisson-Gamma hierarchical model tends to be estimating smaller random effects than the mixed effects models.

Accuracy by importance sampling

Importance sampling is used to assess the accuracy of the log-likelihood by a re-weighing in estimating the integral. For general mixed effects models it simulates from a distribution that is density proportional to a second order Taylor approximation of the marginal likelihood:

$$\widetilde{L}(\boldsymbol{\theta}, \hat{\boldsymbol{u}}_{\theta}, \boldsymbol{Y}) = \exp \left\{ \ell(\boldsymbol{\theta}, \hat{\boldsymbol{u}}_{\theta}, \boldsymbol{Y}) - \frac{1}{2} (\boldsymbol{u} - \hat{\boldsymbol{u}}_{\theta})^T \left(-\ell_{uu}''(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{Y}) | \boldsymbol{u} = \hat{\boldsymbol{u}}_{\theta} \right) (\boldsymbol{u} - \hat{\boldsymbol{u}}_{\theta}) \right\} \tag{19}$$

Excluding a normalization constant, we see that the density above is the density of a multivariate normal distribution $\phi_{\hat{u}_{\theta}, \hat{V}_{\theta}}(\boldsymbol{u})$ with mean $\hat{\boldsymbol{u}}_{\theta}$ and covariance

$$\hat{\boldsymbol{V}}_{\theta} = \boldsymbol{H}^{-1}(\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}) = \left(-\ell_{uu}''(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{Y}) |_{\boldsymbol{u} = \boldsymbol{u}_{\theta}} \right)^{-1} \tag{20}$$

By using this approximation we can approximate the integral in Equation 21 to find the marginal likelihood

$$L_M(\boldsymbol{\theta}, \boldsymbol{Y}) = \int L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{Y})d\boldsymbol{u} = \int \frac{L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{Y})}{\phi_{\hat{u}_\theta, \hat{V}_\theta}(u)} \phi_{\hat{u}_\theta, \hat{V}_\theta}(\boldsymbol{u})d\boldsymbol{u} \qquad (21)$$

This is done by simulating random effects from the multivariate normal distribution with the corresponding mean and variance above. The approximation is then found to be

$$L_M(\boldsymbol{\theta}, \boldsymbol{Y}) = \frac{1}{N}\sum_{i=1}^{N} \frac{L\left(\boldsymbol{\theta}, \boldsymbol{u}^{(i)}, \boldsymbol{Y}\right)}{\phi_{\hat{u}_\theta, \hat{V}_\theta}\left(\boldsymbol{u}^{(i)}\right)} \qquad (22)$$

We use the importance sampling to verify the accuracy given by the Laplace approximation. Using 10000 simulations the results can be seen in Table 7

$N = 10000$	Importance Sampling	glmmTMB (Poisson)	Laplace Approximation
Negative log-likelihood	-132.5682	-131.1625	-131.1003

Table 7: *It is seen that the importance weighted approximations is very close to the true marginal likelihood of the Laplace approximation and the glmmTMB.*

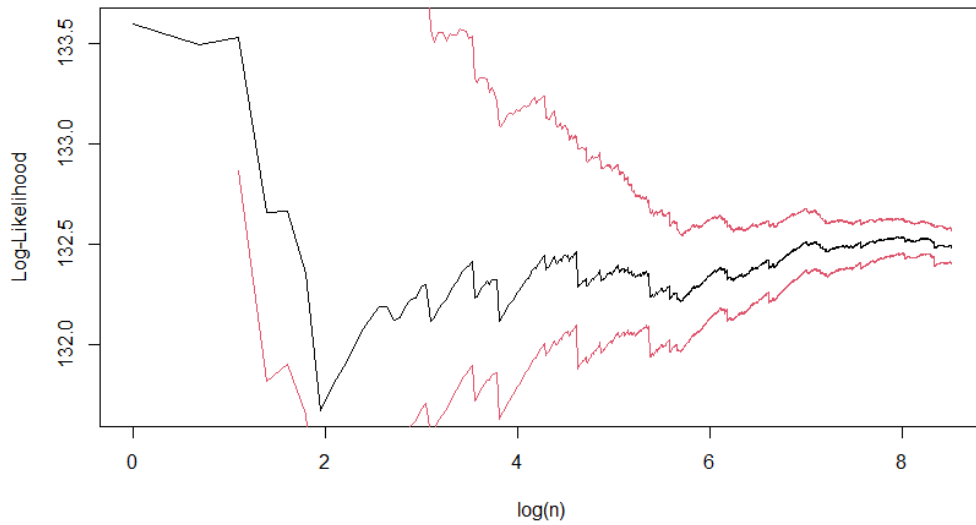Using Monte Carlo the convergence can be seen in Figure 8

Figure 8: *The convergence of the importance sampling. We see it slowly converges to the true likelihood.*

Problem C

Finding the Analytical Distribution of Y_i

The distribution of $Y_i = \left[Y_{i1}, Y_{i2}, \ldots Y_{i,n_i}\right]$ (that is observations from one subject) can be analytically shown to be the negative binomial distribution. To show this the likelihood given in (6) is evaluated using the product form of the likelihood given in (7). As specified in (9) we have

$$f_{Y_{ij}|U_i}(Y_{ij}; \mu_i \mid U_i) = \frac{(\mu_{ij}U_i)^{Y_{ij}}}{Y_{ij}!}\exp(-\mu_{ij}U_i)$$

$$f_{U_i}(U_i; \alpha, \beta) = \frac{1}{\beta^\alpha\Gamma(\alpha)}U_i^{\alpha-1}\exp\left(-\frac{U_i}{\beta}\right).$$

Inserting this in the integral yields the following likelihood for a single observation.

$$
\begin{aligned}
\mathcal{L}_M(Y_{ij}, \mu_{ij}, \alpha, \beta) &= \int_0^\infty \frac{(\mu_{ij} U_i)^{Y_{ij}}}{Y_{ij}!} \exp(-\mu_{ij} U_i) \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp\left(-\frac{U_i}{\beta}\right) dU_i \\
&= \frac{\mu_{ij}^{Y_{ij}}}{Y_{ij}! \beta^\alpha \Gamma(\alpha)} \int_0^\infty \exp(-\mu_{ij} U_i) U_i^{Y_{ij}+\alpha-1} \exp\left(-\frac{U_i}{\beta}\right) dU_i \\
&= \frac{\mu_{ij}^{Y_{ij}}}{Y_{ij}! \beta^\alpha \Gamma(\alpha)} \int_0^\infty U_i^{Y_{ij}+\alpha-1} \exp\left(-\frac{\mu_{ij} U_i \beta}{\beta} - \frac{U_i}{\beta}\right) dU_i \\
&= \frac{\mu_{ij}^{Y_{ij}}}{Y_{ij}! \beta^\alpha \Gamma(\alpha)} \int_0^\infty U_i^{Y_{ij}+\alpha-1} \exp\left(-\left(\frac{\mu_{ij}\beta+1}{\beta}\right) U_i\right) dU_i.
\end{aligned}
$$

Following the argumentation in the proof of theorem 6.1 in the book, the integrand is now the kernel of a gamma distribution with $\tilde{\alpha} = Y_{ij} + \alpha$ and $\tilde{\beta} = \left(\frac{\mu_{ij}\beta+1}{\beta}\right)^{-1}$. The normalization constant of a gamma distribution is

$$
\frac{1}{\Gamma(\tilde{\alpha})\tilde{\beta}^{\tilde{\alpha}}} = \frac{(\mu_{ij}\beta+1)^{\alpha+Y_{ij}}}{\Gamma(Y_{ij}+\alpha)\beta^{\alpha+Y_{ij}}},
$$

Thus

$$
\int_0^\infty U_i^{Y_{ij}+\alpha-1} \exp\left(-\left(\frac{\mu_{ij}\beta+1}{\beta}\right) U_i\right) dU_i = \left(\frac{(\mu_{ij}\beta+1)^{\alpha+Y_{ij}}}{\Gamma(Y_{ij}+\alpha)\beta^{\alpha+Y_{ij}}}\right)^{-1} = \frac{\Gamma(Y_{ij}+\alpha)\beta^{\alpha+Y_{ij}}}{(\mu_{ij}\beta+1)^{\alpha+Y_{ij}}}
$$

The likelihood for a single distribution thus becomes

$$
\mathcal{L}_M(\alpha, \beta, \mu_{ij}, Y_{ij}) = \frac{\mu_{ij}^{Y_{ij}}}{Y_{ij}! \beta^\alpha \Gamma(\alpha)} \frac{\Gamma(Y_{ij}+\alpha)\beta^{\alpha+Y_{ij}}}{(\mu_{ij}\beta+1)^{\alpha+Y_{ij}}}
$$

Using that

$$
\binom{Y_{ij}+\alpha-1}{Y_{ij}} = \frac{\Gamma(Y_{ij}+\alpha)}{\Gamma(Y_{ij}+\alpha-Y_{ij})Y_{ij}!} = \frac{\Gamma(Y_{ij}+\alpha)}{\Gamma(\alpha)Y_{ij}!},
$$

we can rewrite the likelihood to

$$
\begin{aligned}
\mathcal{L}_M(\alpha, \beta, \mu_{ij}, Y_{ij}) &= \binom{Y_{ij}+\alpha-1}{Y_{ij}} \frac{\mu_{ij}^{Y_{ij}} \beta^{Y_{ij}}}{(\mu_{ij}\beta+1)^{\alpha+Y_{ij}}} \\
&= \binom{Y_{ij}+\alpha-1}{Y_{ij}} \left(\frac{1}{\mu_{ij}\beta+1}\right)^\alpha \left(\frac{\mu_{ij}\beta}{\mu_{ij}\beta+1}\right)^{Y_{ij}},
\end{aligned}
$$

which is the density of a negative binomial distribution with $1 - p = \frac{1}{\mu_{ij}\beta+1}$, $k = Y_{ij}$ and $r = \alpha$. Notice that μ_{ij} is given as

$$
\mu_{ij} = \exp(x_{ij}\theta).
$$

To ensure that the distribution of U_i has mean 1, $\alpha = \beta^{-1}$ has to hold. Bearing this in mind only α and θ has to be estimated. The marginal likelihood of one observation thus becomes

$$
\mathcal{L}_M = \binom{Y_{ij}+\beta^{-1}-1}{Y_{ij}} \left(\frac{1}{\mu_{ij}\beta+1}\right)^{\beta^{-1}} \left(\frac{\mu_{ij}\beta}{\mu_{ij}\beta+1}\right)^{Y_{ij}}.
$$

This approach works if the value of the random effects for each observation is independent of all other random effects. The total likelihood can then be found by multiplying the likelihoods of all observations (or summing the log-likelihoods). In cases like this, where some of the observations has the same random effect, this has to be taken into account. The likelihood of one all observations with the same random effect, U_i, can then be found as

$$
\mathcal{L}_M = \int_0^\infty f_{Y_{i1}, Y_{i2}, \dots Y_{in_i}, \mu_i | U_i}(Y_{i1}, Y_{i2}, \dots Y_{in_i}, \mu_i \mid U_i) f_{U_i, \psi}(U_i, \psi) dU_i.
$$

We assume that the values of Y_{ij} follows $Pois(\mu_{ij}U_i)$ distributions independently. As we only use sex as an explanatory variable, and as the sex does not change within a subject $\mu_{ij} = \mu_{ik}$, and we can therefore write $\mu_{ij} = \mu_i$. The marginal likelihood therefore becomes

$$\mathcal{L}_M = \int_0^\infty \left(\prod_{j=0}^{n_i} \frac{(\mu_i U_i)^{Y_{ij}}}{Y_{ij}!} \exp(-\mu_i U_i) \right) \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp\left(-\frac{U_i}{\beta} \right) dU_i.$$

Using the same techniques as the for the marginal likelihood for one observation, this can be greatly simplified

$$\mathcal{L}_M = \int_0^\infty \exp(-\mu_i U_i)^{n_i} \left(\prod_{j=1}^{n_i} \frac{\mu_i^{Y_{ij}} U_i^{Y_{ij}}}{Y_{ij}!} \right) \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp\left(-\frac{U_i}{\beta} \right) dU_i$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \exp(-\mu_i U_i n_i) \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}} U_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\prod_{j=1}^{n_i} Y_{ij}!} U_i^{\alpha-1} \exp\left(-\frac{U_i}{\beta} \right) dU_i$$

$$= \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\beta^\alpha \Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \int_0^\infty U_i^{\sum_{j=1}^{n_i} Y_{ij}} U_i^{\alpha-1} \exp\left(-\frac{\mu_i U_i n_i \beta}{\beta} \right) \exp\left(-\frac{U_i}{\beta} \right) dU_i$$

$$= \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\beta^\alpha \Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \int_0^\infty U_i^{\sum_{j=1}^{n_i} Y_{ij}+\alpha-1} \exp\left(-U_i \frac{\mu_i n_i \beta + 1}{\beta} \right) dU_i$$

Again the integrand can be recognized as the kernel of a Gamma$(\tilde{\alpha}, \tilde{\beta})$-distribution with $\tilde{\alpha} = \sum_{j=1}^{n_i} Y_{ij} + \alpha$ and $\tilde{\beta} = \frac{\beta}{\mu_i n_i \beta + 1}$. The value of the integral therefore becomes

$$\int_0^\infty U_i^{\sum_{j=1}^{n_i} Y_{ij}+\alpha+1} \exp\left(-U_i \frac{\mu_i n_i \beta + 1}{\beta} \right) dU_i = \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}}.$$

Inserting this in the likelihood yields

$$\mathcal{L}_M(Y_i, \mu_i, \alpha, \beta) = \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\beta^\alpha \Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}}$$

$$= \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}},$$

where again $\beta = \alpha^{-1}$. The likelihood for all observations is found by multiplying the marginal likelihoods for all subjId's.

$$\mathcal{L}_M(Y, \mu, \alpha, \beta) = \prod_{i=1}^{47} \mathcal{L}_M(Y_i, \mu_i, \alpha, \beta) = \prod_{i=1}^{47} \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}} \tag{23}$$

This likelihood can be used to find the analytical MLE. First the marginal log-likelihood is determined, by taking the logarithm of the result in (23).

$$l_M(Y; \mu, \alpha, \beta) = \log \left(\prod_{i=1}^{47} \frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}} \right)$$

$$= \sum_{i=1}^{47} \log \left(\frac{\mu_i^{\sum_{j=1}^{n_i} Y_{ij}}}{\Gamma(\alpha) \prod_{j=1}^{n_i} Y_{ij}!} \frac{\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha) \beta^{\sum_{j=1}^{n_i} Y_{ij}}}{(\mu_i n_i \beta + 1)^{\sum_{j=1}^{n_i} Y_{ij}+\alpha}} \right)$$

$$= \sum_{i=1}^{47} \sum_{j=1}^{n_i} Y_{ij} \log(\mu_i) + \log(\Gamma(\sum_{j=1}^{n_i} Y_{ij} + \alpha)) + \sum_{j=1}^{n_i} Y_{ij} \log(\beta) - \log(\Gamma(\alpha))$$

$$- \sum_{j=1}^{n_i} \log(Y_{ij}!) - \left(\sum_{j=1}^{n_i} Y_{ij} + \alpha \right) \log(\mu_i n_i \beta + 1)$$

This is implemented in R. Note again that

$$\mu = \exp(X^T \cdot \theta)$$
$$\beta = \alpha^{-1},$$

where X is the design-matrix. X is parameterized with a common intercept (θ_0) and a factor for sex=male (θ_1). That is the mean-value becomes

$$\mu_i = \exp(\theta_0 + \theta_1 \cdot \mathbf{1}(\texttt{sex}_i = male))$$

The estimates become

$$\theta_0 = -0.1519$$
$$\theta_1 = -1.074$$
$$\alpha = 1.938.$$

Determining the Conditional Mean

The conditional mean of the random effect U_i given the observations within one subjId i can be founds using Bayes-rule.

$$
\begin{aligned}
g_{U|Y} &= \frac{f_{Y_i,U}(Y_i, U)}{g_Y(Y; \alpha, \beta)} \\
&= \frac{f_{Y_i|U_i}(Y_i \mid U_i)g_U(U)}{g_Y(Y; \alpha, \beta)} \\
&= \frac{\left(\prod_{j=0}^{n_i} \frac{(\mu_i U_i)^{Y_{ij}}}{Y_{ij}!} \exp(-\mu_i U_i)\right) \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp\left(-\frac{U_i}{\beta}\right)}{g_Y(Y; \alpha, \beta)} \\
&\propto U_i^{\sum_{j=1}^{n_i} Y_{ij} + \alpha - 1} \exp\left(-U_i \frac{\mu_i n_i \beta + 1}{\beta}\right),
\end{aligned}
$$

where the same steps ans in the previous questions have been used to obtain the last proportionality. This is again the kernel of a $G(\tilde\alpha, \tilde\beta)$ distribution with $\tilde\alpha = \sum_{j=1}^{n_i} Y_{ij} + \alpha$ and $\tilde\beta = \frac{\beta}{\mu_i n_i \beta + 1}$. The conditional mean and variance can then be found as the mean and variance of a $G(\tilde\alpha, \tilde\beta)$-distribution. That is

$$\mathbb{E}[U_i \mid Y_i] = \tilde\alpha\tilde\beta = \left(\sum_{j=1}^{n_i} Y_{ij} + \alpha\right)\left(\frac{\beta}{\mu_i n_i \beta + 1}\right)$$

$$\mathbb{V}[U_i \mid Y_i] = \tilde\alpha\tilde\beta^2 = \left(\sum_{j=1}^{n_i} Y_{ij} + \alpha\right)\left(\frac{\beta}{\mu_i n_i \beta + 1}\right)^2$$

The conditional means can be considered as the values of the random effects. Figure 10 shows a histogram of the random effects along with the Gamma-distribution with the MLE-parameters.

Problem D

To model the given data, we have tried four different approaches; two mixed effects models based on the binomial and Poisson distributions respectively, and a Poisson-Gamma hierarchical model. All the models have been based on sex as the only explanatory variable. Some of the models (glmmTMB and the Laplace-approximation) has included time as an off-set in the Poisson model. This is not included in the analytical model. To compare the effects, we have therefore fit new Poisson-model without the time off-set. Figure 9 shows the random effects for all the methods vs. the analytical, Poisson, random effects The black line is $x = y$. The exponential is taken of the random effects from the two glmmTMB-method, as the random effects are normal-distributed and not gamma-distributed. Because they follow a different distribution, we do not assume

them to be on the exact straight line. The random effects of the Laplace implementation and the analytical random effects are both found from the gamma-distribution, and one would assume that they were equal - the plot however shows that this is not the case. Also one can see from the histograms plotted in Figure 10, that the random effects of the Laplace approximations have what seems like too much representation in the lower values, whereas the analytical random effects seems more likely to be Gamma-distributed.

	Laplace approx. with offset	Laplace approx. without offset	Analytical log-likelihood
α	3.698056	3.593425	1.938804
β	0.2704123	0.278286	0.5157818

Table 8: *The parameters in the Gamma-distributions of the random effects in the different implementation of the Poisson-Gamma hierarchical model.*
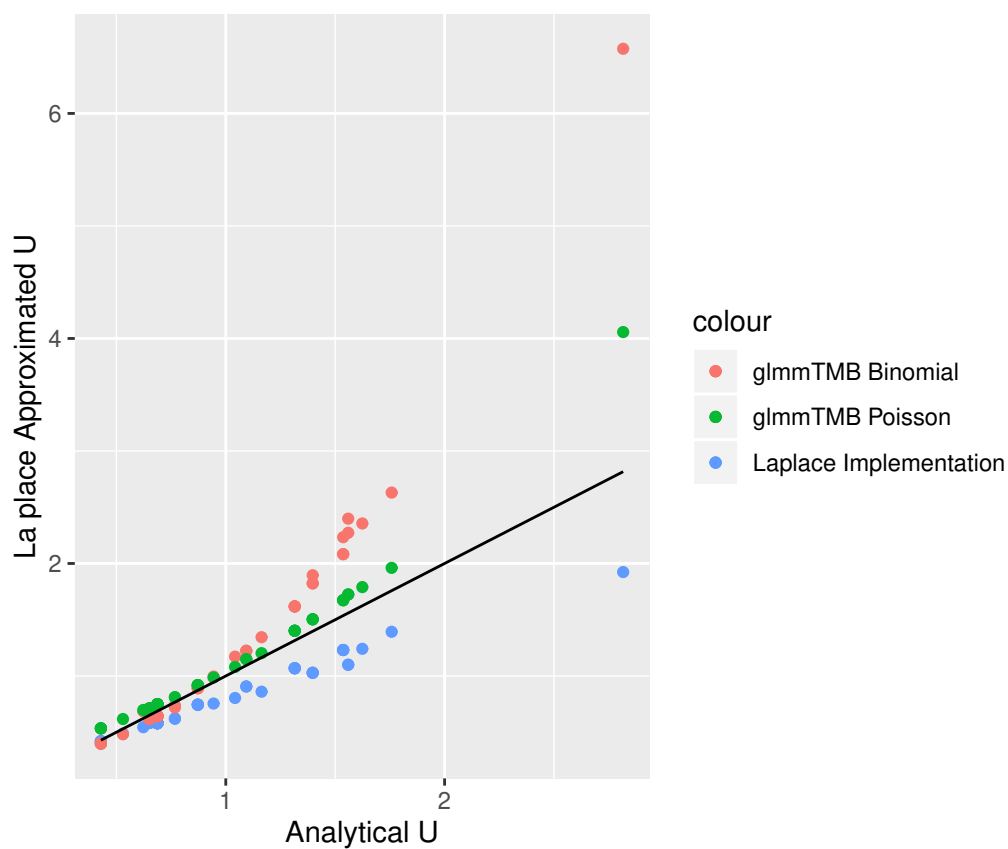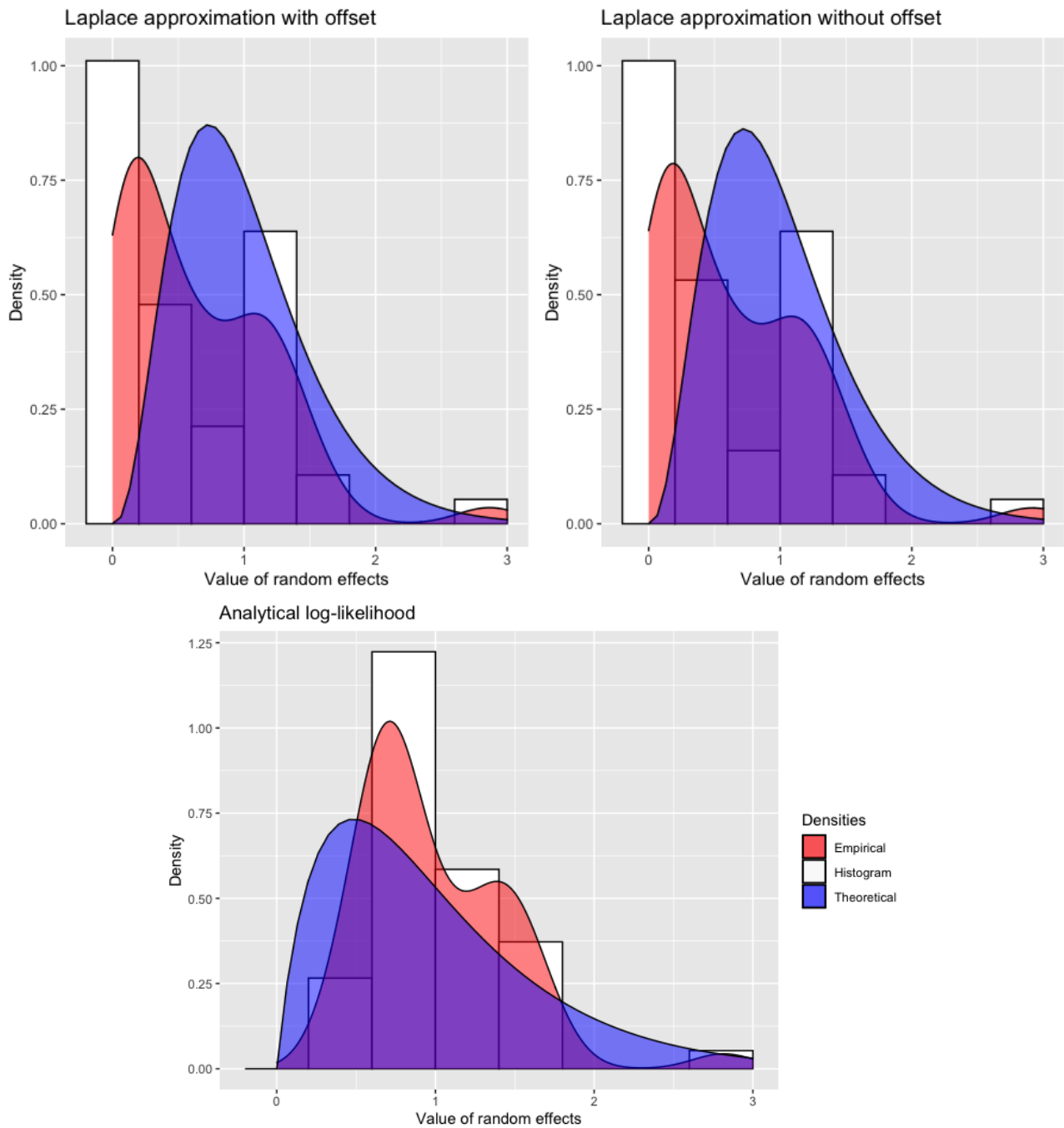

Figure 9: *Comparison of random effect*

Figure 10: *The distribution of the random effects from the three different ways of implementing the Poisson-Gamma hierarchical model. The parameters in used in the theoretical distributions are shown in Table 8.*