Homework 6

Sarah Adilijiang

Problem 2

(a)

```r
# simulate data
set.seed(123)
alpha = 1;  sigma = 1.2
u = rnorm(101, 0, sigma)
x = 2 + 0.01 * (0:100)
y = NULL
for (i in 1:101) {
    y[i] = rpois(1, exp(alpha*x[i]+u[i]))
}

# GLM model
glm_mod = glm(y ~ x-1, family=poisson)

# GLMM model
library(lme4)
n = 0:100
glmm_mod = glmer(y ~ x-1 + (1|n), family=poisson)
```
```r
# compare estimates of the parameters
summary(glm_mod)$coefficients
```

```
##   Estimate  Std. Error  z value Pr(>|z|)
## x 1.275661 0.007449102 171.2503        0
```
```r
summary(glmm_mod)$coefficients
```

```
##   Estimate Std. Error  z value      Pr(>|z|)
## x 1.022431 0.04780302 21.38842 1.71247e-101
```
```r
summary(glmm_mod)$varcor
```

```
##  Groups Name        Std.Dev.
##  n      (Intercept) 1.1555
```
```r
# compare log-likelihoods and other fitting criterions
c(logLik(glm_mod), logLik(glmm_mod))
```

```
## [1] -1798.4194  -420.2122
```
```r
c(AIC(glm_mod), AIC(glmm_mod))
```

```
## [1] 3598.8389  844.4244
```
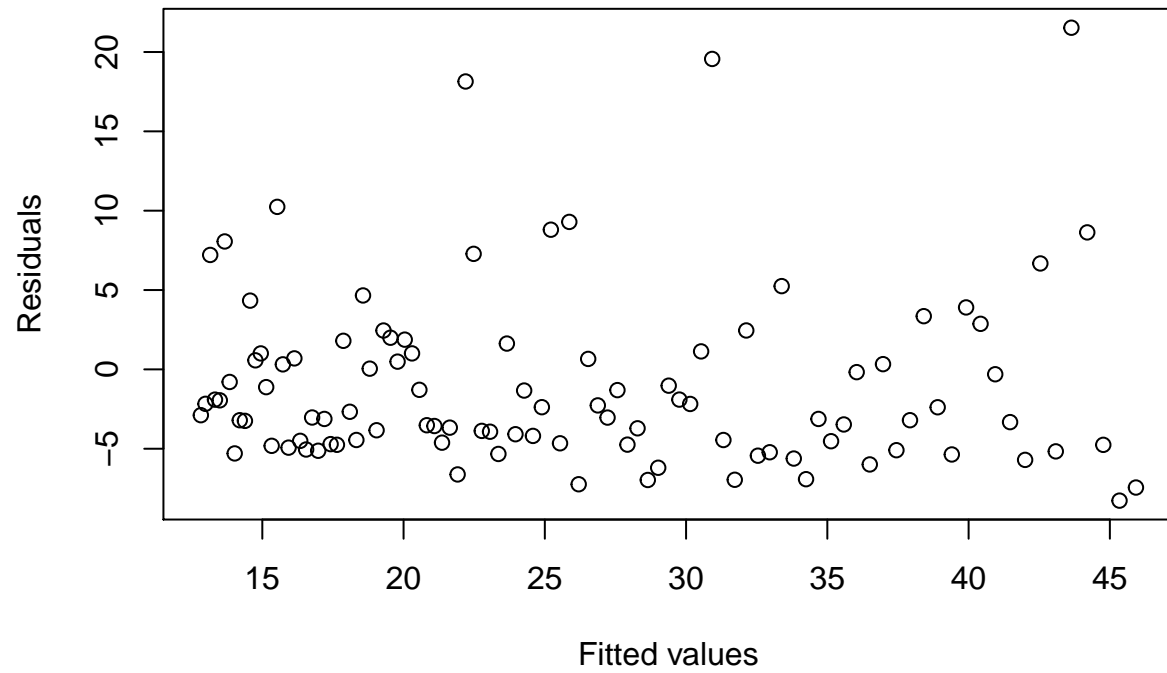```r
c(BIC(glm_mod), BIC(glmm_mod))
```
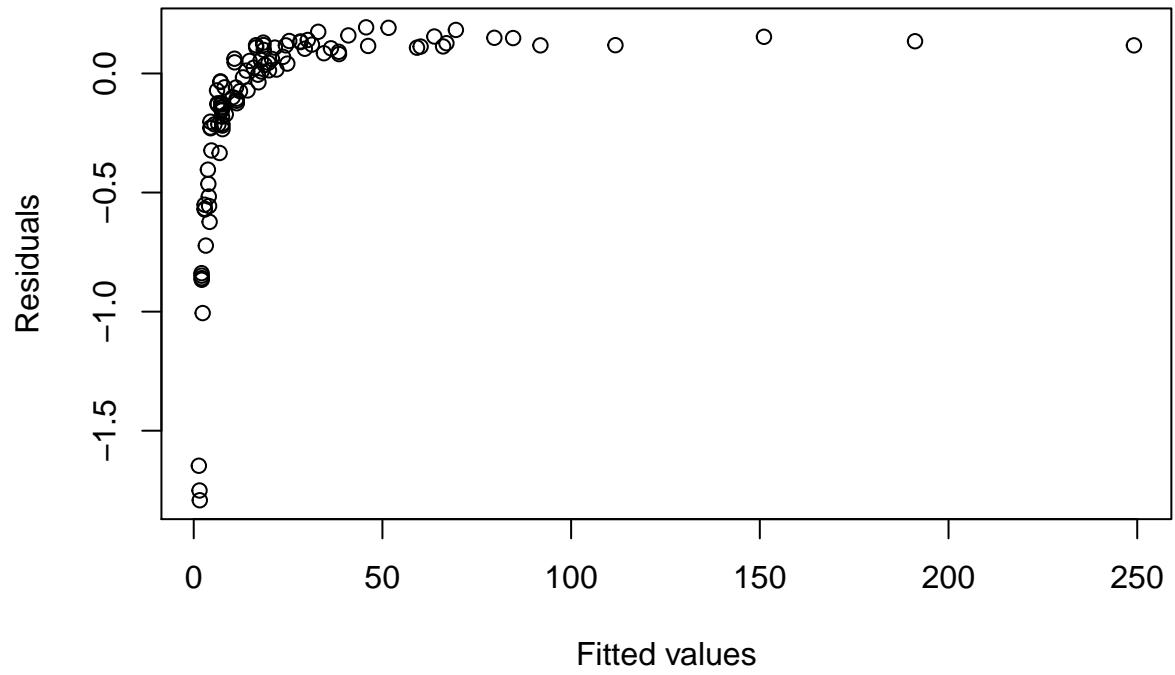
```
[1] 3601.4540 849.6547
```
```r
# compare the absolute values of the residuals
mean(abs(residuals(glm_mod)) >= abs(residuals(glmm_mod)))
```
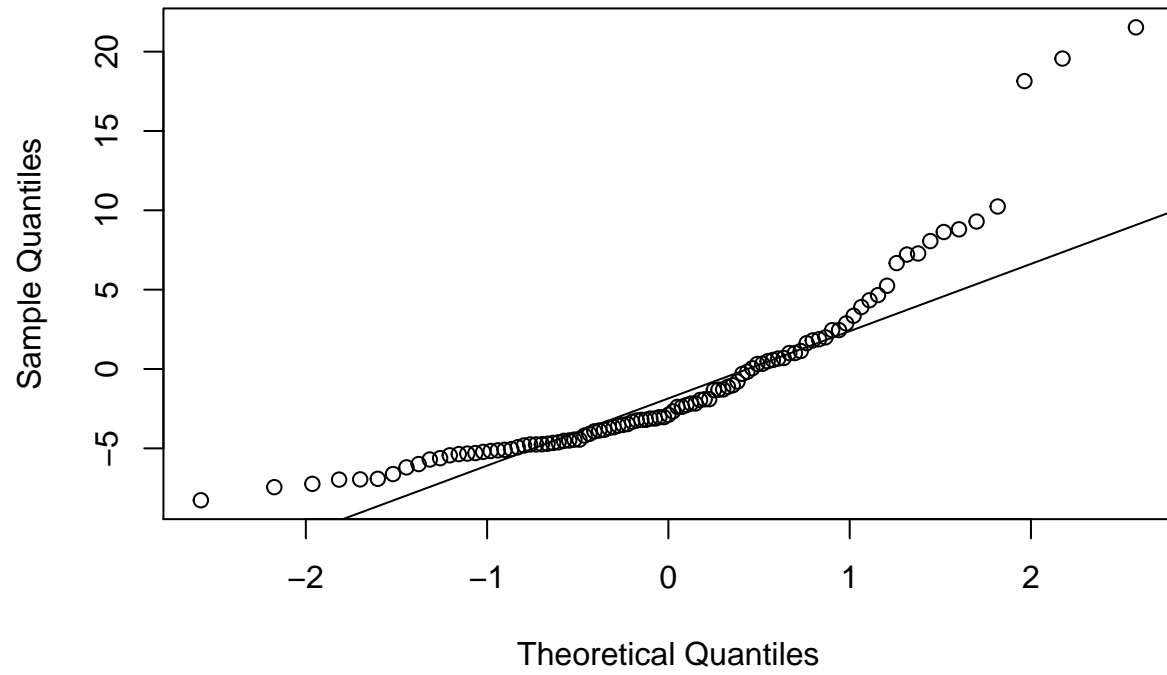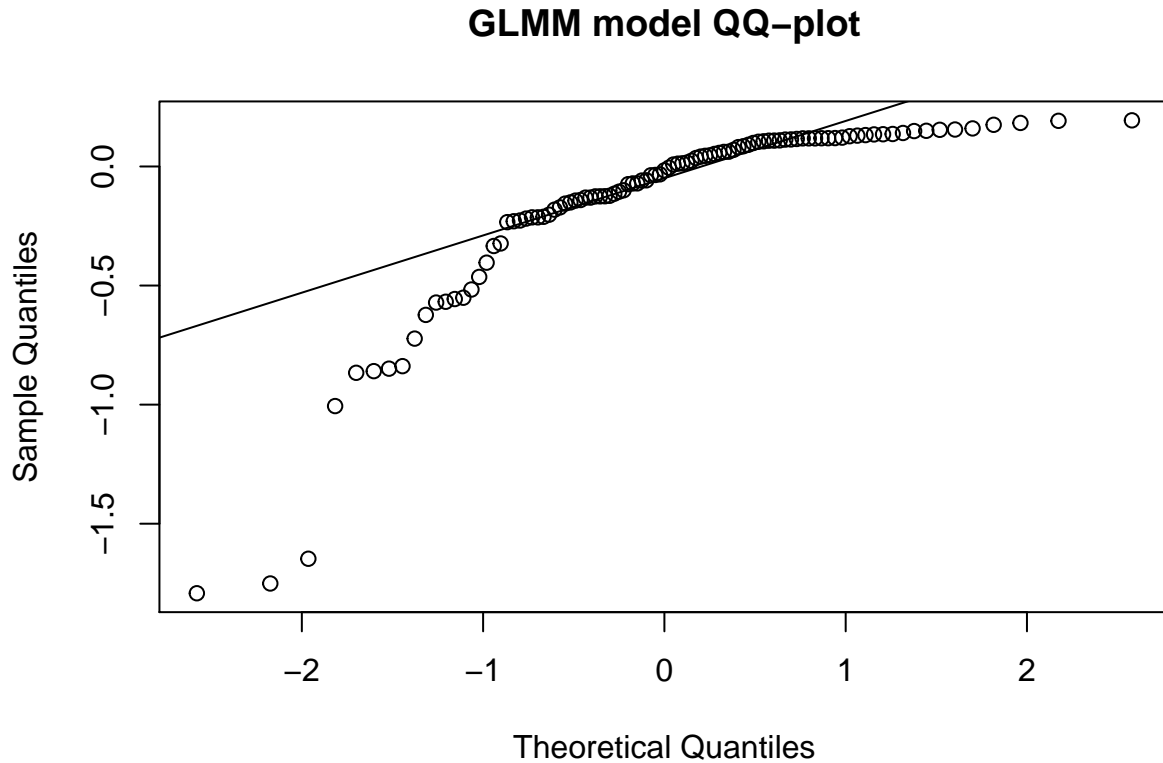
```
## [1] 0.990099
```

GLM model

GLMM model

Residuals vs Fitted values

GLM model QQ-plot

GLMM model QQ−plot

Answer:

The GLM model for Poisson (count) response here is:

likelihood: $P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y!}, \; y_i = 0, 1, 2, ...$

link function (log-link): $\eta_i = \log \mu_i$

linear predictor: $\eta_i = \beta X_i$

The GLMM model for Poisson (count) response here is:

likelihood: $P(Y_i = y_i) = \frac{e^{-\mu_{ij}} \mu_{ij}^{y_i}}{y!}, \; y_i = 0, 1, 2, ...$

link function (log-link): $\eta_{ij} = \log \mu_{ij}$

linear predictor: $\eta_{ij} = \beta X_i + \gamma_j$, where X_i is the fixed effects, γ_j is the random effects, and γ_j i.i.d. $\sim N(0, \sigma_b^2)$

Discusssion:

(1) Estimates of parameters:

The estimate of the coefficient of X_i in the **glmer** model is closer to its true value 1 than that in the **glm** model. But both the estimates in two models are significant.

The **glm** model estimates only one parameter, while the **glmer** model estimates two parameters including the random effect SD. The estimates of the randon effect SD is 1.1555, which has explained much of the variance of the response (true=1.2).

(2) Model fitting:

The log-likelihood of the **glmer** model is much larger than that of the **glm** model. And both the AIC and BIC values are much lower in the **glmer** model. Therefore, the **glmer** model fits the data much better than the **glm** model since the **glmer** model includes a random effect so that it explains the variation of the response better.

(3) Residuals:

Almost all the absolute residual values of the **glmer** model is smaller than those of the **glm** model. It makes sense since the random effect has explained much of the variation of the response.

The residuals ~ fiited values plot for the **glm** model do not show significant anomalous patterns, thus indicating a roughly constant variance in residuals. However, for the **glmer** model, there is a significant increasing and correlated pattern in the residuals, thus the constant variance assumption does not hold.

The QQ-plots for both models show that both of their residuals are not normally distributed.

(b)

See the next page.