Hands-on exercise: gamma models of insurance claim sizes.

You will consider the Danish fire insurance data set used in the notes. However, you are asked to remove some of the most extreme claims, since they cause numerical problems.

```
claims <- read.table(
    "http://www.math.ku.dk/~richard/regression/data/claims.txt",
    sep = ";",
    colClasses = c("character", "numeric", "numeric", "factor"))
claims2 <- subset(claims, claims < 1e7)
```

The response variable Y is the claim size and there are two potential predictors in the data set.

Question 1. Refit the linear, additive model

$$E(\log Y_i) = \beta_0 + \beta_{X_{i,\mathtt{grp}}} + \beta_{\mathtt{sum}} \log X_{i,\mathtt{sum}} \tag{1}$$

as considered on page 34 in the notes, but using the reduced data set. Construct model diagnostic plots for this model.

As show in Example 6.3, the log-normal regression model is also a generalized linear model with a log-link and a quadratic variance function. That is,

$$\log E(Y_i) = \tilde{\beta}_0 + \beta_{X_{i,\mathtt{grp}}} + \beta_{\mathtt{sum}} \log X_{i,\mathtt{sum}} \tag{2}$$

Question 2. Show that with a log-link and a quadratic variance function, $\mathcal{V}(\mu) = \mu^2$, then the weights in IWLS are always 1.

Question 3. Fit the model given by (2). Investigate if the model fits the data. Compare the fitted values with those from the log-normal model by a figure similar to Figures 2.2 and 2.10.

An alternative to the log-normal model is the log-gamma model, where (1) still specifies the mean value, but where the response is assumed to have a gamma distribution. Such a model can be fitted as a generalized linear model with a gamma response, but with the identity link.

Question 4. Fit the log-gamma model and discuss the model fit.

One difficulty with generalized linear models is that the residuals are typically not supposed to follow any specific distribution. Thus it is not obvious how to justify distributional assumptions beside the mean-variance structure. If the response has a continuous distribution with distribution function $F_{\mu,\psi}$ we can, however, construct a pp-plot by comparing

$$F_{\hat{\mu}_i,\hat{\psi}}(Y_i)$$

to the quantiles of the uniform distribution.

Question 5. Construct pp-plots for both the log-normal and log-gamma model and discuss which response distribution appears most appropriate.