# Exercises 3

1. Generate a data set using the following R code:

```
set.seed(33101)
b<-1
n<-500;
x<-rnorm(n,10,sd=2);
shape<-2;
y<-rgamma(n,shape,scale=b*(x^2)/shape)
```

   and fit the models

```
m1<-glm(y~x,gaussian(link = "identity"))
m2<-glm(y ~x, family=quasi(variance="mu^2", link="identity"))
m3<-glm(y~log(x),Gamma(link = "log"))
```

   Plot $x$ and $y$ and add lines for the expected responses from each model. (Useful R commands: `predict` (type `?predict.glm` for usage), `plot`, `order`, `lines`, `legend`)

2. Create an eps (encapsulated postscript) figure with four panels (subfigures). The first panel contains the plot from the previous exercise, the second panel presents the deviance residuals for the first model, the third panel presents the deviance residuals for the second model, and the fourth panel presents the deviance residuals for the third model. The figure should have two rows and two columns. (Useful R commands: `residuals`, `postscript`, `setwd`, `layout`, `plot`, `dev.off`.)

3. Carry out a simulation study on the accuracy of estimated variances of GLM parameters for the sample size 500. Generate covariate $x$ from exponential family with mean 10 and generate response $Y$ from Poisson($x$). Use $\log(x)$ as a covariate and fit a GLM for $y$ with the Poisson family and log-link. Store the regression coefficient for $x$ and the estimated standard deviation of the regression coefficient. These can be taken from the summary of the model. Repeat e.g. 1000 times and calculate the mean and the median of the estimates and the empirical standard deviation of the estimates. Report the results and the details of the simulation. (Useful R commands: `set.seed`,

`for` (type `help("for")`), `vector`, `matrix`, `rexp`, `rpois`, `summary`, `str`, `save`, `mean`, `median`, `sd`.)

4. Expand the simulation of the previous exercise for sample sizes 20, 50, 100, 200, 500, 1000, 2000 and 5000. Plot the mean of the estimates, estimated standard deviation of the estimate and empirical standard deviation of the estimate as functions of the sample size. (Useful R commands: `plot`, `boxplot`, `colMeans`, `apply`.)