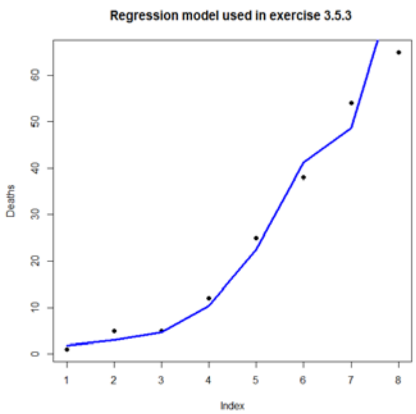
The book's model is log(deaths)~  log(population) + i □

where i = 1 for the age group 30–34 years , . . . , i = 8 for 65–69 years.

Plot below shows the implementation for this model.

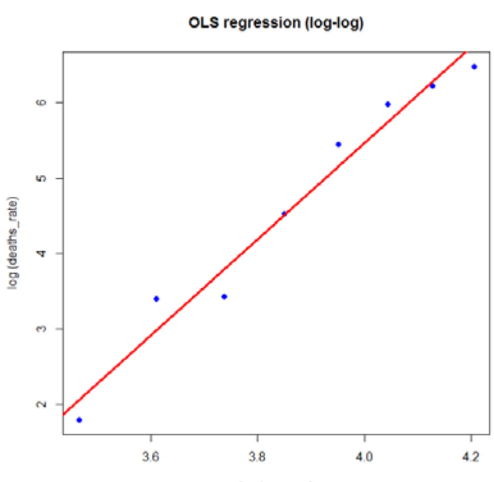


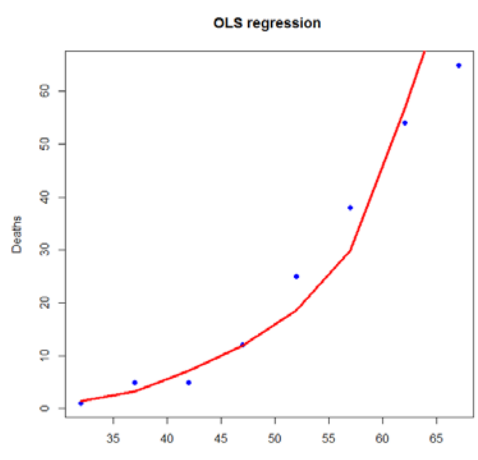

Now we use an OLS model: log(deaths_rate) ~ log(avg_age) where

deaths_rate = (deaths/population)*100000 and

avg_age = avg(30, 34) for the age group 30–34 years, …, = avg(65,69) for 65–69 years.

This model has $R^2$= 0.96 which is a good approximate although it is based on the Normal distribution not the Poisson distribution.



Finally, we use a GLM model for the deaths_rate:

glm(deaths ~ avg_age, offset=log(population +1), family="poisson")

The comparison of GLM and OLS for death rate shows that our OLS model provides a good prediction.

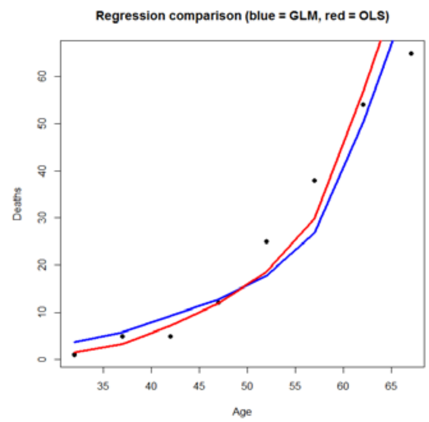



Table below shows the comparisons of the three models with actual data

| age group | actual deaths | book model | OLS | GLM |
|---|---|---|---|---|
| 30-34 | 1 | 2 | 1 | 4 |
| 35-39 | 5 | 3 | 3 | 6 |
| 40-44 | 5 | 5 | 7 | 9 |
| 45-49 | 12 | 10 | 12 | 13 |
| 50-54 | 25 | 22 | 19 | 18 |
| 55-59 | 38 | 41 | 30 | 27 |
| 60-64 | 54 | 49 | 57 | 50 |
| 65-69 | 65 | 83 | 87 | 79 |

## Appendix: R Code

```
#prepration of Data

#install.packages('dobson')

mortality <- dobson::mortality

mortality$age_index <- c(1:8)

age<-seq(from=30 , to=70, by=5)

for(i in 1:length(age)-1)

{avg_age[i]<- floor((age[i] + age[i+1]) /2)}

mortality$avg_age<- avg_age

mortality$deaths_rate<- round(mortality$deaths/mortality$population*100000, 0)


log.deaths <- log(mortality$deaths)

log.avg_age <- log(mortality$avg_age)

log.population <- log(mortality$population)

log.deaths_rate<- log(mortality$deaths_rate)


#regression - OLS

lm1<- lm(log.deaths_rate ~ log.avg_age)

#plots

#plot( lm1$fitted , resid(lm1))

plot( y= log.deaths_rate, x= log.avg_age, type='p', pch=19, col='blue',

     ylab = "log (deaths_rate)" , xlab= "log (avg_age)", main= "OLS regression (log-log)")

abline(lm1,col='red', lwd=3)


plot( mortality$avg_age, mortality$deaths, pch=19,col="blue",
```

```
     xlab= "Age", ylab="Deaths", main= "OLS regression ")
lines(mortality$avg_age, (exp(lm1$fitted)*mortality$population)/100000, col="red",lwd=3)


#regression - book, exercise 3.5.3
lm0_age_index<- lm(log.deaths ~  log.population + mortality$age_index )
#plots
plot( exp(lm0_age_index$fitted ))
plot( mortality$age_index, mortality$deaths, pch=19,col="black",
     xlab= "Index", ylab="Deaths", main= "Regression model used in exercise 3.5.3 ")
lines(mortality$age_index, exp(lm0_age_index$fitted) ,col="blue",lwd=3)


#regression - poisson - rate
glm2 <- glm(mortality$deaths ~ mortality$avg_age, offset=log(mortality$population +1),
         family="poisson",data=mortality)
#plots
plot( mortality$avg_age, mortality$deaths, pch=19,col="black",
     xlab= "Age", ylab="Deaths", main="Regression comparison (blue = GLM, red = OLS)")


lines(mortality$avg_age, glm2$fitted ,col="blue",lwd=3)
lines(mortality$avg_age, (exp(lm1$fitted)*mortality$population)/100000, col="red",lwd=3)
```