Exercise 29 Black-nosed lions
Trophy hunting is an important part of maintaining viable lion populations, if done properly. Removing male lions older than 6 years has little impact on the social structure of a pride, whereas taking younger males is more disruptive. Estimating the age of unknown lions is therefore an important task in managing lion populations. Analyse the "lion.csv" data set to find out whether the pigmentation of a lion's nose can be used to estimate its age.

1. Make a scatter plot of lion's age versus the pigmentation of the nose. Can you see a relationship in the data? What is the explanatory and what is the response variable?
2. Choose a suitable model and fit it to the data.
3. Report the estimated relationship between pigmentation and age in male lions, asses the quality of the fit and check your model assumptions.
4. Predict the age of a lion with 0%, 50% and 100% black pigmentation. What are potential problems with those predictions, which prediction would you trust, which not? Why?

Exercise 30 Regression towards the mean
Suppose a study measured the cholesterol levels of 100 men randomly chosen from a population. After their initial measurement, the men were put on a new drug therapy designed to reduce cholesterol levels. After one year, the cholesterol level of each man was measured again and compared to the first measurement. The researchers were delighted to find that the cholesterol levels had dropped on average in the men who previously had the highest levels. Their excitement dimmed, though, when they realized that the average cholesterol level has increased in the men who previously had the lowest cholesterol levels. What has happened here? Provide an explanation using a linear regression model, assuming that there is no effect of the drug therapy (that is, the two measurements should be perfectly correlated).

1. Use the following R code to simulate a dataset where each pair of measurements is drawn from a normal distribution. Each individual has the same mean and variance in both trials, but the mean of each individual's distribution varies. We would thus expect a perfect 1-1 relationship in the data.

```
# calculate a vector of 100 means, each value represents the mean
# cholesterol of an individual

mean.vec = rnorm(100,5,1)

# draw two independent measurements for each individual note that the
# distrbiution of val1[i] and val2[i] are the same for each individual i

val1 = rnorm(100,mean = mean.vec,sd=0.5)
val2 = rnorm(100,mean = mean.vec,sd=0.5)

# plot the data

plot(val1,val2,xlab="measurement 1",ylab= "measurement 2")
```

2. Fit a linear regression model to the data. What estimate do you get for the slope?
3. Explain how this analysis can help the scientists understand the effect they saw in their study. Hint: try to consider extreme measurements, e.g., when you measure a very large or very small value in the first trial, what do you expected in the second trial.