

STAT 210

Applied Statistics and Data Analysis:

Problem list 7

(Due on week 8)

Exercise 1

The data for this problem is in the file `PL725F_P1.csv`, which has three variables, namely `weight`, `height`, and `species`. Read the data into a data frame named `q1_data`. The data correspond to measurements taken on samples of two species of draxen.

- Do a scatterplot of `weight` as a function of `height`, including the regression line. Fit a regression and print the summary table. Interpret the output in the table. State explicitly the assumptions that underlie this model. Write down the equation for the regression line, and give an interpretation of the parameters. Predict the `weight` for a draxen with `height` = 62.8 cm and include a confidence interval at the 98% level. Include your comments on every step that you take.
- There are two species of draxen in the file, denoted A and B, and this characteristic is available in the categorical variable `species`. If this variable was not read as a `factor`, transform it before you continue. Do a scatterplot of `weight` as a function of `height` and color the dots by species. Comment on what you observe. Fit a different model for each species and add the corresponding regression lines to the scatterplot. Compare the three models that you have fitted and comment. Write down equations for the two new models and compare with the previous one. Predict the value of the `weight` for draxen of both species having height 62.8 cm, including confidence intervals at the 98% confidence level. Compare with the previous prediction. Include your comments on every step that you take.

Solution

Start by reading the data

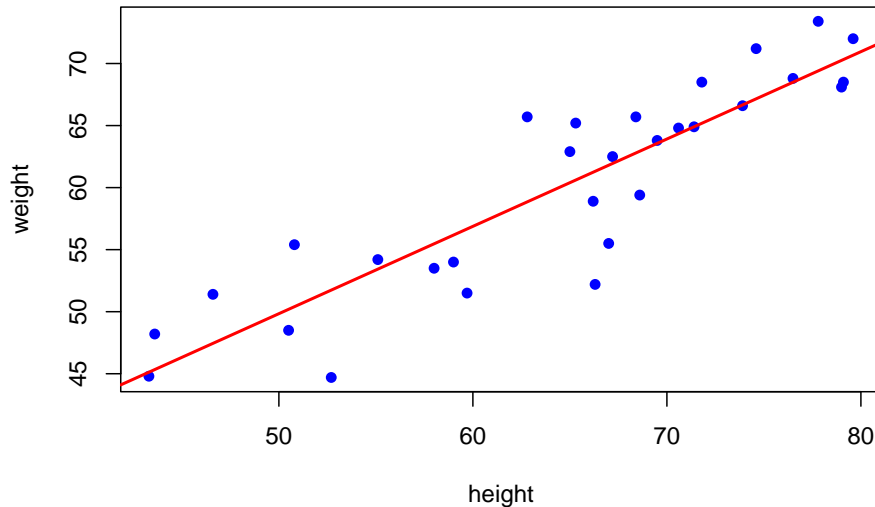
```
dt1 <- read.csv('PL725F_P1.csv')
str(dt1)
```

```
## 'data.frame':   30 obs. of  3 variables:
##  $ weight : num  65.7 58.9 64.8 48.5 54.2 51.5 68.8 62.9 71.2 64.9 ...
##  $ height : num  62.8 66.2 70.6 50.5 55.1 59.7 76.5 65 74.6 71.4 ...
##  $ species: chr   "B" "A" "B" "A" ...
```

```
dt1$species <- factor(dt1$species)
```

- Plot with regression line

```
plot(weight ~ height, data = dt1, pch = 16, col = 'blue')
mod1 <- lm(weight ~ height, data = dt1)
abline(mod1, col = 'red', lwd = 2)
```



In the plot we see an increasing relation between `weight` and `height`. The regression line follows this trend adequately, although there is a good amount of variability that is not captured by the regression. The summary table is printed below

```
summary(mod1)
```

```
##
## Call:
## lm(formula = weight ~ height, data = dt1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.1119 -2.1052  0.2726  2.8987  6.8514
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.64967    4.39109   3.336  0.00241 **
## height      0.70380    0.06703  10.499 3.25e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.847 on 28 degrees of freedom
## Multiple R-squared:  0.7975, Adjusted R-squared:  0.7902
## F-statistic: 110.2 on 1 and 28 DF,  p-value: 3.247e-11
```

The first section of the table gives a summary of the values for the residuals. The median value is reasonably close to zero while the other values on the table are approximately symmetric.

In the central part we have the estimated coefficients for the model, 16.497 for the intercept and 0.704 for the slope. We are also given the standard errors for these estimates, the test statistic and a p -value, which corresponds to testing the null hypothesis that the corresponding coefficient is equal to zero. Both p -values are small. The last part of the summary gives the estimated value for the empirical standard deviation of the residuals (Residual standard error), which is 3.85, the R^2 , which is approximately 79-80%, and, in the last line, the overall test for the significance of the regression, which in the simple linear regression case is equivalent to the t -test for the slope.

The simple linear regression model is based on the assumptions that the errors are independent and have a centered normal distribution with mean zero and common (unknown) variance σ^2 . The equation for the model is

$$\text{weight} = 14.65 + 0.704 \times \text{height}.$$

The intercept corresponds to the value of `weight` when `height` is zero and the slope is the increase in `weight` in kg when `height` increases one cm.

The predicted value for the `weight` is

```
predict(mod1, data.frame(height = 62.8), interval = 'c', level = 0.98)
```

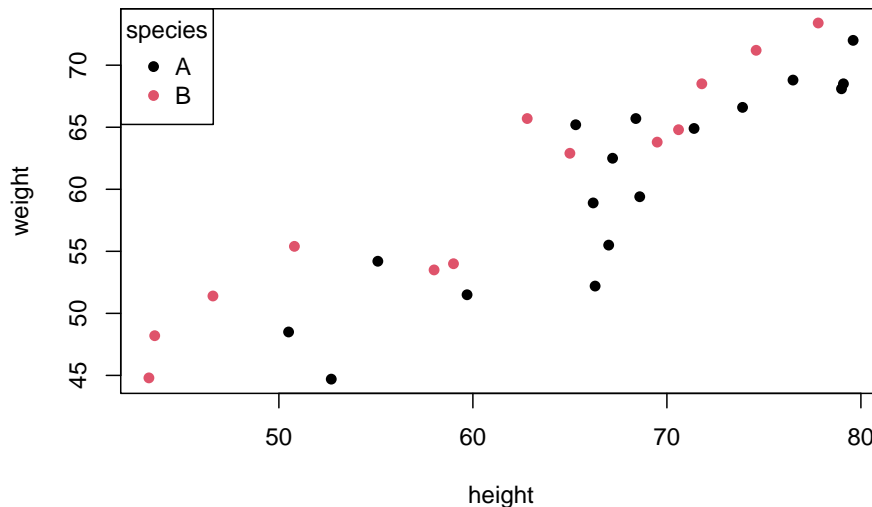
```
##           fit           lwr           upr
## 1 58.84858 57.08848 60.60868
```

(b) Fit separate models for the two species

```
mod2 <- lm(weight ~ height, data = dt1[dt1$species=='A',])
mod3 <- lm(weight ~ height, data = dt1[dt1$species=='B',])
```

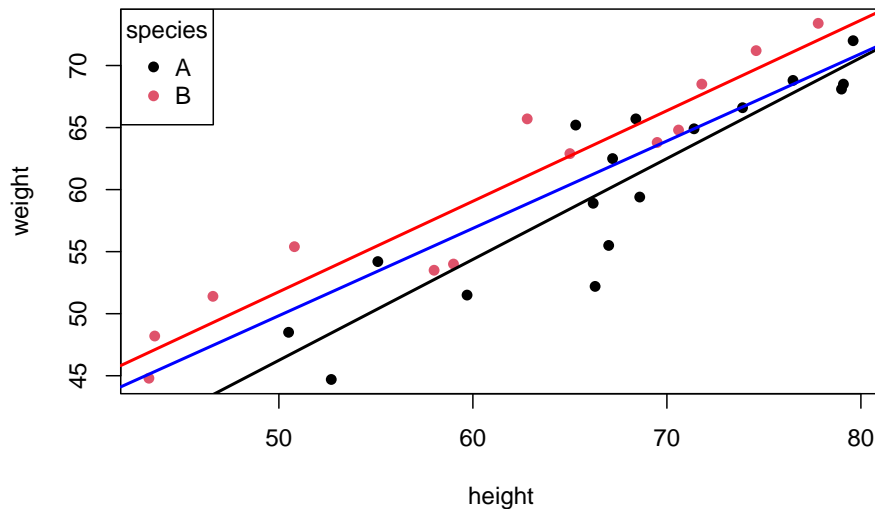
Scatterplot with points colored according to species and corresponding regression lines

```
plot(weight ~ height, data = dt1, col = species, pch = 16)
legend('topleft', c('A', 'B'), col = 1:2, pch = 16, title = 'species')
```



From the colored scatterplot we see that red dots seem to be higher than black dots. This points to having different regression models according to species. We fit separate models and add the lines to the plot, including the line corresponding to the initial model, in blue:

```
plot(weight ~ height, data = dt1, col = species, pch = 16)
legend('topleft', c('A', 'B'), col = 1:2, pch = 16, title = 'species')
abline(mod2, col = 'black', lwd = 2)
abline(mod3, col = 'red', lwd = 2)
abline(mod1, col = 'blue', lwd = 2)
```



And now we see that the difference in slope is small but there is a difference in intercepts. We can check this looking at the coefficients in the summary tables

```
summary(mod2)
```

```
##
## Call:
## lm(formula = weight ~ height, data = dt1[dt1$species == "A",
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.296 -1.966  0.925  1.851  6.517
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.5897     6.8639   0.814   0.428
## height        0.8131     0.1009   8.055 7.92e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.616 on 15 degrees of freedom
## Multiple R-squared:  0.8122, Adjusted R-squared:  0.7997
## F-statistic: 64.88 on 1 and 15 DF,  p-value: 7.916e-07
```

```
summary(mod3)
```

```
##
## Call:
## lm(formula = weight ~ height, data = dt1[dt1$species == "B",
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3337 -2.0783  0.8269  1.4840  4.5937
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.28484     4.28790   3.565  0.00444 **
```

```
## height      0.72964    0.06905  10.567 4.25e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.854 on 11 degrees of freedom
## Multiple R-squared:  0.9103, Adjusted R-squared:  0.9022
## F-statistic: 111.7 on 1 and 11 DF,  p-value: 4.25e-07
```

From the tables we see that the slope is significantly different from zero in both models but the intercept is only significantly different from zero in the second model, for the B species. The table below gives a comparison of the coefficients for the three models fitted:

	Intercept	Slope
All	14.65	0.704
A	5.59	0.813
B	15.28	0.729

We see that the values for the slopes are not too different but the intercept for species A is smaller than for the other two models. Also, The smallest slope corresponds to the full model, including both species.

The equation for the model for species A is

$$\text{weight} = 5.59 + 0.813 \times \text{height}.$$

while for species B we have

$$\text{weight} = 15.28 + 0.729 \times \text{height}.$$

The predicted value for species A is

```
predict(mod2, data.frame(height = 62.8), interval = 'c', level = 0.98)
```

```
##          fit          lwr          upr
## 1 56.64996 54.06229 59.23764
```

while for species B it is

```
predict(mod3, data.frame(height = 62.8), interval = 'c', level = 0.98)
```

```
##          fit          lwr          upr
## 1 61.10629 58.92955 63.28303
```

The predicted value with the initial model was

```
predict(mod1, data.frame(height = 62.8), interval = 'c', level = 0.98)
```

```
##          fit          lwr          upr
## 1 58.84858 57.08848 60.60868
```

We see that the predicted value with the full model is in between the predicted values for the two models fitted for the species, which is reasonable since the regression line is in between. Note also that there is less uncertainty in the prediction with the full model since the confidence interval is thinner. The reason for this is that we have more data to fit this model.

Exercise 2

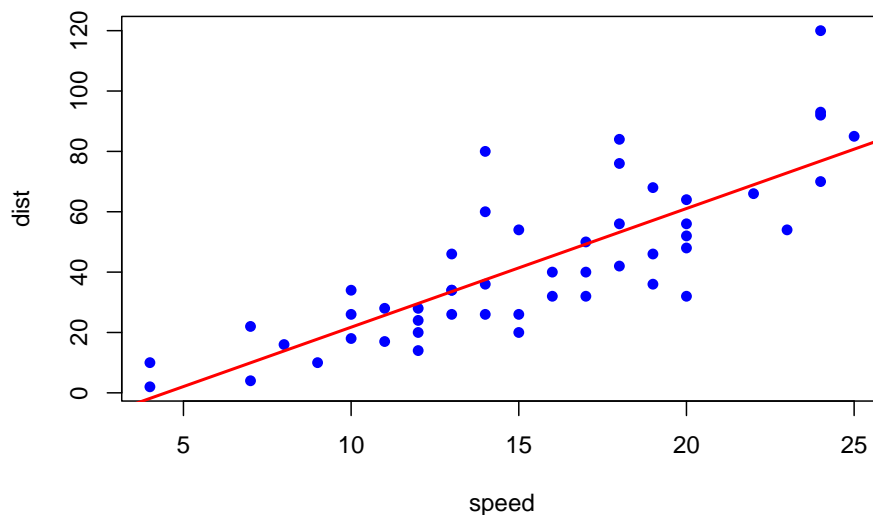
For this problem we use the data set `cars`, available in R. The data give the speed of cars and the braking distances. Note that the data were recorded in the 1920s.

- (a) Plot the data. Fit a regression model and add the regression line. Print a summary of the regression and interpret the output. Write down the equation for the regression line, and give an interpretation of the parameters.
- (b) An important step after fitting a model is to check whether the assumptions on which the model was built are satisfied. This is a topic that will be discussed in next week's videos, but as an advance you will use two statistical tests for normality and homogeneous variances, respectively. The first test is the Shapiro-Wilk test that we have used previously. You should use this test on the standardized residuals, which can be obtained with the command `rstandard(mod)`, where `mod` is the name of the model you fitted using the `lm` function. The Levene test for homogeneous variances that we used for ANOVA is not suitable for a regression model because it requires that the data be grouped, which is not the case here. We use the `ncvTest` function available in the `car` package. The argument for this function is the name of the model that you fitted. Use both tests for the model fitted in (a) and comment. Do you think the assumptions are satisfied?
- (c) Fit a new model for the square root of `dist` as a function of speed. Do a scatterplot including the regression line. Print a summary of the regression and interpret the output. Write down the equation for the regression line, and give an interpretation of the parameters. Use the tests introduced in (b) on the new model and compare with the previous results. Comment.

Solution

- (a) Plot including regression line

```
plot(dist ~ speed, data = cars, pch = 16, col = 'blue')
md1 <- lm(dist ~ speed, data = cars)
abline(md1, col = 'red', lwd = 2)
```



The scatterplot shows that there is an increasing relation between the speed and the breaking distance, which is at least partly captured by the regression model. There are, however, some indications that the model may not be suitable. On the one hand, it is clear that points corresponding to lower speeds tend to be closer to the regression line than points corresponding to higher speed. The plot has a funnel shape, indicating that the variability in the braking values increases with speed. On the other hand, there seem to be more points below the line than above in the central part, suggesting a curvature in the data.

The summary for the regressionline is

```
summary(md1)
```

```
##
## Call:
```

```
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601  0.0123 *
## speed        3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

Looking at the summary for the residuals included at the top, the values for the quartiles are similar but the maximum and minimum are not, and also the median is not close to zero. The estimated value for the intercept is -17.58 and the estimated slope is 3.94. Both parameters are significantly different from zero at the 95% confidence level according to the t-tests in the summary. The equation for the model is

$$\text{dist} = -17.5791 + 3.9324 \times \text{speed}$$

The slope represents the increase in the braking distance for a unit increase in speed, an increase of 3.94 feet per extra mile per hour in speed. The intercept is the braking distance corresponding to a speed of 0 mph. A negative value, as we have in this case, makes no physical sense.

(b) We do both tests. For normality we have

```
shapiro.test(rstandard(md1))
```

```
##
## Shapiro-Wilk normality test
##
## data:  rstandard(md1)
## W = 0.94518, p-value = 0.0217
```

The p -value is small and we should reject the null hypothesis of normality at the 5% significance level.

```
library(car)
```

```
## Loading required package: carData
```

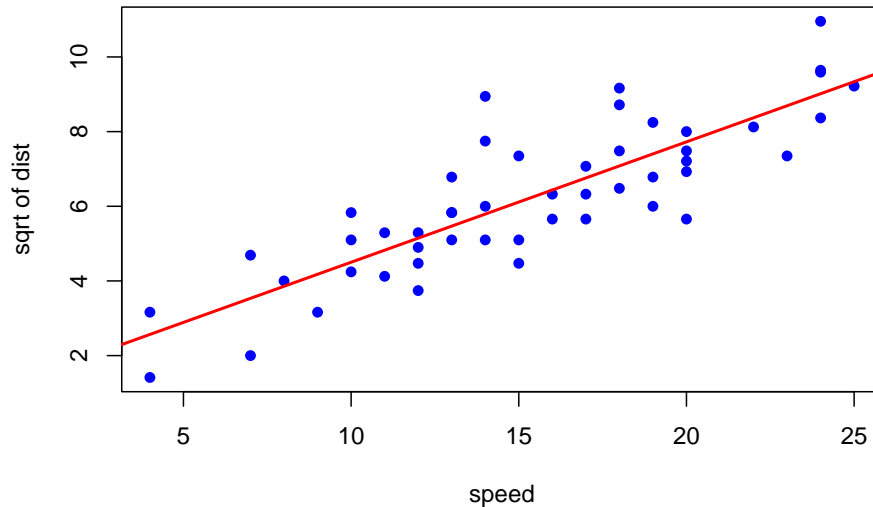
```
ncvTest(md1)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 4.650233, Df = 1, p = 0.031049
```

The p -value in this test is also small and at the 5% significance level we reject the null hypothesis of homogeneous variances. We see that both assumptions are not supported by the data.

(c) We now fit a model for the square root of the braking distance as a function of the speed and plot it:

```
plot(sqrt(dist) ~ speed, data = cars, pch = 16, col = 'blue', ylab = 'sqrt of dist')
md2 <- lm(sqrt(dist) ~ speed, data = cars)
abline(md2, col = 'red', lwd = 2)
```



The model looks better. The issues we remarked on the initial model seem to be absent, the variability around the regression line is more homogeneous and the number of points above and below the regression line are similar. the summary table is

```
summary(md2)
```

```
##
## Call:
## lm(formula = sqrt(dist) ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0684 -0.6983 -0.1799  0.5909  3.1534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.27705     0.48444   2.636  0.0113 *
## speed        0.32241     0.02978  10.825 1.77e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.102 on 48 degrees of freedom
## Multiple R-squared:  0.7094, Adjusted R-squared:  0.7034
## F-statistic: 117.2 on 1 and 48 DF,  p-value: 1.773e-14
```

The summary for the residuals is roughly symmetric. The estimated intercept is 1.277 while the estimated slope is 0.3224. Both parameters are significantly different from zero at the 95% confidence level according to the t-tests in the summary. The equation for the model is

$$\sqrt{\text{dist}} = 1.277 + 0.3224 \times \text{speed}$$

The value for the intercept has the same interpretation as before (but now is positive) while the slope represents the increase of the square root of the braking distance per unit increase of the speed. The model can be also written as

$$\text{dist} = 1.631 + 0.8235\text{speed} + 0.104(\text{speed})^2$$

We now look at the test for normality and homoscedasticity introduced in (b):

```
shapiro.test(rstandard(md2))
```

```
##
## Shapiro-Wilk normality test
##
## data:  rstandard(md2)
## W = 0.97386, p-value = 0.3298
```

```
ncvTest(md2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.01205185, Df = 1, p = 0.91258
```

We see that both p -values are now above the usual significance levels and both assumptions seem to be satisfied.

Exercise 3

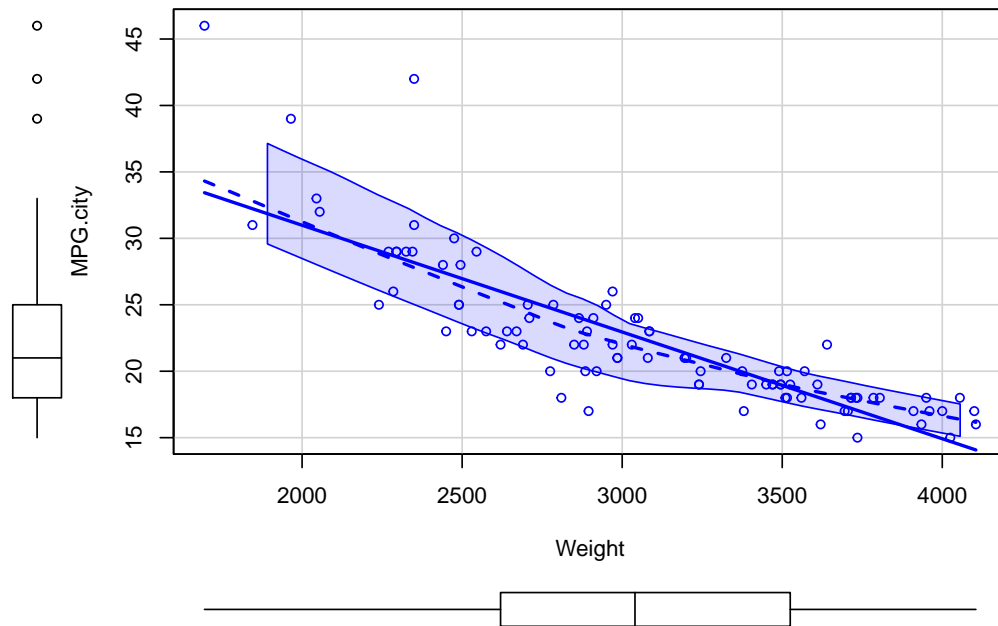
For this exercise we will use the data set `Cars93` in the `MASS` package, that has information about 93 cars on sale in the USA in 1993.

- (i) Draw a scatterplot of `MPG.city` against `Weight`. For this, use the function `scatterplot` in the `car` package. This function draws the points and also a simple regression line for the two variables. Moreover, it also plots a broken line that represents a local smoother function for the points as well as confidence bands for the smoother. The function also graphs boxplots for both variables on the corresponding axes. How would you interpret the differences between the regression line and the local smoother function that you see on the graph?
- (ii) Use the function `lm` to fit a regression line to this data. Use the function `summary` on the output of the regression. Interpret the t -tests in the table. Are the parameters different from zero?
- (iii) Write down explicitly the model that you get and interpret the meaning of the coefficients.
- (iv) Describe the sampling distribution for the estimated parameters in this regression.
- (v) Give confidence intervals at a confidence level of 98% for the parameters of the regression.

Solution

- (i) Scatterplot of `MPG.city` against `Weight`

```
library(MASS)
library(car)
scatterplot(MPG.city ~ Weight, data = Cars93)
```



Comparing the regression line (solid) with the local smoother (broken), we see that they are close, and most of the regression line falls in the confidence band of the smoother. However, we also see that in the central region of the data, roughly between 2200 and 3400, the smoother is always below the regression line, while the situation is reversed outside this interval. This observation suggests that the model is not completely adequate for the data.

(ii) Regression model:

```
model2 <- lm(MPG.city ~ Weight, data = Cars93)
summary(model2)
```

```
##
## Call:
## lm(formula = MPG.city ~ Weight, data = Cars93)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7946 -1.9711  0.0249  1.1855 13.8278
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  47.048353   1.679912   28.01  <2e-16 ***
## Weight       -0.008032   0.000537  -14.96  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.038 on 91 degrees of freedom
## Multiple R-squared:  0.7109, Adjusted R-squared:  0.7077
## F-statistic: 223.8 on 1 and 91 DF,  p-value: < 2.2e-16
```

The null hypothesis for the t -test in the table is that the corresponding coefficient is equal to zero. In both cases the p -value is practically zero, which says that both are significantly different from zero.

(iii) Equation for the model:

$$\text{MPG.city} = 47.048353 - 0.008032 \times \text{Weight}.$$

The slope is the change in the change in `MPG.city` when the weight changes by one unit. In this case

a change of 100 pounds in the car's weight produces a reduction of 0.8032 mpg in city driving. The intercept is the value of the dependent variable (`MPG.city`) when the weight of the car is equal to zero, in this case 47.048353.

(iv) The estimated parameters are $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$, which have a normal distribution:

$$\hat{\beta} = N((\beta_0, \beta_1)', \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ is obtained in R with

```
(invXtX <- summary(model2)$cov.unscaled)
```

```
##           (Intercept)      Weight
## (Intercept)  3.057094e-01 -9.598633e-05
## Weight      -9.598633e-05  3.123637e-08
```

The variance is unknown and is estimated by the mean square. The standard deviation is

```
summary(model2)$sigma
```

```
## [1] 3.03831
```

and the estimated variance is

```
summary(model2)$sigma^2
```

```
## [1] 9.231327
```

The estimated covariance matrix for $\hat{\beta}$ can be obtained with

```
vcov(model2)
```

```
##           (Intercept)      Weight
## (Intercept)  2.8221035327 -8.860813e-04
## Weight      -0.0008860813  2.883531e-07
```

or multiplying $\hat{\sigma}^2$ times $(\mathbf{X}'\mathbf{X})^{-1}$

```
(summary(model2)$sigma^2)*invXtX
```

```
##           (Intercept)      Weight
## (Intercept)  2.8221035327 -8.860813e-04
## Weight      -0.0008860813  2.883531e-07
```

(v) Confidence interval:

```
confint(model2, level = 0.98)
```

```
##           1 %           99 %
## (Intercept) 43.07027740 51.026428609
## Weight      -0.009303987 -0.006760796
```

Exercise 4

For this question we will use the data set `cats` in the library `MASS`, which has the heart and body weights for a sample of cats.

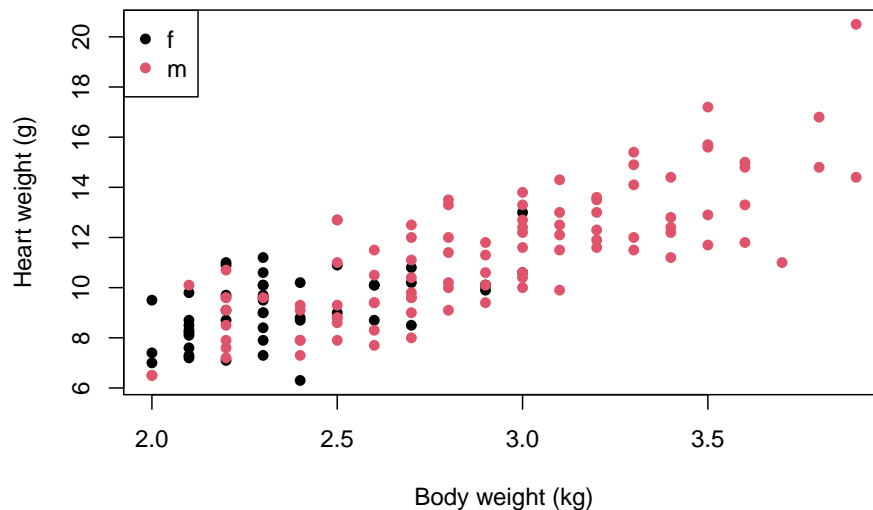
- (i) Draw a scatterplot of `Hwt` against `Bwt`. Color the points according to sex. Comment on what you see on the graph. Count how many data points for each sex there are.

- (ii) Use the function `lm` to fit a regression line to this data. Add the regression line to the plot. Use the function `summary` on the output of the regression. Interpret the t -tests in the table. Are the parameters different from zero? Write down explicitly an equation for the model that you get and interpret the meaning of the coefficients.
- (iii) Fit separate models for each sex. Plot a scatterplot of the data with points colored by sex and add the two regression lines. Comment.
- (iv) Print summary tables for the two models you fitted in (iii) and discuss the results.
- (v) Write down the equations for the two models fitted in (iii). Comment.

Solution.

- (i) Scatterplot

```
library(MASS)
data(cats)
plot(cats$Bwt, cats$Hwt, col = cats$Sex, pch = 16, xlab = 'Body weight (kg)',
     ylab = 'Heart weight (g)')
legend('topleft', c('f', 'm'), pch = 16, col = 1:2)
```



In the plot we see that there is a linear relation between the variables and that values for female cats tend to be smaller than for males, for both body and heart weight. There are many overlapping points.

Number of points:

```
sum(cats$Sex == 'F')
```

```
## [1] 47
```

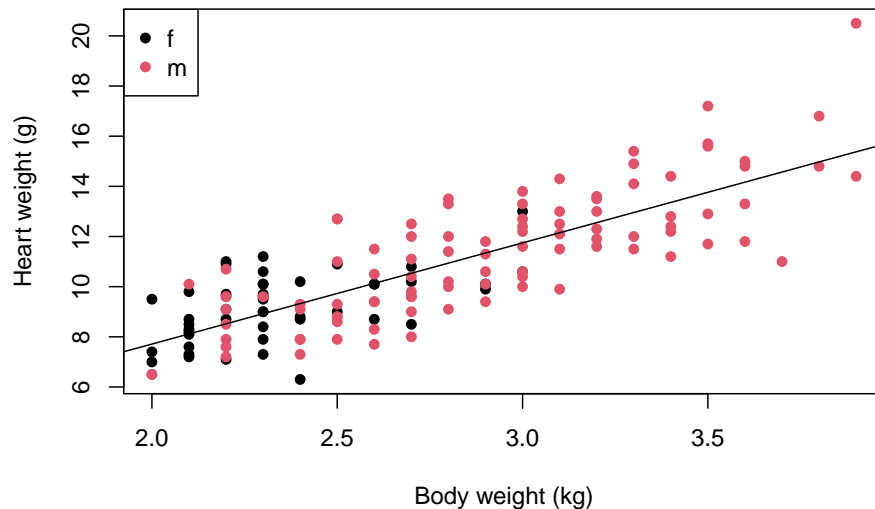
```
sum(cats$Sex == 'M')
```

```
## [1] 97
```

There are about twice as many male cats in the sample as female cats.

- (ii) Regression model:

```
mod1 <- lm(Hwt ~ Bwt, data = cats)
plot(cats$Bwt, cats$Hwt, col = cats$Sex, pch = 16, xlab = 'Body weight (kg)',
     ylab = 'Heart weight (g)')
legend('topleft', c('f', 'm'), pch = 16, col = 1:2)
abline(mod1)
```



```
summary(mod1)
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt, data = cats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5694 -0.9634 -0.0921  1.0426  5.1238
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.3567     0.6923  -0.515   0.607
## Bwt           4.0341     0.2503  16.119 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.452 on 142 degrees of freedom
## Multiple R-squared:  0.6466, Adjusted R-squared:  0.6441
## F-statistic: 259.8 on 1 and 142 DF, p-value: < 2.2e-16
```

The p -value for the intercept is large, indicating that it is not significant different from zero. However, even in cases like this, if we do not have a very strong reason for setting the intercept value to zero, we keep it in the model. The p -value for the slope is highly significant, and the slope is different from zero.

The equation for the model is:

$$\text{Hwt} = -0.3567 + 4.0341 \times \text{Bwt}.$$

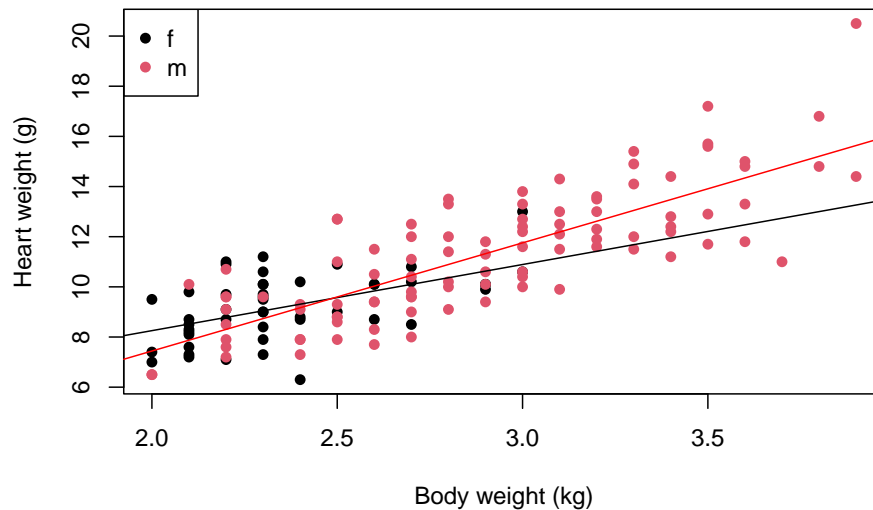
For every kilogram that the body weight of a cat increases, the heart weight increases 4.0341 grams. The intercept is interpreted as the value of the heart weight when the body weight is 0, and has a negative value (which makes no sense) of -0.3567.

(iii) Separate models:

```
mod2 <- lm(Hwt ~ Bwt, cats, subset = (Sex == 'F'))
mod3 <- lm(Hwt ~ Bwt, cats, subset = (Sex == 'M'))

plot(cats$Bwt, cats$Hwt, col = cats$Sex, pch = 16, xlab = 'Body weight (kg)',
     ylab = 'Heart weight (g)')
legend('topleft', c('f', 'm'), pch = 16, col = 1:2)
```

```
abline(mod2)
abline(mod3, col = 'red')
```



The lines have different slope and intercept.

(iv) Summary tables

```
summary(mod2)
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt, data = cats, subset = (Sex == "F"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.00871 -0.68599 -0.04506  0.79583  2.21858
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9813     1.4855   2.007 0.050785 .
## Bwt           2.6364     0.6254   4.215 0.000119 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.162 on 45 degrees of freedom
## Multiple R-squared:  0.2831, Adjusted R-squared:  0.2671
## F-statistic: 17.77 on 1 and 45 DF,  p-value: 0.0001186
```

For the female cats, the slope is significant while the intercept has a p -value of 0.050785, which is close to the significance level of 5%. The value for the slope is down, from 4.0341 for the single model to 2.6364, while the intercept is bigger.

```
summary(mod3)
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt, data = cats, subset = (Sex == "M"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.7728 -1.0478 -0.2976  0.9835  4.8646
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1841     0.9983  -1.186   0.239
## Bwt           4.3127     0.3399  12.688 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.557 on 95 degrees of freedom
## Multiple R-squared:  0.6289, Adjusted R-squared:  0.625
## F-statistic: 161 on 1 and 95 DF, p-value: < 2.2e-16
```

For the male cats, again the intercept is not significant while the slope is highly significant. Comparing with the single model, the intercept is smaller while the slope is bigger. In conclusion, the values for intercept and slope for the single model are in between those of the models by gender.

(v) Equations.

For female cats

$$\text{Hwt} = 2.9813 + 2.6364 \times \text{Bwt}.$$

For male cats

$$\text{Hwt} = -1.1841 + 4.3127 \times \text{Bwt}.$$