

MAS115 R programming, Homework Solutions 4

Bryony Moody (Original solutions by Prof. P. G. Blackwell)

2020-21

1 Producing a plot of bodyweight vs. brainweight

Here is the R code which I used to produce the picture.

```
library(MASS)
Animals
```

	body	brain
Mountain beaver	1.350	8.1
Cow	465.000	423.0
Grey wolf	36.330	119.5
Goat	27.660	115.0
Guinea pig	1.040	5.5
Dipliodocus	11700.000	50.0
Asian elephant	2547.000	4603.0
Donkey	187.100	419.0
Horse	521.000	655.0
Potar monkey	10.000	115.0
Cat	3.300	25.6
Giraffe	529.000	680.0
Gorilla	207.000	406.0
Human	62.000	1320.0
African elephant	6654.000	5712.0
Triceratops	9400.000	70.0
Rhesus monkey	6.800	179.0
Kangaroo	35.000	56.0
Golden hamster	0.120	1.0
Mouse	0.023	0.4
Rabbit	2.500	12.1
Sheep	55.500	175.0
Jaguar	100.000	157.0
Chimpanzee	52.160	440.0
Rat	0.280	1.9
Brachiosaurus	87000.000	154.5
Mole	0.122	3.0
Pig	192.000	180.0

```
pdf("BrainBodyRegression.pdf", width = 8, height = 6)
plot(log(brain) ~ log(body), data = Animals, ylim = c(0,10), asp=1,
     ylab = "Log of brainweight in g",
```

```

xlab = "Log of bodyweight in kg",
main = "Body and brainweights for 28 species of animal",pch = 19)

abline(lm(log(brain) ~ log(body), data = Animals))

#identify(log(Animals$body), log(Animals$brain), labels = row.names(Animals))

Dino <- c(6,16,26)
text(log(Animals$body)[Dino], log(Animals$brain)[Dino],
      labels = row.names(Animals)[Dino], pos = 2)

ModernAnimals <- Animals[-Dino,]
abline(lm(log(brain) ~ log(body), data = ModernAnimals), col = "red", lty = 2)

legend("topleft", legend = c("All Animals", "Modern Animals"), lty = c(1,2),
      col = c("black", "red"))
dev.off()

```

pdf
2

Note: Initially I used the *commented-out* `identify` to identify the three outliers and had no `text` command. When I ran the `identify` version it told me these were the 6th, 16th and 26th rows in the dataframe. I then used these as the labels in the `text` command in the final version. A slicker version is to use

```
Dino <- identify(log(body), log(brain), labels = row.names(Animals), plot = FALSE)
```

to identify the points without labelling them while saving their indices to automate construction of the final plot. (Evaluation is turned off for this chunk, as it would be interactive.)

In case you didn't realise, the three outlying animals were dinosaurs and extinct a long time ago. All of the other animals are 'extant', that is currently alive. The fit of the data to our relationship is actually pretty good and people do think there is a relationship of the form described between bodyweight and brainweight.

Why might it be worth specifying `asp=1`? That means that the scales on the two axes are the same, and so the apparent slope of a plotted regression line corresponds to its 'actual' slope—the value of β in the equation in the handout. For example, if brain weight were proportional to body weight (in reality it isn't that simple, even for modern animals), then we would have $\beta = 1$ in the equation, and `asp=1` would ensure that the line in the figure had slope 1 too.

Body and brainweights for 28 species of animal

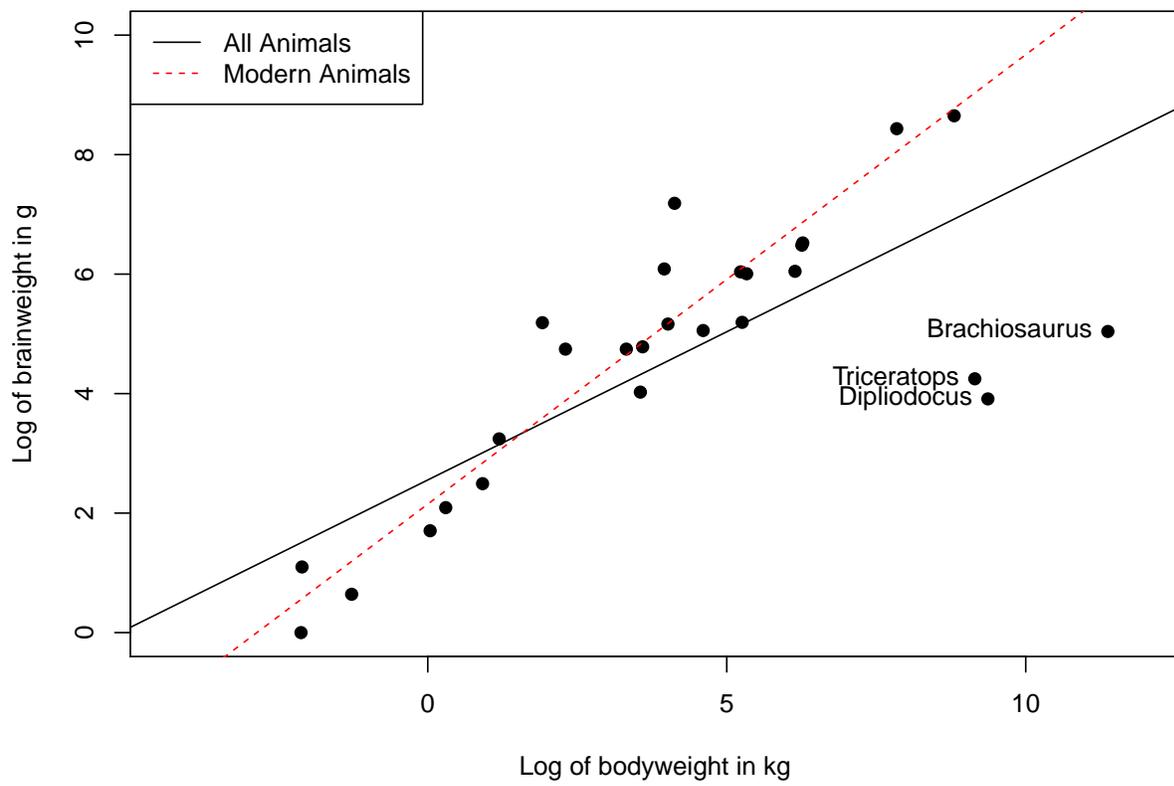


Figure 1: The PDF plot produced by the earlier code