STATS 100C: HW 6

Eric Chuu (604406828)

May 21, 2017

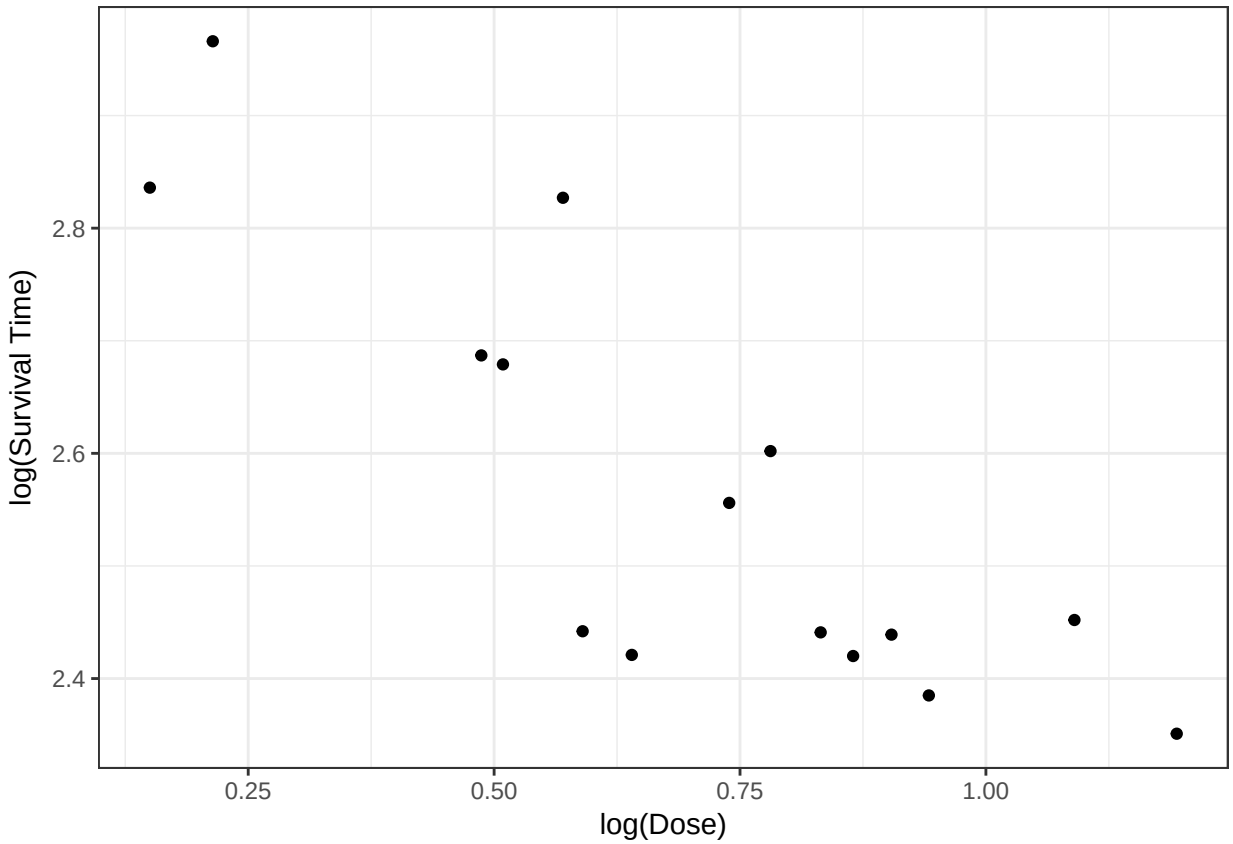
Problem 4.15

```
silk = read.table("https://www.biz.uiowa.edu/faculty/jledolter/RegressionModeling/Data/Chapter4/silkw.t
n    = dim(silk)[1];
p    = dim(silk)[2] - 1;
X    = as.matrix(cbind(rep(1, n), silk[,c(2:3)]))
y    = as.matrix(silk[,1])
```

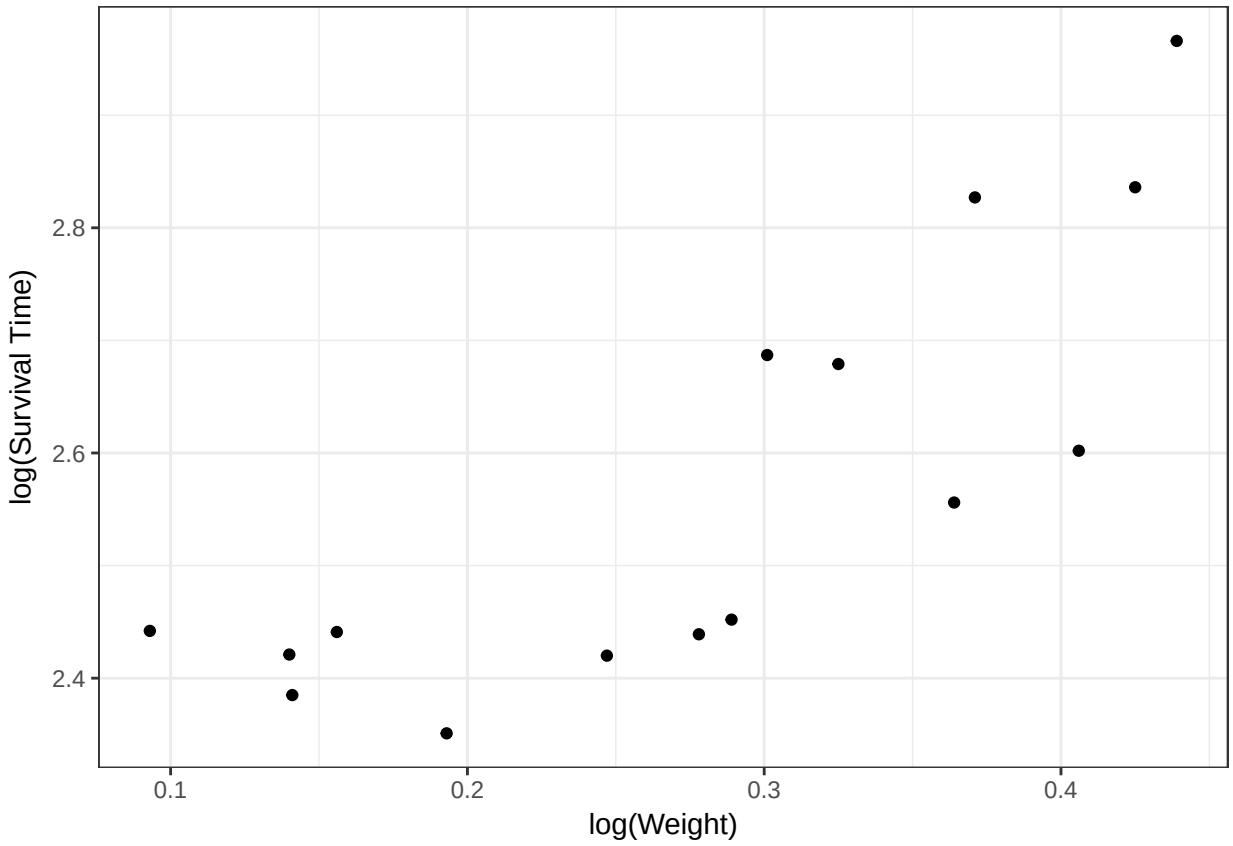
(a) Plot the response y versus each predictor variable.

From the first plot, there appears to be a inverse linear relationship between log(Survival Time) and log(Dose). From the second plot, there is a positive correlation between log(Survival Time) and log(Weight), but the growth is not linear.

```
ggplot(silk, aes(x = X1, y = Y)) + geom_point() +
    labs(x = "log(Dose)", y = "log(Survival Time)") + theme_bw()
```




```
ggplot(silk, aes(x = X2, y = Y)) + geom_point() +
    labs(x = "log(Weight)", y = "log(Survival Time)") + theme_bw()
```



(b) Obtain the least squares estimates for β and give the fitted equation.

```
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y    # 3 x 1
```

Calculating $\hat{\beta}$, as shown above, we can write the fitted equation as follows

$$\hat{y} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x_1} + \hat{\beta}_2\mathbf{x_2} = 2.59 \cdot \mathbf{1} + -0.38\mathbf{x_1} + 0.88\mathbf{x_2}$$

(c) Construct the ANOVA table and test for a significant linear relationship between y and the other two predictor variables.

(d) Which independent variable do you consider the better predictor of log(Survival Time). What are your reasons?

(e) Of the models involving one or both of the independent variables, which do you prefer, and why?