lab 7 Rat Data Analysis

An experiment was conducted to investigate the amount of a particular drug present in the liver of a rat. Nineteen rats were randomly selected, weighed, placed under light ether anesthesia and given an oral dose of the drug. Because large livers would absorb more of a given dose than smaller livers, the actual dose an animal received was approximately determined as 40 mg of the drug per kilogram of body weight. Liver weight is known to be strongly related to body weight. After a fixed length of time, each rat was sacrificed, the liver weighed, and the percent of the dose in the liver determined. The experimental hypothesis was that, for the method of determining the dose, there is no relationship between the percentage of the dose in the liver (Y) and the body weight BodyWt, liver weight LiverWt, and relative Dose.

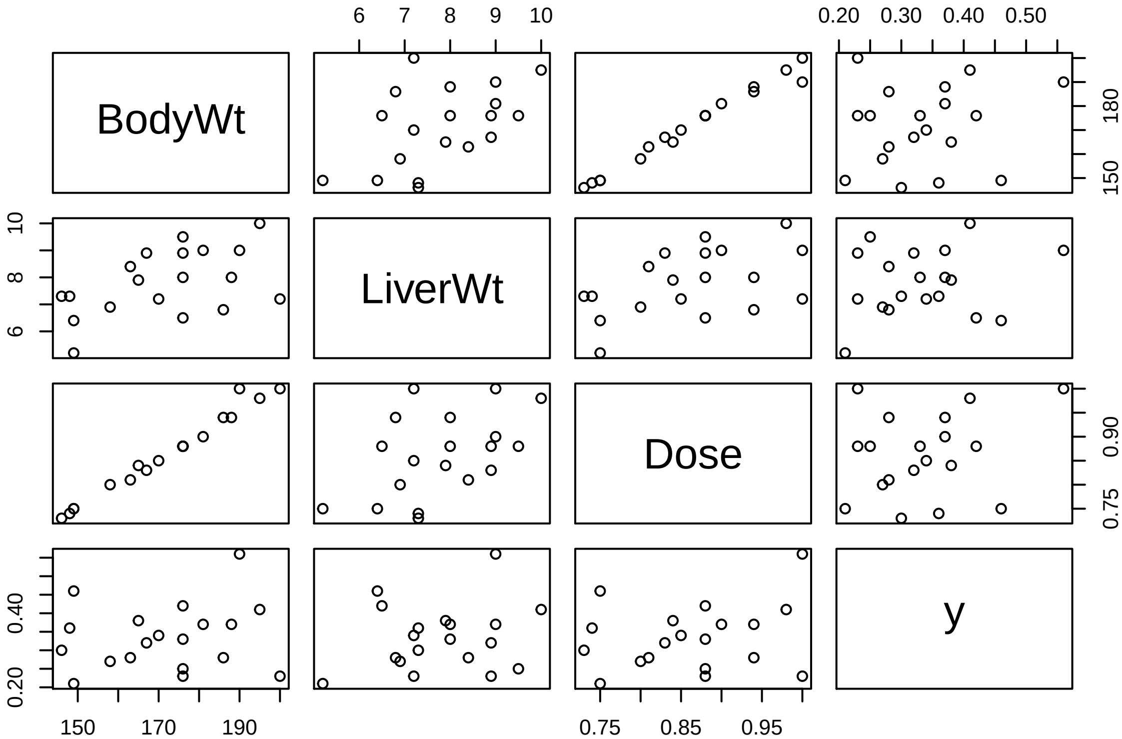
```r
#install.packages("alr4")
library(alr4)
```

```
## Loading required package: car

## Loading required package: effects

##
## Attaching package: 'effects'

## The following object is masked from 'package:car':
##
##     Prestige
```

```r
summary(rat)
```

```
##      BodyWt         LiverWt            Dose              y
##  Min.   :146.0   Min.   : 5.200   Min.   :0.7300   Min.   :0.2100
##  1st Qu.:160.5   1st Qu.: 7.050   1st Qu.:0.8050   1st Qu.:0.2750
##  Median :176.0   Median : 7.900   Median :0.8800   Median :0.3300
##  Mean   :171.5   Mean   : 7.811   Mean   :0.8621   Mean   :0.3353
##  3rd Qu.:183.5   3rd Qu.: 8.900   3rd Qu.:0.9200   3rd Qu.:0.3750
##  Max.   :200.0   Max.   :10.000   Max.   :1.0000   Max.   :0.5600
```

```r
pairs(rat)
```



As had been expected, the marginal summary plots for Y versus each of the predictors suggests no relationship.

```
lm1 <- lm(y~.,data=rat)
summary(lm1)
```

```
##
## Call:
## lm(formula = y ~ ., data = rat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.100557 -0.063233  0.007131  0.045971  0.134691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.265922   0.194585   1.367   0.1919
## BodyWt      -0.021246   0.007974  -2.664   0.0177 *
## LiverWt      0.014298   0.017217   0.830   0.4193
## Dose         4.178111   1.522625   2.744   0.0151 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07729 on 15 degrees of freedom
## Multiple R-squared:  0.3639, Adjusted R-squared:  0.2367
## F-statistic:  2.86 on 3 and 15 DF,  p-value: 0.07197
```

BodyWt and Dose have significant t-tests, with $p < 0.05$ in both cases, indicating that the two measurements combined are a useful indicator of Y. The analysis so far, based only on summary statistics, might lead to the conclusion that while neither BodyWt or Dose are associated with the response when the other is ignored, in combination they are associated with the response. But from scatterplots, Dose and BodyWt are almost perfectly linearly related, so they measure the same thing!

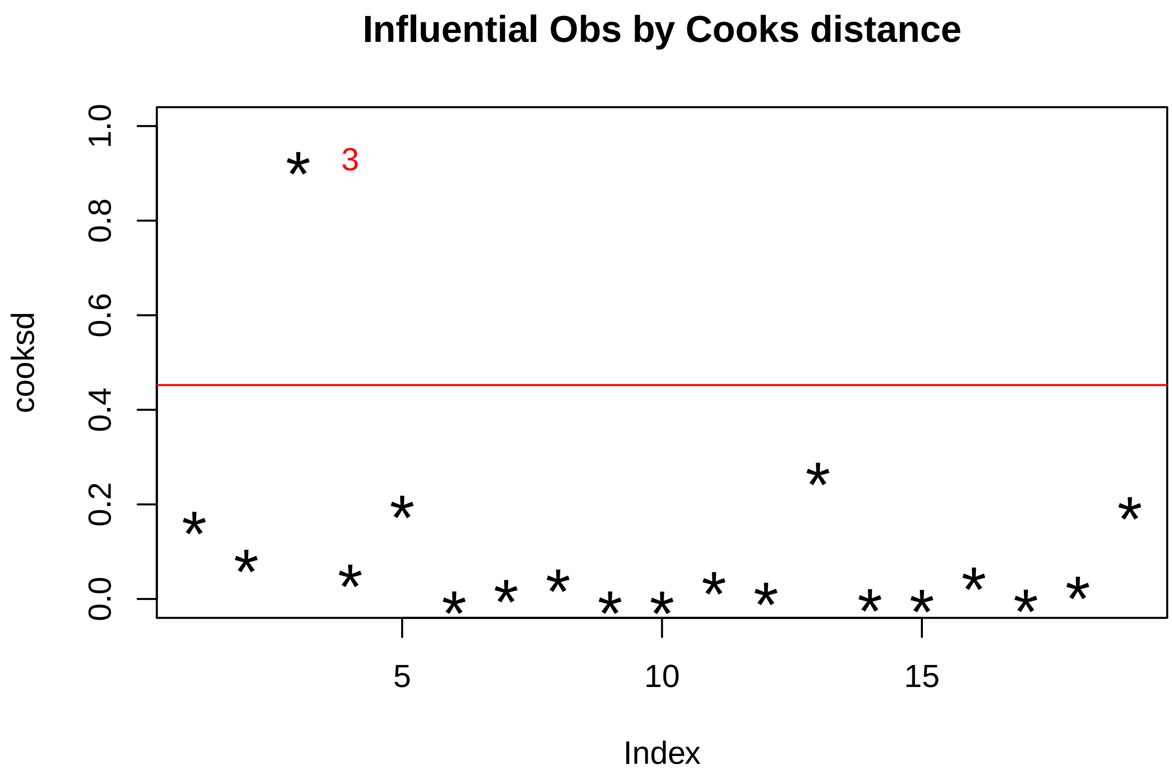
We turn to case analysis to attempt to resolve this paradox.

```
### outlier test
car::outlierTest(lm1)
```

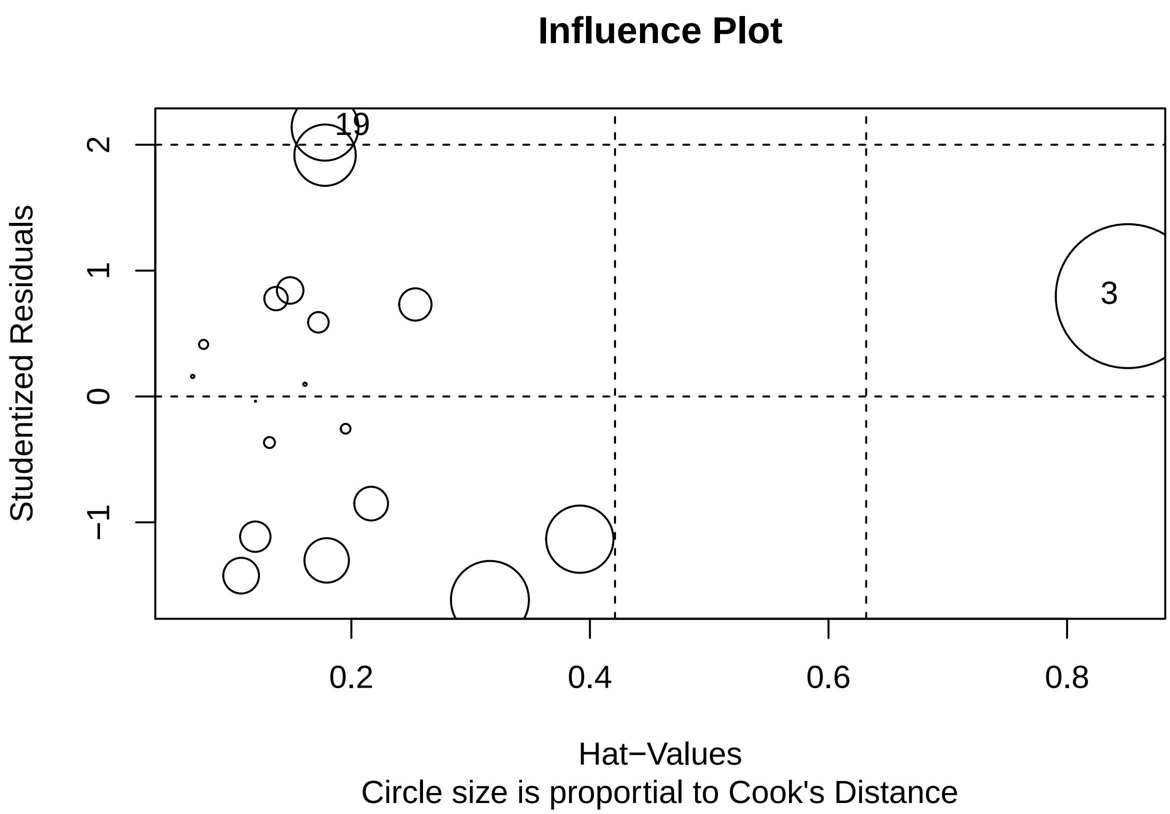
```
##
## No Studentized residuals with Bonferonni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferonni p
## 19 2.138833           0.050557      0.96058
```

From outlier test, there's no outlier. Let's look at the cook distances of the observations.

```
# plot cook distance
#plot(lm1, which = 4)

### calculate cook distance
cooksd <- cooks.distance(lm1)
###plot cook distance
plot(cooksd, pch="*", cex=2, main="Influential Obs by Cooks distance", ylim=c(0,1))
# add cutoff line
abline(h = 4*mean(cooksd, na.rm=T), col="red")
# add labels
text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>4*mean(cooksd, na.rm=T),names(cooksd),""), col
```




```
### influence plot
influencePlot(lm1, main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )
```




```
##      StudRes       Hat     CookD
## 3  0.7972915 0.8509146 0.9296160
## 19 2.1388332 0.1779618 0.1999403
```

Cook's distance immediately locates a possible cause: case three has D3 = .93; no other case has Di bigger than 0.27, suggesting that case number three alone may have large enough influence on the fit to induce the anomaly. One suggestion at this point is to delete the third case and recompute the regression. Let's see what happens if we delete this observation.

```
rat_3 <- rat[-3,]

lm2 <- lm(y~.,data=rat_3)
summary(lm2)
```

```
##
## Call:
## lm(formula = y ~ ., data = rat_3)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.102154 -0.056486  0.002838  0.046519  0.137059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.311427   0.205094   1.518    0.151
## BodyWt      -0.007783   0.018717  -0.416    0.684
## LiverWt      0.008989   0.018659   0.482    0.637
## Dose         1.484877   3.713064   0.400    0.695
##
```

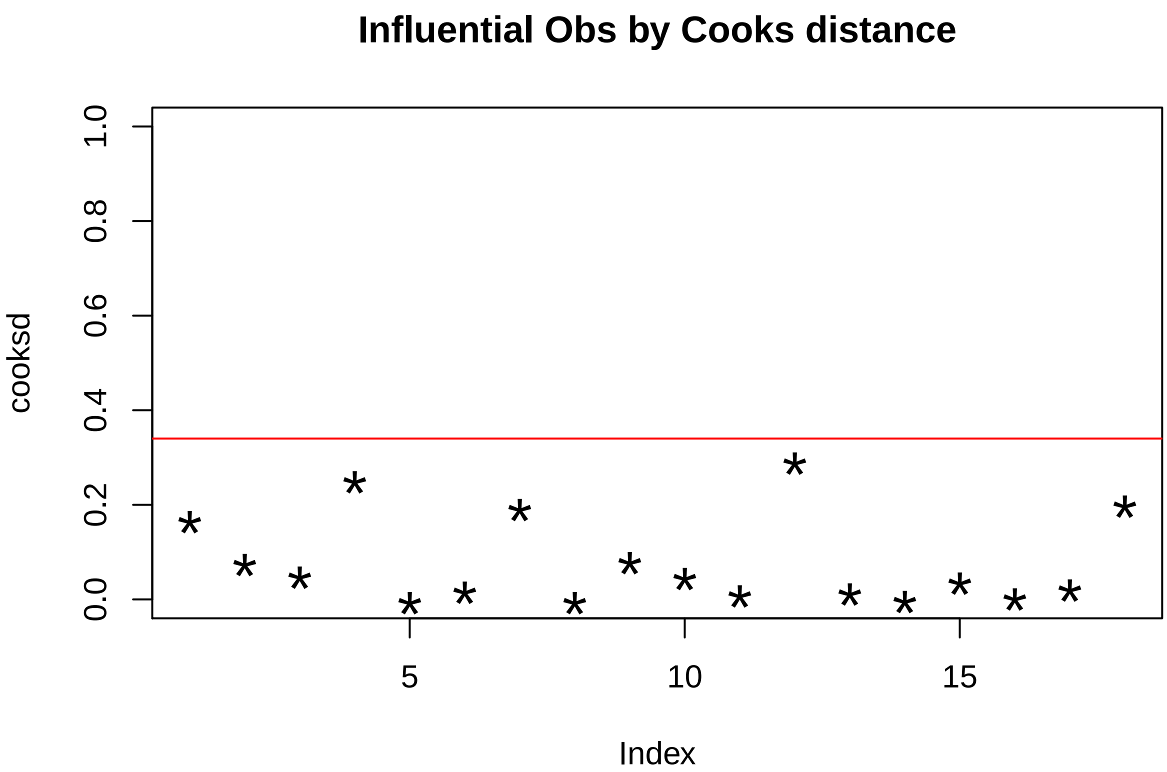
```
## Residual standard error: 0.07825 on 14 degrees of freedom
## Multiple R-squared:  0.02106,    Adjusted R-squared:  -0.1887
## F-statistic: 0.1004 on 3 and 14 DF,  p-value: 0.9585
```

The paradox dissolves and the apparent relationship found in the first analysis can thus be ascribed to the third case alone.

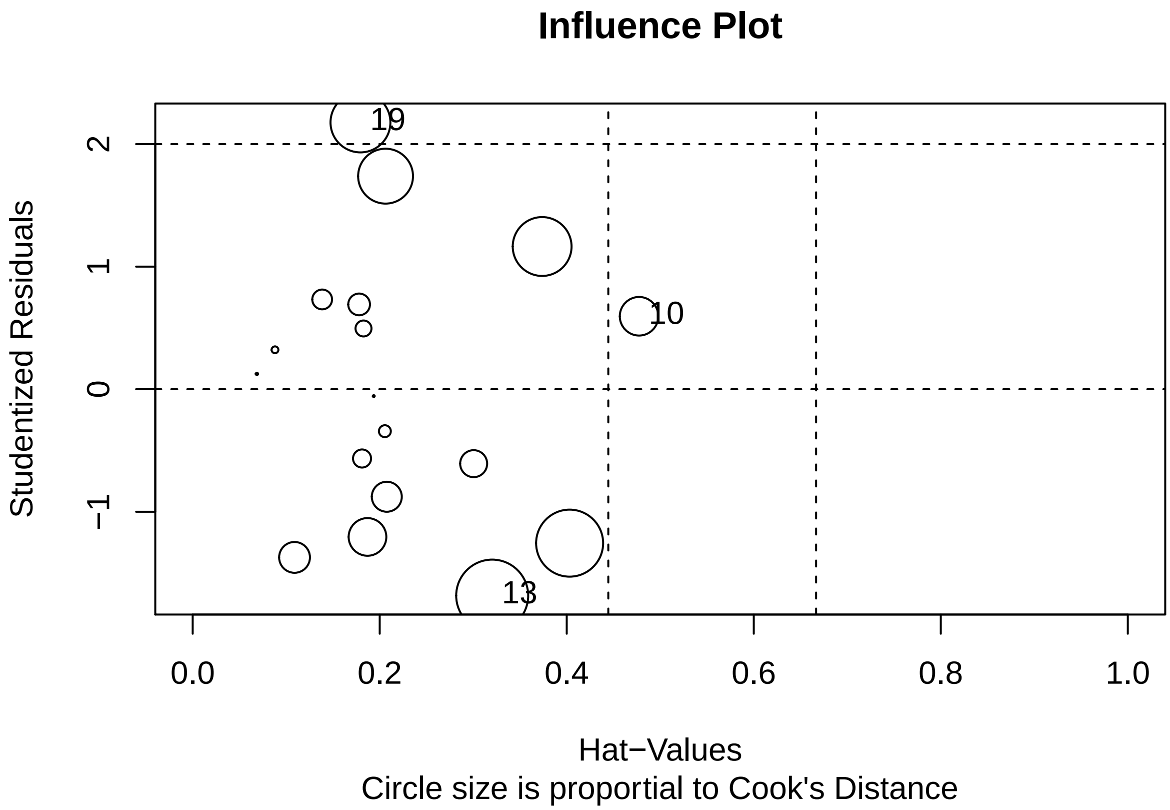
```
car::outlierTest(lm2)
```

```
##
## No Studentized residuals with Bonferonni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferonni p
## 19 2.176449            0.04855       0.8739
```

```
#plot(lm2, which = 4)

cooksd <- cooks.distance(lm2)
plot(cooksd, pch="*", cex=2, main="Influential Obs by Cooks distance", ylim=c(0,1))  # plot cook's dist
abline(h = 4*mean(cooksd, na.rm=T), col="red")  # add cutoff line
text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>4*mean(cooksd, na.rm=T),names(cooksd),""), co
```




```
influencePlot(lm2, main="Influence Plot", sub="Circle size is proportial to Cook's Distance", xlim = c(
```




```
##        StudRes       Hat      CookD
## 10  0.5950429 0.4773475 0.08475638
## 13 -1.6841868 0.3202994 0.29541209
## 19  2.1764494 0.1794026 0.20435499
```

Rat number three, with weight 190 g, was reported to have received a full dose of 1.000, which was a larger dose than it should have received according to the rule for assigning doses; for example, rat eight with weight of 195 g got a lower dose of 0.98. A number of causes for the result found in the first analysis are possible: (1) the dose or weight recorded for case 3 was in error, so the case should probably be deleted from the study, or (2) the regression fit in the second analysis is not appropriate except in the region defined by the 18 points excluding case 3. This has many implications concerning the experiment. It is possible that the combination of dose and rat weight chosen was fortuitous, and that the lack of relationship found would not persist for any other combinations of them, since inclusion of a data point apparently taken under different conditions leads to a different conclusion. This suggests the need for collection of additional data, with dose determined by some rule other than a constant proportion of weight.