R Handout - Simple Linear Regression

Dr. Derek Ogle
Northland College

Mar 2014, Vermont CFWRU Workshop

Preliminaries

```
> library(FSA)      # for Subset(), residPlot(), fitPlot()
```

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/MnFats.csv",header=TRUE)
> d <- Subset(d,sex!="UNK")  # removed one unknown sex individual (for simplicity)
> d <- within(d, { fyear <- factor(year)
                   loglen <- log(len)
                   logwt <- log(wt)
                 } )
> view(d)
```

```
    unit year len   wt sex age logwt loglen fyear
8   MN-1 2000 401  490   M  15 6.194  5.994  2000
82  MN-1 2000 399  488   F  13 6.190  5.989  2000
94  MN-1 2000 526 1490   F  17 7.307  6.265  2000
156 MN-1 2003 442  690   M  16 6.537  6.091  2003
234 MN-1 2006 566 1280   M  20 7.155  6.339  2006
302 MN-1 2006 559 1678   F  17 7.425  6.326  2006
```

```
> clr <- rgb(0,0,0,1/4)
```

Model Fitting

```
> lm1 <- lm(wt~len,data=d)
> names(lm1)
```

```
 [1] "coefficients"  "residuals"     "effects"       "rank"          "fitted.values"
 [6] "assign"        "qr"            "df.residual"   "xlevels"       "call"
[11] "terms"         "model"
```

```
> coef(lm1)
```

```
(Intercept)         len
  -3108.526       8.643
```

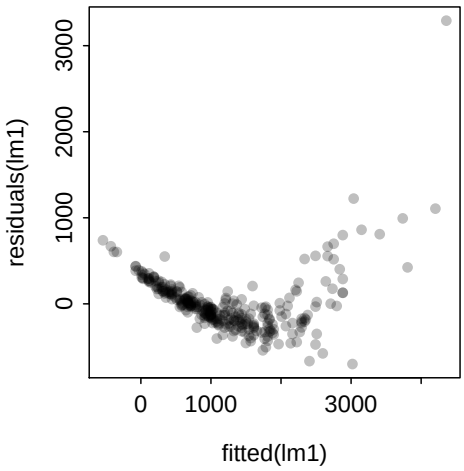
```
> residuals(lm1)[1:10]   # only show first 10 residuals
```

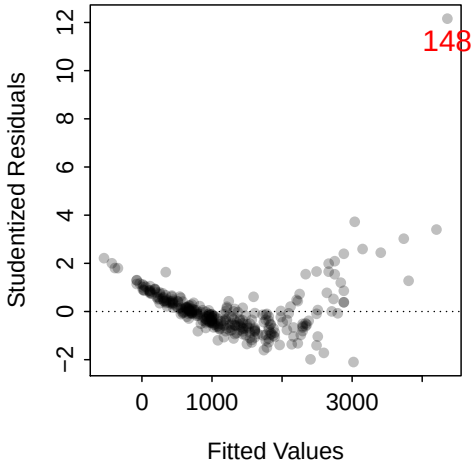
```
    1     2     3     4     5     6     7     8     9    10
669.1 301.0 254.5 305.4 173.0 193.0 145.7 132.5 139.3 223.4
```

```
> fitted(lm1)[1:10]      # only show first 10 fitted values
```

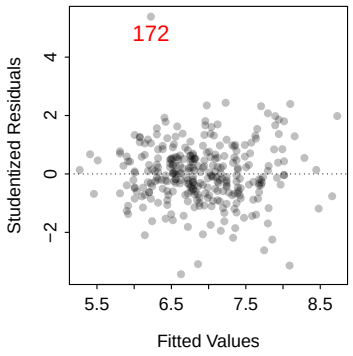
```
      1       2       3       4       5       6       7       8       9      10
-429.08   29.02  115.45  184.60  296.96  296.96  314.25  357.47  400.68  426.61
```

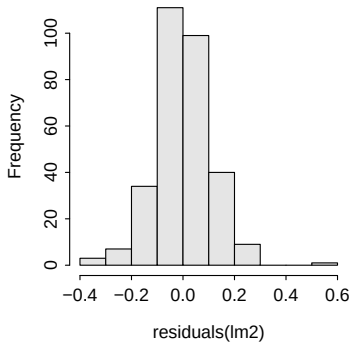
```
> plot(residuals(lm1)~fitted(lm1),pch=16,col=clr)   # manual construction -- Left
> residPlot(lm1,col=clr)                            # auto construction -- Right
```

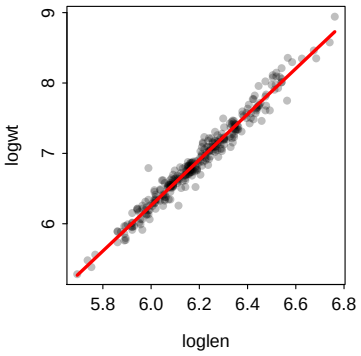





```
> lm2 <- lm(logwt~loglen,data=d)
> residPlot(lm2,col=clr)    # Left
> hist(~residuals(lm2))     # Middle
> fitPlot(lm2,col.pt=clr)   # Right
```





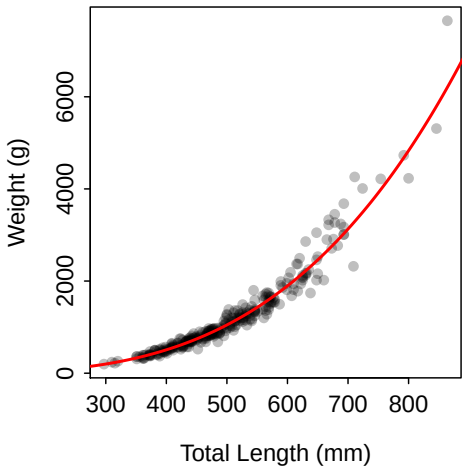



```
> d[172,]                    # Outlier?
    unit year len  wt sex age logwt loglen fyear
172 MN-1 2003 399 890   M  14 6.791  5.989  2003
> d[d$len>395 & d$len<405,]  # Fish w/ similar len
    unit year len  wt sex age logwt loglen fyear
7   MN-1 2000 396 460   F  11 6.131  5.981  2000
8   MN-1 2000 401 490   M  15 6.194  5.994  2000
55  MN-1 2000 396 510   M  11 6.234  5.981  2000
56  MN-1 2000 399 460   M  10 6.131  5.989  2000
82  MN-1 2000 399 488   F  13 6.190  5.989  2000
151 MN-1 2003 396 460   M   7 6.131  5.981  2003
172 MN-1 2003 399 890   M  14 6.791  5.989  2003
227 MN-1 2006 401 430   M  12 6.064  5.994  2006
242 MN-1 2006 399 567   F  17 6.340  5.989  2006
277 MN-1 2006 401 522   M   7 6.258  5.994  2006
> d1 <- d[-172,]            # Remove the fish
```

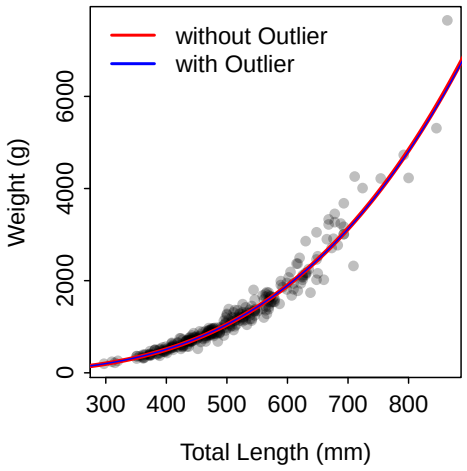
Model Fitting

```
> lm3 <- lm(logwt~loglen,data=d1)
> anova(lm3)
Analysis of Variance Table

Response: logwt
           Df Sum Sq Mean Sq F value Pr(>F)
loglen      1  112.5     112   10097 <2e-16
Residuals 301    3.4       0
> summary(lm3)


Call:
lm(formula = logwt ~ loglen, data = d1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3687 -0.0639 -0.0046  0.0639  0.2668

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.2516     0.2006   -66.1   <2e-16
loglen        3.2514     0.0324   100.5   <2e-16

Residual standard error: 0.106 on 301 degrees of freedom
Multiple R-squared:  0.971,Adjusted R-squared:  0.971
F-statistic: 1.01e+04 on 1 and 301 DF,  p-value: <2e-16
> coef(lm3)
(Intercept)      loglen
    -13.252       3.251
> confint(lm3)
              2.5 %  97.5 %
(Intercept) -13.646 -12.857
loglen        3.188   3.315
> # Predict weight for 400 mm individual
> ( p1 <- predict(lm3,data.frame(loglen=log(400)),interval="prediction") )
    fit   lwr   upr
1 6.229 6.021 6.437
> exp(p1)
    fit   lwr   upr
1 507.2 411.8 624.8
```

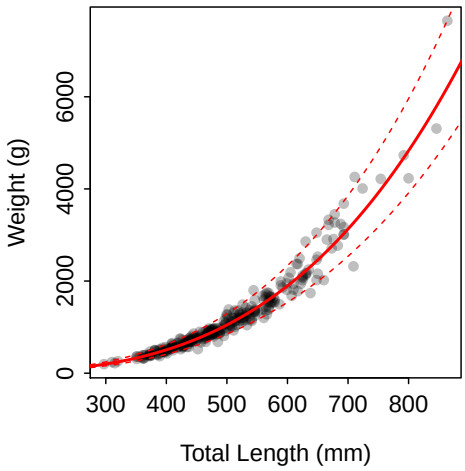
```
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> ( cf <- coef(lm3) )
(Intercept)      loglen
    -13.252       3.251
> curve(exp(cf[1])*x^cf[2],from=275,to=900,col="red",lwd=2,add=TRUE)
```




```
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> curve(exp(cf[1])*x^cf[2],from=275,to=900,col="red",lwd=3,add=TRUE)
> ( cfOut <- coef(lm2) )
(Intercept)      loglen
     -13.18        3.24
> curve(exp(cfOut[1])*x^cfOut[2],from=275,to=900,col="blue",lwd=1,add=TRUE)
> legend("topleft",c("without Outlier","with Outlier"),col=c("red","blue"),lwd=2,bty="n")
```




```
> # Predict weight for all lengths b/w 275 and 900 mm
> xs <- seq(275,900,1)
> pW <- exp(predict(lm3,data.frame(loglen=log(xs)),interval="prediction"))
> pW[1:5,]  # first five rows
    fit   lwr   upr
1 150.0 121.4 185.3
2 151.8 122.9 187.5
3 153.6 124.3 189.7
4 155.4 125.8 191.9
5 157.2 127.3 194.2
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> lines(pW[,"fit"]~xs,col="red",lwd=2)
> lines(pW[,"lwr"]~xs,col="red",lty=2)
> lines(pW[,"upr"]~xs,col="red",lty=2)
```


Weight (g)
0
2000
4000
6000
300
400
500
600
700
800
Total Length (mm)