Regression example

Colin Gillespie

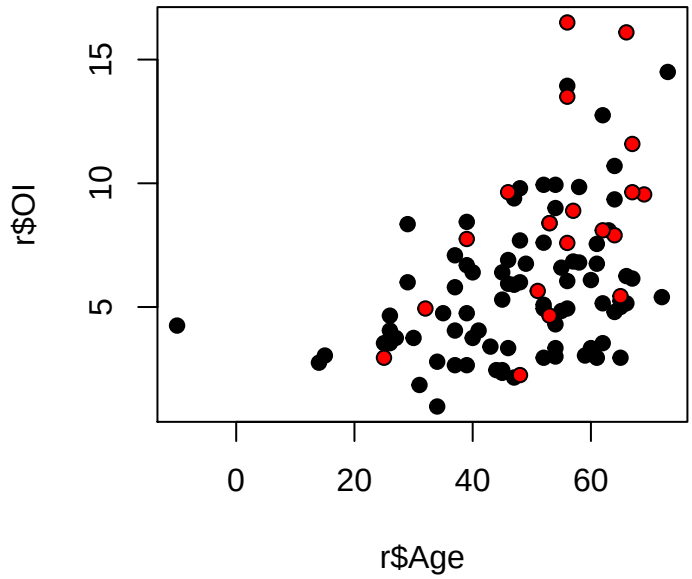
Fri Oct 24 11:19:12 2014

Load data set

```
r = read.csv("../data/regression.csv", header=TRUE)
```

Plot the data and colour the points by gender

```
plot(r$Age, r$OI, pch=21, bg=r$Sex)
```



Bad data, so remove the negative age

```
r = r[r$Age > 0,]
```

Fit a multiple linear regression model. Predict OI using Age and Sex However, we need to log transform OI

```
(m = lm(log(OI) ~ Age + Sex, data=r))
```

```
##
## Call:
## lm(formula = log(OI) ~ Age + Sex, data = r)
##
## Coefficients:
## (Intercept)          Age      SexMale
##      0.8292       0.0162       0.3189
```

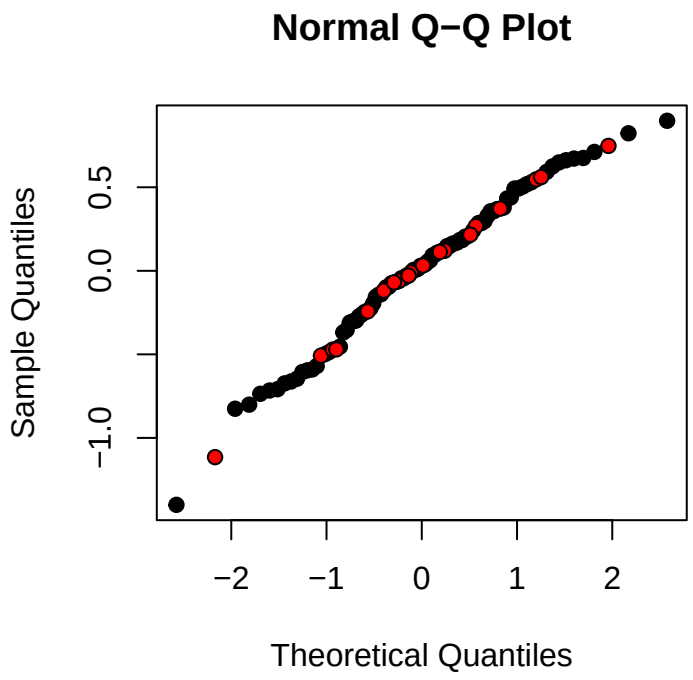
Standard ANOVA output

```
summary(m)
```

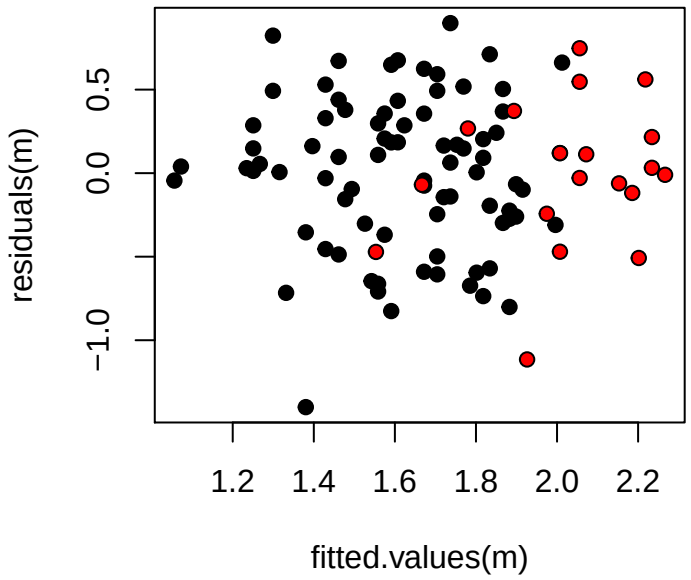
```
##
## Call:
## lm(formula = log(OI) ~ Age + Sex, data = r)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4005 -0.2795  0.0308  0.3055  0.8979
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.82920    0.17771    4.67  9.8e-06
## Age          0.01621    0.00352    4.60  1.3e-05
## SexMale      0.31890    0.11568    2.76    0.007
##
## Residual standard error: 0.455 on 97 degrees of freedom
## Multiple R-squared:  0.262,  Adjusted R-squared:  0.247
## F-statistic: 17.2 on 2 and 97 DF,  p-value: 3.96e-07
```

Standard residual plots Could use rstandard(m) for standardised residuals

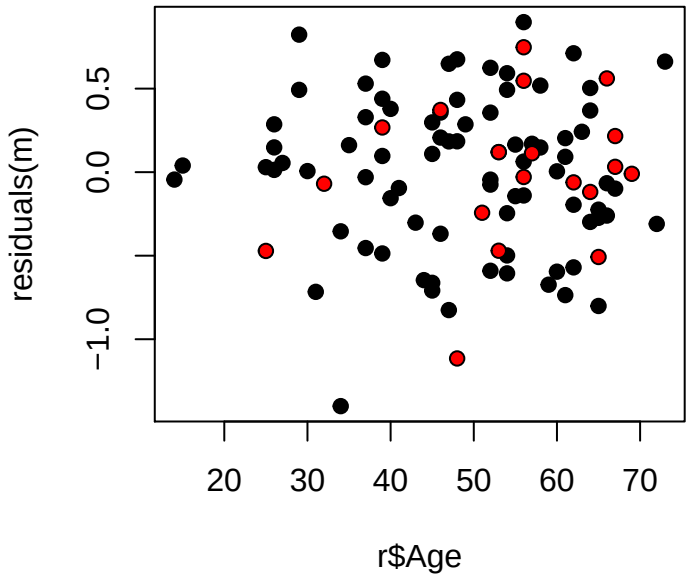
```
qqnorm(residuals(m), pch=21, bg=r$Sex)
```



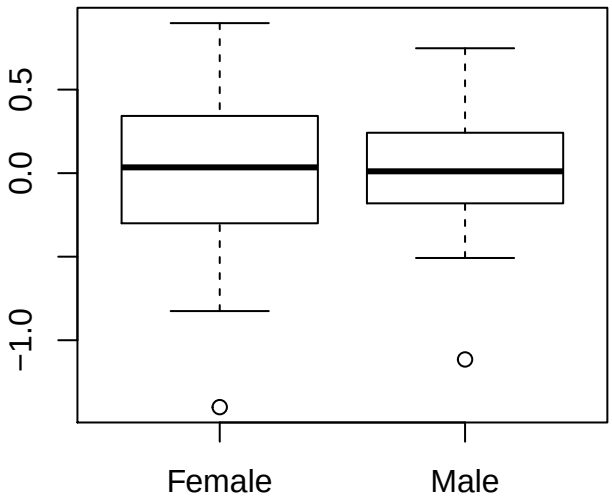

```
plot(fitted.values(m), residuals(m), bg=r$Sex, pch=21)
```




```
plot(r$Age, residuals(m), bg=r$Sex, pch=21)
```




```
boxplot(residuals(m) ~ r$Sex)
```


0.5
0.0
−1.0
Female
Male