MATH 423/533 – SIMPLE LINEAR REGRESSION: RESIDUALS IN R

For the simple linear regression model

$$\mathbb{E}[Y_i|x_{i1}] = \beta_0 + \beta_1 x_{i1} \qquad i = 1, 2, \ldots, n$$

where $n = 20$ and

- x_{i1} is Age, the single continuous predictor;
- Y_i is the Shear Strength outcome random variable;
- y_i is the observed version of Y_i.

The **residuals** e_i and **fitted values** $\widehat{y}_i$ are defined for $i = 1, \ldots, n$ as

$$e_i = y_i - \widehat{y}_i \qquad \text{and} \qquad \widehat{y}_i = \mathbf{x}_i\widehat{\beta} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1}.$$

where $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)^\top$ are the least squares estimates.

Simple Linear Regression: Residuals

```
> x<-RocketProp$Age
> y<-RocketProp$Strength
> fit.RP<-lm(y ~ x)
>
> RP.resids<-residuals(fit.RP)
> RP.fitted<-fitted(fit.RP)
>
> #Sum of residuals
> sum(RP.resids)
[1] -1.367795e-13
>
> #Sum of x times residuals
> sum(x*RP.resids)
[1] -3.595346e-12
> t(x) %*% RP.resids
               [,1]
[1,] -3.410605e-12
>
> #Sum of fitted values times residuals
> sum(RP.fitted*RP.resids)
[1] 5.09317e-11
> t(RP.fitted) %*% RP.resids
     [,1]
[1,]    0
>
> #Sum of y values times residuals
> sum(y*RP.resids)
[1] 166254.9
> t(y) %*% RP.resids
         [,1]
[1,] 166254.9
> ##Non-zero !!

```

```
> #Produce some plots
> par(mfrow=c(2,1))
> plot(x,RP.resids,xlab='x',ylab='Residuals',pch=19,ylim=range(-250,250))
> abline(h=0,lty=2)
> title('Residuals vs X (Age)')
> plot(RP.fitted,RP.resids,xlab=expression(hat(y)),ylab='Residuals',
  + pch=19,ylim=range(-250,250))
> abline(h=0,lty=2)
> title('Residuals vs Fitted values')
```

The resulting plot of the data is given below: