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

For the simple linear regression model

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

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** \hat{y}_i are defined for i = 1, ..., n as

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

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

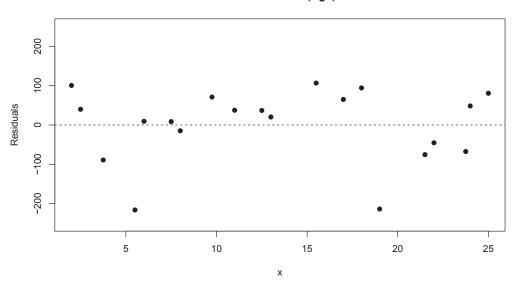
Simple Linear Regression: Residuals

```
1 > x<-RocketProp$Age
2 > y<-RocketProp$Strength
3 > \text{fit.RP} < -\text{lm}(y \sim x)
4 >
 5 > RP.resids<-residuals(fit.RP)</pre>
6 > RP.fitted<-fitted(fit.RP)
7 >
8 > \#Sum of residuals
9 > sum(RP.resids)
10 [1] -1.367795e-13
11 >
12 > \#Sum of x times residuals
13 > sum(x*RP.resids)
14 [1] -3.595346e-12
15 > t(x) % RP.resids
16
17 [1,] -3.410605e-12
18 >
19 > \#Sum of fitted values times residuals
20 > sum(RP.fitted*RP.resids)
21 [1] 5.09317e-11
22 > t(RP.fitted) %*% RP.resids
23
        [,1]
24 [1,]
25 >
26 > \#Sum of y values times residuals
27 > sum(y*RP.resids)
28 [1] 166254.9
29 > t(y) %*% RP.resids
30
             [,1]
31 [1,] 166254.9
32 > \# \text{Non-zero} !!
33
```

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

The resulting plot of the data is given below:

Residuals vs X (Age)



Residuals vs Fitted values

