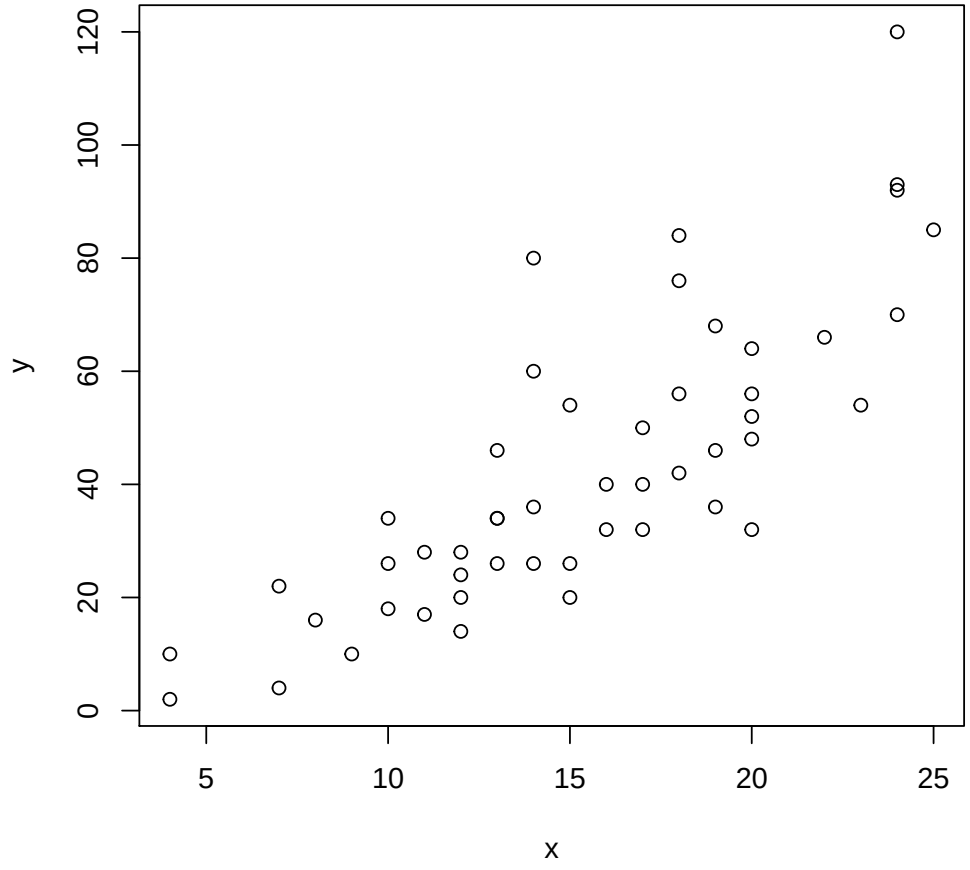
Homework 3 - STAT 511

Amal Agarwal

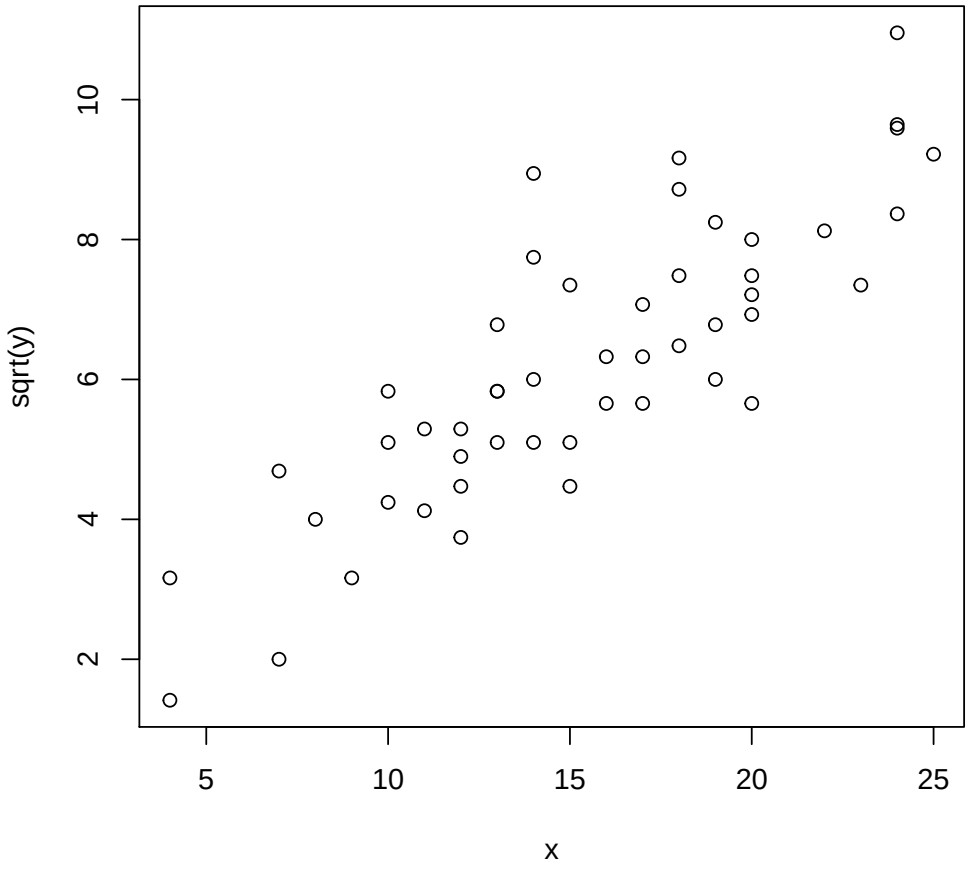
Answer 1

(a) Let us call y as the stopping distance (response variable) and x as the speed (predictor variable). For Exploratory Data Analysis (EDA) following scatter plots were obtained:

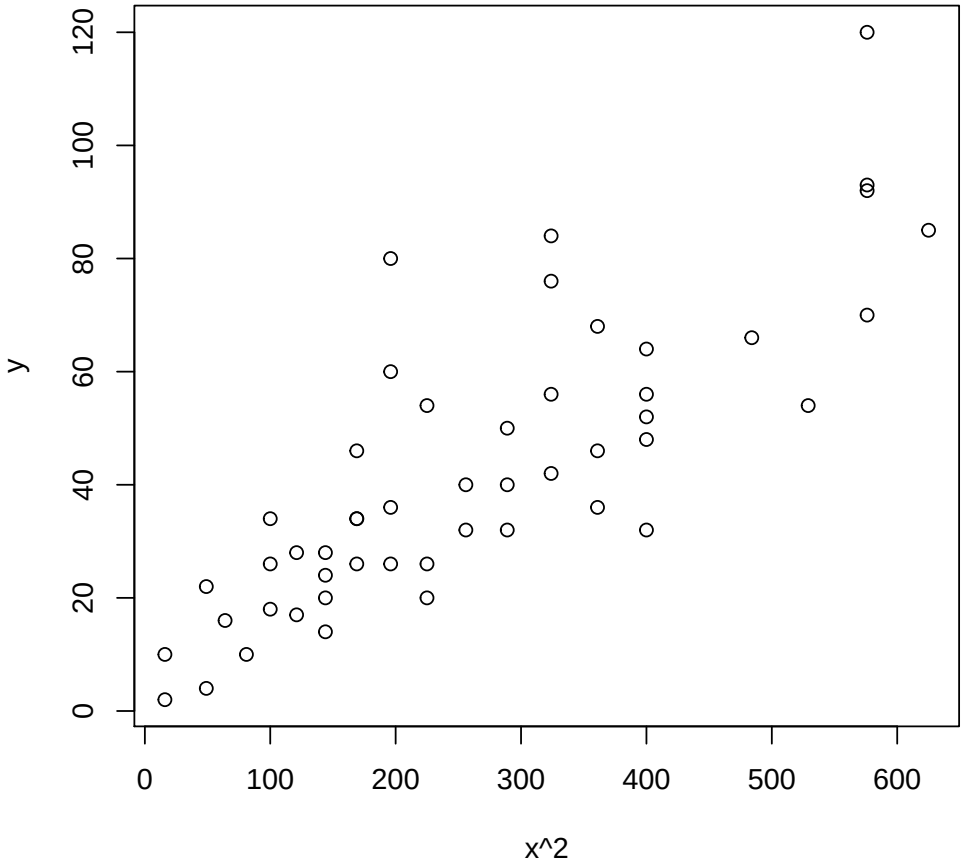
- Plot of y vs. x



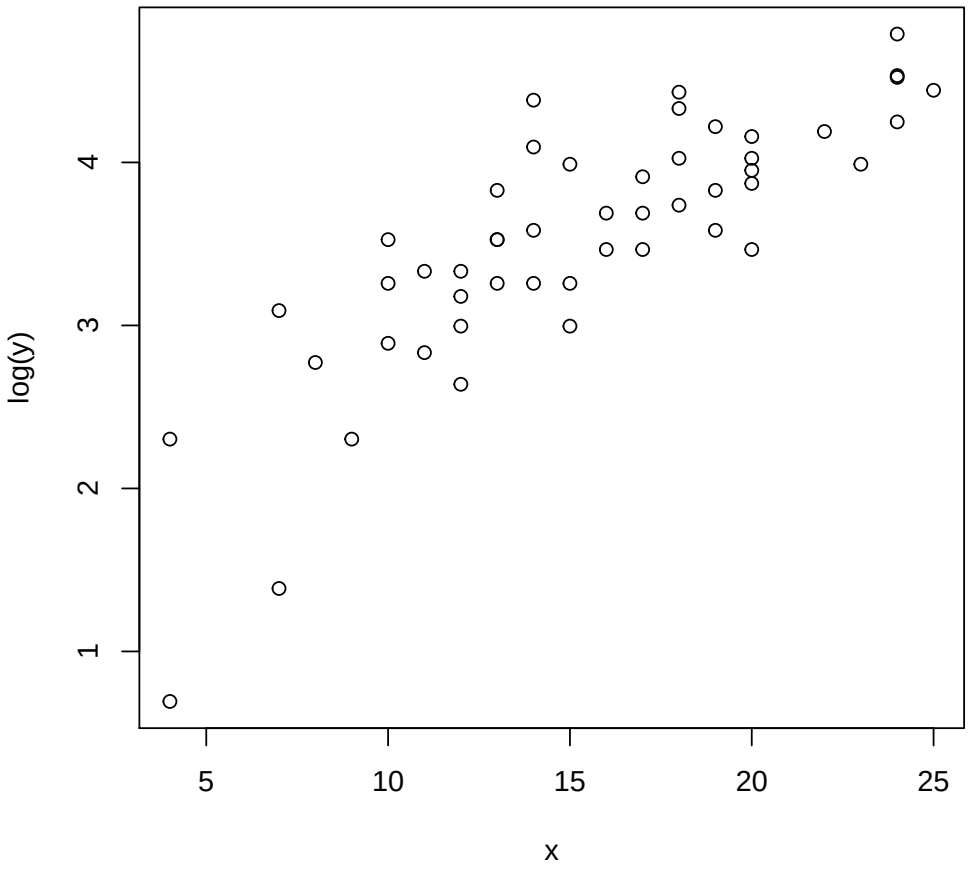
- Plot of $\sqrt{y}$ vs. x



- Plot of y vs. x^2



- Plot of $log(y)$ vs. x



It can be observed that the scatter plot of y vs. x does indicates a non linear relationship between y and x. Further tha density of data points for $x < 7$ and $x > 21$ is relatively lower than for $8 < x < 20$. The transformation $log(y)$ vs. x can be ruled out since it does not look linear. The transformation y vs. x^2 looks promising in terms of fitting a simple linear model. However, if I have to choose, the scatter plot that shows the most closest linear relationship after the transformation is $\sqrt{y}$ vs. x.

(b) The simple linear regression model in R that can be used to fit the given data is

$$y = \beta_0 + \beta_1 x + \epsilon$$

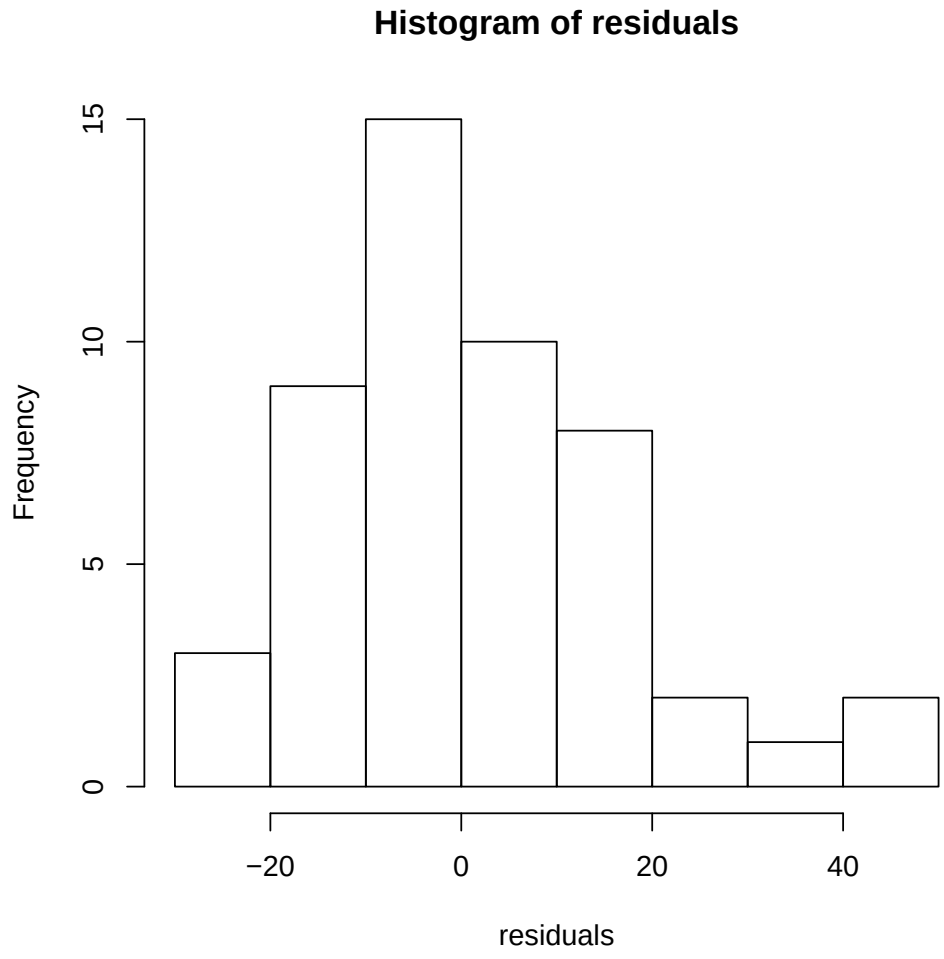
(c) Using lm in R, to estimate model parameters and residuals.

```
> fit1=lm(y~x)
> residuals=fit1$resid
> hist(residuals)
> ## the table given in summary(fit)
> summary(fit1)$coef
```

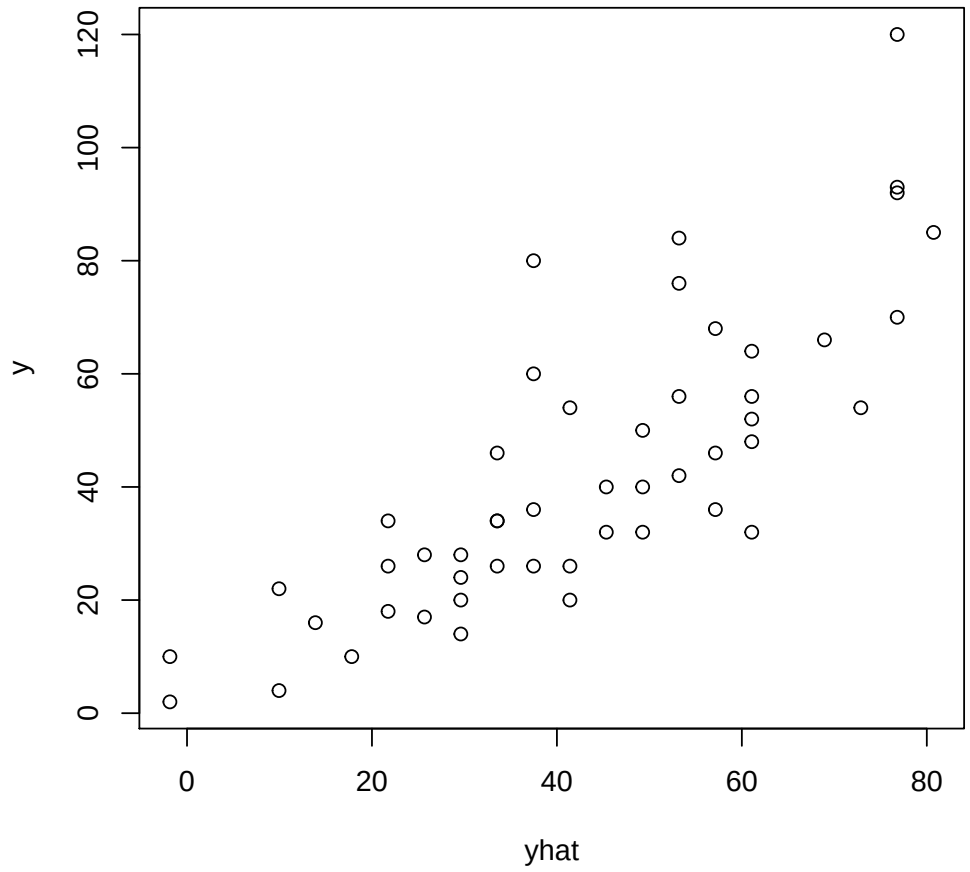
```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) -17.579095  6.7584402 -2.601058 1.231882e-02
x             3.932409  0.4155128  9.463990 1.489836e-12
```

```
> summary(fit1)$sigma
```

```
[1] 15.37959
```



(d)
```
> yhat=fit1$fitted.values
> plot(yhat,y)
```



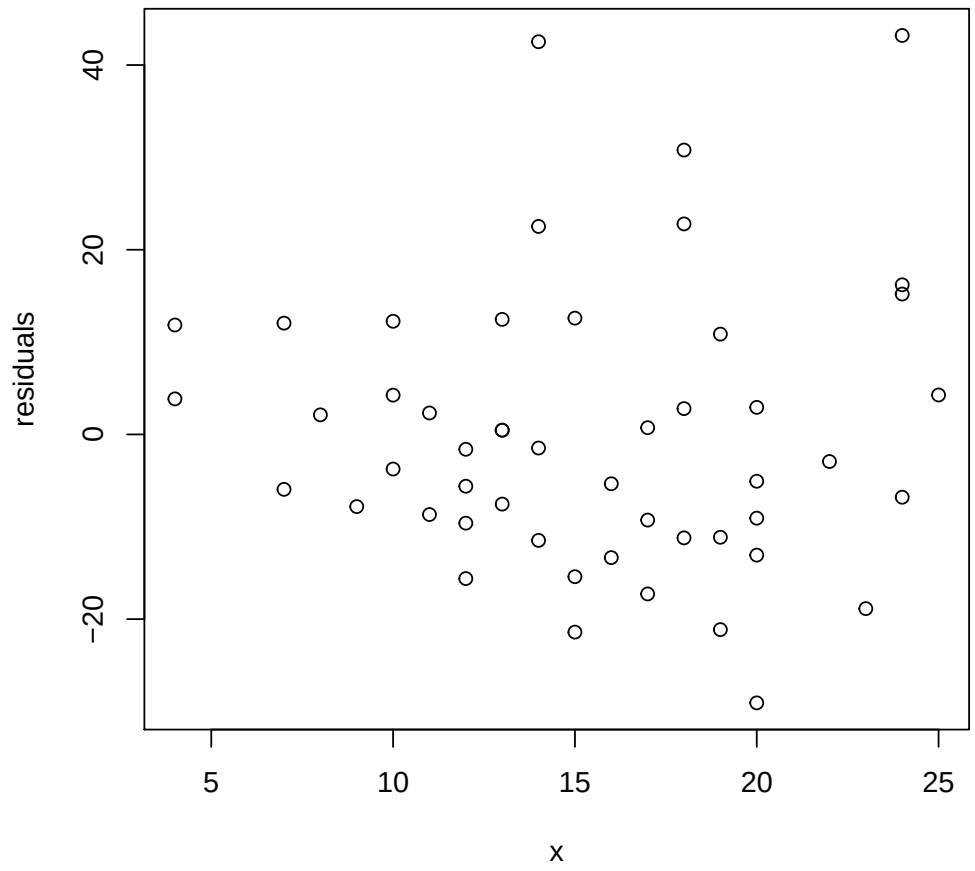
The mean of residuals is given as

```
> mean(residuals)

[1] 8.65974e-17
```

This is very close to zero which indicates that our assumption of $E(\epsilon) = 0$ is satisfied. To check for homoscedasticity, residuals can be plotted against x as

```
> plot(x,residuals)
```



Clearly, as x increases, the fluctuations of residuals around zero increases which implies that the estimated variance of the residuals increase with the covariate speed. This behaviour is called heteroscedascity which clearly violates our assumption of homoscedascity in the classical linear model.