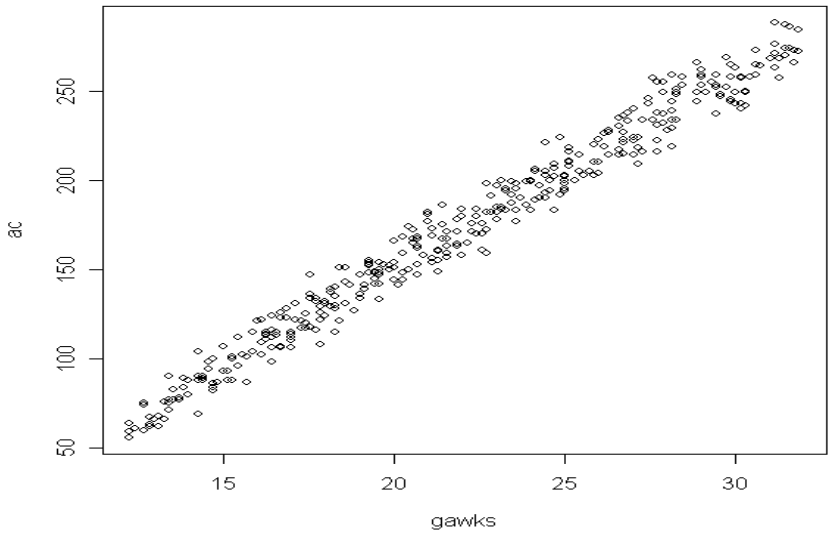
R Codes	Description
`> gest = read.table("http://ehsan.karim.googlepages.com/gestation.txt", head=T)` `> gest`	Read data from external web file
`> names(gest)` `[1] "gawks" "ac"`	Check names of the variables in the dataset: **ac** and **gawks**. Purpose of our analysis is to see if there is a relationship (hopefully linear) between the abdominal circumference (ac) and gestation period (gawks).
`> attach(gest)`	Attach the data to identify the variables as independent vectors
`> plot(gawks, ac)` 	We would see the relationship between ***gawks (x)*** and ***ac (y)***. **Scatterplot:** this shows that there is a denite linear relationship the first variable is the independent variable, x, and the second variable is the dependent variable, y.
`> linreg = lm(ac~gawks)`	gawks is our independent variable, x, and ac is our dependent variable, y. We fit **regression** line.

`> summary(linreg)`
 	
`Call:`
`lm(formula = ac ~ gawks)`
 	
`Residuals:`
`Min 1Q Median 3Q Max`
`-21.1477 -6.5674 -0.3889 6.9741 23.8949`
	Details about the fitted regression:
	a) coefficients
	b) R-squared
	c) p-value
	d) S.E.

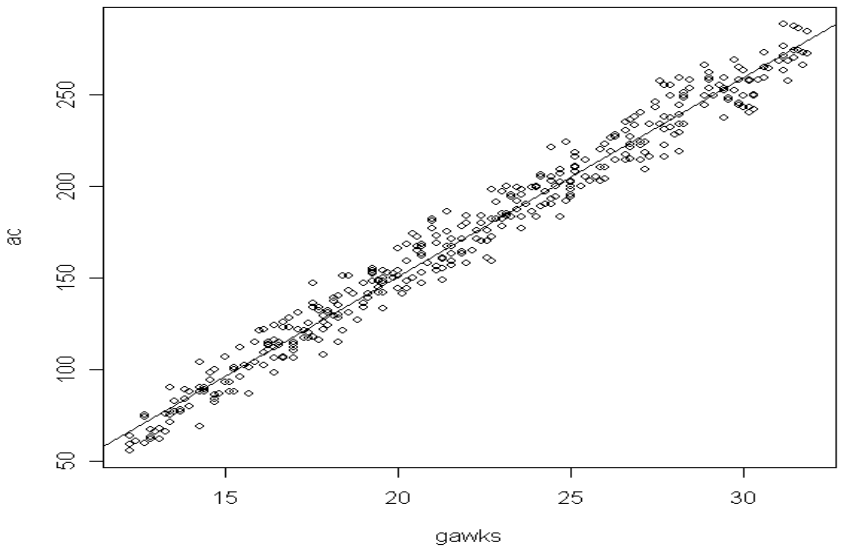
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -67.57293    1.95712  -34.53   <2e-16 ***
gawks        10.90542    0.08598  126.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.112 on 389 degrees of freedom
Multiple R-Squared: 0.9764,     Adjusted R-squared: 0.9763
F-statistic: 1.609e+04 on 1 and 389 DF,  p-value: < 2.2e-16
```

```
> names(linreg)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"
```

Check the values that the object stores

```
> plot(gawks, ac)
> abline(linreg)
```



Plot the regression line

```
> pred = predict(linreg)          # method 1
> linreg$fitted.values            # method 2
```

Find predicted value, Y-hat

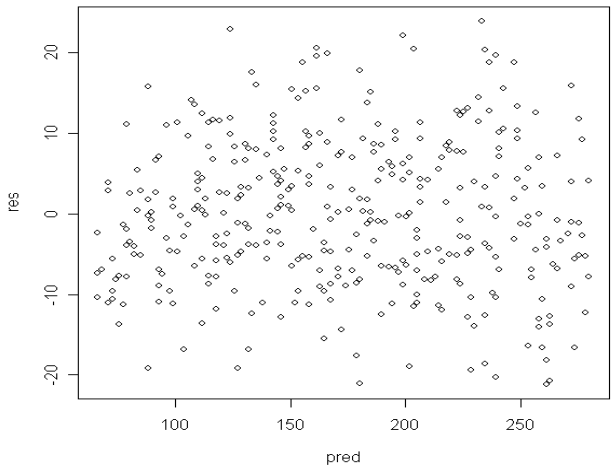
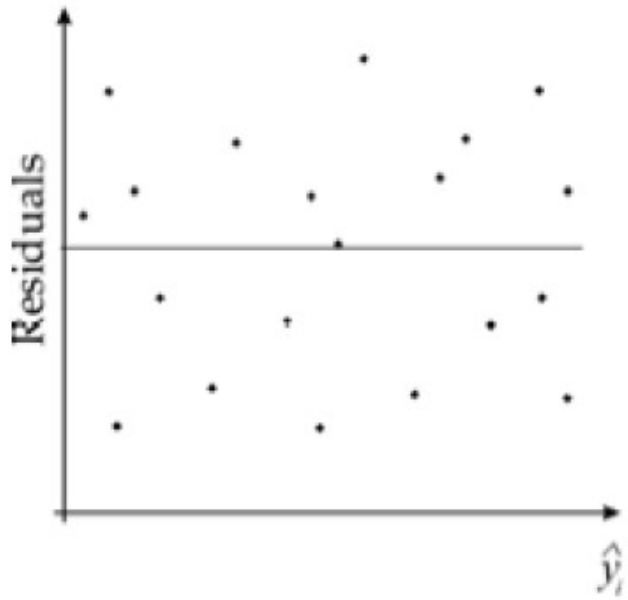
```
> res = ac - pred                 # method 1
> res = residuals(linreg)         # method 2
> res = linreg$residuals          # method 3
```

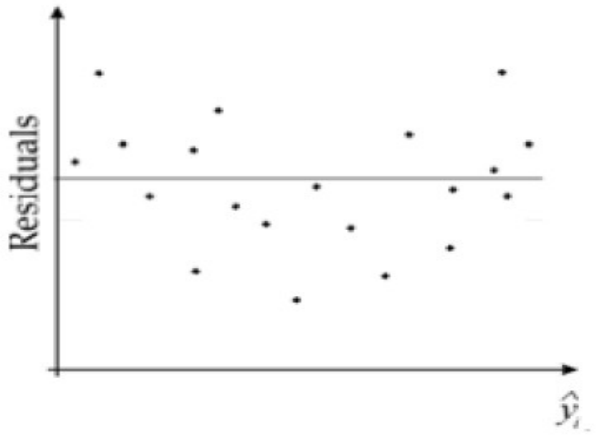
Find residual values=Y - Y-hat

Regression Diagnostics: Assumptions:

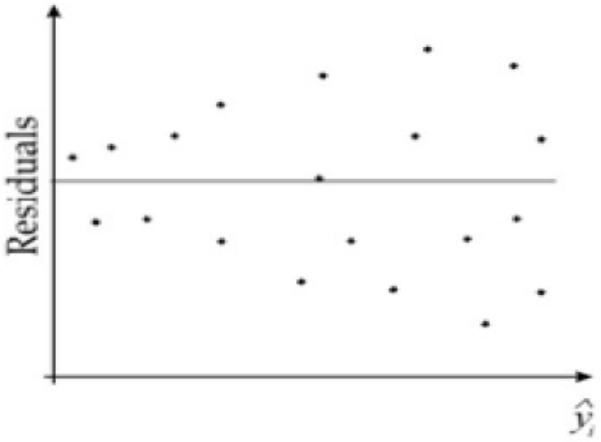
(i) mean of 0 (linear model holds)
(ii) constant variance
(iii) identically normally distributed

Assumption Checking:

R Codes	Description
`> plot(pred, res)` 	Check of assumption (I) No trend detected. Seems random. Overall mean seems like zero by eye inspection.
	**Not from this example: For Illustration:** Ideal case: hovers around zero.



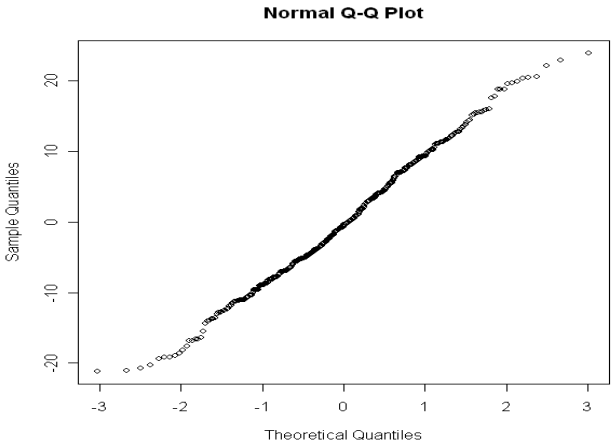
Not from this example: For Illustration:
Curvature present, not ideal,
mean could be non-zero.



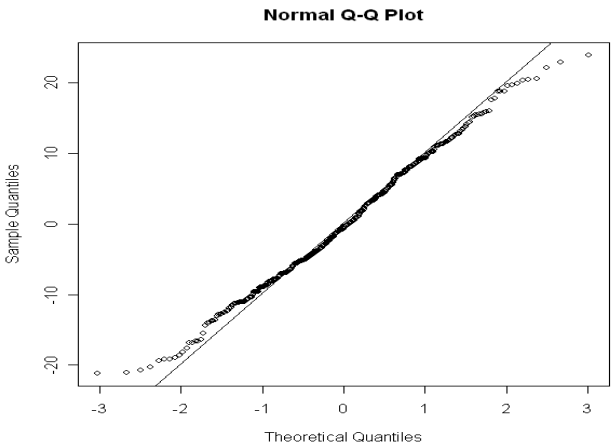
Not from this example: For Illustration:
Non-constant variance is possibly present and the violation of assumption (ii).

Such violation does not happen in our current example.

```
> qqnorm(res)
> qqline(res)
```



Check of assumption (iii) : seems ok.




```
> plot(linreg, las = 1)
```

Automated graph (*advanced*)