

ORIGIN ≡ 1

Diagnostics for Simple Linear Regression

W. Stein

Listed here are guidelines to useful R functions, based in part on Chapters 3 of *Applied Linear Statistical Models* 5th Edition (KNNL).

Example: Anderson's Iris data

```
iris := READPRN("c:/2008LinearModelsData/iris.txt")
```

```
SL := iris<2> < assigning variable SL
```

```
n := length(SL) n = 150 < n = number of observations X
```

```
i := 1..n < constructing index variable i
```

```
XbarSL := mean(SL) XbarSL = 5.8433 < mean of X
```

```
SDSL :=  $\sqrt{\text{Var}(SL)}$  SDSL = 0.8281 < sample standard deviation of X
```

```
SESL :=  $\frac{SD_{SL}}{\sqrt{n}}$  SESL = 0.0676 < standard error of the sample mean of X
```

Prototype in R:

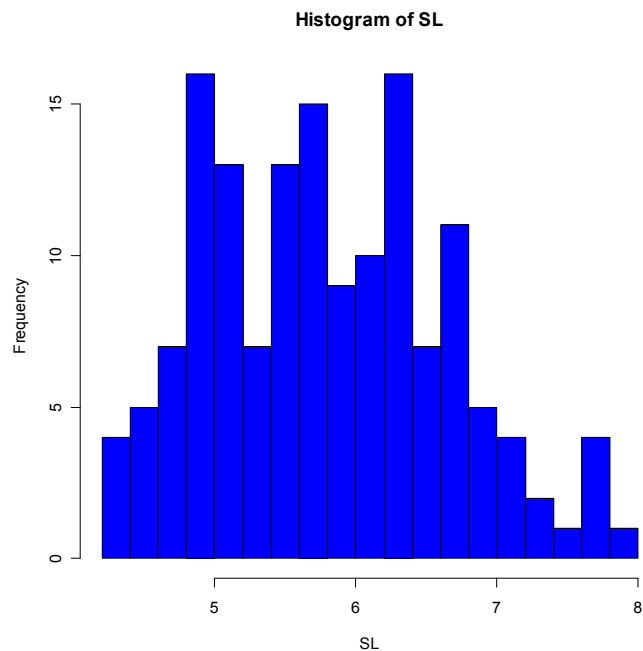
Anderson's Iris data set is part of R's base system and may be used at anytime without loading it.

ANDERSON'S IRIS DATASET:

```
iris
attach(iris)
SL=Sepal.Length
SW=Sepal.Width
PL=Petal.Length
PW=Petal.Width
```

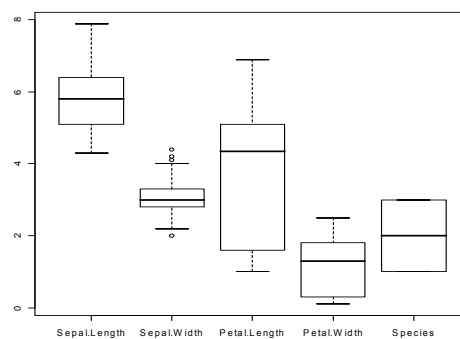
HISTOGRAM:

```
? hist()
hist(SL,breaks=15,col="blue")
```



BOX PLOT:

```
? boxplot()
boxplot(PL)
boxplot(iris)
boxplot(SL~Species)
```



STEM & LEAF PLOT:

```
? stem()
stem(SL)
```

SCATTER PLOTS OF RESIDUALS

DEFINING EXAMPLE X,Y

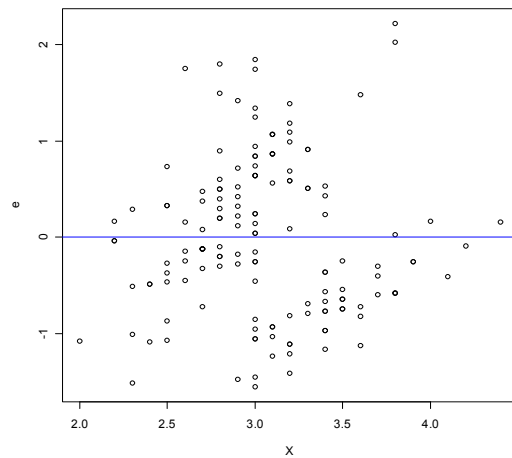
```
Y=SL
X=SW
```

The decimal point is 1 digit(s) to the left of the |

```
42 | 0
44 | 0000
46 | 000000
48 | 00000000000
50 | 00000000000000000000
52 | 00000
54 | 000000000000000
56 | 0000000000000000
58 | 00000000000
60 | 00000000000000
62 | 00000000000000
64 | 00000000000000
66 | 00000000000
68 | 00000000
70 | 00
72 | 0000
74 | 0
76 | 00000
78 | 0
```

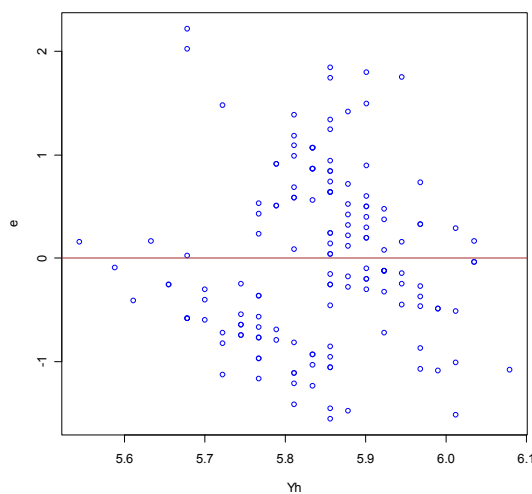
PLOT OF RESIDUALS:

```
e=resid(lm(Y~X))
plot(X,e,col="black")
abline(0,0,col="blue")
```



RESIDUALS VS FITTED VALUES:

```
e=resid(lm(Y~X))
Yh=fitted(lm(Y~X))
plot(Yh,e,col="blue")
abline(0,0,col="brown")
```



PLOT OF SEMI-STUDENTIZED RESIDUALS:**K=summary(lm(Y~X),digits=10)****sqrtMSE=K\$sigma****sqrtMSE****sqrtMSE**

[1] 0.8250966

< compare with Residual
Standard Error in table below.
Calculation is confirmed.

print(summary(lm(Y~X)),digits=10)

Call:

lm(formula = Y ~ X)

Residuals:

	Min	1Q	Median	3Q	Max
	-1.5561393675	-0.6332907839	-0.1119796328	0.5579488182	2.2225494814

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.5262225509	0.4788963398	13.62763	< 2e-16 ***
X	-0.2233610611	0.1550809299	-1.44029	0.15190

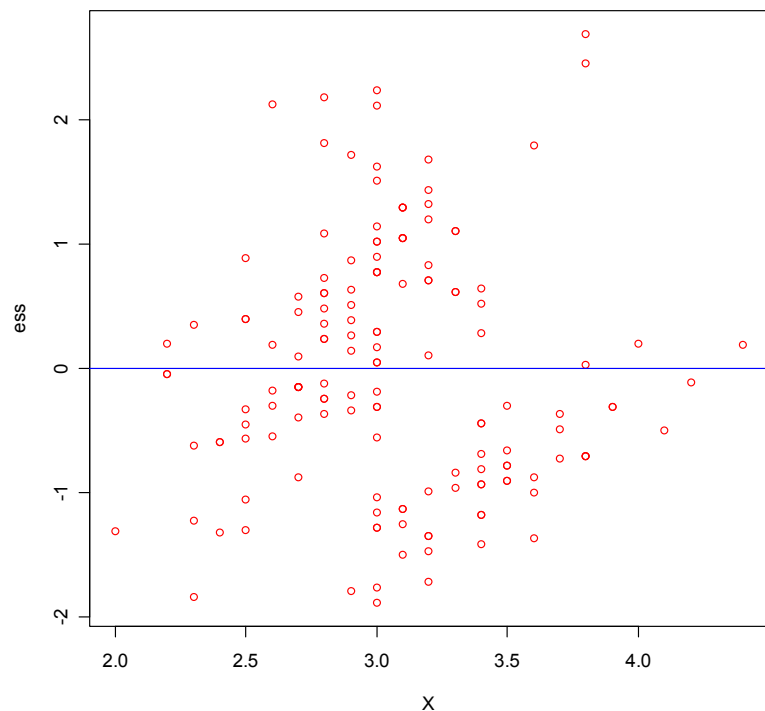
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8250966 on 148 degrees of freedom

Multiple R-Squared: 0.01382265414, Adjusted R-squared:
0.007159293696

F-statistic: 2.074426898 on 1 and 148 DF, p-value: 0.1518982607

ess=e/sqrtMSE
plot(X,ess,col="red")
abline(0,0,col="blue")



```
# LOWESS LINES:  
# USING TOLUCA DATA  
K=read.table("c:/2008LinearModelsData/toluca.txt")  
attach(K)  
X=V1  
Y=V2  
plot(lowess(Y~X),type="l",col="red")  
points(X,Y,col="blue")
```

