S-plus Demonstration: September 13, 2001

The following output demonstrates how to use S-plus to create scatterplots with regression lines, retrieve output containing the value of b0 and b1, residuals, R-squared, etc.

```
> motif()
> conc_read.table(”concord1.dat”, header=T)
> names(conc)
```

[1]	”case”	”water81”	”water80”	”water79”	”income”	”educat”	”retire”
[8]	”peop81”	”cpeop”	”peop80”				

```
> plot(conc$income, conc$water81, xlab=”Income (in thousands)”, ylab=”Summer 1981 Water Use”)
> title(”Income (in thousands) vs. Summer 1981 Water Use in Concord Water Study”)
> abline(lm(conc$water81   conc$income, data=conc))
> waterlm_lm(conc$water81   conc$income, data=conc)

> summary(waterlm)
```

Call: lm(formula = conc$water81   conc$income, data = conc)
Residuals:

Min	1Q	Median	3Q	Max
-2765	-889.8	-239.8	536.8	7010

Coefficients:

	Value	Std. Error	t value	Pr(>| t |)
(Intercept)	1201.1244	123.3245	9.7395	0.0000
conc$income	47.5487	4.6523	10.2205	0.0000

Residual standard error: 1352 on 494 degrees of freedom
Multiple R-Squared: 0.1745
F-statistic: 104.5 on 1 and 494 degrees of freedom, the p-value is 0

Correlation of Coefficients:

	(Intercept)
conc$income	-0.8705

```
> source(”ps.q”)
> ps(”ConcRegLine”)
```

Generated postscript file ”ConcRegLine.ps” (see Figure 1).

```
> names(waterlm)
```

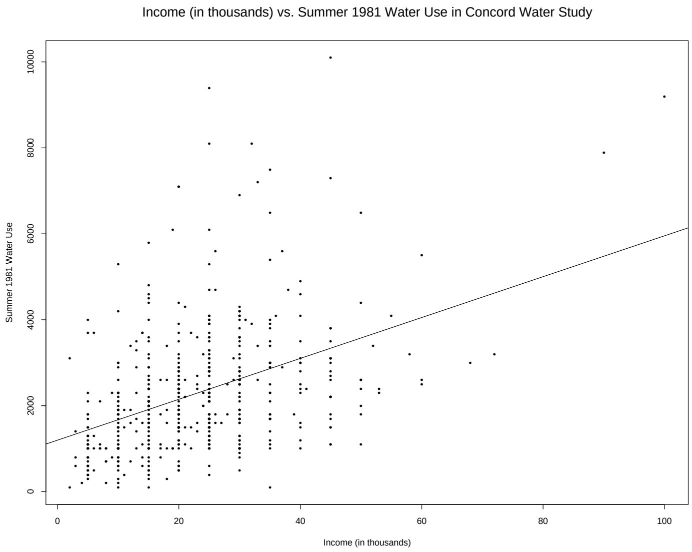

Figure 1: Scatterplot with associated regression line

[1]	"coefficients"	"residuals"	"fitted.values"	"effects"
[5]	"R"	"rank"	"assign"	"df.residual"
[9]	"contrasts"	"terms"	"call"	

Create residual plot to check normality assumption
```
> source("scatbox.q")
> scatbox(waterlm$fit, waterlm$res, xlab="Fit", ylab="Residual")
> abline(h=0) # h denotes y-coordinate for horizontal line on plot
> title("Yhat vs. e Plot for Water-use Regression")
> ps("YhatVsEPlot")
```
Generated postscript file "YhatVsEPlot.ps" (see Figure 2).

NOTE: residuals demonstrate influence and heteroscedasticity. We should transform the data.

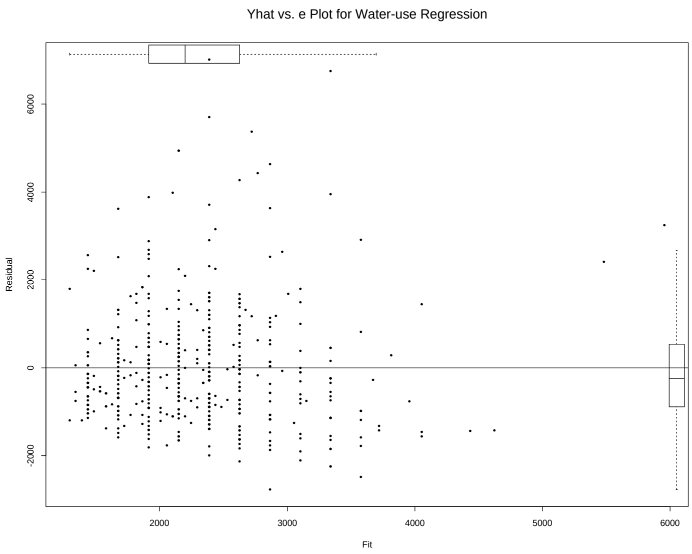

Figure 2: Residual scatterplot to check for normality