Regression Software

Paul E. Johnson[1] [2]

[1]Department of Political Science

[2]Center for Research Methods and Data Analysis, University of Kansas
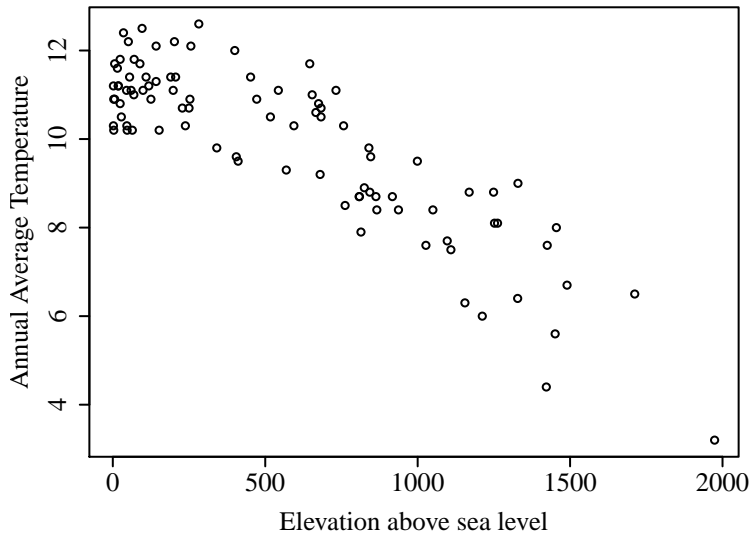
2011

Let's compare some Programs

- Regression Examples

R Code: Get Data, Create a Plot, Run a Regression Model

```
dat ← read.table("/home/pauljohn/ps/SVN−repo/stat
   /DataSets/OregonTemps/ortann.csv",
     header = T, sep = ",")

plot(tann ∼ elevation, data = dat, xlab = "
   Elevation above sea level", ylab = "Annual
   Average Temperature",
     main = "")
mod ← lm(tann ∼ elevation, data = dat)
```

Plot Temperature and Altitude

summary(mod)

```
Call:
lm(formula = tann ~ elevation, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6841 -0.6026 -0.1081  0.7613  2.1034

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.6881385  0.1502978   77.77   <2e-16 ***
elevation   -0.0032377  0.0002016  -16.06   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
    1

Residual standard error: 0.9582 on 90 degrees of freedom
Multiple R^2: 0.7413,   Adjusted R^2: 0.7385
F-statistic: 257.9 on 1 and 90 DF,  p-value: < 2.2e-16
```
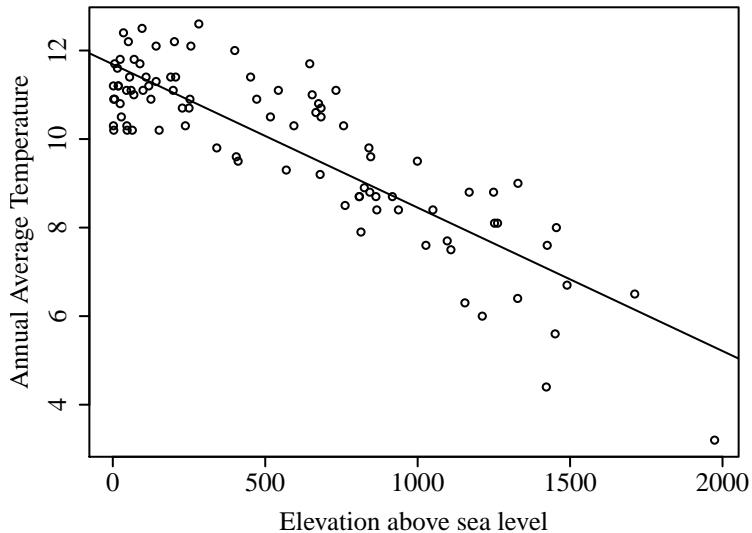
Draw the Fitted Line

Make a Nice Looking Table

	Estimate	S.E.
(Intercept)	11.688*	(0.15)
elevation	-0.003*	(0)
N	92	
RMSE	0.958	
R^2	0.741	

* $p \leq 0.05$

How About SAS?

NOTE: Thumb position at 6.

Dependent Variable: tann tann

Number of Observations Read   92
Number of Observations Used   92

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr >
Model	1	236.83743	236.83743	257.93	<.000
Error	90	82.64127	0.91824		
Corrected Total	91	319.47870			

Root MSE	0.95825	R-Square	0.7413
Dependent Mean	9.88478	Adj R-Sq	0.7385
Coeff Var	9.69416		

SAS (p2)

```
                        The REG Procedure
                          Model: MODEL1
                   Dependent Variable: tann tann

                        Parameter Estimates

                              Parameter      Standard
Variable      Label      DF    Estimate         Error    t Value    Pr > |t|

Intercept     Intercept   1    11.68814       0.15030      77.77     <.0001
elevation     elevation   1    -0.00324       0.00020160   -16.06    <.0001
```

How About Stata? I

```
. use "/home/pauljohn/ps/SVN-repo/stat/Regression/ortann.dta"
(Written by R.              )

. reg tann elevation
```

Source	SS	df	MS		Number of obs =	92
					F(1, 90) =	257.93
Model	236.837428	1	236.837428		Prob > F =	0.0000
Residual	82.6412674	90	.918236304		R-squared =	0.7413
					Adj R-squared =	0.7385
Total	319.478696	91	3.5107549		Root MSE =	.95825

tann	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
elevation	-.0032377	.0002016	-16.06	0.000	-.0036382	-.0028372
_cons	11.68814	.1502978	77.77	0.000	11.38955	11.98673