SIMPLE LINEAR REGRESSION: EXAMPLES

EXAMPLE 1: Coleman Report Data

Data were collected at 20 US schools, and used to examine the relationship between performance of students in the school in a verbal reasoning test and the socioeconomic status of the catchment area.

School	Status	Score	School	Status	Score
	x	y		x	y
1	7.20	37.01	11	-12.86	23.30
2	-11.71	26.51	12	0.92	35.20
3	12.32	36.51	13	4.77	34.90
4	14.28	40.70	14	-0.96	33.10
5	6.31	37.10	15	-16.04	22.70
6	6.16	33.90	16	10.62	39.70
7	12.70	41.80	17	2.66	31.80
8	-0.17	33.40	18	-10.99	31.70
9	9.85	41.01	19	15.03	43.10
10	-0.05	37.20	20	12.77	41.01

Reference: Mosteller and Tukey (1977) *Data Analysis and Regression*

SPSS Results:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-50.682	5.193		-9.760	.000	-61.591	-39.772
	status	1.534	.146	.927	10.499	.000	1.227	1.841

a. Dependent Variable: testscore

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.927[a]	.860	.852	3.70509

a. Predictors: (Constant), status

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1513.213	1	1513.213	110.230	.000[a]
	Residual	247.099	18	13.728		
	Total	1760.312	19			

a. Predictors: (Constant), status

b. Dependent Variable: testscore

Here, to test for significant correlation, we use the test statistic

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.927}{\sqrt{(1-0.927^2)/(20-2)}} = 10.486$$

which we must compare against the Student$(n-2) \equiv$ Student(18) distribution. For a two-tailed test at the significance level $\alpha = 0.05$, the critical values are $C_R = \pm 2.101$ (McClave and Sincich, page 896, column headed $t_{0.025}$), so the hypothesis H_0 that the true correlation is zero is **rejected**.

EXAMPLE 2: Hooker's Temperature and Pressure Data

The following data record the boiling point temperature (in degrees Celsius) of water under different atmospheric pressures. The data were collected in a Himalayan expedition by botanist Joseph Hooker.

x	y	x	y	x	y	x	y
210.8	29.211	196.4	21.928	189.5	18.869	184.1	16.817
210.2	28.559	196.3	21.654	188.8	18.356	183.2	16.385
208.4	27.972	195.6	21.605	188.5	18.507	182.4	16.235
202.5	24.697	193.4	20.480	185.7	17.267	181.9	16.106
200.6	23.726	193.6	20.212	186.0	17.221	181.9	15.928
200.1	23.369	191.4	19.758	185.6	17.062	181.0	15.919
199.5	23.030	191.1	19.490	184.1	16.959	180.6	15.376
197.0	21.892	190.6	19.386	184.6	16.881		

Reference: Forbes, J. (1957). Further experiments and remarks on the measurement of heights by boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

SPSS Results:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	146.673	.776		188.911	.000	145.085	148.261
	Pressure	2.253	.038	.996	59.143	.000	2.175	2.330

a. Dependent Variable: Boiling point of Water (C)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996[a]	.992	.991	.8060

a. Predictors: (Constant), Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2272.474	1	2272.474	3497.902	.000[a]
	Residual	18.840	29	.650		
	Total	2291.315	30			

a. Predictors: (Constant), Pressure

b. Dependent Variable: Boiling point of Water (C)

Here, to test for significant correlation, we use the test statistic

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.996}{\sqrt{(1-0.996^2)/(31-2)}} = 60.027$$

which we must compare against the Student$(31-2) \equiv$ Student(29) distribution. For a two-tailed test at the significance level $\alpha = 0.05$, the critical values are $C_R = \pm 2.045$ (McClave and Sincich, page 896, column headed $t_{0.025}$), so the hypothesis H_0 that the true correlation is zero is **rejected**.