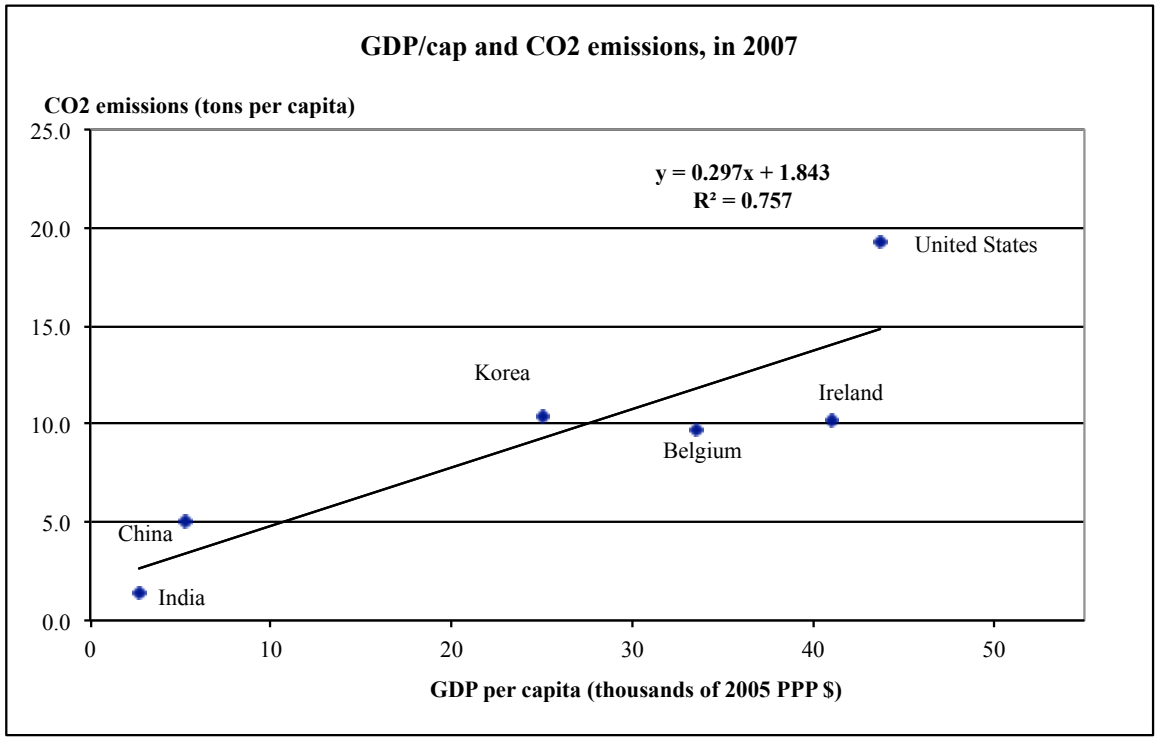
Example of a simple regression



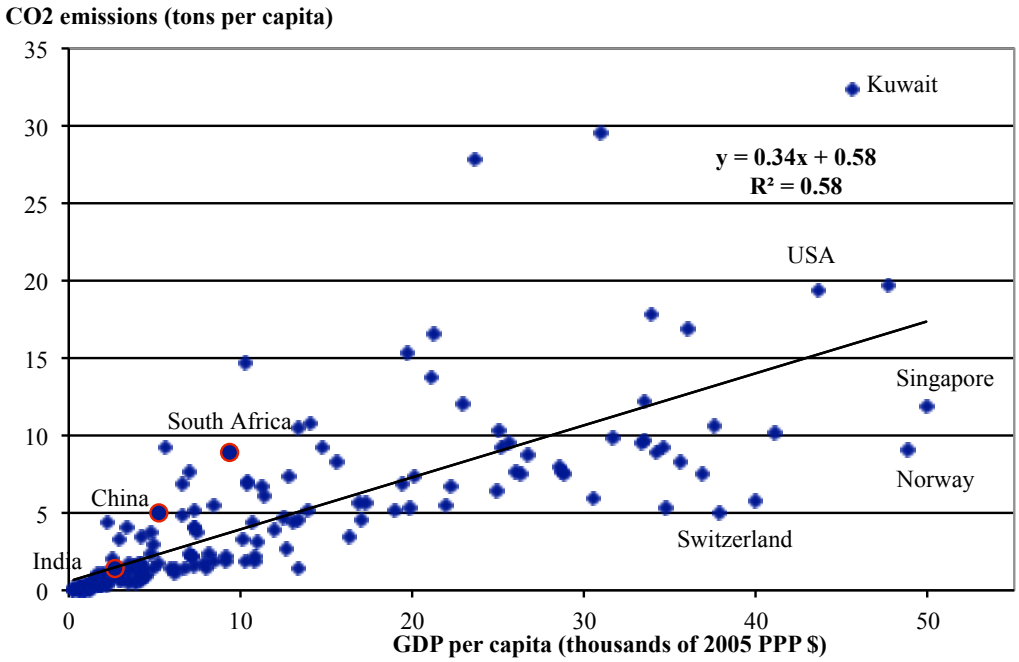
Excel spreadsheet:

	GDP/cap (PPP$ 1000) x	CO2/cap (tons) y	(x-xbar)	(y-ybar)	(x-xbar)^2	(x-xbar)(y-ybar)	yhat	uhat^2	(y-ybar)^2
United States	43.7	19.3	18.5	10.0	342.9	184.5	14.84	19.9	99.3
Ireland	41.0	10.2	15.8	0.9	250.2	13.7	14.04	14.7	0.8
Belgium	33.5	9.7	8.3	0.4	69.2	3.0	11.81	4.4	0.1
Korea, Rep.	25.0	10.4	-0.2	1.1	0.0	-0.2	9.28	1.3	1.1
India	2.7	1.4	-22.5	-7.9	505.5	178.4	2.65	1.6	62.9
China	5.2	5.0	-20.0	-4.3	399.3	86.6	3.39	2.6	18.8
Sum	151.1	56.0	0.0	0.0	1567.1	466.1	56.0	44.5	183.1
Sample mean	25.2	9.3	0.0	0.0					
Sample variance/covariance					313.4	93.2			

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} = \frac{93.2}{313.4} = 0.297 \text{ tons/\$1000} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 9.3 - 0.297 * 25.2 = 1.843 \text{ tons}$$

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{44.5}{183.1} = 0.757$$

Per Capita Carbon Dioxide Emission, 2007



Math/Stat Review

Sample mean of the characteristic x from n observations $x_1, x_2, \ldots, x_n$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_i x_i$$

Sample variance:

$$S_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

Sample standard deviation: $S_x = \sqrt{S_x^2}$

Sample covariance between two characteristics of the same n observations $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$

$$S_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \ldots + (x_n - \bar{x})(y_n - \bar{y})}{n-1} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

Sample correlation: $\dfrac{S_{xy}}{S_x S_y}$

Two useful relationships

$$\overline{ax+b} = \frac{ax_1 + b + ax_2 + b + \ldots + ax_n + b}{n} = a\bar{x} + b$$

$$S_{ax+b}^2 = \frac{(ax_1 - a\bar{x})^2 + (ax_2 - a\bar{x})^2 + \ldots + (ax_n - a\bar{x})^2}{n-1} = a^2 S_x^2$$