Regression

```
> ### Descriptive Statistics

> lapply(CorrelationData, function(x) c(length(x), mean(x), sd(x)))

$Outcome1
[1] 4.00000 2.00000 2.44949

$Outcome2
[1] 4.00000 6.00000 2.44949

> cov(Outcome1,Outcome2)

[1] 3

> ### Inferential Statistics

> model <- lm(Outcome2 ~ Outcome1)
> summary(model)

Call:
lm(formula = Outcome2 ~ Outcome1)

Residuals:
   1    2    3    4
-1.0  2.0 -2.5  1.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0000     1.7854   2.801    0.107
Outcome1      0.5000     0.6124   0.816    0.500

Residual standard error: 2.598 on 2 degrees of freedom
Multiple R-squared:   0.25,     Adjusted R-squared:  -0.125
F-statistic: 0.6667 on 1 and 2 DF,  p-value: 0.5
```

These statistics calculated separately for each variable using procedures described previously.

The Covariance ("COV") is not easily determined from the summary statistics of the output, but rather from the data (and the calculations are therefore not shown here).

The Unstandardized Regression Coefficients ("Estimate") are also a function of the Covariance and the descriptive statistics:

$$B_1 = \frac{COV}{(SD_X)^2} = \frac{3.000}{(2.449)^2} = 0.500$$

$$B_0 = M_Y - (B_1)(M_X) = 6.000 - (0.500)(2.000) = 5.000$$

The Standardized Regression Coefficient for the predictor can be similarly determined:

$$\beta_1 = B_1\left(\frac{SD_X}{SD_Y}\right) = 0.500\left(\frac{2.449}{2.449}\right) = 0.500$$

"R" is a function of the covariance and the standard deviations of both variables:

$$R = \frac{COV}{(SD_X)(SD_Y)} = \frac{3.000}{(2.45)(2.45)} = 0.500$$

$$R^2 = 0.500^2 = 0.250$$