

Multiple Regression

M. Drew LaMar
September 12, 2016

Introduction to Quantitative Biology, Fall 2016

Class announcements

- No blog this week!
- Homework is due Wednesday

Linear Models

Definition: A *model* is a simplified abstract or concrete representation of objects and their representations or processes in the real world.

Is a linear regression model representing a relationship or process occurring between two variables?

 When poll is active, respond at Pollev.com/mdlama  Text **MDLAMA** to **37607** once to join

Simulated data - Two explanatory variables

Let's create two uncorrelated continuous variables ($\rho = 0$), each with a gaussian/normal distribution.

```
S <- matrix(c(1,0.0,0.0,1),2,2) #Correlation
matrix
X <- rmvnorm(mean=c(0,0), sig=S, n=100)
head(X)
```

Activities: Check that each column of X is normally distributed. What is the resulting correlation coefficient? Use the cor.test function to find a confidence interval.

Simulated data - Response

Let's create a response variable Y that depends on the two variables X_1 and X_2 .

```
X1 <- X[,1]
X2 <- X[,2]
Y <- 1 + 2*X1 + 4*X2
ucor.data <- data.frame(X1 = X1, X2 = X2, Y = Y)
```

Activities: Construct 3 linear models: (1) using X_1 and X_2 to predict Y , (2) using only X_1 , and (3) using only X_2 .

Activities: Repeat the above with a correlation coefficient of $\rho = 0.9$ between X_1 and X_2 . What do you notice?