BIO 682
Multivariate Statistics (Lite)
Spring 2010

Steve Shuster

http://www4.nau.edu/shustercourses/BIO682/index.htm

Lecture 10

Outline for This Section

1. Multiple regression in ecological and behavioral analyses.
2. A summary of multivariate methods for ecological analysis (PCA, DCA, NMDS)
3. Additional variance component approaches for understanding ecological data.
4. Note: this will not be an exhaustive consideration of multivariate methods.

Multiple Regression

The general purpose of multiple regression (c.f., Pearson, 1908) is to learn more about the relationship between ***several independent*** or predictor variables (X_1-X_k) and a ***dependent*** or criterion variable ($\hat{Y}$).

In general, multiple regression allows the researcher to determine ***which independent variable is most successful*** at predicting variation in the independent variable.

Multiple Regression

The conventional partial regression equation from Sokal and Rohlf (1995), estimates the value of the dependent variable, $\hat{Y}$, as a function of k independent variables, $X_1 \ldots X_k$.

This approach is based on the standard least squares approach of linear regression.

$$\mathbf{Y = a + bX}$$

Multiple Regression

$$\mathbf{Y = a + bX}$$

The Y variable can be expressed in terms of a ***constant*** (**a**) and a ***slope*** (**b**) multiplied by the X variable.

The constant is also referred to as the ***intercept***, and the slope as the ***regression coefficient*** or β coefficient.

In The Multivariate Case,

Each Y and X_j variable is standardized as y' or x',

$$y' = (Y - \boldsymbol{\bar{Y}}) / s_Y$$

and

$$x_j' = (X_j - \boldsymbol{\bar{X}}_j) / s_{Xj}$$

where $\boldsymbol{\bar{Y}}$ and $\boldsymbol{\bar{X}}_j$ represent sample averages, and s_Y and s_{Xj} represent the standard deviations of each sample.

Partial Regression Coefficients

Each standard partial regression coefficient is expressed as,

$$b'_{Yj.} = bY_{j.}\,(s_{Xj} / s_Y).$$

Thus, the standardized multiple regression equation is,

$$\hat{y}' = b'_{Y1}.x'_1 + b'_{Y2.}x'_2 + \ldots + b'_{Yk.}x'_k$$

Partial Regression Coefficients

$$\hat{y}' = b'_{Y1}.x'_1 + b'_{Y2.}x'_2 + \ldots + b'_{Yk.}x'_k$$

This expression gives the ***rate of change*** of the dependent variable, Y, in standard deviation units, per one standard deviation unit of independent variable X_j, with all other independent variables ***held constant***.

.

Partial Regression Coefficients

$$y^{\wedge\prime} = b'_{Y1}.x'_1 + b'_{Y2.}x'_2 + \ldots + b'_{Yk.}x'_k$$

Another way to express this fact is to say that, for example, variable X_1 is correlated with the Y variable, after ***controlling for*** all other independent variables.

Partial Regression Coefficients

$$\hat{y}' = b'_{Y1.}x'_1 + b'_{Y2.}x'_2 + \ldots + b'_{Yk.}x'_k$$

The magnitude of the standard partial regression coefficients tells you something about the ***relative importance*** of different variables.

X variables with ***larger*** standardized partial regression coefficients have a ***stronger*** relationship with the Y variable.

For Example,

A significant negative correlation between hair length and height exists in many human populations (i.e., short people have longer hair).

However,

IF we were to add the variable *Gender* into the multiple regression equation, this correlation disappears.

This is because women, on the average, have ***longer hair*** than men;

They also are ***shorter*** on the average than men.

After We Remove Gender,

The relationship between hair length and height ***disappears*** because hair length does not make a unique contribution to the prediction of height, beyond what it shares in the prediction with variable *Gender*.

Stated differently, after controlling for the variable *Gender*, the partial correlation between hair length and height is nearly zero.

height' = $\mathbf{b'}_{\mathbf{Ygender.}}x'_{gender} + \mathbf{b'}_{\mathbf{Yhair.}}x'_{hair}$

How Much Is Explained?

Each added X_i causes R^2 to increase (unless the variable has exactly the same values as another).

How well the equation fits the data is expressed by R^2, the "coefficient of multiple determination;“ $R^2 = 0$ to1.

Significance is a function of the R^2, the number of observations, and the number of X variables.

Additional Notes

The procedures for interpreting R^2 and examining “F to Remove.”

http://udel.edu/~mcdonald/statmultreg.html

Applications

Examining the relative effect of spatial and temporal variation in mate availability on mating system evolution.

Examining the relative effects of selection acting in different contexts or at different levels of organization.

Factors Influencing the Intensity of Sexual Selection

1. The mean spatial crowding of mates (m^*)

2. The mean temporal crowding of mates (t^*).

Shuster & Wade 2003

The Mean Crowding of Females in Space

The mean crowding of females on resources defended by males can be expressed as,

$$m^* = m + [(V_m / m) - 1]$$

Where m = the mean number of females per patch, V_m = the variance about this mean, and m^* is the mean crowding of females per patch.

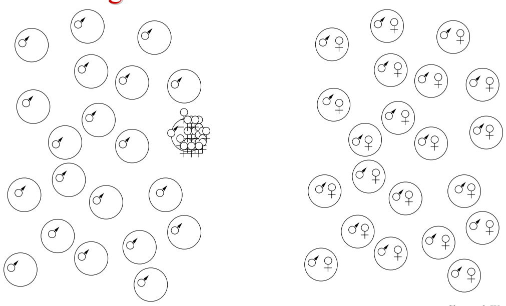
In this context, m^* represents the number of other females the average female experiences on her resource patch.

The Spatial Distribution

$N_{males} = N_{females} = 20$

m^* = large

m^* = small



Shuster & Wade 2003

The Mean Crowding of Females in Time

The mean crowding of females within intervals during the breeding season can be expressed as,

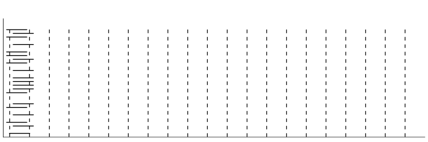
$$t^* = t + [(V_t / t) - 1]$$

Where t = the mean number of females per interval, V_t = the variance about this mean, and t^* is the mean crowding of females per interval.

t^* = the number of other receptive females the average receptive female experiences within her period of receptivity.

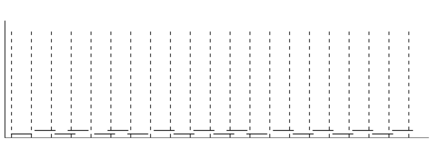
The Temporal Distribution

$N_{males} = N_{females} = 20$



t^* = large

t^* = small



Shuster & Wade 2003

The Opportunity for Selection

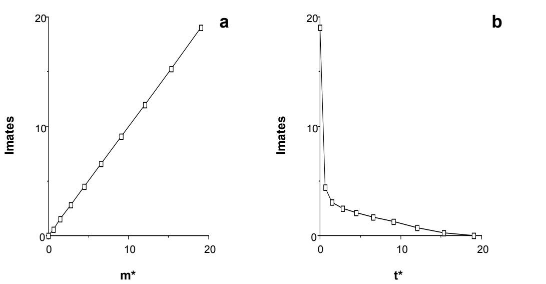
(Crow 1958, 1962; Wade 1979)

$$I = V_W / W^2 = V_w$$

Compares the fitness of breeding parents *relative* to the population before selection.

The variance in fitness, V_w, provides an empirical estimate for selection's strength.

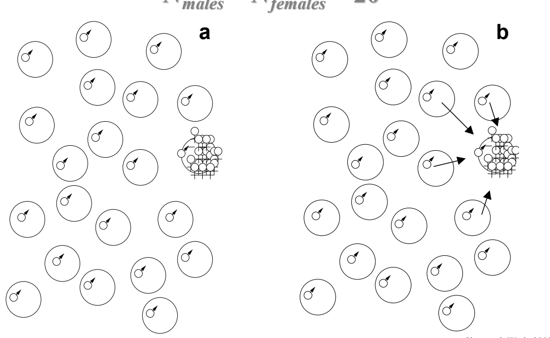
The Relationships Between m^*, t^* and I_{mates}



Shuster & Wade 2003

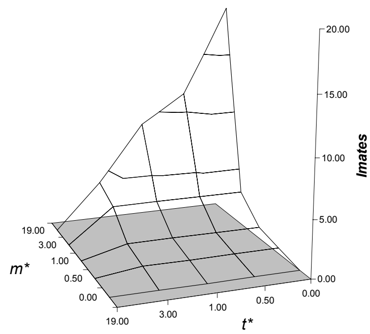
Mating System Dynamics

$N_{males} = N_{females} = 20$



Shuster & Wade 2003

The I_{mates} Surface



Shuster & Wade 2003

Considering Regression Coefficients for *m** and *t**

Thus, the standardized multiple regression equation is,

$$\hat{y}' = b'_{Y1.}x'_1 + b'_{Y2.}x'_2$$

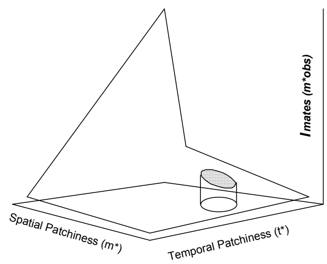
where x'_1 equals the standardized spatial patchiness of receptive females, *m**, and x'_2 equals the standardized temporal patchiness of receptive females, *t**.

And,

Each $b'_{Yi.}$ is the standardized partial regression coefficient of the relationship between the spatial or temporal distribution of females and the opportunity for sexual selection, I_{mates}.

That is, each *b*-prime gives the expected rate of change of the dependent variable, I_{mates}, in standard deviation units per one standard deviation unit of each independent variable, while holding the other independent variable constant.

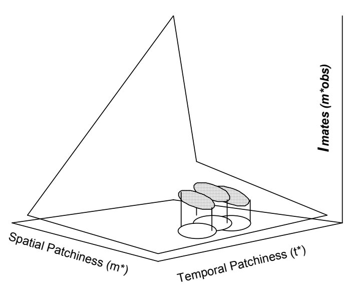
Within Season Variation in m^* and t^*



Will describe an ellipsoid on the I_{mates} surface.

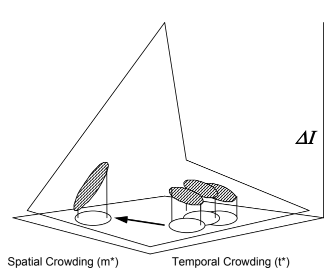
The slope of the ellipsoid will depend on how m^* and t^* interact within a single breeding season.

Among-Species Variation in m^* and t^*



Each species is expected to generate a unique cloud of points along the m^*, t^* and I_{mates} axes.

Divergence in m^* and t^*



Related species are expected to show similar mating systems.

Also, divergent species within larger taxa should be *predictably divergent*.