Multiple regression:
fitting the best fitting plane

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

Where, a = intercept (when X_1=0 and X_2=0)

Y = income

X_1 = education

X_2 = % female in occupation

[unit of analysis = occupation]

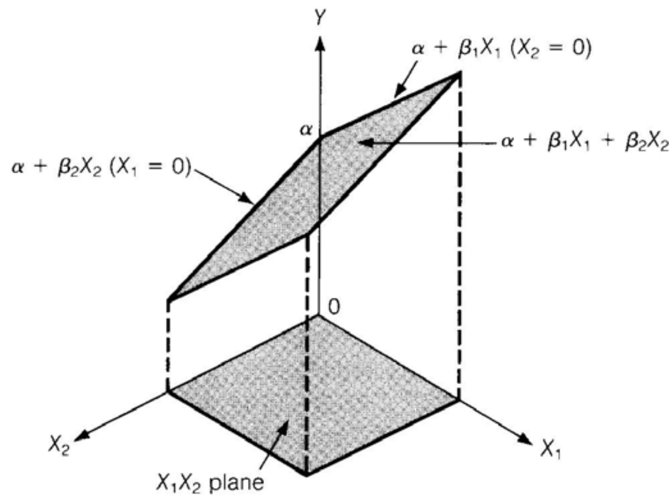
Multiple regression: model

Educ

Income

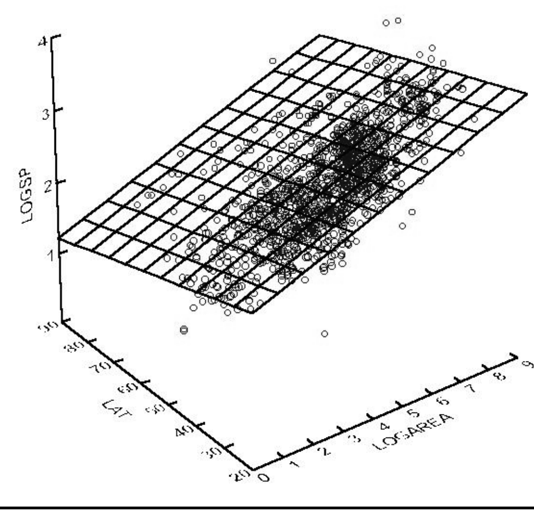
% female

Multiple regression: 3D model



Source: Agresti and Finlay, 1986, p. 317

Multiple regression: 3D model



Multiple regression:
interpretation of b's

- b_1 = slope for education (net effect of education on income controlling for percent female; how much income in dollars for each year of education)
- b_2 = slope for % female (net effect of percent female on income controlling for education; how much income in dollars for each percent female)

Multiple regression:
interpretation of b's

Coefficients called:

- Metric coefficient
- Net regression coefficient
- Unstandardized regression coefficient

Interpretation:

- Net effect
- Independent effect
- Controlling for

Multiple regression:
standardized regression coefficients

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 \implies$$

$$\hat{y} = B_1 x_1 + B_2 x_2$$

where, $B_{yx} = b_{yx} (s_x/s_y)$

Creating standard scores:

$x_1 = (X_1 - \overline{X}_1)/sX_1$

Multiple regression:
standardized regression coefficients

- B's range from -1 to +1
- Interpretation: a one s.d. change in the independent variable, produces a predicted change of "Beta" s.d.'s in the dependent variable, net of other variables
- More common interpretation: if $B_1 > B_2$ then education is much more important in predicting income than is % female
- "considerably larger, more than twice as important"

Murdock data:

Y = stratification, X_1=political integration, X_2=money exchange

Society	X_1	X_2	Y	Society	X_1	X_2	Y
001	2	0	1	011	1	1	1
003	3	2	2	013	0	0	0
005	3	3	3	015	2	0	1
007	4	0	2	017	2	3	1
009	0	0	0	019	4	3	2

Calculating standardized and unstandardized coefficients

Computational formula:

$$r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2]\,[n \sum Y_i^2 - (\sum Y_i)^2]}}$$

Calculating standardized and unstandardized coefficients

- $r_{yx1} = .865$
- $r_{yx2} = .620$
- $r_{x1x2} = .482$

Calculating standardized coefficients

$$Byx.z = \frac{r_{yx} - r_{yz}r_{xz}}{1 - r^2_{xz}}$$

[General formula]

Calculating standardized coefficients

- $Byx_1.x_2 = .737$
- $Byx_2.x_1 = .265$

$$\hat{y} = .737x_1 + .265x_2$$

Interpretation: political integration is substantially more important than money in determining level of stratification

Calculating unstandardized coefficients

$b_{yx} = B_{yx}\,(s_y/s_x)$, using s.d. computational formula:

$$s_y = \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2}$$

$\overline{Y}$

Calculating unstandardized coefficients

Calculate s.d.'s:

- $s_y = .9$
- $s_{x1} = 1.37$
- $s_{x2} = 1.33$

Calculate b's:

$b_{yx1.x2} = B_{yx1.x2}\ (s_y/s_{x1})$
$= (.737)\ (.9/1.37) = .484$

$b_{yx2.x1} = B_{yx2.x1}\ (s_y/s_{x2})$
$= (.265)\ (.9/1.33) = .179$

Calculating intercept

$\overline{Y} = a + b_1\overline{X}_1 + b_2\overline{X}_2$

$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$

$a = (1.3) - (.484)(2.1) - (.179)(1.2) = .0688$

$a = .069$

Prediction equation

$\hat{Y} = .069 + .484X_1 + .179X_2$

Interpret!

Calculating R^2

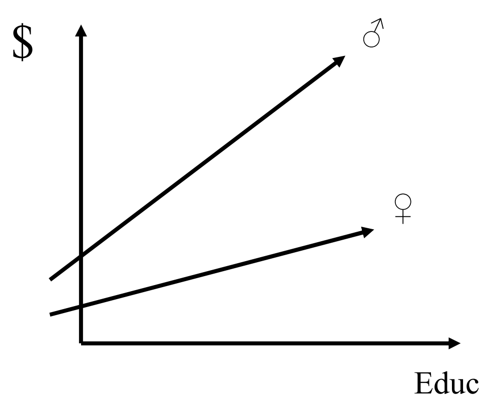
$R^2_{y.x1x2} = B_{yx1}r_{yx1} + B_{yx2}r_{yx2}$
$= (.737)(.865) + (.265)(.620)$
$= .802$

Interpretation: 80 percent of the variation in stratification is explained by political integration and money

Standardized vs. unstandardized coefficients

- Use standardized to compare variables ***within*** equations
- Use unstandardized to compare same variable ***across*** equations

Unstandardized coefficients



Caveats

1) Don't interpret regression lines beyond where you have data
2) Report to three significant nonzero digits (retain larger # of digits in intermediate calculations)
3) Multicollinearity: problem with high correlations