Week 1 - Math Facts
Standard Multiple Regression

Stat 209
D Rogosa
[redacted]

I Two predictor model $Y \quad X_1 \; X_2$

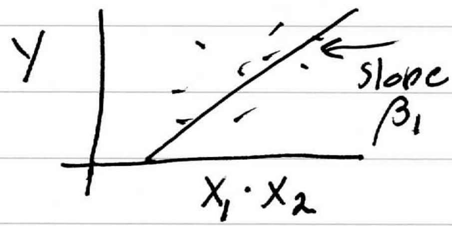
$$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

adjusted (partial) variable interpretation

$$\beta_1 = \beta_{YX_1 \cdot X_2} = \beta_{Y(X_1 \cdot X_2)}$$

adjusted var $X_1 \cdot X_2 = X_1 - \beta_{X_1X_2} X_2$

same for β_2



fantasy: "holding constant"
reality: "reaming out"

M-B sec 6.2 nihills data

extends to p predictors: $\beta_1 = \beta_{Y(X_1 \cdot \underset{\sim}{X}_{(1)})}$ $\quad \underset{\sim}{X}_{(1)}$ all but X_1
(see MT, Coleman ex, Berk, NWK)

II Errors in Variables

a. Familiar 1 predictor case $\quad X = \xi + \epsilon$ (obs, true)

$$E(Y|\xi) = \beta_0 + \beta_1 \xi$$

$$E(Y|X) = \gamma_0 + \gamma_1 X \qquad \text{result: } \gamma_1 = \beta_1 R_X$$

[derive on reverse] c.f. MB sec 6.7 diet data $\lambda = R_X$

attenuation: proportional bias $\quad R_X = \dfrac{Var(\xi)}{Var(X)}$ reliability

correlation analog: $Y = \eta + \delta$

$$\rho_{XY} = \rho_{\eta\xi}\sqrt{R_X R_Y}$$

δ, ϵ indep.

IIa. Single predictor result:

$$\gamma_1 = \frac{Cov(Y, X)}{Var(X)} = \frac{Cov(Y, \xi) + \cancel{Cov(Y, \epsilon)}^{\,0}}{Var(X)}$$

$$= \frac{Cov(Y, \xi)}{Var(\xi) * [Var(X)/Var(\xi)]} = \beta_1 \left(\frac{Var(\xi)}{Var(X)} \right) = \beta_1 R_X$$

MB p 208 $\quad \tau^2 = \sigma_\epsilon^2 \quad s_2^2 = Var(\xi) \quad \lambda = R_X \; (=.86)$

Week 1 math facts cont'd

b. Two predictor case (Cochran 1968, 1970)

upgrade notation $X_1 = \xi_1 + \epsilon_1$, $X_2 = \xi_2 + \epsilon_2$, $\rho = \mathrm{Corr}(\xi_1, \xi_2)$

ϵ_1, ϵ_2 independent, R_i = reliability $X_i = \sigma^2_{\xi_i} / \sigma^2_{X_i}$

true score Regression $E(Y \mid \xi_1, \xi_2) = \beta_0 + \beta_1 \xi_1 + \beta_2 \xi_2$

observed score regr. $E(Y \mid X_1, X_2) = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$

eq (11.2)

$$\gamma_1 = \frac{\beta_1 R_1 (1 - \rho^2 R_2) + \beta_2 \beta_{\xi_2 \xi_1} R_1 (1 - R_2)}{1 - \rho^2 R_1 R_2}$$

bias pos, neg, even flip signs

for γ_2 permute 1,2 subscripts in above

Special Case: $R_2 = 1$ ($X_2 = \xi_2$, no meas error)

$$\gamma_1 = \beta_1 R_1 \left[\frac{1 - \rho^2}{1 - \rho^2 R_1} \right]$$

more severe attenuation

$$\gamma_2 = \beta_2 + \frac{\beta_1 \beta_{\xi_1 \xi_2} (1 - R_1)}{1 - \rho^2 R_1}$$

[cf M-B eq 6.4]

M-B p.206

bias pos or neg for γ_2, even though X_2 measured perfectly

can flip relative magnitudes, even signs

same for "variance explained" measures

DAAG errorsINx() function

Comments

1) analogs of I, II for logistic regression

2) regression is good for fitting, interpret coeffs w/ caution