Ch 4 : Categorical Independent variables

Indicator Dummy variables with intercept.

EV Simple regression

$x_i = 1$ if Drug
$\quad\;\; 0$ if Placebo

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

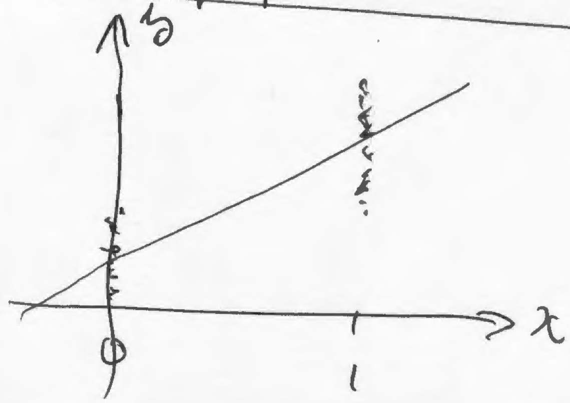
	x	$E(y \mid x)$	
Drug	1	$\beta_0 + \beta_1$	$= \mu_1$
Placebo	0	β_0	$= \mu_2$

test of $H_0 : \beta_1 = 0$

Same as traditional 2-sample test

Estimated regression coefficients

	x	$\hat{y} = b_0 + b_1 x$
Drug	1	$b_0 + b_1 = \bar{y}_1$
Placebo	0	$b_0 = \bar{y}_0$



Drug A, Drug B, Placebo

Condition	x_1	x_2	$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
Drug A	1	0	$\beta_0 + \beta_1 = \mu_1$
Drug B	0	1	$\beta_0 + \beta_2 = \mu_2$
Placebo	0	0	$\beta_0 = \mu_3$

Category with no indicator is the <u>REFERENCE</u> category.

Is there a difference in mean response for the 3 experimental treatments?

$H_0: \beta_1 = \beta_2 = 0$

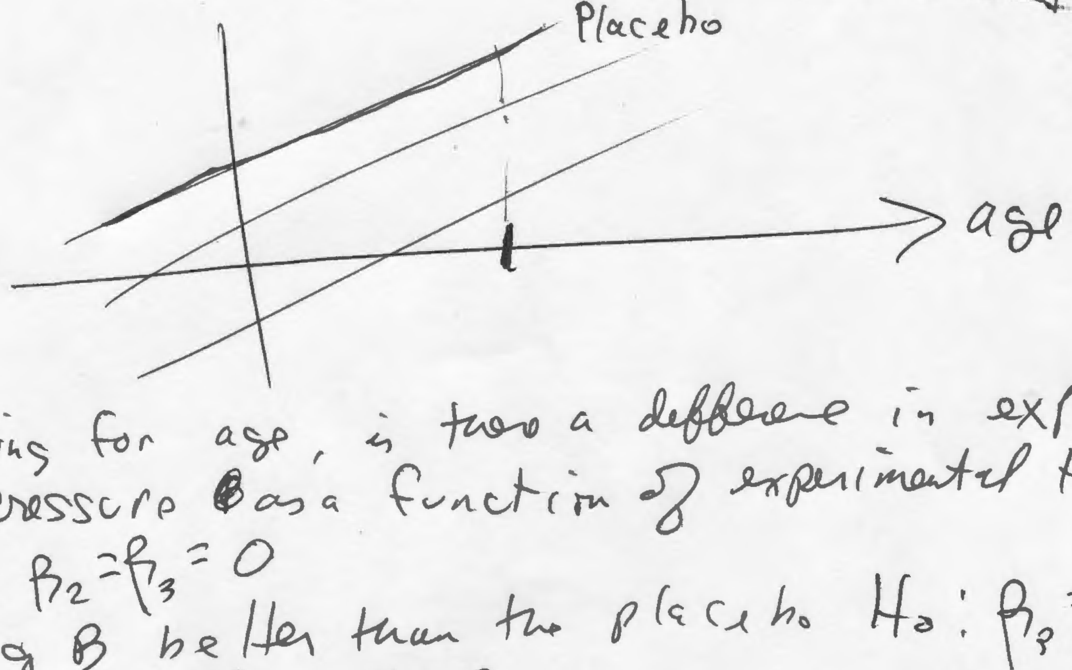
Compare Drug A & B $H_0: \beta_1 = \beta_2$

For p categories, need $p-1$ dummy variables
(If the model has an intercept)

Add a quantitative independent variable
("covariate")

x_1 = Age
x_2 = 1 if Drug A, 0 otherwise
x_2 = 1 if Drug B, 0 otherwise

	x_2	x_3	$E(y\|x_1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
A	1	0	$(\beta_0 + \beta_2) + \beta_1 x_1$
B	0	1	$(\beta_0 + \beta_3) + \beta_1 x_1$
Placebo	0	0	$\beta_0 + \beta_1 x_1$



controlling for age, is there a difference in expected blood pressure as a function of experimental treatment?

$H_0: \beta_2 = \beta_3 = 0$

Is Drug B better than the placebo $H_0: \beta_3 = 0$

$$F = \frac{SSR_F - SSR_R}{m s^2} = \left(\frac{n-q-1}{m}\right)\left(\frac{a}{1-a}\right)$$

Indicator dummy variables
with no Intercept

p dummy variables for p categories

$E(y) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

	X_1	X_2	X_3	
A	1	0	0	$\beta_1 = \mu_1$
B	0	1	0	$\beta_2 = \mu_2$
Placebo	0	0	1	$\beta_3 = \mu_3$

"Cell means" coding

Add a covariate

$E(y) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 \underset{\text{Age}}{X_4}$

A	1	0	0	$\beta_1 + \beta_4 X_4$
B	0	1	0	$\beta_2 + \beta_4 X_4$
Placebo	0	0	1	$\beta_3 + \beta_4 X_4$

$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

	X_1	X_2	
A	1	0	$(\beta_0 + \beta_1) + \beta_3 X_3$
B	0	1	$(\beta_0 + \beta_2) + \beta_3 X_3$
Pla	0	0	$\beta_0 + \beta_3 X_3$

One-to-one linear transformations of independent variables

$$y = X\beta + \varepsilon = \underset{n\times(k+1)}{X}\,\underset{(k+1)\times(k+1)}{A}\,A^{-1}\beta + \varepsilon = \underset{XA}{X^*}\,\underset{A^{-1}\beta}{\beta^*} + \varepsilon$$

$$\overset{X}{\begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & & & \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix}} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} x^*_{11} & x^*_{12} & & x^*_{14} \\ x^*_{21} & \vdots & & \vdots \\ \vdots & & & \\ x^*_{n1} & x^*_{n2} & & x^*_{n4} \end{pmatrix}$$

Transform X by $XA = X^*$

Fit the model $y = X^*\beta + \varepsilon$

$$b^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y$$

$$= ((XA)'XA)^{-1}(XA)'y$$

$$= (A'X'XA)^{-1}A'X'y$$

$$= A^{-1}(X'X)^{-1}\underbrace{A'^{-1}A'}_{I}X'y = A^{-1}b$$

$$\beta^* = A^{-1}\beta$$

$$\hat{y}^* = X^* b^* = X \underbrace{A A^{-1}}_{I} b = \hat{y}$$

$$e^* = e, \quad s^{2*} = s^2$$

$$H_0 : C\beta = \gamma \iff \underbrace{CA}_{C^*}\,\underbrace{A^{-1}\beta}_{\beta^*} = \gamma$$

$$F = \frac{(Cb-\gamma)'(C(X'X)^{-1}C')^{-1}(Cb-\gamma)}{m s^2}$$

$$F^* = \frac{(C^*b^*-\gamma)'(C^*(X^{*\prime}X^*)^{-1}C^{*\prime})(C^*b^*-\gamma)}{m s^2}$$

$$= \frac{(C\underbrace{A A^{-1}}_{I} b-\gamma)'(CA((XA)'XA)^{-1}(CA)')^{-1}(Cb-\gamma)}{m s^2}$$

$$= \frac{(Cb-\gamma)'(CA(A'X'XA)^{-1}A'C')^{-1}(Cb-\gamma)}{m s^2}$$

$$= \frac{(Cb-\gamma)'(C\underbrace{A A^{-1}}_{I}(X'X)^{-1}\underbrace{A'^{-1}A'}_{I}C')(Cb-\gamma)}{m s^2}$$

$$= F$$

$$E(y\mid x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

	x_1	x_2	$E(y\mid x_3)$
A	1	0	$\beta_0 + \beta_1 + \beta_3 x_3$
B	0	1	$\beta_0 + \beta_2 + \beta_3 x_3$
Pl	0	0	$\beta_0 + \beta_3 x_3$

$$H_0: \beta_1 = \beta_2 = 0$$

or

$$E(y\mid x_4^*) = \beta_1^* x_1^* + \beta_2^* x_2^* + \beta_3^* x_3^* + \beta_4^* x_4^*$$

	x_1^*	x_2^*	x_3^*	$E(y\mid x_4^*)$
A	1	0	0	$\beta_1^* + \beta_4^* x_4^*$
B	0	1	0	$\beta_2^* + \beta_4^* x_4^*$
Pl	0	0	1	$\beta_3^* + \beta_4^* x_4^*$

$$H_0: \beta_1^* = \beta_2^* = \beta_3^*$$

$$\begin{array}{c} \;\; x_1 \;\; x_2 \;\; x_3 \\ \begin{pmatrix} 1 & 1 & 0 & a_1 \\ 1 & 0 & 1 & a_2 \\ 1 & 0 & 0 & a_3 \\ \vdots & & & \\ 1 & & & \end{pmatrix} \end{array} \underbrace{\begin{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}^{-1} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{pmatrix}}_{4\times 4} = \begin{array}{c} x_1^* \;\; x_2^* \;\; x_3^* \;\; x_4^* \\ \begin{pmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \\ & & & \end{pmatrix} \end{array}$$

A

X^*

$$X \begin{pmatrix} 1 & 1 & 0 & \vdots & 1 & 0 & 0 \\ 1 & 0 & 1 & \vdots & 0 & 1 & 0 \\ 1 & 0 & 0 & \vdots & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{cc} X & A \\ \begin{pmatrix} 1 & 1 & 0 & a_1 \\ 1 & 0 & 1 & a_2 \\ 1 & 0 & 0 & a_3 \\ & \vdots & & \end{pmatrix} & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & \alpha \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$= \begin{pmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \\ & \vdots & & \end{pmatrix}$$

<u>It worked</u>

Interactions: It Depends

Quantitative by Quantitative

$$y_i = \beta_0 + \beta_1 \overset{\text{weight}}{x_{i1}} + \beta_2 \overset{\text{Alch}}{x_{i2}} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$$

Hold x_1 constant $\cancel{+\varepsilon_i}$

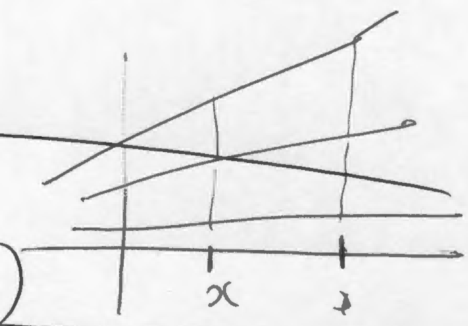
$$= (\beta_0 + \beta_1 x_1) + (\beta_2 + \beta_3 x_1) x_2 + \varepsilon$$

$$= (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3 x_2) x_1 + \varepsilon$$

The relationship between x_2 & $E(y)$ <u>depends</u> on value of x_1

The rel bet x_1 & $E(y)$ depends on x_2

Interactions of Categorical by Quantitative Independent Variables



Treat	d_1	d_2	$E(y \mid x)$
A	1	0	$(\beta_0+\beta_2)+(\beta_1+\beta_4)x$
B	0	1	$(\beta_0+3)+(\beta_1+\beta_5)x$
C	0	0	$\beta_0 + \beta_1 x$

$$y_i = \beta_0+\beta_1 x+\beta_2 d_1+\beta_3 d_2 + \beta_4 d_1 x+\beta_5 d_2 x+\varepsilon_i$$

What Null hypothesis would you test for

Equal Slopes: $H_0: \beta_4=\beta_5=0$

Compare slopes for A vs C: $H_0: \beta_4=0$

" " " A vs B $H_0: \beta_4=\beta_5$

Interaction between x & treatment

$H_0: \beta_4=\beta_5=0$

Equal Regressions: