Adding interactions to regression models

Consider two predictors: $Y = \beta_0' + \overset{(hs)}{\beta_1' X_1} + \overset{(iq)}{\beta_2' X_2} + \varepsilon$

$\Longrightarrow$ association between iq scores <u>does not</u> depend on hs

How can we relax this assumption?

$$\underbrace{E[Y \mid X_1 = 1, X_2] - E[Y \mid X_1 = 0, X_2]}_{\text{want this to depend on } X_2} = \beta_1 + \beta_2 X_2$$

assume linear dependence

new model: $Y = \beta_0 + (\beta_1 + \beta_2 X_2) X_1 + \beta_2 X_2 + \varepsilon$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \underbrace{\beta_{1,2} X_1 X_2}_{\text{(interaction)}} + \varepsilon$$

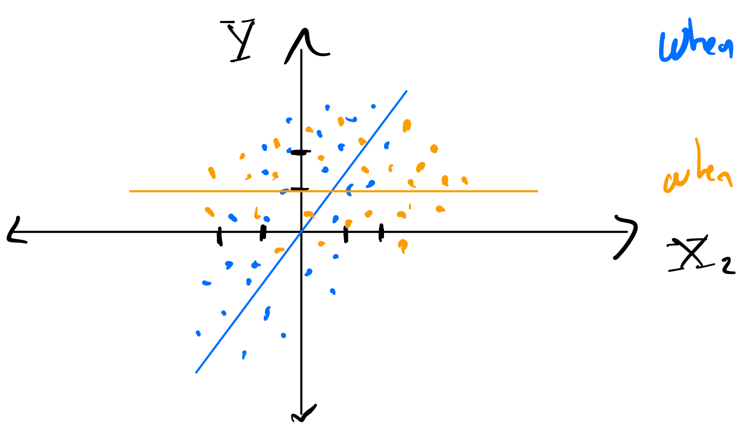
In order to fit coefficients w/ least squares write $X_3 = X_1 X_2$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \underbrace{\beta_3 X_3}_{\text{(interaction)}} + \varepsilon$$

Example Suppose $X_1 \sim \text{Bernoulli}(1/2)$ $\quad \beta_0 = 0,\ \beta_1 = 1$

$X_2 \sim \text{Normal}(0,1)$ $\quad \beta_2 = 1,\ \beta_{1,2} = -1$

Sketch data consistent with these parameters



when $X_1 = 0$, $Y = \beta_0 + \beta_2 X_2 + \varepsilon$

$= X_2$

when $X_1 = 1$, $Y = \beta_0 + \beta_1 + (\beta_1 + \beta_2) X_2 + \varepsilon$

$= 1 + 0 X_2 + \varepsilon$

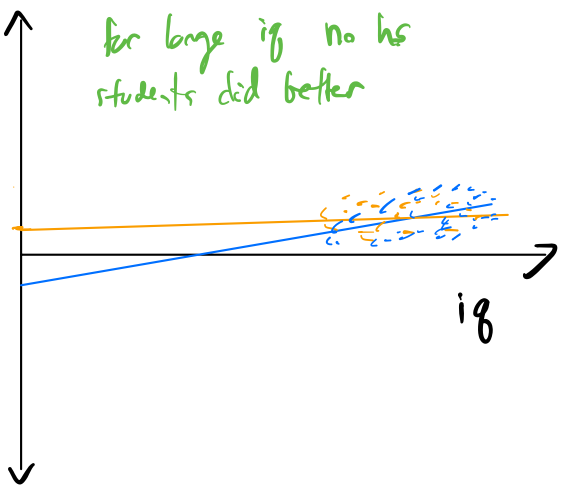
Example (test scores)

	$\hat{\beta}$	p-val
Cost	-11.5	0.4
hs	51.2	0.001
iq	0.96	0.000
hs-iq	-0.48	0.003

Sketch data that is consistent w/ these values, focus on relation between regression lines, not exact #s

when $X_{hs} = 0$

$Y = -11.5 + 0.96 X_{iq} + \varepsilon$



when $X_{hs} = 1$

$Y = -11.5 + 51.2 + (0.96 - 0.48) X_{iq} + \varepsilon$

$\approx 41 + 0.48 X_{iq} + \varepsilon$

typical iq $\approx 80 \Rightarrow \mu_Y \approx$

Fourier models - see typed notes section 4/5