STAT 2008/4038/6038 Regression Modelling 12/5/2017 (1)

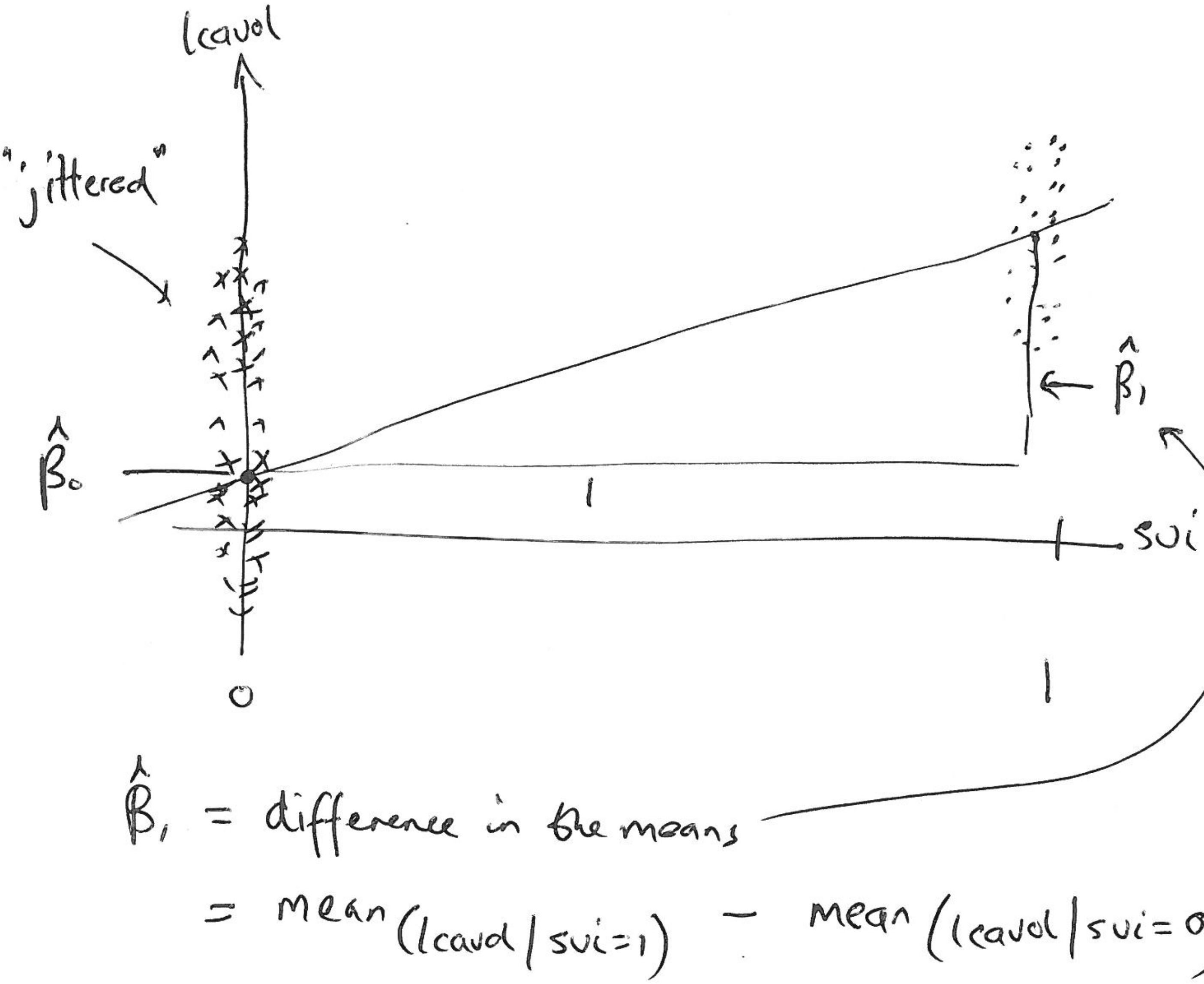
Indicator Variables

$$sui_i = \begin{cases} 1 & \text{if "seminal vesicle invasion" present for observation } i \\ 0 & \text{otherwise} \end{cases}$$

Model

$$lcavol_i = \beta_0 + \beta_1 sui_i + \varepsilon_i, \quad i = 1, \dots n$$

$$\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$



$\hat\beta_1$ = difference in the means

= mean(lcavol | sui=1) − mean(lcavol | sui=0)

Indicator variables (cont^d)

New model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$; $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$

where Y = lcavol

$X_1 = \text{svi}_i = \begin{cases} 1 & \text{if svi = Yes for obs } i \\ 0 & \text{otherwise} \end{cases}$

factor or treatment variable

X_2 = lpsa (a continuous covariate)

→ an "analysis of covariance" model

when svi = 0, $X_1 = 0$

fitted model

$$\hat{Y} = \hat{\beta}_0 + 0 + \hat{\beta}_2 X_2 = \hat{\beta}_0 + \hat{\beta}_2 X_2$$

when svi = 1, $X_1 = 1$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 1 + \hat{\beta}_2 X_2 = (\hat{\beta}_0 + \hat{\beta}_1) + \hat{\beta}_2 X_2$$

two parallel lines — one for svi = 0
& one for svi = 1