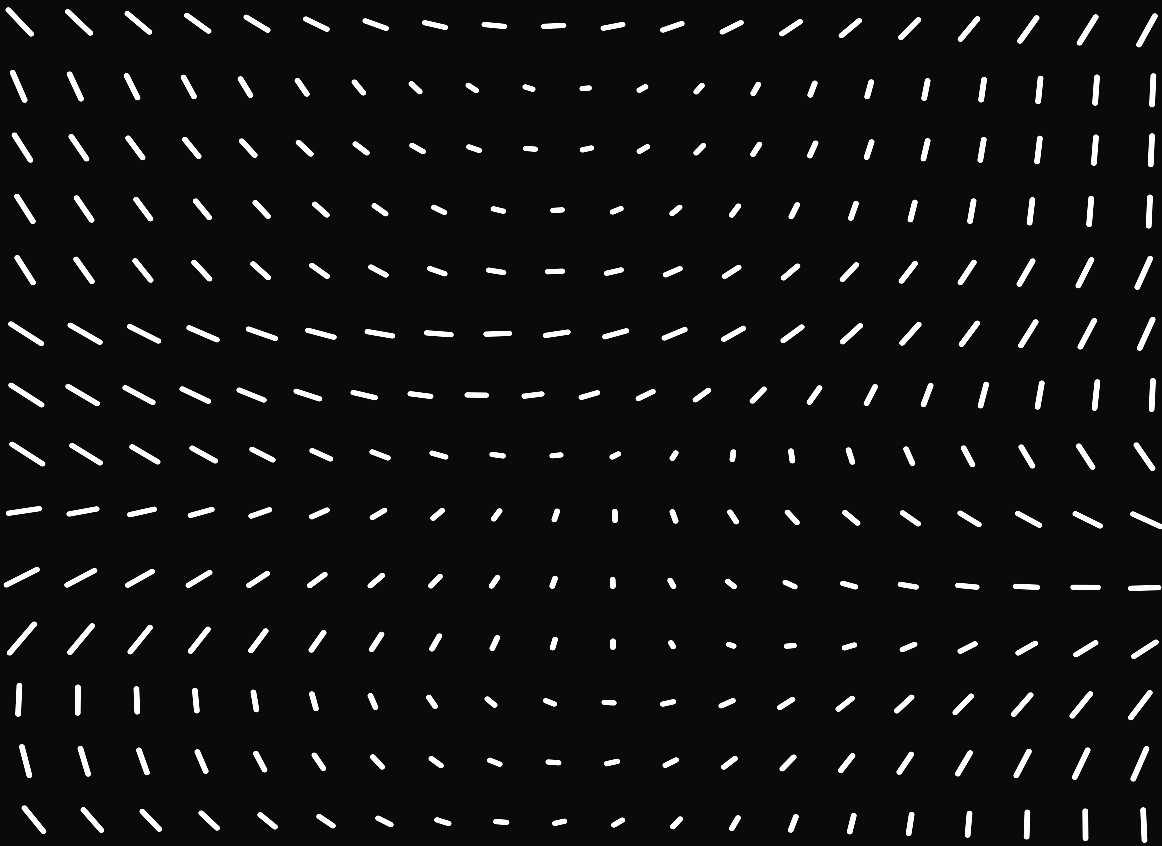
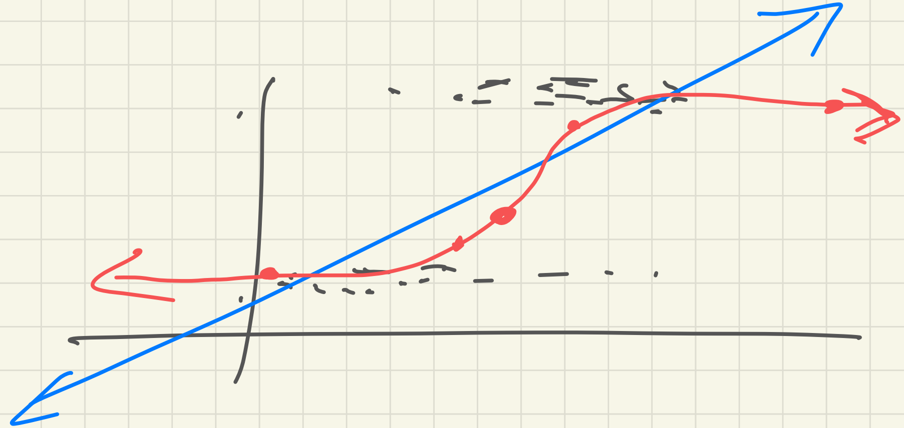


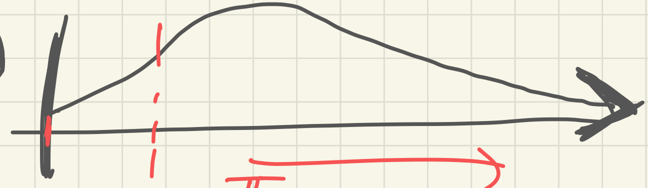
6 10 (m)

Linear probability model

$y_i \in \{0, 1\}$, π_i



$$\frac{\pi_i}{1-\pi_i} \in [0, \infty)$$



odds

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) \in (-\infty, \infty)$$

log odds

Logistic Reg.

"Problems" with Linear Regressions

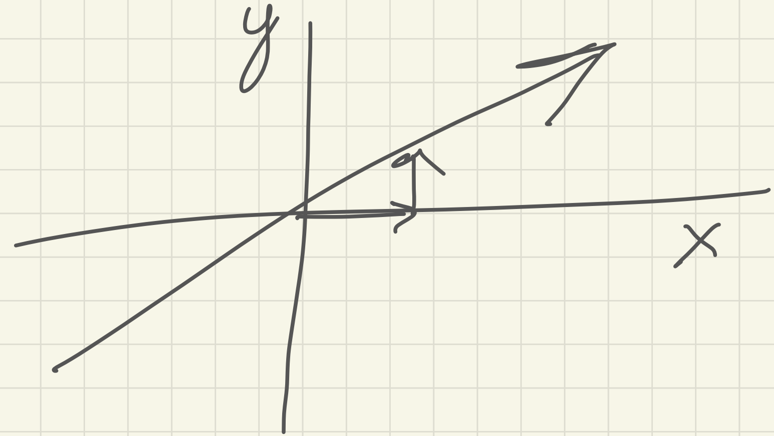
Bias: even though the are <ins>unbiased</ins> on av., they can be biased for special groups.

Heteroskedasticity: even though the model assumes uncertainty is consistent, <ins>it is not.</ins>

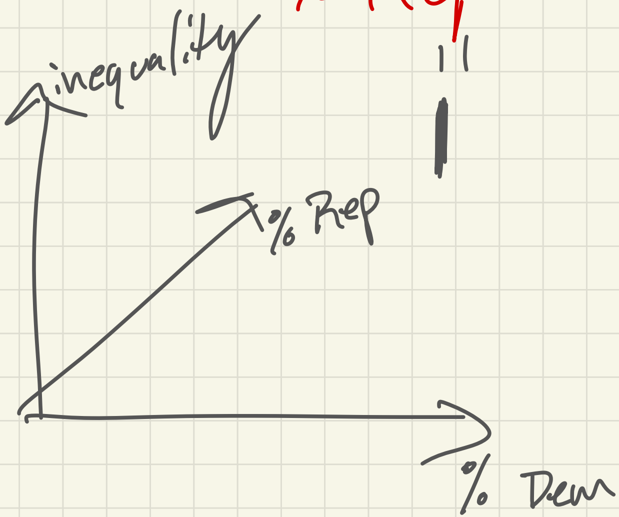
Multicollinearity: even though the model assumes you can change variables separately from one another, <ins>you can't!</ins>

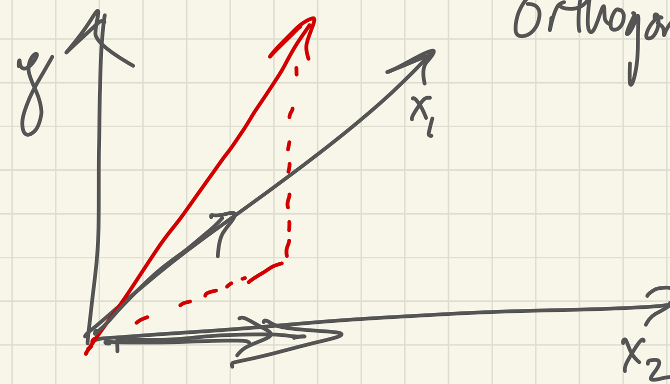
$y = \alpha + x_1\beta_1 + x_2\beta_2 + \ldots + e$

Multicollinearity



inequality ~ % Democrat + % Republicans





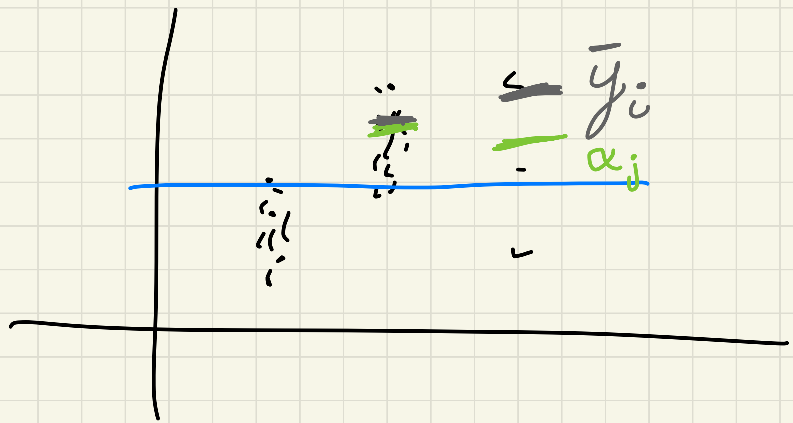
orthogonal

$$\hat{\alpha}_j \propto \frac{n_j}{\sigma_e^2}\bar{y}_j + \frac{1}{\sigma_u^2}\bar{y}$$

1.) group size

2.) W/in group variab.

3.) Betw. group variab

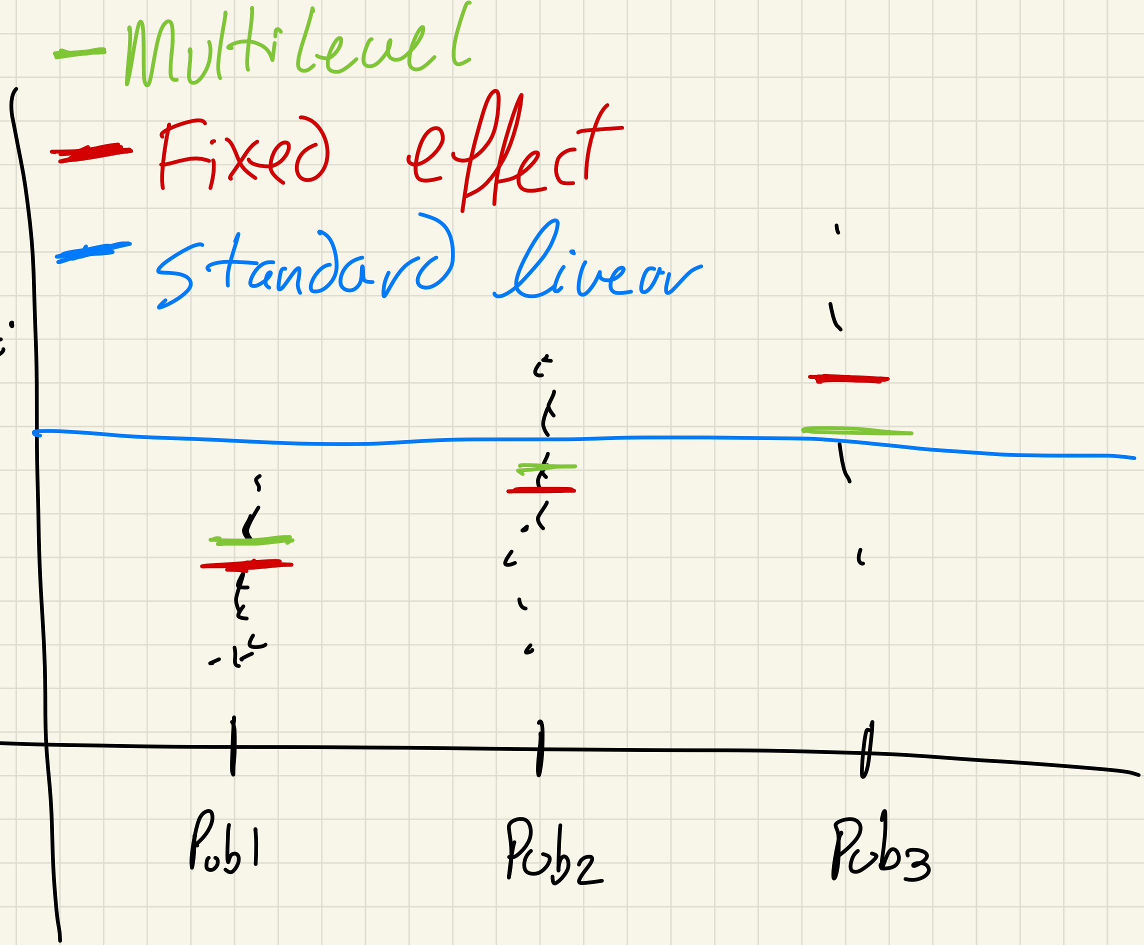


Big groups are close to $\bar{y}_j$

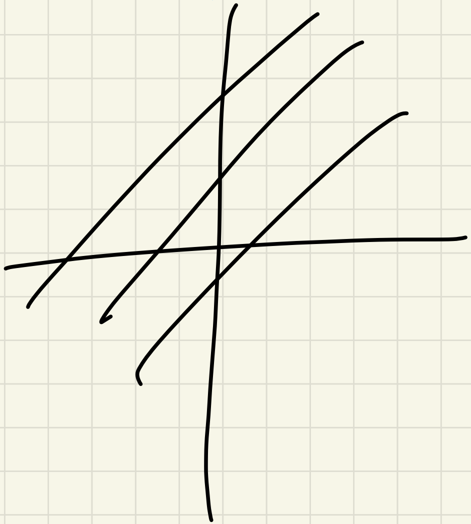
If a problem is noisy w/in groups, close to $\bar{y}$

If groups are very different, close to $\bar{y}_j$

$$y_i = \underbrace{a + U_j} + e_i$$

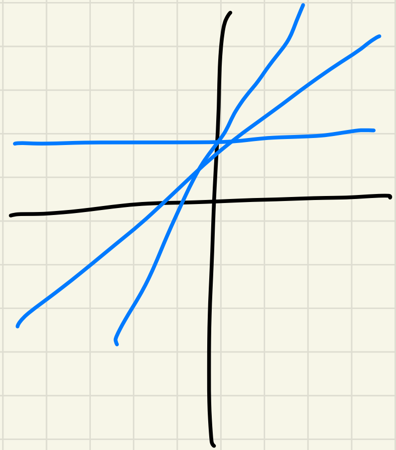


$1+X+(1 \mid group)$



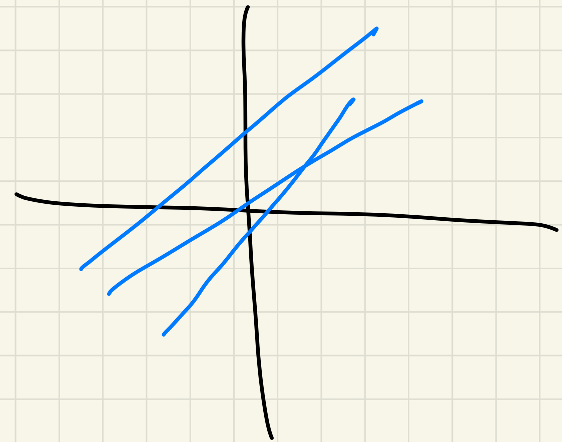
Varying intercept Model

$1+X+(0+X \mid group)$



Varying Slope Model

$1+X+(1+X \mid group)$



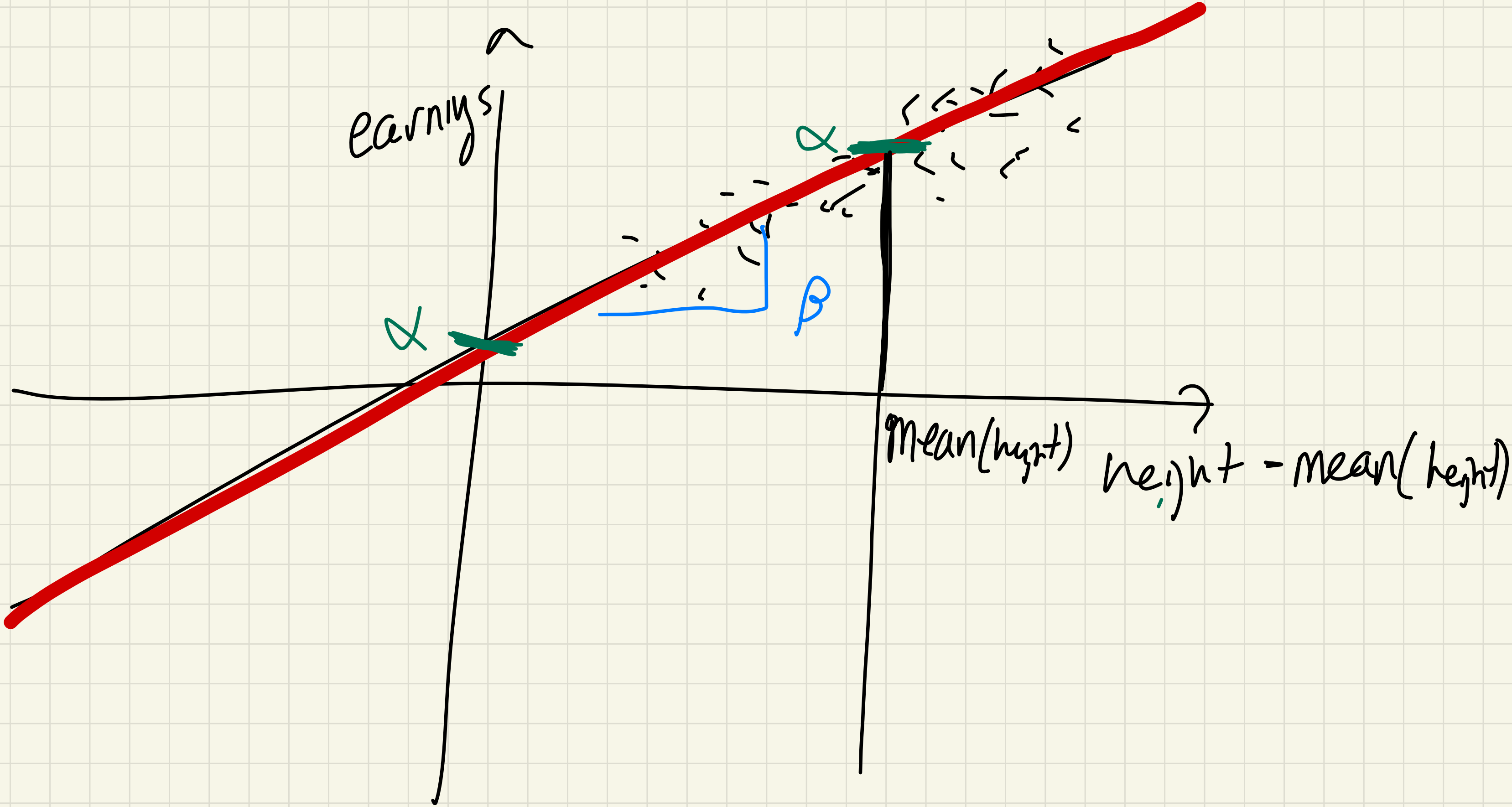
Varying-intercept Varying-slope model

```
outcome ~ fixed + (vary | group)

waiting ~ 1 + (1 | pub)

waiting ~ 1 + n_patrons + (1 + n_patrons | pub)

waiting ~ 1 + n_patrons + (0 + n_patrons | pub)
```



earnings
α
α
β
mean(height)
height - mean(height)

Error formulation: u_j, w_j, e_i

$$y_i = a + u_j + b + w_j + e_i$$

Varying effect

$$y_i = \alpha_j + \beta_j + e_i$$

$$\alpha_j = a + u_j$$

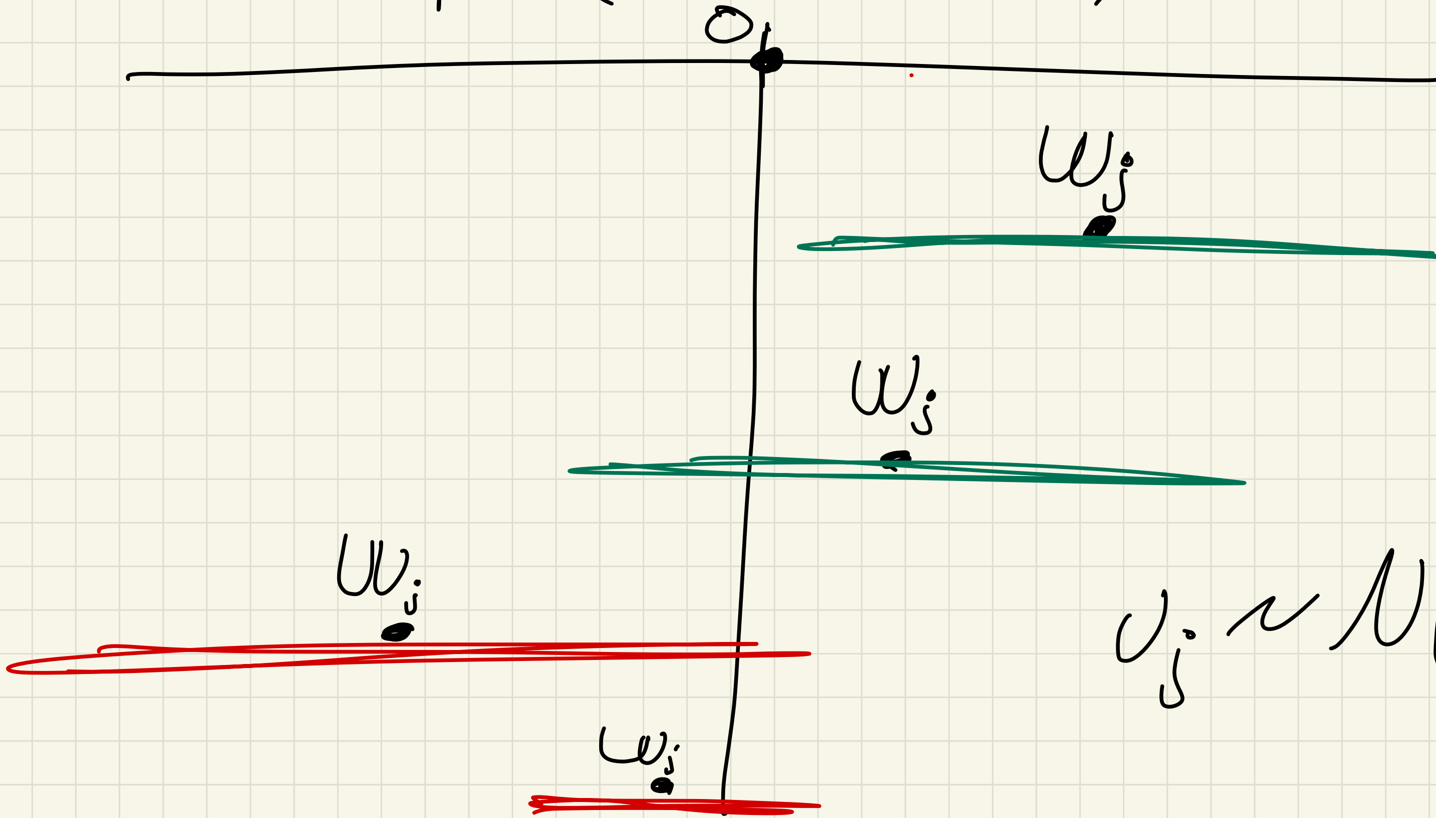
$$\beta_j = b + w_j$$

fixef : a, b

ranef : u_j, w_j

coefficients: α_j, β_j
(mixef)

lattice::dotplot(ranef(model))



$$u_j \sim N(0, \sigma_u^2)$$

earnings ~ 1 + height + (1 + height || sex)

$$y = a + u_j + b + w_j + e_i$$

σ_u σ_w σ_e

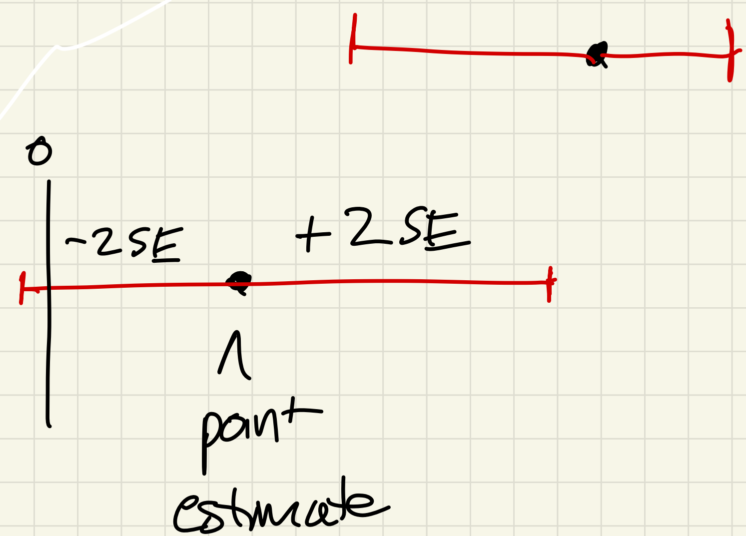
(Intercept)

height

sd__Intercept

sd__height

sd__Observation



Ghost!

$$\text{price} = \alpha_j + e_i$$

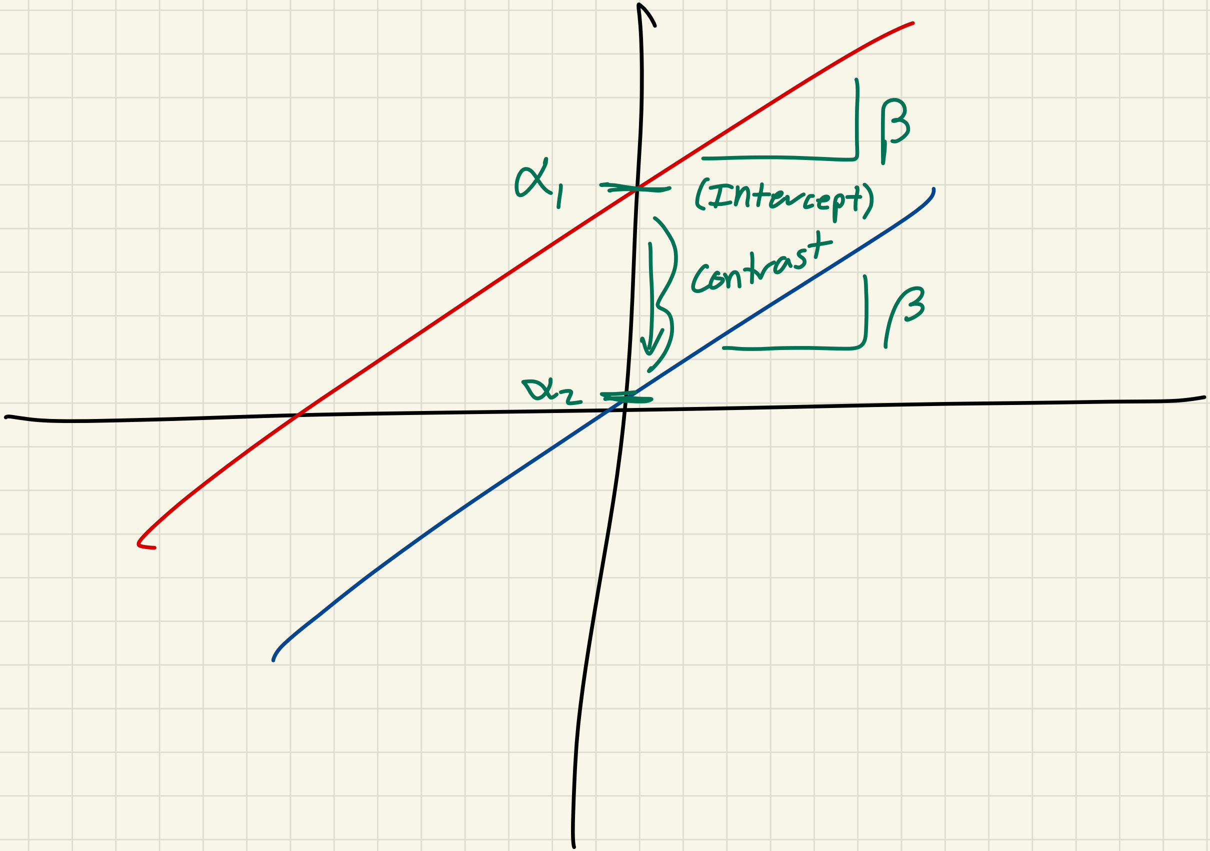
1%

1%

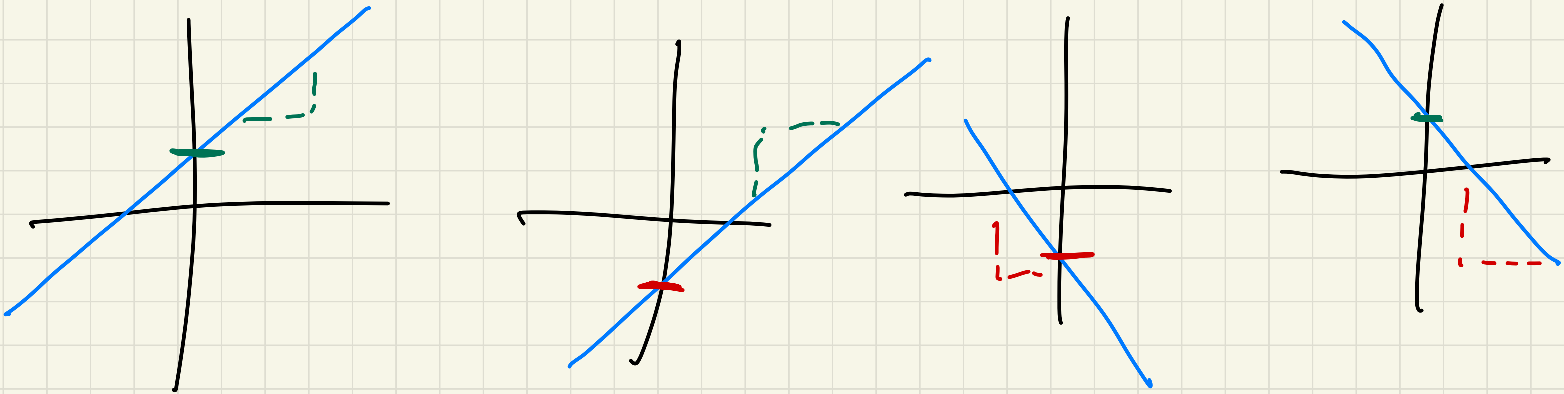
1%

$$\alpha_j = a + \underline{\text{tax rate}} \cdot \gamma + v_j$$

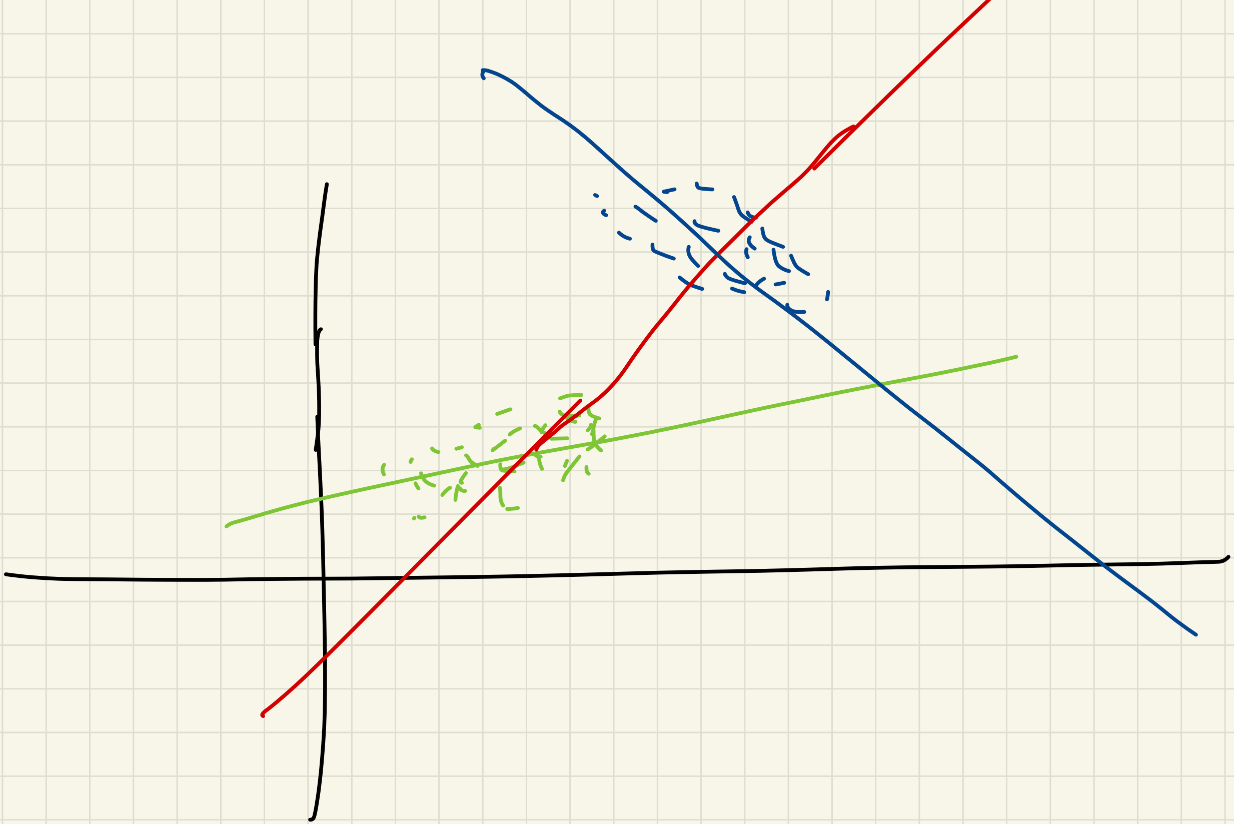
$$\text{price} \sim 1 + (1 + \text{tax rate} \mid La)$$



β
α_1
(Intercept)
contrast
β
α_2



$\alpha +$	$\alpha -$	$\alpha -$	$\alpha +$
$\beta +$	$\beta +$	$\beta -$	$\beta -$



A Grammer (+ Vocabulary)

of

Graphics

ggplot2

Grammer of Graphics

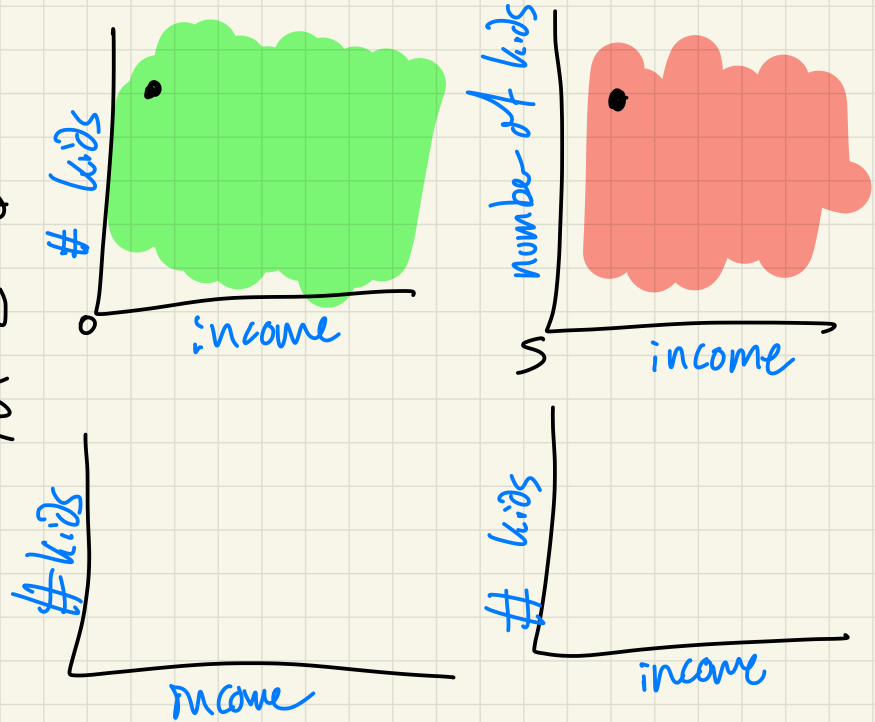
ggplot

Facet

facet_grid(~country)

Voted biden in 2020

Voted trump in 2016



Small multiples of visualizations

```
ggplot(data, aes(x= , Y=, ...))
  + geom_....(aes(...), stat )
  + scale_x_logarithmic( )
  + coord_polar( )
  + facet_grid(~ group)
```

Vega

Altair

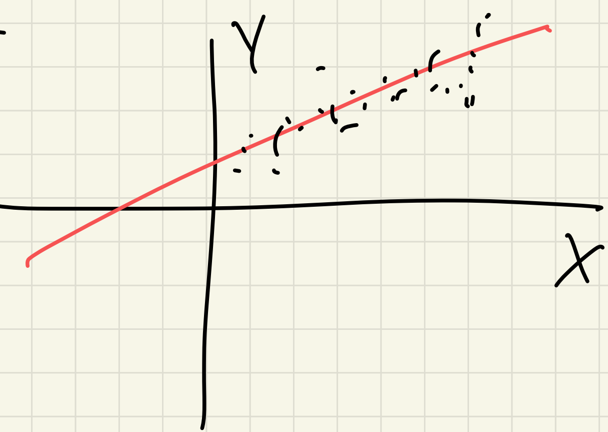
Data

from

JPZ

let $z \equiv x - \bar{x}$ // $x = z + \bar{x}$

$\alpha \equiv y$ when x is 0

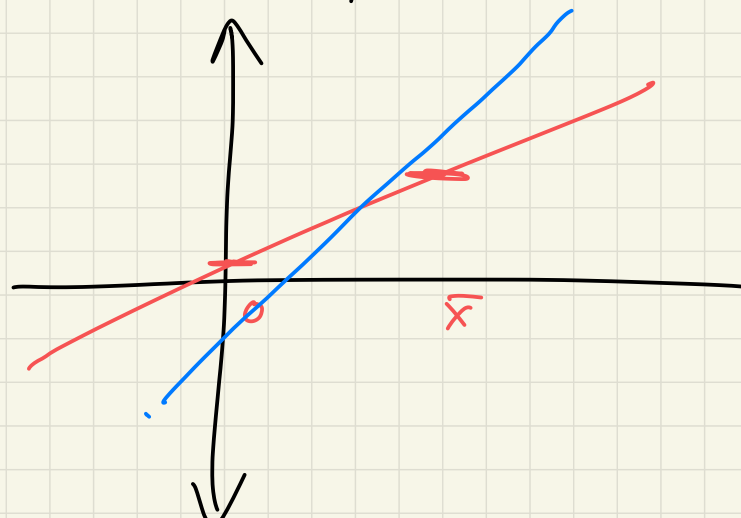


$$y = \alpha + x\beta$$

$$y = \alpha + (z + \bar{x})\beta$$

$$y = \alpha + z\beta$$

$$y = \alpha + (x - \bar{x})\beta$$



Final Exam

1. Logistics

Post on blackboard
10 AM GMT Thursday

2. Concepts

Questions: Slides.app.goo.gl/c9xjk

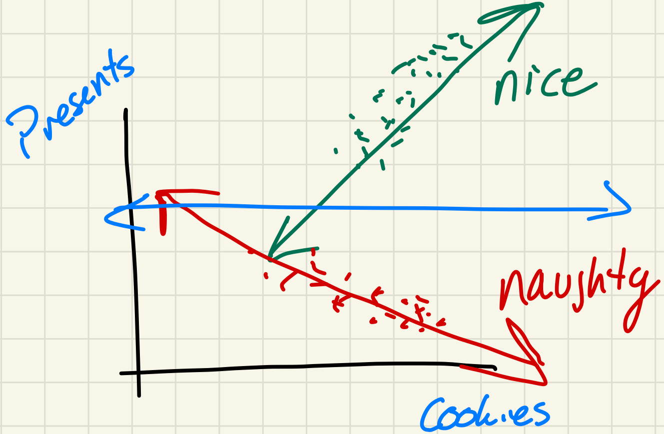
1.) Exploration + Plotting

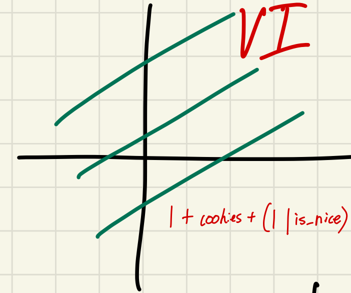
2.) Simple Empirical Question
(maybe using a Lm)

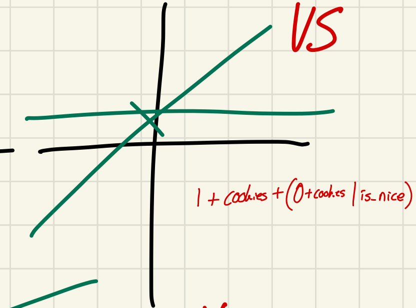
3.) More Complex empirical Q.
(this will be a mLm)

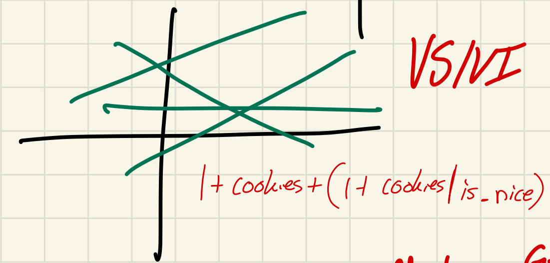
4.) GLM to predict something

MLM









$$y_i = \beta_j \# cookies + \alpha_j + e_i$$

$$\beta_j = b + w_j$$

$$\alpha_j = a + u_j$$

Effect Group

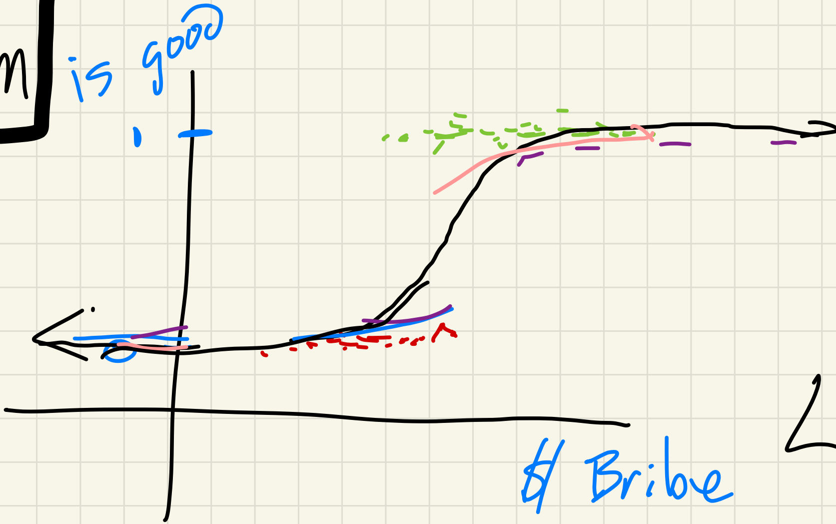
lmer(presents ~ 1 + cookies + (1 + cookies | is_nice),
data=xmas)

Fixed effects

Random effects

$$y_i = (b + w_j) \# cookies + (a + u_j) + e_i$$

GLM




```
glm(is_good ~ Bribe_money,
    data = xmas,
    family = "binomial")
```

Logistic regression

∴ Linear regression on log odds of y.

$$y_i \begin{cases} 0, & \text{if naughty} \\ 1, & \text{if nice} \end{cases}$$

$p_i \in [0, 1]$ probability a person is nice

$$\log\left(\frac{p}{1-p}\right) = \alpha_j + \beta_j \cdot \text{Bribe money} + e_i$$