22 Feb 13
(extra notes)
①

(Bayesian non-identifiability for a parameter θ) $\leftrightarrow$ (post for θ has same uncertainty as prior for θ)

---

structural non-identifiability: cannot learn about from data no matter what data set is

nonstructural non-identifiability: can learn about θ if n is large enough, cannot if n too small //

heterogeneity between studies yielding conclusion that naive pooling is wrong is *not* the same thing as Simpson's Paradox (another setting in which naive pooling can be silly)

ex. regression with observational data (not controlled experiments)

+ sign

$\overset{iid}{\sim} N(0, \sigma_1^2)$

$$Y_i = \beta_0 + \boxed{\beta_1} x_{i1} + e_i$$

naive

↑ outcome

↑ supposedly causal factor (SCF)

− sign

potential confounding factor (PCF)

$$Y_i = \beta_0^* + \boxed{\beta_1^*} x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

better

$\overset{iid}{\sim} N(0, \sigma_2^2)$

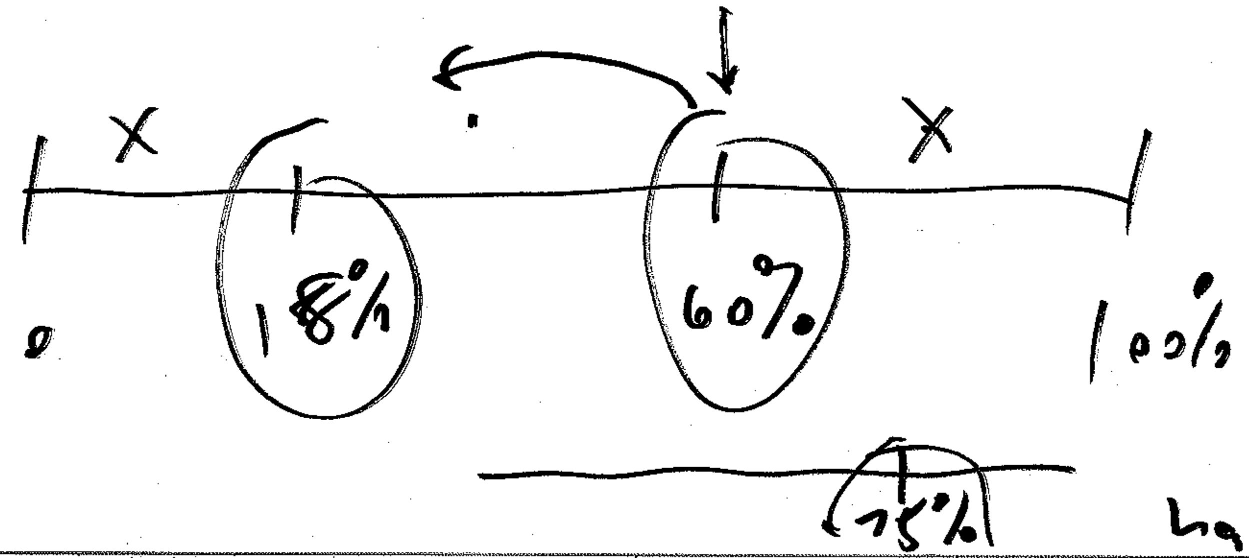
this is Simpson's Paradox

I hospitals : measure observed mortality rate $\hat{p}_i$ for (eg.) heart attack at hosp. i

---

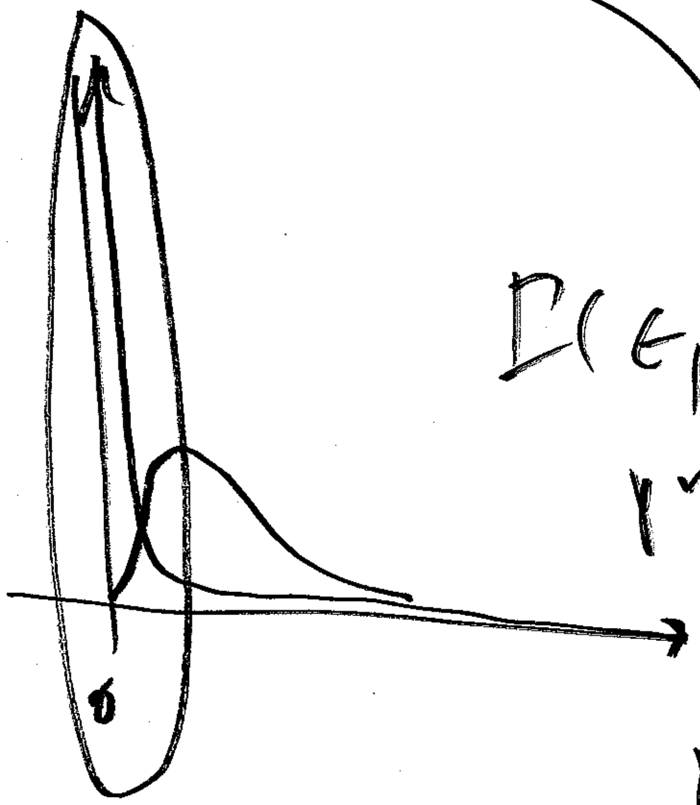
at hosp. 6 $\quad n_6 = \boxed{10} \quad \hat{p}_6 = \boxed{60\%}$

overall mort rate $= \hat{\mu} = \boxed{18\%}$



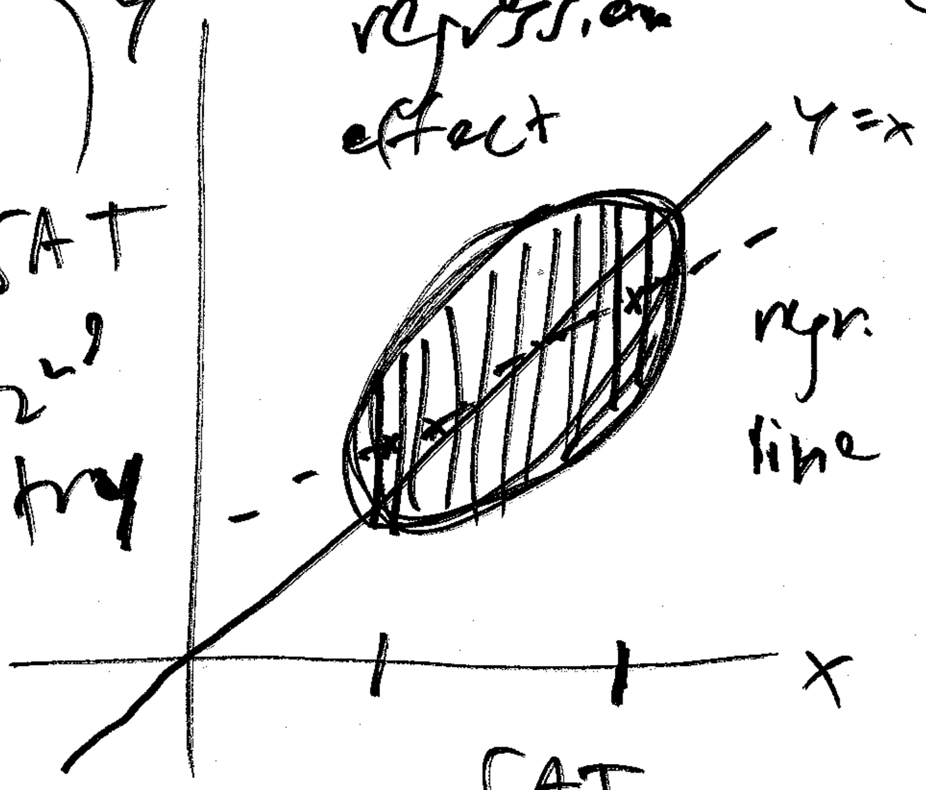
$n_9 = 4$

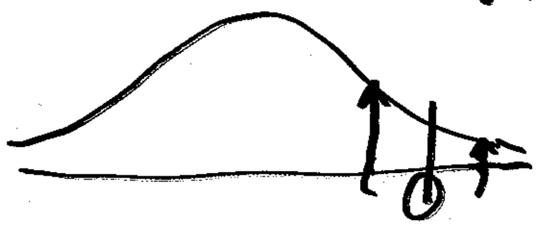
test-retest



$\Gamma(\epsilon, \epsilon)$ prior for precision or variance

(Galton 1890)





dist. of SAT scores

$(Y_i \mid \mu, \sigma^2, \nu) \overset{iid}{\sim} t_\nu(\mu, \sigma^2)$ $i = 1, \dots, 6$

$V_{RS}(Y_i) = \sigma^2 \boxed{\frac{\nu}{\nu - 2}}$

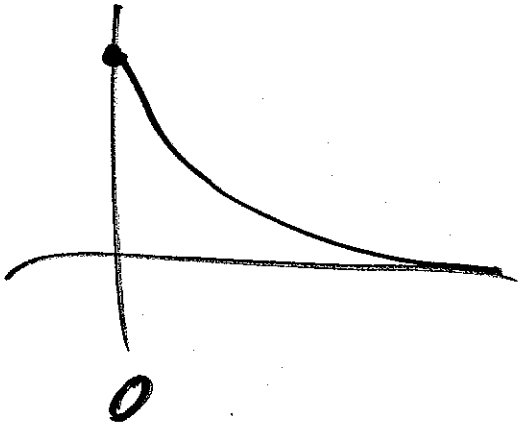
$\boxed{\text{big variance}}$

df ν small or SD big

$\theta_i = \alpha + \beta x_i + e_i \leftarrow$ iid $N(0, \sigma_\theta^2)$

equivalent to $\updownarrow$ fixed known constants

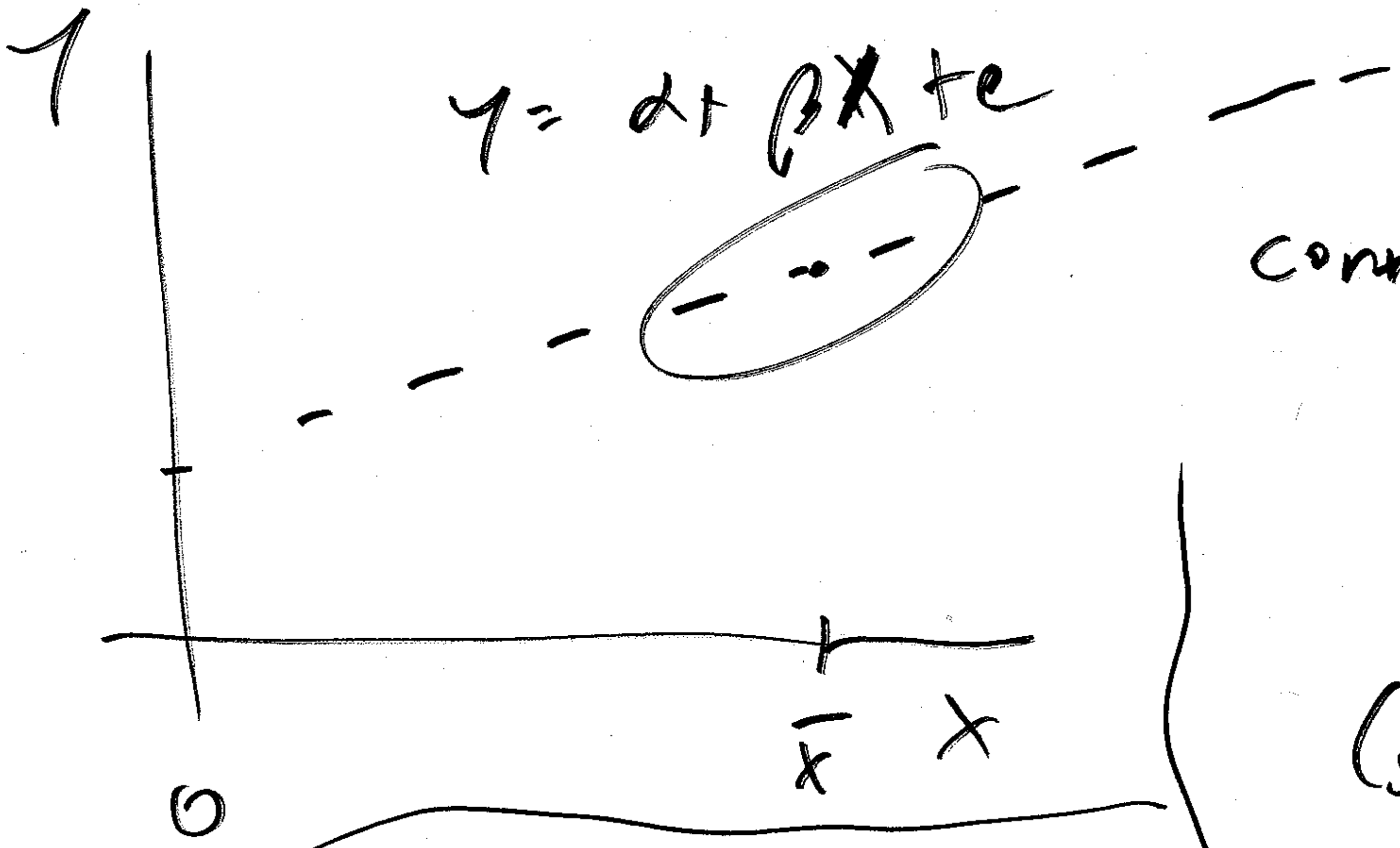
$(\theta_i \mid \alpha, \beta, \sigma_\theta^2, x_i) \overset{\text{indep}}{\sim} N(\alpha + \beta x_i, \sigma_\theta^2)$



$\ell(\sigma_\theta^2 \mid y) \doteq p(\sigma_\theta^2 \mid y)$

(Bayesianly)

$\hat{\sigma}_\theta^2 = 0$ but entire dist is to right of 0



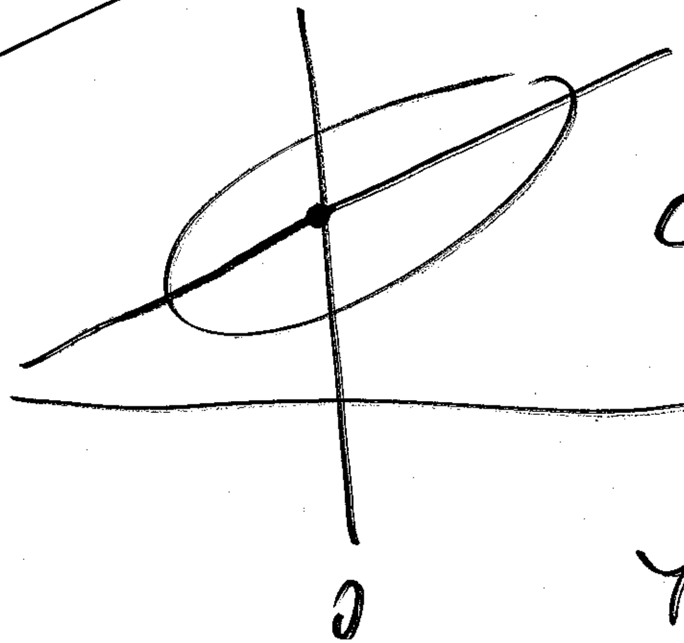
$y = \alpha + \beta x + e$

$\mathrm{corr}(\hat{\alpha}, \hat{\beta}) < 0$

↑

quite strong

(stronger as $\bar{x}$ gets farther from 0)



$\mathrm{corr}(\hat{\beta}, \hat{\alpha}) = 0$

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

model 1

$$(\mu, \sigma^2) \sim \text{diffuse}$$

$$(\theta_i \mid \mu, \sigma^2) \overset{\text{iid}}{\sim} N(\underline{\mu}, \sigma^2) \quad \alpha$$

$$(Y_i \mid \theta_i) \overset{\text{indep}}{\sim} N(\theta_i, V_i)$$

here θ_i get shrunken toward μ

(model 33)

$$(\alpha, \beta, \sigma_\theta^2) \sim \text{diffuse}$$

$$(\theta_i \mid \alpha, \beta, \sigma_\theta^2) \overset{\text{iid}}{\sim} N(\alpha + \beta(x_i - \bar{x}), \sigma_\theta^2)$$

$$(Y_i \mid \theta_i) \overset{\text{indep}}{\sim} N(\theta_i, V_i)$$

here

θ_i get shrunken toward the line $\boxed{\alpha + \beta(x_i - \bar{x})}$

research results

$U(0, c)$ prior on σ_θ better calibrated than $\Gamma(\epsilon, \epsilon)$ prior on $\frac{1}{\sigma_\theta^2} = \tau_\theta$

simulation study: generate many (1,000 or 10,000) datasets from model (33) with various choices of known data-generating values $\boxed{(\alpha, \beta, \sigma_\theta)}$; construct "95%" ← nominal inferential intervals for $(\alpha, \beta, \sigma_\theta)$ on each data set; key score: what actual % of time these intervals include known data-generating parameter values?

actual coverage rate] if nominal $\doteq$ actual, good calibration model

(33)

$$\alpha, \beta, \sigma_\theta \sim (*)$$

$$(\theta_i \mid \alpha, \beta, \sigma_\theta^2) \overset{ind}{\sim} N(\alpha + \beta(x_i - \bar{x}), \sigma_\theta^2)$$

$$(y_i \mid \theta_i) \overset{ind}{\sim} N(\theta_i, V_i) \quad i = 1, \ldots, I$$

known

(*) well-calibrated diffuse priors on $(\alpha, \beta, \sigma_\theta^2)$

$\alpha \sim N(0$, huge variance$)$

$\beta \sim N(0$, huge variance$)$ ← tiny precision e.g. 1.0 E-6

$\sigma_\theta \sim U(\underbrace{0, c})$ ← particularly important when I is small

c just large enough to avoid truncation of likelihood function

$$(\alpha, \beta, \sigma_\theta^2) \sim p(\alpha, \beta, \sigma_\theta^2)$$

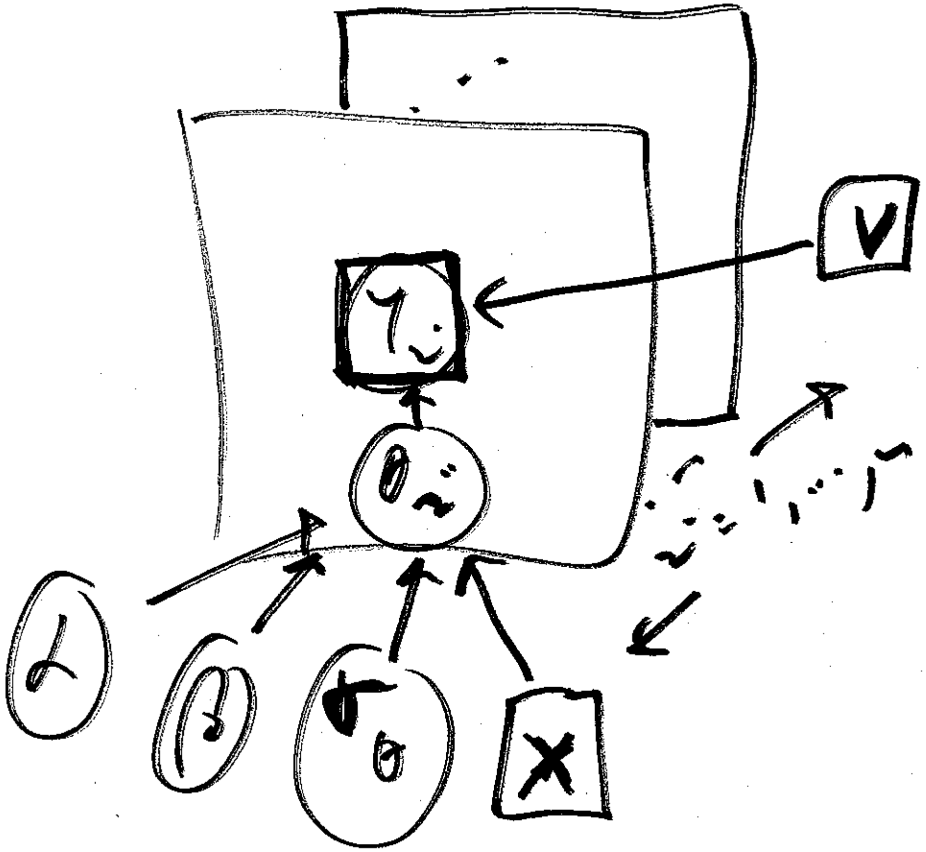
$$(\theta_i \mid \alpha, \beta, \sigma_\theta) \overset{ind}{\sim} N(\alpha + \beta(x_i - \bar{x}), \sigma_\theta^2)$$

$$(Y_i \mid \theta_i) \overset{ind}{\sim} N(\theta_i, V_i)$$

known

---

graphical model



⑫

$\bar{E}$ sample mean in E group

$\bar{C}$ ——— C ———

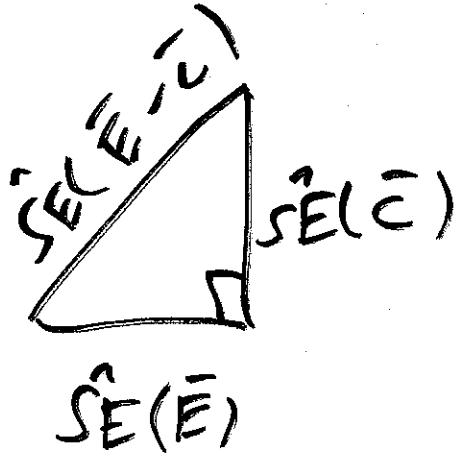
$\hat{\Delta} = \bar{E} - \bar{C}$

(2 indep samples)

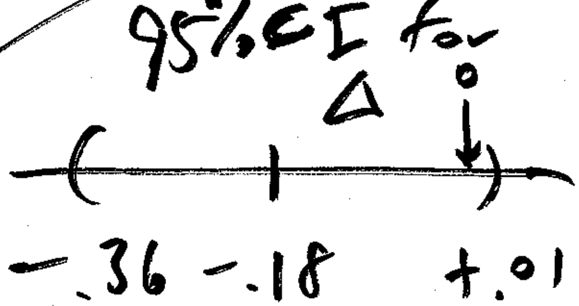
$\hat{SE}(\hat{\Delta}) = \hat{SE}(\bar{E} - \bar{C})$

$= \sqrt{\left(\hat{SE}(\bar{E})\right) + \left(\hat{SE}(\bar{C})\right)^2}$

$= \sqrt{\frac{S_E^2}{n_E} + \frac{S_C^2}{n_C}}$

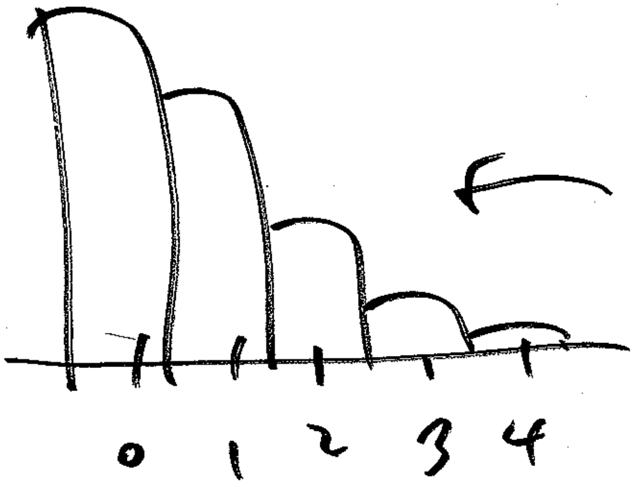


$\hat{SE}(\bar{E}) = \frac{S_E}{\sqrt{n_E}}$

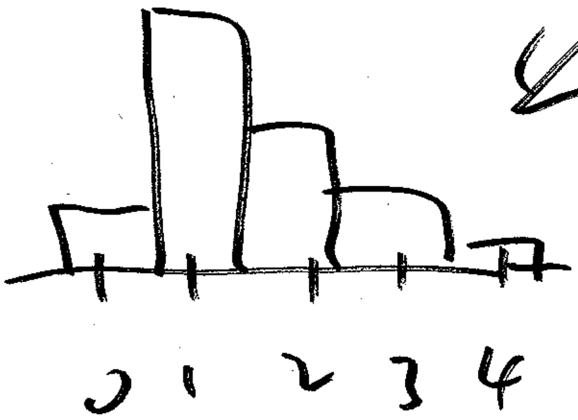


$\left(\begin{matrix} H_0 : \Delta = 0 \\ H_A : \Delta \neq 0 \end{matrix} \right)$

(not stat sig)



← could be Poisson

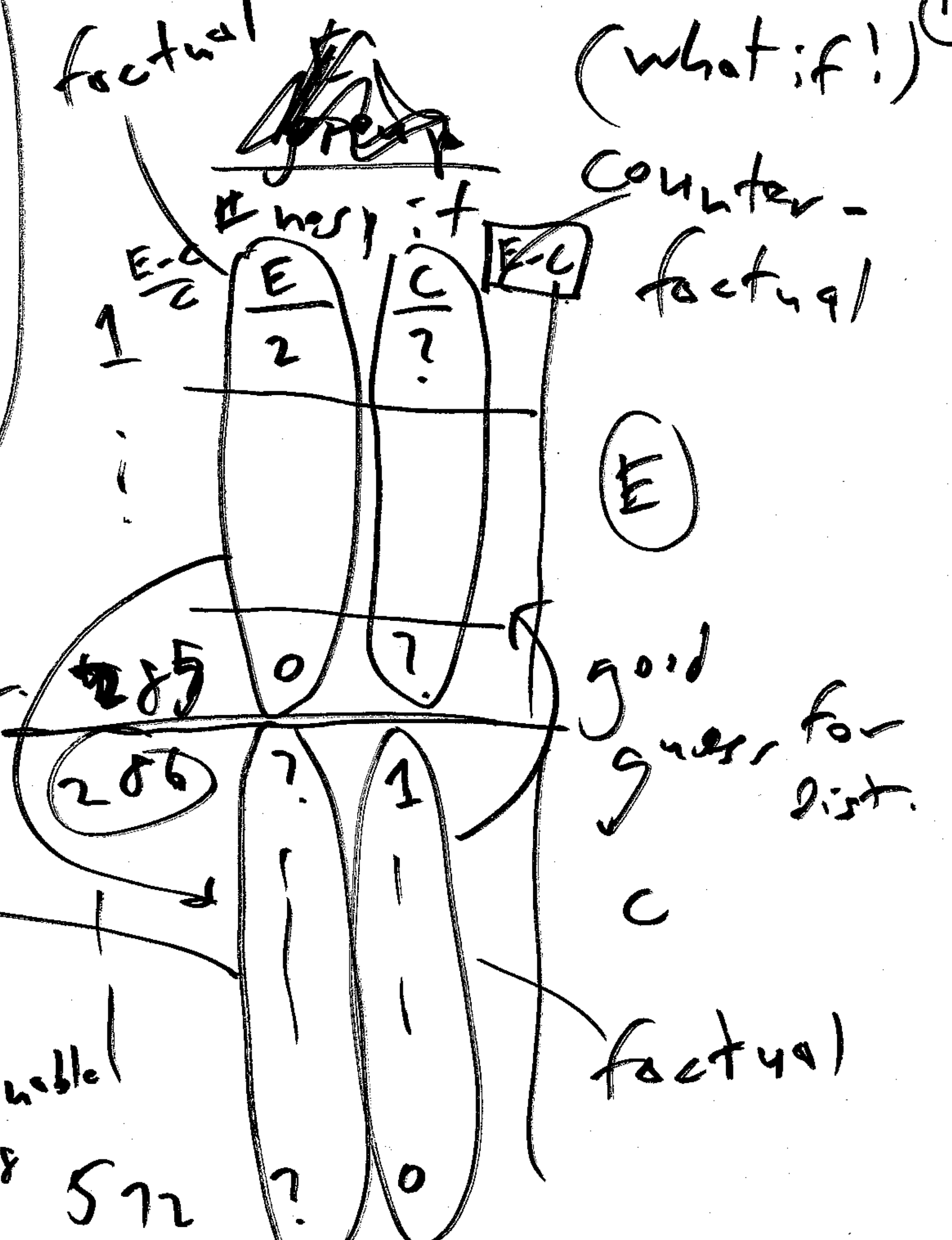


$$(C_i \mid \lambda_C) \overset{iid}{\sim} P(\lambda_C)$$

$$(E_j \mid \lambda_E) \overset{iid}{\sim} P(\lambda_E)$$

Neyman counterfactual model for causal inference

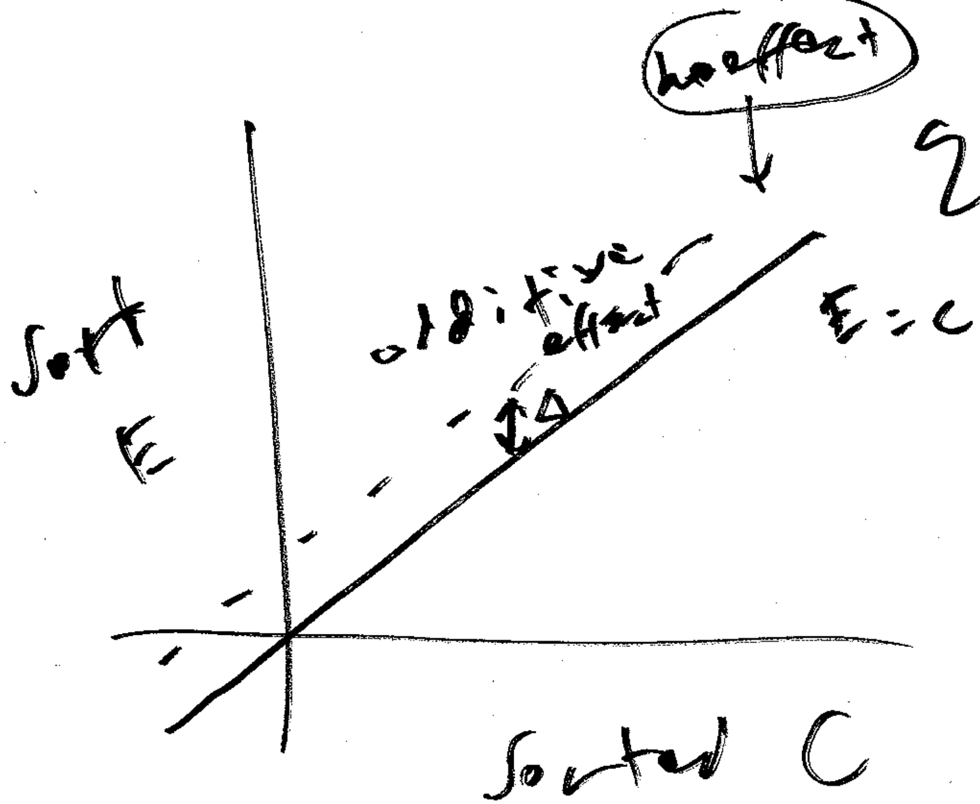
(what if!)

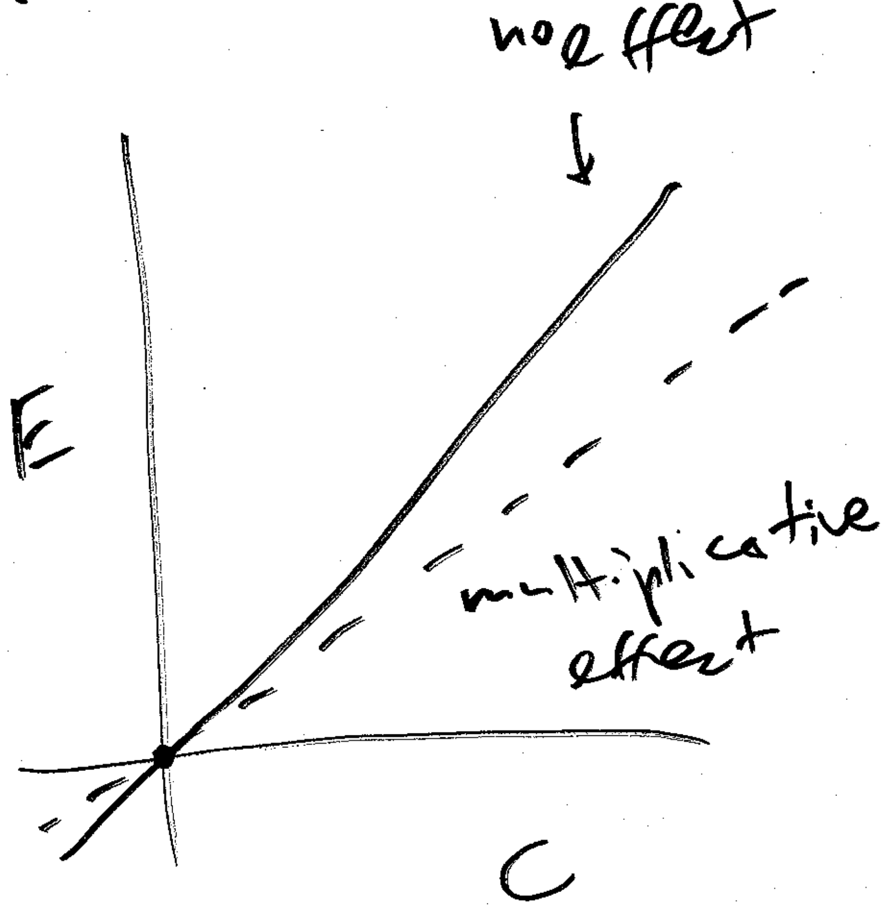


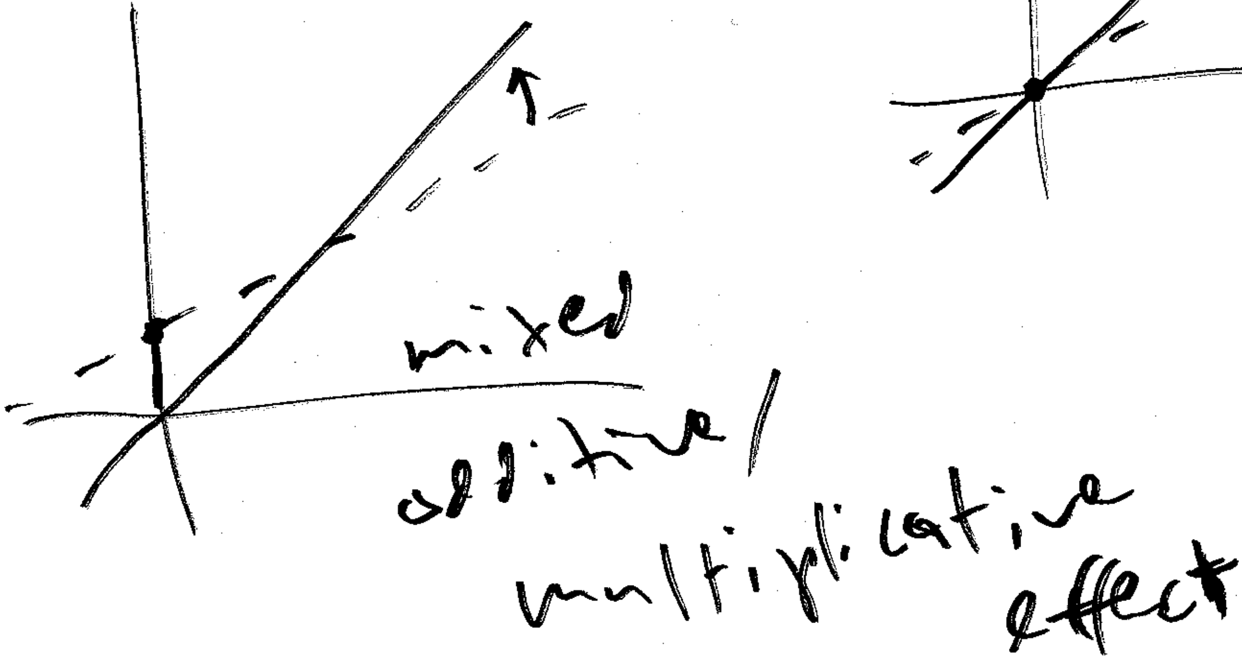
additive: $\cancel{x}E \overset{st}{=} C + \Delta$ not reasonable

(multiplicative) $\cancel{x}E_{\cancel{x}} \overset{st}{=} C_{\cancel{x}}(1+\Delta)_{\cancel{x}}$ ← r.18 unit fix

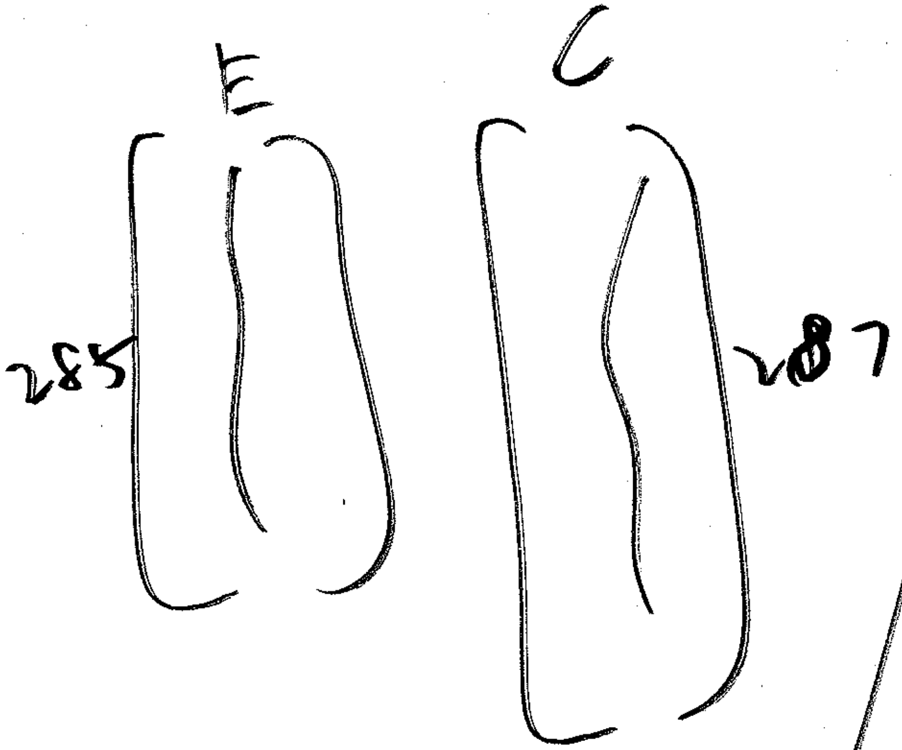
qq plot


no effect
Sort E
additive effect
Δ
E=C
Sorted C


no effect
E
multiplicative effect
C

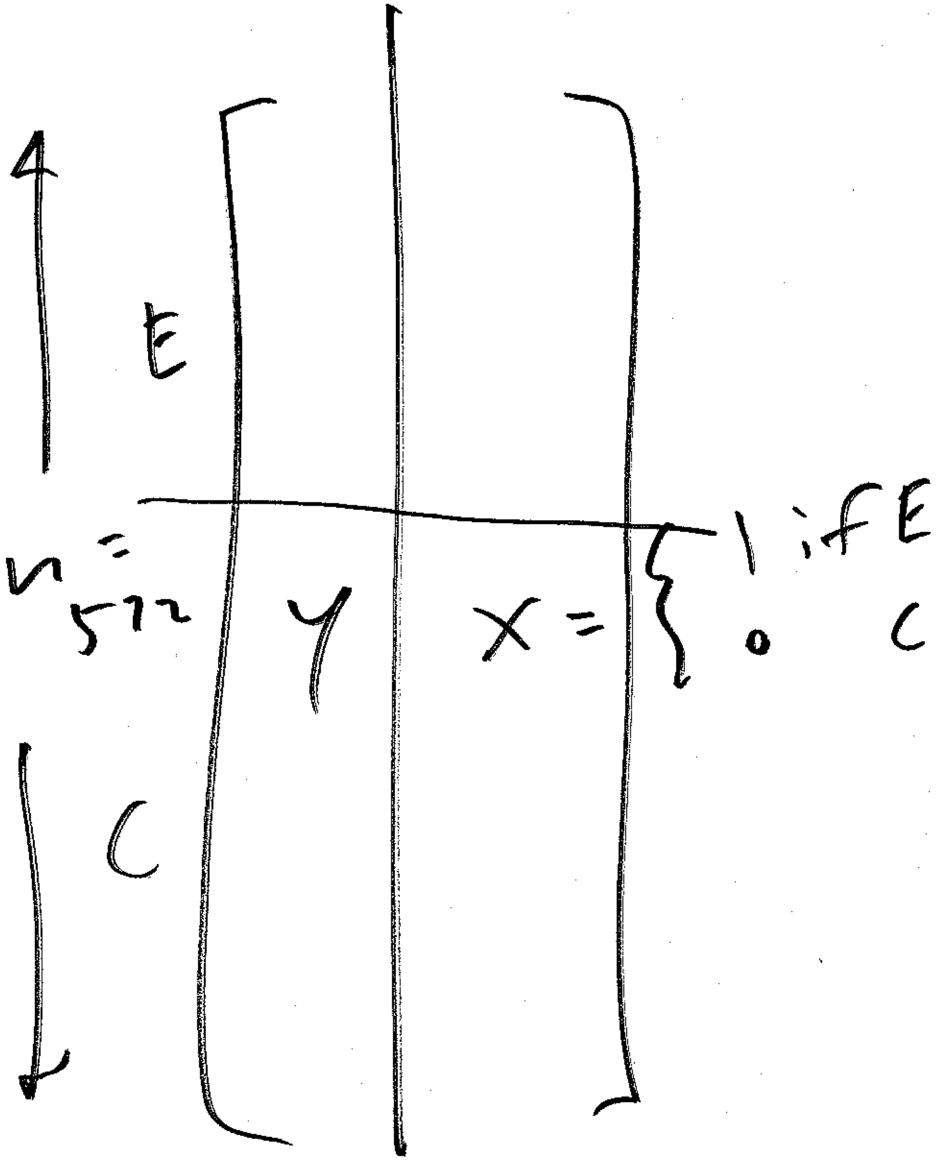

mixed
additive/
multiplicative
effect

2-indep. samples data



$$(Y_i \mid \lambda_i) \overset{\text{indep}}{\sim} P(\lambda_i)$$

$$\log \lambda_i = \gamma_0 + \gamma_1 x_i$$



(17)

$(y_i \mid \lambda_i) \overset{\text{indep}}{\sim} P(\lambda_i)$

fixed effects

FE

$$\log(\lambda_i) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \dots + \gamma_k x_{ik} + e_i \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$$

x_{i1}: $\begin{cases} 1 & \text{if E} \\ 0 & \text{C} \end{cases}$

x_{i2}: age

x_{i3}: sickness score

model expansion

$(\gamma_0, \gamma_1, \dots) \sim$ diffuse

(hierarchical) mixture representation

unobserved → latent variables

lognormal mixture of poissons

random effects → e_i

if this 0 → get FE model