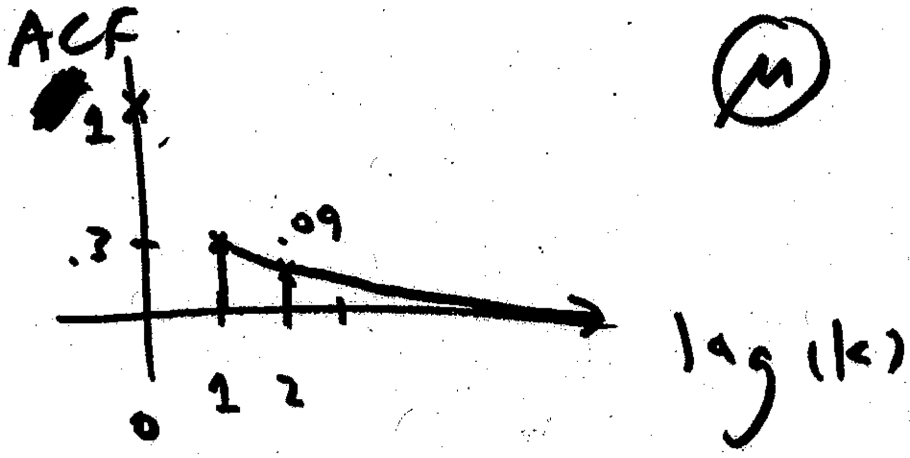
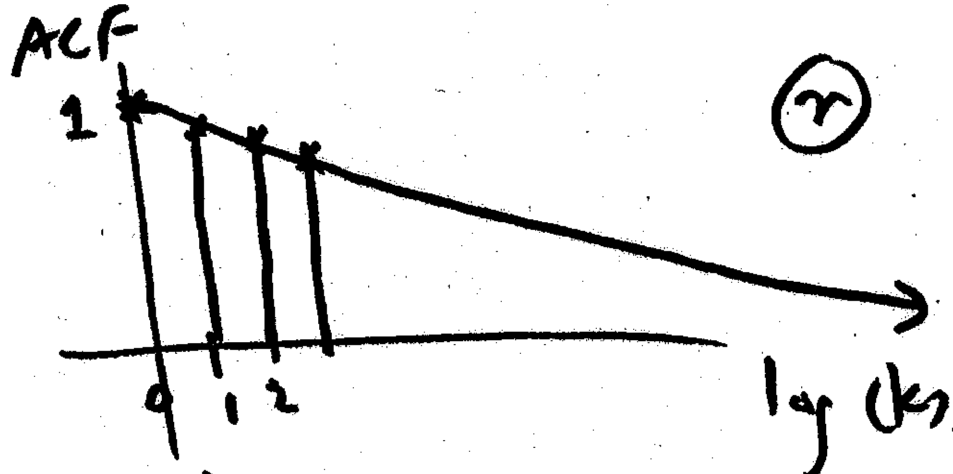
(handwritten notes, 15 Feb) ①



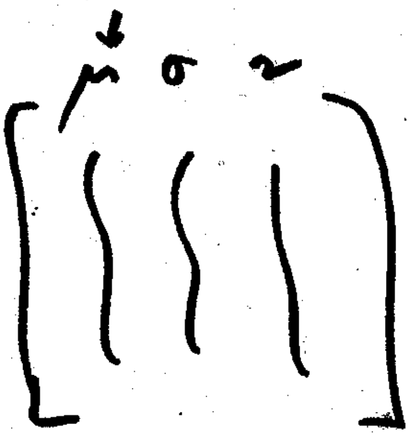
μ

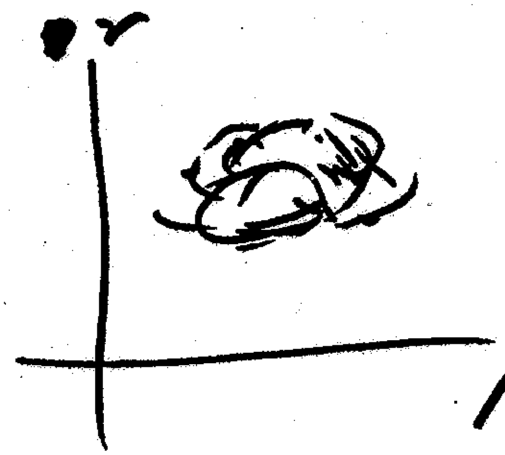
(mixing well)



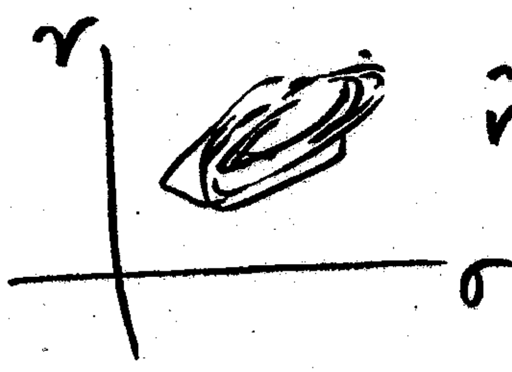
r

(mixing badly)

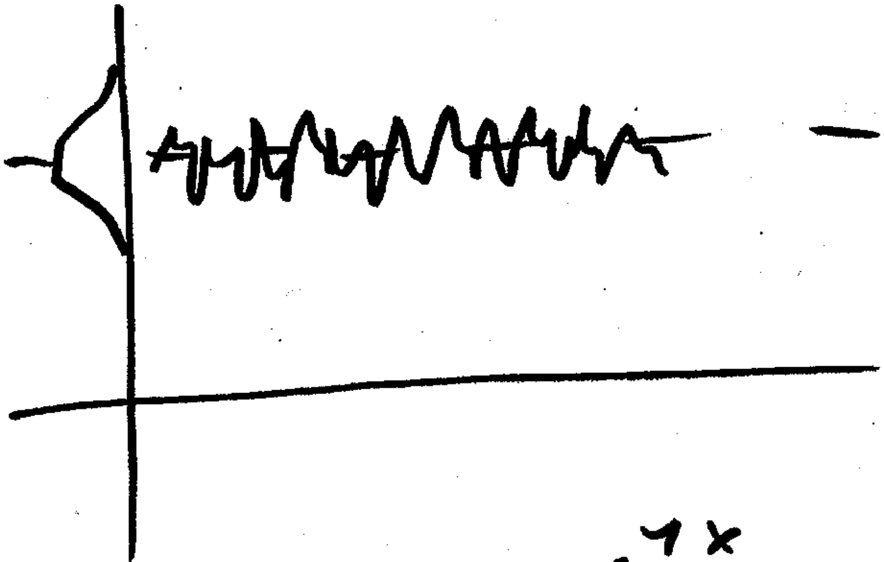


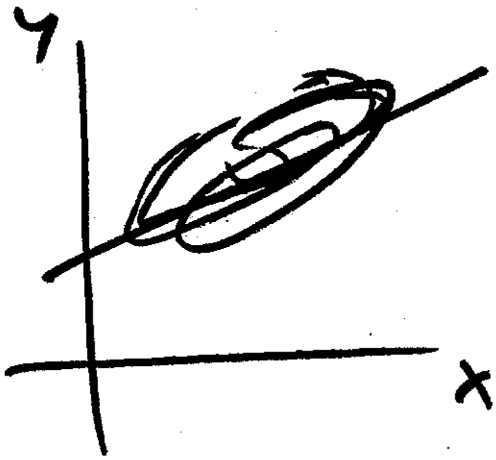


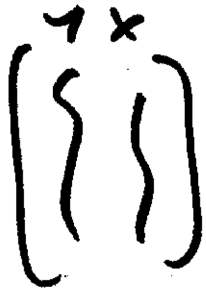
$\hat{r}_{\mu r}$ = $\underline{\underline{+.09}}$



$\hat{r}_{\sigma r}$ = +.55







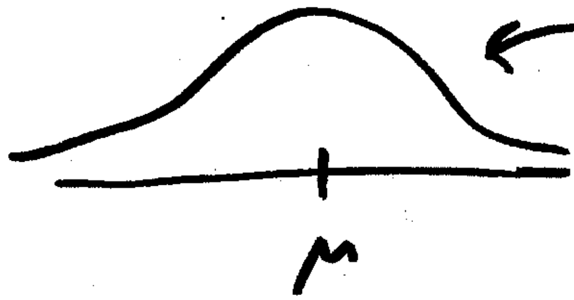
$\overset{iid}{N}(0, \sigma_e^2)$

$$Y_i = \underbrace{\beta_0 + \beta_1 x_i} + e_i$$

standard simple linear regression model

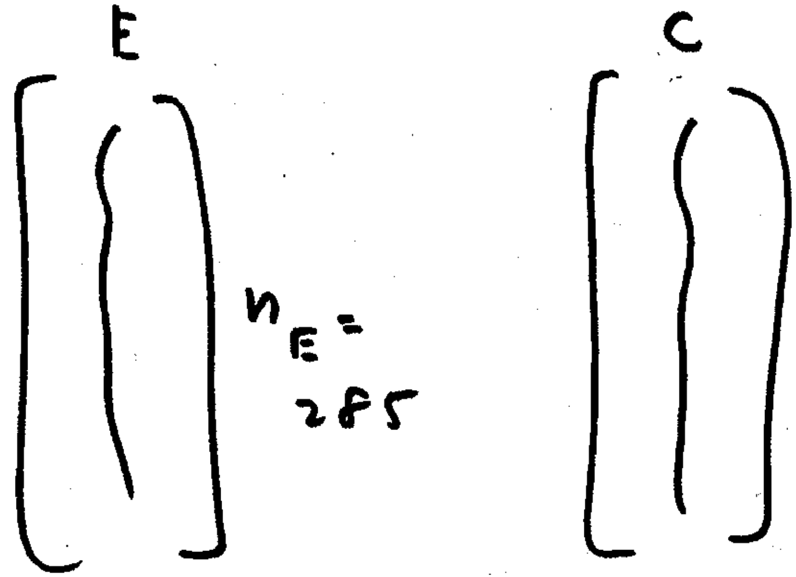
(multiple)

$$Y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}} + e_i$$



symmetric
unimodal
finite variance

among all such dist., normal has minimal Fisher information for the location parameter μ (related to maximum-entropy property of Gaussian)



(2-indep. samples problem)

mean $0.77 = \bar{E}$ mean $0.94 = \bar{C}$

SD $1.01 = s_E$ SD $1.24 = s_C$

$\bar{E}$ as an est. of μ_E

$\bar{C}$ ——— μ_C (in repeated sampling)

$(\bar{E} - \bar{C})$ ——— $(\mu_E - \mu_C)$

$SE(\bar{E} - \bar{C}) = ?$; $V(\bar{E} - \bar{C}) = ?$

$$V(\bar{E}) = \frac{\sigma_E^2}{n_E} \quad \& \quad V(\bar{C}) = \frac{\sigma_C^2}{n_C}$$

$$V(\bar{E} - \bar{C}) = V(\bar{E}) + (-1)^2 V(\bar{C}) - \cancel{2c(\bar{E}, \bar{C})} \to 0$$

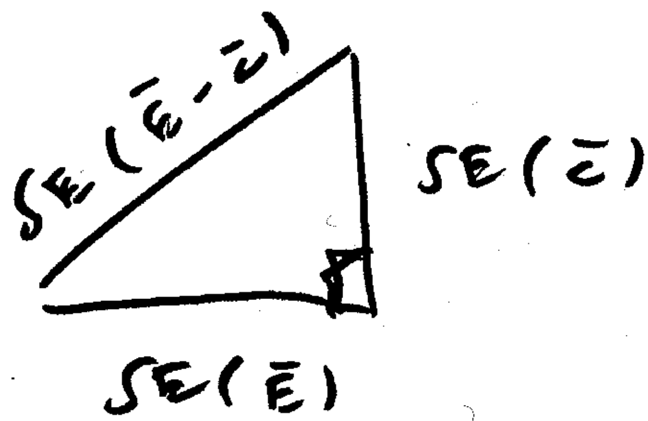
indep

$$= V(\bar{E}) + V(\bar{C})$$

$$SE(\bar{E} - \bar{C}) = \sqrt{V(\bar{E} - \bar{C})}$$

$$= \sqrt{V(\bar{E}) + V(\bar{E})}$$

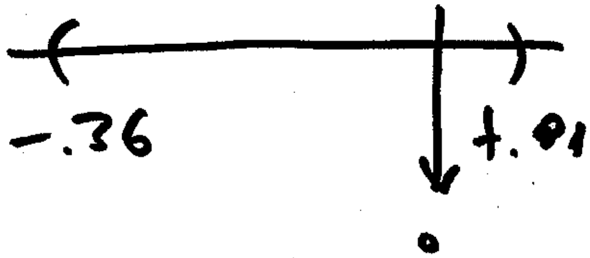
$$= \sqrt{\left(SE(\bar{E})\right)^2 + \left[SE(\bar{C})\right]^2}$$



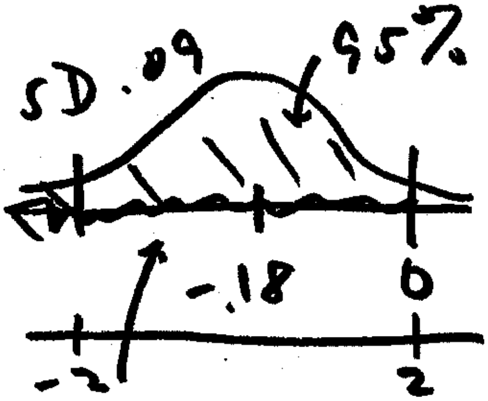
$$\hat{SE}(\bar{C}) = \frac{s_c}{\sqrt{n_c}}$$

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

$$\hat{SE}(\bar{E} - \bar{C}) = \sqrt{\frac{s_c^2}{n_c} + \frac{s_E^2}{n_E}}$$



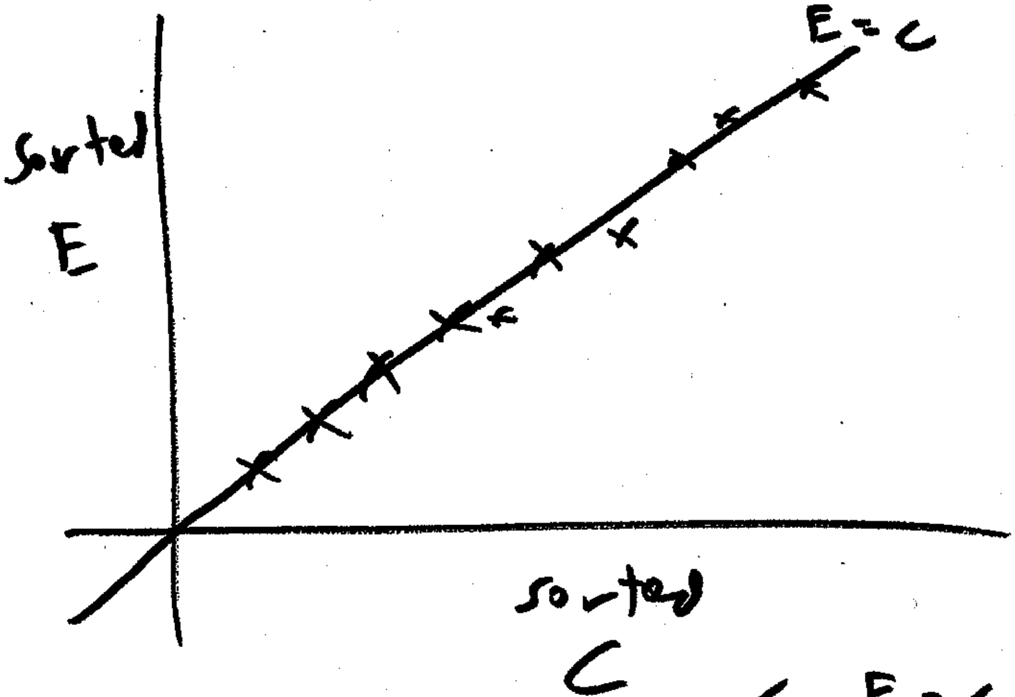
0 in this int so "insufficient evidence to reject $\boxed{H_0: \mu_E = \mu_C}$ at 5% level", but



$p(\mu_E - \mu_C \mid y)$

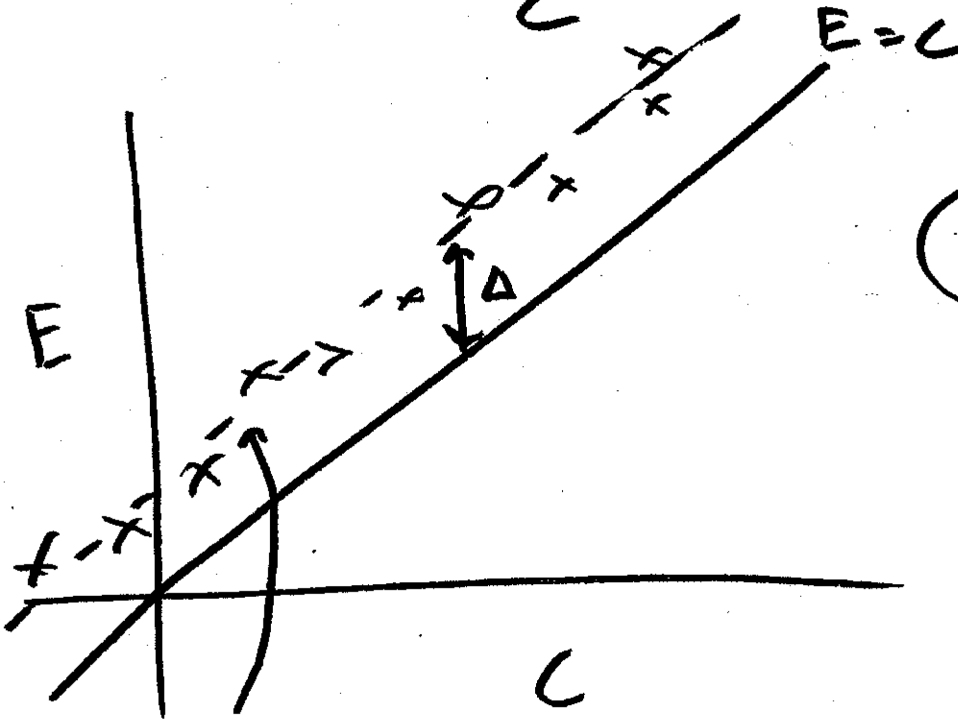
$p(\text{IHGA helps} \mid y) = .975$

39 to 1 posterior odds in favor of {IHGA helps}



$E \stackrel{st}{=} C$

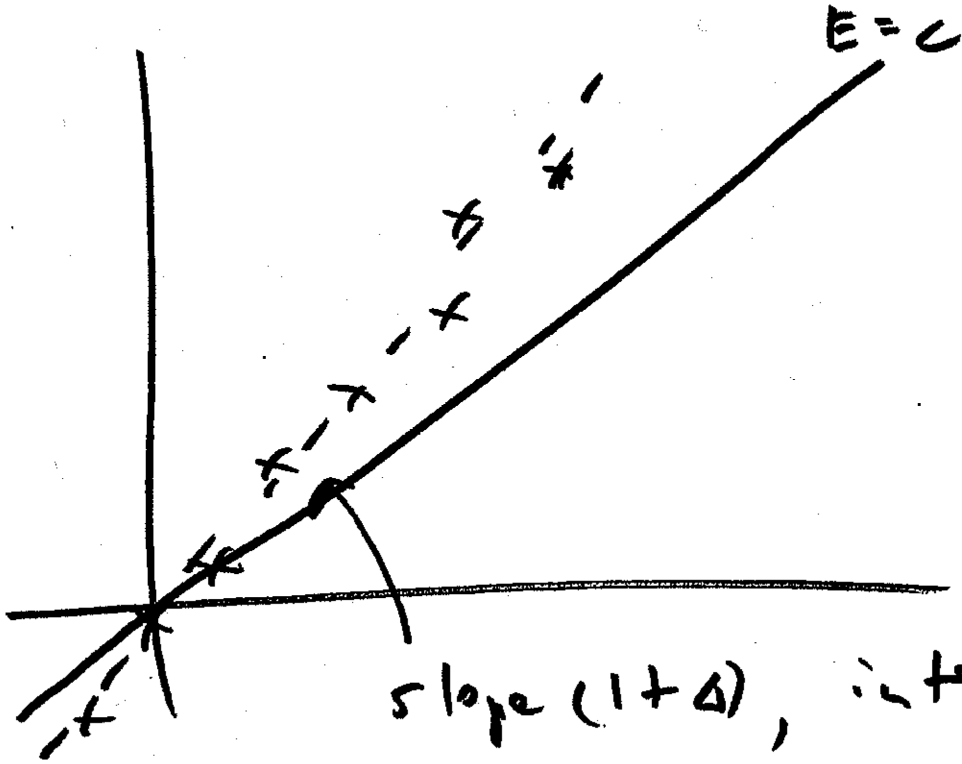
no effect



$E \stackrel{st}{=} C + \Delta$

additive effect

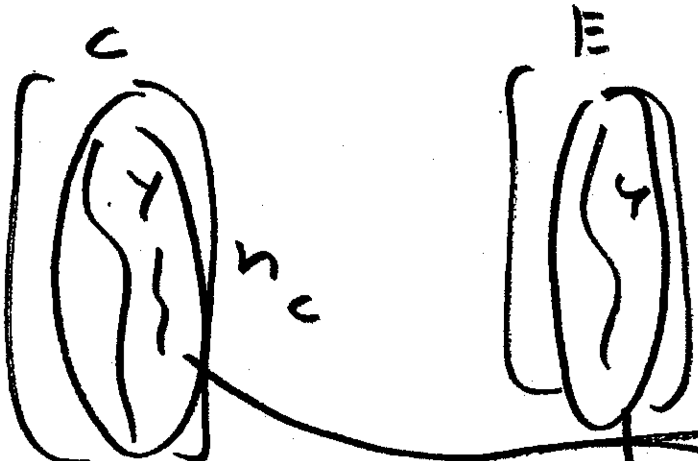
slope 1,
intercept Δ



$E \overset{st}{\doteq} c(1+\Delta)$

mult. effect

slope $(1+\Delta)$, intercept 0

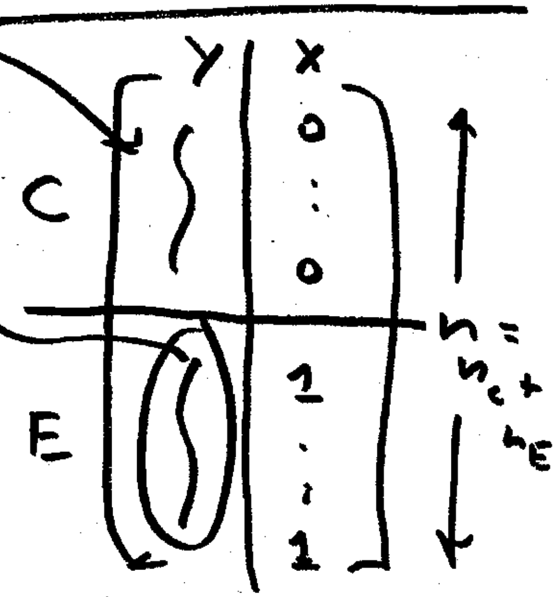


2-indep-samples structure

y: univariate outcome

x: predictor:

regression formulation →



$(Y_i \mid \lambda_i) \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_i)$ $(i = 1, \dots, n)$

← outcome T/C

$\log(\lambda_i) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \dots + \gamma_k x_{ik}$

random effects

(baseline health status) ... (age)

supposedly causal factor (SCF)

$e_i \overset{iid}{\sim} N(0, \sigma_e^2)$ (might be correlated with both Y & x_1)

potential confounding factors (PCFs) (confounders)

randomization is expected to have uncorrelated $(x_2, \dots, x_k)$ with x_1

<u>fact</u>: If $\hat{\theta}_{MLE}$ (dimension k) is the MLE of θ, then for any reasonable function f of θ, the MLE of $f(\theta)$ is $f(\hat{\theta}_{MLE})$ (functional invariance of MLE) (hwk 4)
