AMS 206

25 Jan 18

1

this time: (frequentist inference,
next time: Bayes) with unknown $\theta \in (0, 1)$

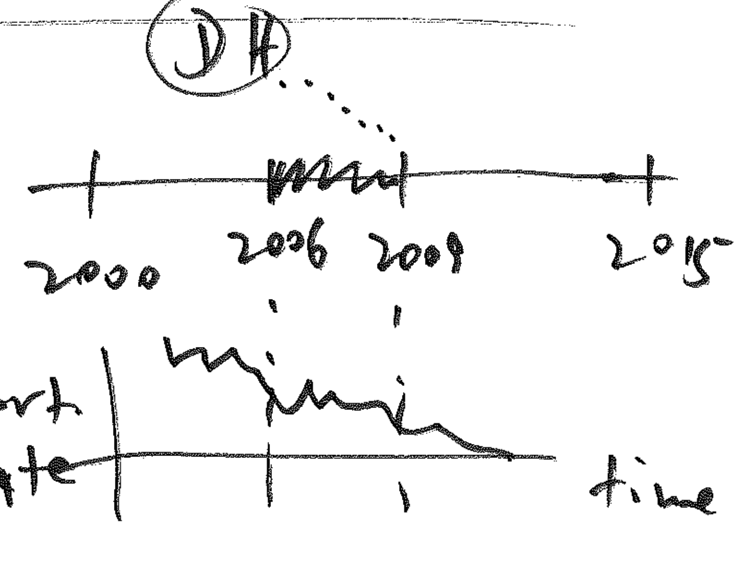
read: (J) ch. 1–3, Appendices A, B; (G) ch 1–2

can use (R + edit window)

or: fancier: install R-studio

observational data: you just look at data-generating processes in the world, without any attempt to randomly sample them

1 (P) like at random from



2

pop. Neyman sample & repeated ~~sampling~~

all AMI pts. similar to sampled patients in all relevant ways

freq. inf.

the observed AMI patients

dataset $\bar{y}$ values

die? (1 = Y, 0 = N)

N

1s & 0s

mean $\theta = ?$

SD $\sigma = \sqrt{\theta(1-\theta)} = ?$

pop. dist. (hist.)

0 1

(actual) like ~~at random~~ SRS = IID

die?

$y_1 \vdots y_n$ 1s & 0s $n = 400$

estimate of $\hat{\theta} \doteq 18\%$

mean $\bar{y} = \frac{S_n}{n} = \frac{72}{400}$

sum $S_n = \sum_{i=1}^{n} y_i$

hypothetical IID

1s & 0s $n = 400$

mean $\bar{y} = ?$

(ex. 18.7%)

$\vdots$

18%

18.7

$\vdots$

a $\bar{y}$

another $\bar{y}$

$m \to \infty$

long run mean

expected value of $\bar{y}$ = θ

estimated long run SD

standard error ($\hat{SE}$) of $\bar{y} = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \doteq 1.9\%$

long run $\hat{SE}$ 1.9%

hist (dist.) of $\bar{y}$

θ

[equivalent frequentist model]

$(Y_1, \ldots, Y_n \mid \theta\, B)$

$(Y_i \mid \theta\, B) \overset{IID}{\sim} \text{Bernoulli}(\theta)$ $(i = 1, \ldots, n)$ ③

$(i = 1, \ldots, n)$

discrete random variable

$E(Y_i \mid \theta) = \theta$

$V(Y_i \mid \theta) = \theta(1-\theta)$

$SD(Y_i \mid \theta) = \sqrt{\theta(1-\theta)}$

$E(\bar{Y} \mid \theta) = \theta$

$EV(\bar{Y}) =$

expected value

$(i = 1, \ldots, n)$

if

$E(Y_i) = \mu$

$V(Y_i) = \sigma^2$

$E(\bar{Y}) = \mu$

$V(\bar{Y}) = \frac{\sigma^2}{n}$

$SE_{RS}(\bar{Y}) = \sqrt{V(\bar{Y})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

standard error

repeated sampling

$= \sqrt{\frac{\theta(1-\theta)}{n}}$

(4)

frequentist inferential conclusions:)
On the basis of this sample of size n with $\bar{Y} = 0.18$, we estimate θ to be about $0.18 \; (\bar{Y}) = 18\%$, give or take about

$$\widehat{SE}(\bar{Y}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

$$= \sqrt{\frac{(0.18)(0.82)}{400}} \doteq 0.019 \doteq 1.9\%$$

unlucky random sampling noise

Central Limit Theorem (CLT): (de Moivre 172?)

as long as $\sigma^2 < \infty$, the dist. of $\bar{Y}$ will look a lot like normal curve if n is large