~~②~~ statsig: ① confidence intervals

STAT 206
29 Jan 21

Fri AM
disc sec.

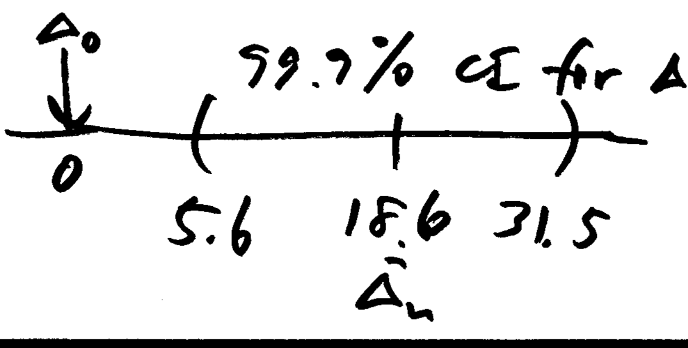
①

② hypothesis/significance testing

BMJ big false positive:
~~BMJ article~~ claimed that MMR vaccine <u>causes</u> autism in children; not true;

Fisher's implicit alternative hypothesis was usually $\Delta \neq 0$ if null: $\Delta = 0$

① CIs: 99.9% = $100(1-\alpha)$% with $\alpha = .001$ CI for Δ: $(5.6, 31.5)$ mmHg!



(0 not in CI) $\leftrightarrow$ dif. between $\hat{\Delta}_n$ & Δ_0 is statsig at 99.9% level

②

$SE_{IID}(\bar{Y}_n) = \frac{\sigma}{\sqrt{n}}$

$\frac{1}{n}\sum_{i=1}^{n} Y_i$

$E_{IID}(\bar{Y}_n) = \mu =$ pop. mean

pop. SD $\overset{\circ}{=} \sigma < \infty$

$\hat{SE}_{IID}(\bar{Y}_n) = \frac{S_D}{\sqrt{n}}$

$= \frac{10.1 \text{ mmHg}}{\sqrt{12}}$

$\doteq 2.9$ mmHg

since 2-tailed

p - value $\doteq 5.3 \cdot 10^{-5}$,

reject null (i.e., diff. betw.

$\hat{\Delta}_n = 18.6$ & $\Delta_0 = 0$ is statsig

small theorem	rule ①	reject null (declare $\hat{\theta}_n$ statsig diff. from θ_0) iff θ_0 is

not in $100(1-\alpha)\%$ CI

rule ② iff 2-tailed p - value $\leq \alpha$

rule ① & rule ② are algebraically identical

claim: $100(1-\alpha)\%$ CI, with α small enough for good reproducibility of our results, is far more informative than 2-tailed p-value

ex. Captopril

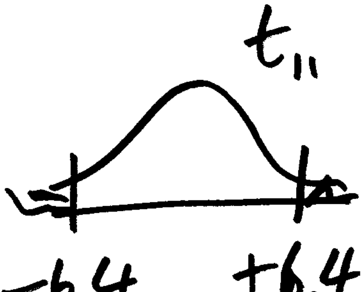
testing

null : $\Delta = \Delta_0 = 0$

alt : $\Delta \neq \Delta_0$ $\quad (n = 12)$

p-value $= 5.3 \cdot 10^{-5}$

2-tailed from t-test



$$t = \pm 6.4 = \frac{\text{signal}}{\text{noise}} \quad (*)$$

can't even tell from 2-tailed p whether Captopril is good or bad

1-tailed

$$p \doteq 2.6 \cdot 10^{-5} \text{ from } t = +6.4 = \frac{\text{signal}}{\text{noise}} = \frac{\bar{\Delta}_n - 0}{\widehat{SE}(\bar{\Delta}_n)}$$

[still can't tell whether diff is ~~practsig~~ signal/noise]

because $t = +6.4$ does not uniquely identify either

signal or noise

practsig statsig

more informed way of reporting

---

if all I tell you is $t = +6.4$, $n = 12$, 1-tailed $p \doteq$ [illegible]$.6 \cdot 10^{-5}$, you can't reconstruct 99.5% CI; but if I give you the 99.5% CI for Δ (practsig), $\hat{\Delta}_n = 18.6$ mmHg

5.6 — $\hat{\Delta}_n$ (18.6) — 31.6

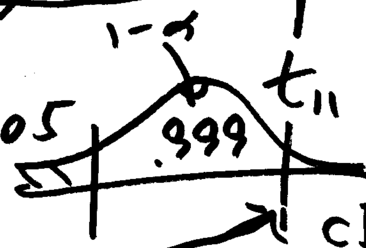
$\widehat{SE}_{IID}(\hat{\Delta}_n) =$

$$\hat{\Delta}_n \pm t^{-1}_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \widehat{SE}_{IID}(\hat{\Delta}_n)$$

$$\underbrace{31.6 - 18.6}_{\text{half-width of CI}} = 13.0 = \overset{4.44}{t_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)} \cdot \hat{SE}_{IID}(\hat{\Delta}_n)$$

(n = 12)

$\alpha = .001 \rightarrow 99.9\%$



density → d t

CDF → p t

Inverse CDF → q t

sampling → r t

$$t_{11}^{-1}(0.9995) = +4.44$$

$$t_{11}^{-1}(.0005) = \text{qt}(.0005, 11) \doteq -4.44$$

$$\hat{SE}_{IID}(\hat{\Delta}_n) = \frac{13.0 \text{ mmHg}}{4.44} \doteq 2.93 \text{ mmHg} = \frac{s_D}{\sqrt{12}}$$

$$s_D = 2.93 \cdot \sqrt{12} \doteq 10.1 \text{ mmHg}$$

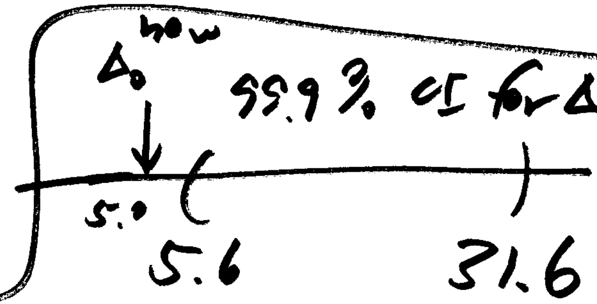
given 99.9% CI for Δ, based on sample of size $n = 12$ → $\hat{\Delta}_n = \bar{D}_n$ (prototyp), s_D, $\hat{SE}_{IID}(\hat{\Delta}_n)$, statistic yields.

3 more advantages of CIs:

(A) suppose instead of null: $\Delta = 0$ units (Δ_0)

somebody wants to test null: $\Delta = 5$ units $= \Delta_0^{new}$

significance testing people have to start over again & calculate a new $t = \frac{\text{signal}}{\text{noise}}$



5.0 is statsig different from $\hat{\Delta}_n = 18.6$ at 99.9% level

(B) given 99.9% CI & n, I can compute a conf. int at any other conf. level
(99.9% = conf. level)

(C) for many unknown θ, $100(1-\alpha)\%$ CIs are approximate Bayesian methods practsig $\neq$ statsig

in general

ex.) some other 1 mg for lowering blood pressure

Captopril-like study

$d = B - A$

$n = 1{,}000$

mean $\hat{\Delta}_n = +1$ mm Hg

SD $s_d = 5$ mm Hg

practsig? no

but statsig? yes

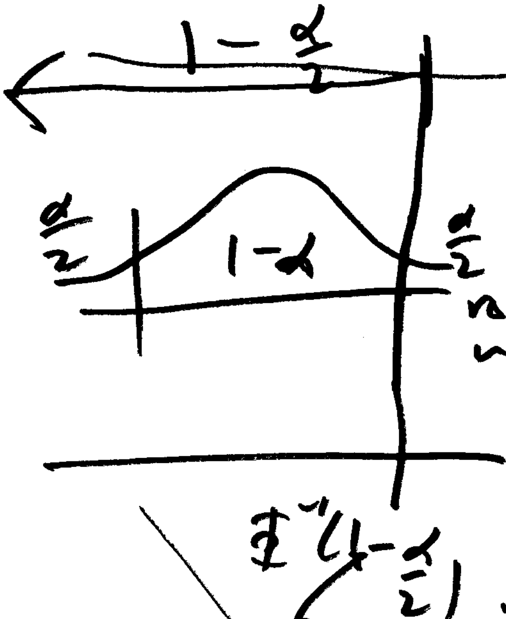
reason: too much data

99.9% CI

1 mmHg $\pm 3.29 \cdot 0.16 \leftarrow s_d / \sqrt{n}$

$\hat{\Delta}_n \pm z_\alpha \cdot \widehat{SE}(\hat{\Delta}_n)$

$\Phi^{-1}(1 - \frac{\alpha}{2})$ $\quad \alpha = .001$



$\Phi^{-1}(1 - \frac{\alpha}{2}) = 3.29$

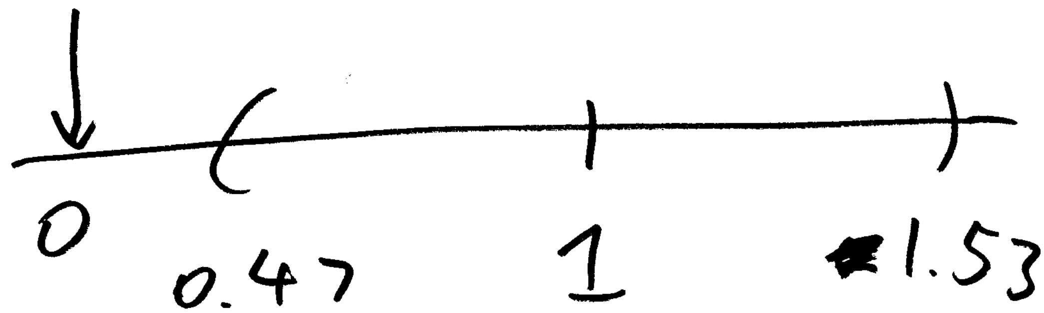
$\widehat{SE}(\hat{\Delta}_n) = \frac{5}{\sqrt{1000}} = 0.16$

qnorm$(1 - \frac{\alpha}{2})$ = qnorm(.9995)

d	norm
p	
q	
r	

$$99.9\%\ CI = \left(1 - (3.29)(0.16),\ 1 + (3.29)(0.16)\right) = (0.47,\ 1.53)\ \text{mmHg}$$

null: $\Delta = 0$



statsig yes, practsig no (!)

basic design rule: try to design data-gathering so that statsig $\doteq$ practsig

how achieve? do sample size determination

practsig yes but statsig no

$\hat{\Delta}_n = 15$ mmHg

$s_D = 15$

$n = 10$

$$\hat{\Delta}_n \pm 3.29 \cdot \frac{15}{\sqrt{10}} = (-1.6,\ 31.6) \text{ mmHg}$$

$\hat{\Delta}_n = 15$

not statsig diff from 0

here: (practsig yes, statsig no) ↔ (not enough data)

sample size determination: choose n so that statsig ≐ practsig

ex. practsig threshold: any

$\Delta \geq \Delta_{practsig}$ ↔ practsig yes

$\Delta_{practsig}$ = 15 mmHg