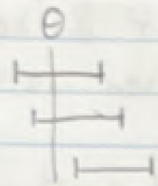
Lecture 14 10/26/20

Many people say that a single statistical inference (all 3 goals) is meaningless in the following way. Since we see only one dataset from the DGP, and thus only one $\hat{\hat{\theta}}$ from $\hat{\theta}$,

(1) Point estimation is silly because we have no idea how wrong we are,

(2) hypothesis testing is silly because if you reject, you don't know if you made a Type I error and if you retain, you don't know if you made a Type II error.

(3) Confidence intervals are silly because you don't know if θ is inside of the CI you computed.



all you see is: $\vdash\!\!\dashv$ $\quad P(\underset{\substack{\uparrow \\ \text{constant}}}{\theta} \in CI)$ $[0,1]$

"Wrong" in point estimation = large loss. "Wrong" in testing is a type I or type II error. "Wrong" for a confidence interval means it doesn't include θ.

"Do you have a better idea?" Their answer may be to do nothing.
"Statistics is like real life, you need to be okay with making mistakes."

This is the point in the class where Math 341 should begin. Math 341 also looks at the three goals of inference from a "Bayesian" perspective (we've looked at it in this class from a "Frequentist" perspective) which means you allow θ to be modeled as a r.v. We use MLE's, Fisher information and maybe other things from this class.

Recall the AF heart surgery study. For those subjects that didn't take the PUFAs, their AF incidence was $\hat{\hat{\theta}} = 27/81 = 0.333$

What if I care about the "odds against" getting AF?

$\phi := \dfrac{1-\theta}{\theta} = g(\theta)$

to create a point estimate, I'll plug in my estimate into g: $\hat{\hat{\phi}} = \dfrac{1-\hat{\hat{\theta}}}{\hat{\hat{\theta}}} = \dfrac{0.667}{0.333} = 2:1$

Why do we care? Because thinking in terms of odds-against is a useful way of thinking about risk (differently than probability).

What if I want to test odds-against or create a CI for odds against?

$$H_0: \phi = \phi_0 \quad , \quad CI_{\phi,1-\alpha} = [\cdots, \cdots]$$

What do we need to accomplish both testing and CI construction? We need the sampling distribution, $\hat{\phi}$.

C&B p 240-243 and it's called the "Delta Method." Let g be a differentiable function with no critical points and let $\hat{\theta}$ be an asymptotically normal estimator and $\hat{\phi} = g(\hat{\theta})$, then ...

$$\frac{g(\hat{\theta}) - g(\theta)}{|g'(\theta)| SE[\hat{\theta}]} \xrightarrow{d} N(0,1) \overset{\text{Richard}}{\Longrightarrow} \frac{g(\hat{\theta}) - g(\theta)}{|g'(\hat{\theta})| \hat{SE}[\hat{\theta}]} \xrightarrow{d} N(0,1)$$

$$\Downarrow$$

$$g(\hat{\theta}) \sim N\left(g(\theta), \left(|g'(\theta)| SE[\hat{\theta}]\right)^2\right)$$

$$\Rightarrow CI_{\phi,1-\alpha} \approx \left[g(\hat{\theta}) \pm Z_{1-\frac{\alpha}{2}} |g'(\hat{\theta})| SE[\hat{\theta}]\right]$$

$$\Rightarrow RET_{\phi,1-\alpha} \approx \left[g(\theta_0) \pm Z_{1-\frac{\alpha}{2}} \left[g'(\theta_0) SE[\hat{\theta}]\right]_{\hat{\theta}=\theta_0}\right]$$

↑ 2-sided test $H_a: \theta \neq \theta_0$

Proof: let $\hat{\theta}$ be asymptotically normal and $g'(\theta)$ nonzero everywhere. Consider the quantity:

$$\frac{g(\hat{\theta}) - g(\theta)}{g'(\theta) SE[\hat{\theta}]} \approx \frac{(\hat{\theta} - \theta) \cancel{g'(\theta)}}{\cancel{g'(\theta)} SE[\hat{\theta}]} \xrightarrow{d} N(0,1)$$

By a first order Taylor series approximation,

$$g(\hat{\theta}) \approx g(\theta) + (\hat{\theta} - \theta) g'(\theta) \Longrightarrow g(\hat{\theta}) - g(\theta) \approx (\hat{\theta} - \theta) g'(\theta)$$

$\theta \in [0,1]$

Let's do our odds-against example now. $\phi = \frac{1-\theta}{\theta} = g(\theta) \Rightarrow g'(\theta) = -\theta^{-2}$ ($= \theta^{-1} - 1$)

$$CI_{\phi,1-\alpha} \approx \left[\frac{1-\hat{\theta}}{\hat{\theta}} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{1}{\hat{\theta}^2} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\right]$$

in our data ...

$$CI_{\phi,95\%} \approx \left[2 \pm 1.96 \cdot \frac{1}{(0.333)^2} \sqrt{\frac{(.333)(.667)}{81}}\right] = [1.07, 2.93]$$

Prob 11 on midterm ... iid DGP, mean θ, variance σ^2, both unknown, $\hat{\theta} = \bar{X}$

$$\frac{\hat{\theta} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

$\Rightarrow g(\theta) \Rightarrow g'(\theta) = \frac{1}{\theta} > 0$

I want a $CI_{\phi, 1-\alpha}$ where $\phi := \ln(\theta)$, log survival

$$CI_{\phi,1-\alpha} \approx \left[\ln(\hat{\theta}) \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{1}{\hat{\theta}} \frac{S}{\sqrt{n}}\right]$$

For our data, $\bar{x} = 2.57$, $s = 1.00$, $n = 30$

$$CI_{\phi,95\%} \approx \left[\ln(2.57) \pm 1.96 \frac{1}{2.57} \cdot \frac{1.00}{\sqrt{30}}\right] = [0.805, 1.083]$$

In the AF study, the first group didn't get PUFAs, the second group did get PUFAs (control group, experimental group). The incidence estimates were: $\hat{\hat{\theta}}_1 = 0.333$, $n_1 = 81$, $\hat{\hat{\theta}}_2 = 0.152$, $n_2 = 79$

How much more likely is someone to get AF without PUFAs than with PUFAs?

$$RR := \frac{P(\text{AF without PUFAs})}{P(\text{AF with PUFAs})} = \frac{\theta_1}{\theta_2}, \quad \hat{\hat{RR}} = \frac{\hat{\hat{\theta}}_1}{\hat{\hat{\theta}}_2} = \frac{0.333}{0.152} = 2.192$$

"RR" is "relative risk" or "risk ratio" and it's another way to think about the relationship between two incidence (proportion) metrics. $\theta_1 - \theta_2$ is sometimes called "risk difference." The difference between these two concepts is large. For examples,

Scenario #1: $\theta_1 = 0.6$, $\theta_2 = 0.5$, $\theta_1 - \theta_2 = 0.1$, $RR = 1.2$ "20% more likely"

Scenario #2: $\theta_1 = 0.11$, $\theta_2 = 0.01$, $\theta_1 - \theta_2 = 0.1$, $RR = 11$ "1100% more likely"

How do we do testing and confidence interval construction for the RR? "Multivariate Delta Method" and it's beyond the scope of this course but we will use a result of it which you'll need to know and we won't prove it.

$$g: \mathbb{R}^K \to \mathbb{R}, \quad \Sigma = \text{Var}\begin{bmatrix}\hat{\theta}_1 \\ \vdots \\ \hat{\theta}_K\end{bmatrix}$$ Math 368 variance - covariance matrix

and $$\Sigma^{-1}\left(\begin{bmatrix}\hat{\theta}_1 \\ \vdots \\ \hat{\theta}_K\end{bmatrix} - \begin{bmatrix}\theta_1 \\ \vdots \\ \theta_K\end{bmatrix}\right) \xrightarrow{d} N_K(\vec{0}_K, I_K)$$

$$\Rightarrow \frac{g(\hat{\theta}_1 \ldots \hat{\theta}_K) - g(\theta_1, \ldots, \theta_K)}{\sqrt{\vec{\nabla}g^T \Sigma \nabla_g}} \xrightarrow{d} N(0,1)$$

If $K=2$, and $\hat{\theta}_1$ is independent of $\hat{\theta}_2$ then...

$$\Sigma = \begin{bmatrix} Var[\hat{\theta}_1] & 0 \\ 0 & Var[\hat{\theta}_2] \end{bmatrix}, \quad \phi = g(\theta_1, \theta_2)$$

$$\Rightarrow \frac{g(\hat{\theta}_1, \hat{\theta}_2) - g(\theta_1, \theta_2)}{\underbrace{\begin{bmatrix} \frac{\partial g}{\partial \theta_1} & \frac{\partial g}{\partial \theta_2} \end{bmatrix} Var \begin{bmatrix} \hat{\theta}_1 & 0 \\ 0 & Var[\hat{\theta}_2] \end{bmatrix}}_{\begin{bmatrix} \frac{\partial g}{\partial \theta_1} & Var[\hat{\theta}_1] \\ \frac{\partial g}{\partial \theta_2} & Var[\hat{\theta}_2] \end{bmatrix}} \begin{bmatrix} \frac{\partial g}{\partial \theta_1} \\ \frac{\partial g}{\partial \theta_2} \end{bmatrix}} = \frac{g(\hat{\theta}_1, \hat{\theta}_2) - g(\theta_1, \theta_2)}{\sqrt{\left(\frac{\partial g}{\partial \theta_1}\right)^2 Var[\hat{\theta}_1] + \left(\frac{\partial g}{\partial \theta_2}\right)^2 Var[\hat{\theta}_2]}} \xrightarrow{d} N(0,1) \quad *$$

$$\Downarrow$$

$$CI_{\phi, 1-\alpha} \approx \left[g(\hat{\theta}_1, \hat{\theta}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\left(\frac{\partial g}{\partial \theta_1}\Big|_{\substack{\theta_1 = \hat{\theta}_1 \\ \theta_2 = \hat{\theta}_2}}\right)^2 Var[\hat{\theta}_1] + \left(\frac{\partial g}{\partial \theta_2}\Big|_{\substack{\theta_1 = \hat{\theta}_1 \\ \theta_2 = \hat{\theta}_2}}\right)^2 Var[\hat{\theta}_2]} \right]$$

Back to our case of the RR. This case fits the corollary. We have two independent estimators (from two independent populations).

$$\phi = RR = \frac{\theta_1}{\theta_2} = g(\theta_1, \theta_2) \Rightarrow \frac{\partial g}{\partial \theta_1} = \frac{1}{\theta_1}, \quad \frac{\partial g}{\partial \theta_2} = \frac{-\theta_1}{\theta_2^2}$$

$$CI_{RR, 1-\alpha} \approx \left[\frac{\hat{\theta}_1}{\hat{\theta}_2} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{\hat{\theta}_2^2} \frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_1^2}{\hat{\theta}_2^4} \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}} \right]$$

with our data

$$CI_{RR, 95\%} \approx \left[2.192 \pm 1.96 \sqrt{\frac{1}{(0.152)^2} \frac{0.333(0.667)}{81} + \frac{(0.333)^2}{(0.152)^4} \frac{0.152(.848)}{79}} \right]$$

$$= [1.020, 3.362]$$