DD office 1.5 hour Thu 16 Apr 20

Quiz 3

STAT 131 16 Apr 20

①

$$P(U \mid D) = \frac{P(U)\,P(D \mid U)}{P(D)}$$

$P(D) > 0$ assume

(a) suppose we put

$P(U) = 0$ → $\frac{P(U)\,P(D \mid U)}{P(D)} = 0$

~~good idea? no~~

so if $P(U) = 0$ then $P(U \mid D) = 0$

no matter what D is

(b) suppose we put $P(U) = 1$

$$P(U \mid D) = \frac{\cancel{P(U)}^{\,1}\,P(D \mid U)}{P(D)} = \frac{P(D \mid U)}{P(D)} = 1$$

but if $P(U) = 1$, $P(D \mid U) = P(D)$

for any T/F prop. D,

$$P(D \mid \underline{1=1}) = P(D)$$

$$Y \text{ (outcome)} = \begin{cases} 1 & \text{if dead in 1992} \\ 0 & \text{not} \end{cases}$$

$$X = (\text{"treatment"}) = \begin{cases} 1 & \text{if current smoker in 1972} \\ 0 & \text{not} \end{cases}$$

(supposedly causal factor)

$$Z_1 = (CF) = \begin{cases} 1 & \text{if 18-64 in 1972} \\ 0 & \text{not} \end{cases}$$

(confounding factor)

tests to see if Z_1 (age) is a CF

(1) plausible that Z_1, Y assoc.? <u>yes</u>

as Z_1 goes from 0 (<u>65+</u>) to 1 (18-64), $P(\text{dead}) \downarrow$

(2) plausible that z_1, X assoc.? yes

as z_1 goes from 0 (65+) to 1 (18-64), P(smoker in 1972) ↑

$$P(S \mid \cancel{18-}64) = \frac{533}{1072} \doteq 49.7\%$$

($z_1 = 1$)

$$P(S \mid \underbrace{65+}_{z_1 = 0}) = \frac{49}{242} \doteq 20.2\%$$

$$\boxed{o_A = \frac{p_A}{1 - p_A}}$$

if $p_A \doteq 0$ then $o_A \doteq p_A$

$$p_A \doteq \frac{1}{38.32} \doteq 0.026 \rightarrow o_A = \frac{\frac{1}{\cancel{38.32}}}{\frac{37.32}{\cancel{38.32}}} = \frac{1}{37.32}$$

0.027

$\boxed{\text{have to know 3 things:}}$ $P(C) = 0.01$

① **prevalence** of COVID-19 in U.S. pop.:

case 1: $\frac{641,166}{328,200,000} \doteq 0.2\%$

case 2: 1.0%

② $P(+ \mid C) = 93.8\%$ = sensitivity of Cellex test

③ $P(- \mid \text{not } C)$ = specificity of Cellex test = 95.6%

culprit

method I

truth

sens.

false positive

Cellex test says	C	not C	
+	9,380	43,560	52,940
−	620	946,440	947,060
	10,000	990,000	1,000,000

false negative

case 2

$P(C) = 1\%$

$P(\text{not } C \mid -) = \frac{946,440}{947,060} \doteq .9993$

$P(C \mid +) = \frac{9380}{52,940} = 0.177 \doteq 18\%$ (!)

antibody concentration

cellex says $\ominus$ | $\boxed{C}$ | cellex says $\oplus$

case 2

truth

		C	not C	
cellex test says	$\oplus$	1,876	43,912	45,788
	$\ominus$	124	954,088	954,212
		2000	99,800	1,000,000

$P(C) = 0.002$

$$P(C \mid \oplus) = \frac{1{,}876}{45{,}788} = 0.04\,(!)$$

$$P(\text{not } C \mid \ominus) = \frac{954{,}088}{954{,}212} = .99987$$

case	$P(C \mid \oplus)$	$P(\text{not } C \mid \ominus)$	prevalence
1	4%	99.99%	0.2%
2	18%	99.93%	1.0%

cellex test almost perfect when $\ominus$

but it's cheap when it says (+)

truth

cellex test says	c	not c
(+)	true positive	false positive (bad)
(−)	false negative (bad)	true negative

person told (+) but really (−) (bad)

person told (−) but really (+) | much worse

decision theory

action space $\mathcal{a}$

$$\mathcal{a} = \{ a_1, a_2, a_3 \}$$

a_1: give cellex to everybody & pretend it's right (quicker, cheaper)

a_2: give gold standard to everybody & pretend it's right (slower, more costly)

a_3: cellex; if (−) conclude not c; if (+), gold standard & believe it know:

$P(c)$

$P(+ \mid c)$ = sens.

$P(- \mid \text{not } c)$ = spec.

want

$$\boxed{P(c \mid +)}$$

$$P(\text{not } c \mid -)$$

$$P(A \mid B) \stackrel{?}{=} P(B \mid A) \quad \text{no}$$

$P(\text{rain} \mid \text{clouds overhead})$ = low to in between

$P(\text{clouds overhead} \mid \text{rain})$ = high

prevalence　　sensitivity

$$P(c \mid +) = \frac{P(c)\,P(+ \mid c)}{\boxed{P(+)}} \qquad (1)$$

(annoying) normalizing constant

method II

$$p(\text{not } C \mid +) = \frac{p(\text{not } C)\, p(+ \mid \text{not } C)}{p(+)} \qquad (2)$$

$$\frac{(1)}{(2)} = \frac{p(C \mid +)}{p(\text{not } C \mid +)} = \left[\frac{p(C)}{p(\text{not } C)}\right]\left[\frac{p(+ \mid C)}{p(+ \mid \text{not } C)}\right]$$

(likelihood ratio)

(posterior odds ratio in favor of C given data) = (prior odds ratio in favor of C) (Bayes factor)

~~posterior~~ information

prior information

(likelihood information)

$$\boxed{p(C \mid +)} = \frac{\boxed{p(C)}\ \boxed{p(+ \mid C)}}{p(+)}$$

data

unknown

(normalizing constant)

(before data) (a $\underline{\text{priori}}$)

(after data) (a $\underline{\text{posteriori}}$)

$$\begin{pmatrix}\text{posterior}\\ \text{info}\end{pmatrix} = \frac{\begin{pmatrix}\text{prior}\\ \text{info}\end{pmatrix} \cdot \begin{pmatrix}\text{likelihood}\\ \text{info}\end{pmatrix}}{\begin{pmatrix}\text{normalizing}\\ \text{constant}\end{pmatrix}}$$

$$p(\underline{\text{unknown}} \mid \text{data}) = \frac{p(\underline{\text{unknown}}) \cdot p(\underline{\text{data} \mid \text{unknown}})}{\begin{pmatrix}\text{normalizing}\\ \text{constant}\end{pmatrix}}$$

$$\frac{P(C \mid +)}{P(\text{not } C \mid +)} = \overbrace{\left(\frac{.01}{.99}\right)}^{\text{prevalence (case 2)}} \left(\frac{\overset{\text{sensitivity}}{0.938}}{\underset{1 - \text{specificity}}{1 - .956}}\right)$$

10

$$\boxed{P(+ \mid \text{not } C) = 1 - P(- \mid \text{not } C)}$$

$$\left(\frac{1}{99}\right)\left(\frac{469}{22}\right)$$