Data-individual

Kian Parnis 0107601L

22nd January 2021

Queen's Gambit

(i)

P(W_1) = P(W_1|beginner)P(beginner)+P(W_1|intermediate)P(intermediate) + P(W_1|master)P(master) = $\frac{1}{3}$(0.9+0.5+0.3)=$\frac{17}{30}$ Probability of Beth winning the first game is $\frac{17}{30}$

(ii)

$W_1 \cap W_2$ = P(W_1,W_2|beginner)P(beginner) + P(W_1,W_2|intermediate)P(intermediate) + P(W_1,W_2|master)P(master) = [$W_1 \cap W_2$ conditionally independent] = $\frac{1}{3}(0.9^2 + 0.5^2 + 0.3^2)$= $\frac{23}{60}$

Probability of A given B:

P(W_1|W_2) = $\frac{W_1 \cap W_2}{P(W_1)} = \frac{\frac{23}{60}}{\frac{17}{30}} = \frac{23}{34}$ Probability of Beth winning the second game is $\frac{23}{34}$

(iii)

Independent means that the outcome of one game doesn't affect the chance of winning the game after. Conditional Independence means that outcome of one game does not affect the probability of winning in the next. With this said, Conditional Independence is more reasonable here because in general independence, winning the first game can give information about the second game.

Conditionitis

Given "Conditionitis" effects 0.05% then P(D) = 0.0005 and P(D^C) = 0.9995 Given screening identity's disease in 90% of cases then P(T|D) = 0.9 and P(T^C|D) = 0.1 Given generated false positives 1% then P(T|D^C) = 0.001 and P(T^C|D^C) = 0.999

To Find:

(Using Bayas Rule)
P(D|T) = $\frac{P(D)P(T|D)}{P(D)P(T|D)+P(D^C)P(T|D^C)} = \frac{(0.0005)(0.9)}{(0.0005)(0.9)+(0.9995)(0.001)} = \frac{900}{2899} \approx 0.31$

(Using Conditional Probability)
P(D|T^C) = $\frac{P(D \cap T^C)}{P(T^C)} = \frac{P(D)P(T^C|D)}{P(T^C)} = \frac{(0.0005)(0.1)}{P(T^C)}$

(Using Law of Total Probability)
$P(T^C) = P(D \cap T^C) + P(D^C \cap T^C) = (P(D)P(T^C|D) + P(D^C)P(T^C|D^C)$ = (0.0005)(0.1)+(0.9995)(0.999) = 0.9985505

$$\therefore P(D|T^C) = \frac{(0.0005)(0.1)}{0.9985505} \approx 5.007 * 10^{-5}$$

Hypo-headaches

Given sample size: 400
Number of observations: 70
Rejection: p = 14% / 0.14

i)

Parameter of interest in given sample:

Standard Deviation $\sigma = \sqrt{\frac{p(1-p)}{N}} = \sqrt{\frac{(0.14)(1-(0.14))}{400}} = 0.0173$
Where The Standard Deviation is a measure of how spread out numbers are.

ii)

Null Hypothesis = $H_0 : p = 0.14$
Alternative Hypothesis = $H_1 : p > 0.14$

iii)

Given that in the problem there is no reference to α we can assume the most popular value i.e., $\alpha = 0.05$
Critical value = +1.645

Sample proportion: $\hat{p} = \frac{x}{N} = \frac{7}{40} = 0.175$

Z-test: $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{N}}} = \frac{0.175-0.14}{0.0173} = 2.023$

$2.023 > 1.645 \therefore$ reject null and accept alternative hypothesis. EMA should reject the drug.

Spam Filters

Given Spam 80% then P(S) = 0.8 and legitimate $(S^C) = 0.2$
Given 10% of spam has phrase "free credit" then P(F|S) = 0.1
Given 1% of non-spam has phrase "free credit" then $P(F|S^C) = 0.01$

i)

To proof P(S|F) (Given "free credit" the probability of spam) (Using Bayas Rule)

P(S|F) = $\frac{P(S)P(F|S)}{P(S)P(F|S)+P(S^C)P(F|S^C)} = \frac{(0.8)(0.1)}{(0.8)(0.1)+(0.2)(0.01)} = \frac{40}{41} = 0.2$

ii)

To proof $P(L|F_1, F_2)$ (Mail is legitimate given both filters are legitimate) Given $P(F_1, F_2|L) = P(F_1, F_2|L^C)$ = 0.9 (Using Bayas Rule)

$P(L|F_1, F_2) = \frac{P(L)P(F_1,F_2|L)}{P(L)P(F_1,F_2|L)+P(L^C)P(F_1,F_2|L^C)} = \frac{(0.2)(0.9)}{(0.2)(0.9)+(0.8)(0.9)} = 0.2$

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Declaration

Plagiarism is defined as "the unacknowledged use, as one's own work, of work of another person, whether or not such work has been published" (<u>Re lations Governin Conduct at Examinations</u>, 1997, Regulation I (viii), University of Malta),

I the undersigned, declare that the [assignment] submitted is my work, except where acknowledged and referenced.

I understand that the penalties for making a false declaration may include, but are not limited to, loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Work submitted without this signed declaration will not be corrected, and will be given zero marks.

* Delete as appropriate.

(N.B. If the assignment is meant to be submitted anonymously, please sign this form and submit it to the Departlnental Officer separately from the assignment).

Kian Parnis	Kian
Student Name	Signature
Student Name	Signature
Student Name	Signature
Student Name	Signature

ARI1101 — Data Science Individual — Course

Code — Title of work submitted

12/02/2021

Date