Assignment #2

Thursday, September 22, 2022 10:43 AM

Apprentice

1.

race	move up	move down	remain
white	16	15	74
Black	7	14	45
Hispanic	5	13	43
Asian	13	10	48

a.

race	move up (+)	move down (-)	remain (R)	
white (W)	0.052805	0.04950495	0.244224	0.346535
Black (B)	0.023102	0.04620462	0.148515	0.217822
Hispanic (H)	0.016502	0.04290429	0.141914	0.20132
Asian (A)	0.042904	0.0330033	0.158416	0.234323
	0.135314	0.17161716	0.693069	1

   i. Letters and symbols in parentheses are the symbols for the event e.g. let + be the move up event and W be the person being white

b. $P(H) = 0.20132 = 20.132\%$

c. $P(-) = 0.17161716 \approx 17.16\%$

d. $P(+|B) = \dfrac{0.023102}{0.217822} \approx 0.1061 = 10.61\%$

e. $P(W\ and\ R) = 0.244224 \approx 24.42\%$

f. If we made a table for the joint probabilities using the marginal probabilities, we get this table (look below). Although the numbers are pretty consistent, there are some categories that don't match up well, meaning that it is probably somewhat dependent.

   i.

race	move up	move down	remain	
white	0.046891	0.05947129	0.240173	0.346535
Black	0.029474	0.03738196	0.150966	0.217822
Hispanic	0.027241	0.03454999	0.139529	0.20132
Asian	0.031707	0.04021392	0.162402	0.234323
	0.135314	0.17161716	0.693069	1

g. Not quite sure what this question means. If it wants joint probabilities based on margin probabilities, look above. If it wants that marginal probability is the sum of joint probabilities, look below:

   i. $P(A) = P(A\ and\ B) + P(A\ and\ !B)$
   ii. $P(W) = P(W\ and\ +) + P(W\ and\ !+) = P(W\ and\ +) + P(W\ and\ -) + P(W\ and\ R)$
   iii. $= 0.046891 + 0.05947129 + 0.240173 = 0.346535 = P(W)$

h. $P(+|W) + P(!+|W) = P(+|W) + P(-|W) + P(R|W)$

   i. 0.052805/0.346535+0.04950495/0.346535+0.244224/0.346535=1
   ii. Adds up to 100%

2. 1003 participants, COVID test correctly identifies positives 95% of the time, COVID in population is 1.5% and 3.5% the test incorrectly identifies negative.

   a. Let C be having COVID and T be testing positive, $P(C) = 1.5\%, P(!C) = 98.5\%$

      i. True positive: $P(T|C) = 95\%$
      ii. False positive: $P(T|!C) = 3.5\%$

iii. True negative: $P(!T|!C) = 100\% - 3.5\% = 96.5\%$
iv. False negative: $P(!T|C) = 100\% - 95\% = 5\%$

b. $P(T) = P(T|C) * P(C) + P(T|!C) * P(!C) = 95\% * 1.5\% + 3.5\% * 98.5\% = 4.8725\%$

c. $P(C|T) = \dfrac{P(T|C) * P(C)}{P(T)} = \dfrac{95\% * 1.5\%}{4.8725\%} \approx 29.2458\%$

i. Yeah, that is surprisingly low, but makes sense since the percent of the population that has COVID is very low and the false positive chance is pretty high.

Master

3.

race	move up (+)	move down (-)	remain (R)	
white (W)	0.052805	0.04950495	0.244224	0.346535
Black (B)	0.023102	0.04620462	0.148515	0.217822
Hispanic (H)	0.016502	0.04290429	0.141914	0.20132
Asian (A)	0.042904	0.0330033	0.158416	0.234323
	0.135314	0.17161716	0.693069	1

a. $P(!W \text{ and } !R) = 1 - (0.346535 + 0.693069 - 0.244224) \approx 0.2046 = 20.46\%$

b. $P(B \text{ or } H|+) = \dfrac{0.023102 + 0.016502}{0.135314} \approx 0.2927 = 29.27\%$

c. $P(!W \text{ or } !R) = 1 - P(W \text{ and } R) = 1 - 0.244224 \approx 0.7557 = 75.58\%$

4. 1003 participants, COVID test correctly identifies positives 95% of the time, COVID in population is 1.5% and 3.5% the test incorrectly identifies negative.

a. Highest P(C|T) with lower false negatives

i. Assuming false negatives become 0, then true positives would be 100%

1) $\mathrm{P}(!T|C) = 0, \mathrm{P}(T|C) = 100\%$

ii. $P(C|T) = \dfrac{P(T|C) * P(C)}{P(T)} = \dfrac{100\% * 1.5\%}{4.9475\%} \approx 30.3183\%$

1) $P(T) = 100\% * 1.5\% + 3.5\% * 98.5\% = 4.9475\%$

2) We can't raise it higher because the chance of false positives and the number of people without COVID make up the rest of the percentage. The only want to make it higher is for a higher percentage of the population to have COVID or to have chance of false positive drop a lot.

b. Lowest P(C|T) with higher false negatives

i. Assuming false negatives become 100%, then true positives would be 0

1) $\mathrm{P}(!T|C) = 100\%, \mathrm{P}(T|C) = 0$

ii. $\mathrm{P}(C|T) = \dfrac{\mathrm{P}(T|C) * \mathrm{P}(C)}{\mathrm{P}(T)} = \dfrac{0\% * 1.5\%}{3.4475\%} = 0\%$

1) $\mathrm{P}(T) = 0\% * 1.5\% + 3.5\% * 98.5\% = 3.4475\%$

2) It becomes 0% because all people who have the disease test negative because all false negatives so nobody who tests positive actually has COVID.

Wizard

5. 0.33% of women >40 develop breast cancer, 10.5% of patients with breast cancer receive false negative. False positives to 7.5% of patients without breast cancer. Let B be breast cancer and T be positive test. $P(B) = 0.33\%. P(!T|B) = 10.5\%, P(T|!B) = 7.5\%, P(T|B) = 89.5\%, P(!T|!B) = 92.5\%$

a. $P(B|T) = \dfrac{P(T|B) * P(B)}{P(T)} = \dfrac{89.5\% * 0.33\%}{7.7706\%} \approx 3.8\%$

i. $P(T) = 89.5\% * 0.33\% + 7.5\% * 99.67\% = 7.7706\%$

b. $P(T|B) = 90\%, P(!T|B) = 10\%, P(T|!B) = 8\%, P(!T|!B) = 92\%$

i. $P(T) = 90\% * 0.33\% + 8\% * 99.67\% = 8.2706\%$

ii. $P(B|T) = \dfrac{90\% * 0.33\%}{8.2706\%} = 3.591\%$

iii. Although the number of true breast cancer given a positive test is lower, I still would say it's better since it is more important to know someone has breast cancer earlier on than make sure the number of positives are correct since it would allow further testing and earlier treatment if there are any problems.

6. $P(!T_1 \text{ and } !T_2|C) = P(!T_1|C) * P(!T_2|C) = 5\% * 5\% = 0.25\%$
   a. $P(!T_1 \text{ and } !T_2|!C) = P(!T_1|!C) * P(!T_2|!C) = 96.5\% * 96.5\% = 93.1225\%$
   b. $P(!T_1 \text{ and } !T_2) = P(!T_1 \text{ and } !T_2|C) * P(C) + P(!T_1 \text{ and } !T_2|!C) * P(!C)$
      $= 0.25\% * 1.5\% + 93.1225\% * 98.5\% \approx 91.7294\%$
   c. $P(C|!T_1 \text{and } !T_2) = \frac{P(!T_1 \text{and } !T_2|C) * P(C)}{P(!T_1 \text{and } !T_2)} = \frac{0.25\% * 1.5\%}{91.7294\%} = 0.0041\%$
   d. Having COVID-19 given 2 negative tests is 0.0041%, which is extremely low.