MAT 2379, Introduction to biostatistics

Assignment 1 - solutions

The assignment is marked on 24 points.

Question 1: Problem 1.6 from the textbook. **[1]**

(a) A flower is purple only if it genotype is $CCPP$. The flower is white in all the other cases, i.e. when its genotype is $CcPP, ccPP, CCPp, CcPp, ccPp, CCpp, Ccpp, ccpp$.
(b) We cross $CcPc \times CcPp$.

Female Gamete	Male Gamete $\frac{1}{4}CP$	$\frac{1}{4}Cp$	$\frac{1}{4}cP$	$\frac{1}{4}cp$
$\frac{1}{4}CP$	$\frac{1}{16}CC\ PP$ (purple)	$\frac{1}{16}CC\ Pp$ (white)	$\frac{1}{16}Cc\ PP$ (white)	$\frac{1}{16}Cc\ Pp$ (white)
$\frac{1}{4}Cp$	$\frac{1}{16}CC\ Pp$ (white)	$\frac{1}{16}CC\ pp$ (white)	$\frac{1}{16}Cc\ Pp$ (white)	$\frac{1}{16}Cc\ pp$ (white)
$\frac{1}{4}cP$	$\frac{1}{16}Cc\ PP$ (white)	$\frac{1}{16}Cc\ Pp$ (white)	$\frac{1}{16}cc\ PP$ (white)	$\frac{1}{16}cc\ Pp$ (white)
$\frac{1}{4}cp$	$\frac{1}{16}Cc\ Pp$ (white)	$\frac{1}{16}Cc\ pp$ (white)	$\frac{1}{16}cc\ Pp$ (white)	$\frac{1}{16}cc\ pp$ (white)

The probability that the offspring has purple flowers is 1/16.

Question 2: Problem 2.8 from the textbook. **[6]**

Let B be the event that the student got an A+ in Biology and let C be the event that the student got an A+ in Chemistry. We have $P(B) = 375/1150$, $P(C) = 50/1150$ and $P(B \cap C) = 45/1150$.

(a) We want $P(B \cap C') = P(B) - P(B \cap C) = 375/1150 - 45/1150 = 330/1150 = 0.2870$.
(b) We want $P(C \cap B') = P(C) - P(B \cap C) = 50/1150 - 45/1150 = 5/1150 = 0.0043$.
(c) We want $P(B' \cup C') = 1 - P(B \cap C) = 1 - 45/1150 = 1105/1150 = 0.9609$.
(d) We want $P(B \cup C) = P(B) + P(C) - P(B \cap C) = 380/1150 = 0.3304$.
(e) We want $P(B' \cap C') = 1 - P(B \cup B) = 1 - 380/1150 = 770/1150 = 0.6696$.

Question 3: Problem 3.2 from the textbook. **[4]**

We denote by D the event that a patient has sleep apnea, by A the event that the patient has symptom A and by B the event that the patient has symptom B. The four events $G_1 = A \cap B$, $G_2 = A \cap B'$, $G_3 = A' \cap B$ and $G_4 = A' \cap B'$ form a partition of the sample space S. We know that

$$P(G_1) = 0.56, \quad P(G_2) = 0.21, \quad P(G_3) = 0.19, \quad P(G_4) = 0.04$$

and

$$P(D|G_1) = 0.60, \quad P(D|G_2) = 0.45, \quad P(D|G_3) = 0.35, \quad P(D|G_4) = 0.03.$$

(a) By the total probability rule,

$$\begin{aligned} P(D) &= P(D|G_1)P(G_1) + P(D|G_2)P(G_2) + P(D|G_3)P(G_3) + P(D|G_4)P(G_4) \\ &= (0.60)(0.56) + (0.45)(0.21) + (0.35)(0.19) + (0.03)(0.04) = 0.4982 \end{aligned}$$

(b) We have to calculate $P(A|D)$. Note that

$$P(A|D) = P(G_1|D) + P(G_2|D).$$

By Bayes' rule,

$$P(G_1|D) = \frac{P(D|G_1)P(G_1)}{P(D)} = \frac{(0.60)(0.56)}{0.4982} = 0.6744$$

and

$$P(G_2|D) = \frac{P(D|G_2)P(G_2)}{P(D)} = \frac{(0.45)(0.21)}{0.4982} = 0.1896.$$

Hence, $P(A|D) = 0.6744 + 0.1896 = 0.864$.

Question 4: Let A="has fatty liver syndrome" and B="suffers from cage layer fatigue". We have **[5]**

$$P(A) = 0.04, \quad P(B) = 0.03, \quad \text{and} \quad P(A \cap B') + P(B \cap A') = 0.04.$$

(a) Since $0.04 = P(A) = P(A \cap B) + P(A \cap B')$ and $0.03 = P(B) = P(B \cap A) + P(B \cap A')$, then

$$0.07 = P(A) + P(B) = 2\,P(A \cap B) + P(A \cap B') + P(B \cap A').$$

Isolating $P(A \cap B)$ gives

$$P(A \cap B) = \frac{0.07 - [P(A \cap B') + P(B \cap A')]}{2} = \frac{0.07 - 0.04}{2} = 0.015,$$

since $P(A \cap B') + P(B \cap A') = 0.04$.

(b) The probability of have at least one of the conditions is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.04 + 0.03 - 0.015 = 0.055.$$

Thus, the probability that it has none of the conditions is

$$P(A' \cap B') = 1 - P(A \cup B) = 1 - 0.055 = 0.945.$$

Question 5: Let D be the event that the person has the disease and $+$ be the event that the diagnostic test is positive. We have **[4]**

$$P(+|D) = 0.99, \quad P(-|D') = 0.97, \quad P(D) = 0.015.$$

The probability that a random person from this population will get a positive result is (by the total probability rule):

$$\begin{aligned} P(+) &= P(+|D)P(D) + P(+|D')P(D') \\ &= (0.99)(0.015) + (0.03)(0.815) = 0.0393. \end{aligned}$$

Thus, the positive predictive value is

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{P(+|D)P(D)}{P(+)} = \frac{(0.99)(0.015)}{0.0393} = 0.3779.$$

Question 6: Let D="Down Syndrome is present" and $T+$="test is positive". **[4]**

(a) The false positive rate is $P(+|D') = 55/992 = 0.0554$ and the false negative rate is $P(-|D) = 3/8 = 0.375$.

(b) The sensitivity is $P(+|D) = 5/8 = 0.625$ and the specificity of the test is $P(-|D') = 937/992 = 0.9446$.

Marking Scheme:
Question 1:

- 1 point for giving the prob. of 1/16 in part (b)

Question 2:

- 1 point for properly identifying the given information, i.e. defining events and displaying the given probabilities.
- 1 points for each part. It is okay if the students used a Venn diagram instead of the rules.

Question 3:

- 1 point for properly identifying the given information, i.e. defining events and displaying the given probabilities.

(a) 1 point for using the total prob. rule; 1/2 for final answer of 0.4982.

(b) 1 point for noticing that $P(G_1|D) = P(D|G_1)P(G_1)/P(D)$.

(b) 1 point for noticing that $P(A|D) = P(G_1|D) + P(G_2|D)$.

(b) 1/2 point for final answer $P(A|D) = 0.864$.

Question 4:

- 1 point for properly identifying the given information, i.e. defining events and displaying the given probabilities.
- 2 points for part (a)
- 1 point for using $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ in (b).
- 1 point for using $P(A' \cap B') = 1 - P(A \cup B)$

Question 5:

- 1 point for properly identifying the given information, i.e. defining events and displaying the given probabilities.
- 1 point for using the total prob. rule to compute $P(+)$

- 1 point for noticing that PPV = $P(+|D)P(D)/P(+)$.
- 1 point for the final answer 0.3779

Question 6:

- 1 point for each of the four measures.