STT465 HW1

Candy Li

2018/9/10

Question 1

Since $\theta = 0.9$, the genotype frequencies in population I are

$$P(AA) = \theta^2 = 0.81$$

$$P(AB) = 2\theta(1-\theta) = 0.18$$

$$P(BB) = (1-\theta)^2 = 0.01$$

Thus,

$$P(D_i = 1, G_i = AA) = P(D_i = 1|G_i = AA) * P(AA) = 0$$

$$P(D_i = 1, G_i = AB) = P(D_i = 1|G_i = AB) * P(AB) = 0.018$$

$$P(D_i = 1, G_i = BB) = P(D_i = 1|G_i = BB) * P(BB) = 0.009$$

Therefore, the prevalence of disease $P(D_i = 1) = \sum P(D_i = 1, G_i) = 0.027$ in population I.

Question 2

Given that an individual in population I has developed the disease, the probability that the genotype of that individual is AA is

$$P(G_i = AA|D_i = 1) = \frac{P(D_i = 1, G_i = AA)}{P(D_i = 1)} = 0$$

Similarly, the probability that the genotype of that individual is AB is

$$P(G_i = AB|D_i = 1) = \frac{P(D_i = 1, G_i = AB)}{P(D_i = 1)} = 0.667$$

and the probability that the genotype of that individual is BB is

$$P(G_i = BB|D_i = 1) = \frac{P(D_i = 1, G_i = BB)}{P(D_i = 1)} = 0.333$$

Question 3

Since $\theta = 0.95$, the genotype frequencies in population II are

$$P(AA) = \theta^2 = 0.9025$$

$$P(AB) = 2\theta(1-\theta) = 0.095$$

$$P(BB) = (1-\theta)^2 = 0.0025$$

Thus,

$$P(D_i = 1, G_i = AA) = P(D_i = 1|G_i = AA) * P(AA) = 0$$

$$P(D_i = 1, G_i = AB) = P(D_i = 1|G_i = AB) * P(AB) = 0.0095$$

$$P(D_i = 1, G_i = BB) = P(D_i = 1|G_i = BB) * P(BB) = 0.00225$$

Therefore, $P(D_i = 1) = \sum P(D_i = 1, G_i) = 0.01175$ in population II.

Since individuals are equally likely to come from population 1 or 2, $P(Population = 1) = P(Population = 2) = 0.5$.

Since $P(D_i = 1, Population = 1) = 0.027$ and $P(D_i = 1, Population = 2) = 0.01175$,

$$P(D_i = 0, Population = 1) = P(Population = 1) - P(D_i = 1, Population = 1) = 0.473$$

$$P(D_i = 0, Population = 2) = P(Population = 2) - P(D_i = 1, Population = 2) = 0.48825$$

Thus, $P(D_i = 0) = P(D_i = 0, Population = 1) + P(D_i = 0, Population = 2) = 0.96125$.

Therefore, given that an individual is healthy, the probability that this individual comes from population I is

$$P(Population = 1|D_i = 0) = \frac{P(D_i = 0, Population = 1)}{P(D_i = 0)} = 0.4921$$

Question 4

	$Y = 0$	$Y = 1$	$P(X)$
$X = 0$	0.16	0.24	0.4
$X = 1$	0.24	0.36	0.6
$P(Y)$	0.4	0.6	1

Since $P(X, Y) = P(X)P(Y)$, the two Bernoulli random variables above are independent.

Since $P(X = 0, Y = 1) = P(X = 1, Y = 0)$, the two Bernoulli random variables above are exchangeable, which implies that they are identically distributed.

Thus, they satisfy IID.

	$Y = 0$	$Y = 1$	$P(X)$
$X = 0$	0.04	0.36	0.4
$X = 1$	0.36	0.24	0.6
$P(Y)$	0.4	0.6	1

Since $P(X, Y) \neq P(X)P(Y)$, the two Bernoulli random variables above are not independent.

Since $P(X = 0, Y = 1) = P(X = 1, Y = 0)$, the two Bernoulli random variables above are exchangeable.

Thus, they satisfy exchangeability but not IID.

	$Y = 0$	$Y = 1$	$P(X)$
$X = 0$	0.08	0.18	0.26
$X = 1$	0.32	0.42	0.74
$P(Y)$	0.4	0.6	1

Since $P(X, Y) \neq P(X)P(Y)$, the two Bernoulli random variables above are not independent.

Since $P(X = 0, Y = 1) \neq P(X = 1, Y = 0)$, the two Bernoulli random variables above are not exchangeable.

Thus, they do not satisfy IID and are not exchangeable.

Question 5

a)

Since $P(Z=1)=0.6$, $P(Z=0)=1-P(Z=1)=0.4$.

Let $X=0$, $Y=0$, and $Z=0$, then

$$P(X=0,Y=0|Z=0)=\frac{P(X=0,Y=0)P(Z=0)}{P(Z=0)}=\frac{0.06*0.4}{0.4}=0.06$$

$$P(X=0|Z=0)=\frac{P(X=0)P(Z=0)}{P(Z=0)}=\frac{0.3*0.4}{0.4}=0.3$$

and

$$P(Y=0|Z=0)=\frac{P(Y=0)P(Z=0)}{P(Z=0)}=\frac{0.2*0.4}{0.4}=0.2$$

Since $P(X,Y|Z)=P(X|Z)P(Y|Z)$, (X,Y) are conditionally independent.

b)

Let $X=0$ and $Y=0$, then

$$P(X=0)=P(X=0|Z)P(Z)=(0.06+0.24)*0.4+(0.12+0.28)*0.6=0.36$$

$$P(Y=0)=P(Y=0|Z)P(Z)=(0.06+0.14)*0.4+(0.12+0.18)*0.6=0.26$$

and

$$P(X=0,Y=0)=P(X=0,Y=0|Z)P(Z)=0.06*0.4+0.12*0.6=0.096$$

Since $P(X)P(Y)\neq P(X,Y)$, (X,Y) are not independent.

c)

Let $X=0$ and $Y=1$, then

$$P(X=0,Y=1)=P(X=0,Y=1|Z)P(Z)=0.24*0.4+0.28*0.6=0.264$$

$$P(X=1,Y=0)=P(X=1,Y=0|Z)P(Z)=0.14*0.4+0.18*0.6=0.164$$

Since $P(X=0,Y=1)\neq P(X=1,Y=0)$, (X,Y) are not exchangeable.