

Homework 1

Chengbo Gu

Problem 1

(a) Let A denotes the event that none has high blood lead level in a randomly selected group of 16 children.

$$\text{Then } P(A) = \left(1 - \frac{1}{8}\right)^{16} = 0.1181$$

(b) Let B denotes the event that one child has high blood lead level in a randomly selected group of 16 children.

$$\text{Then } P(B) = C_{16}^1 \left(\frac{1}{8}\right) \times \left(\frac{7}{8}\right)^{15} = 0.2699$$

(c) Let C denotes the event that two children have high blood lead level in a randomly selected group of 16 children.

$$\text{Then } P(C) = C_{16}^2 \left(\frac{1}{8}\right)^2 \times \left(\frac{7}{8}\right)^{14} = 0.2891$$

(d) Let D denotes the event that three or more have high blood lead level in a randomly selected group of 16 children.

$$\text{Then } P(D) = 1 - P(A) - P(B) - P(C) = 0.3229$$

Problem 2

Let X denotes the height of corn plants, then $X \sim N(145, 22)$

$$(a) P(135 \leq X \leq 155) = P\left(\frac{135-145}{22} \leq \frac{X-145}{22} \leq \frac{155-145}{22}\right) = P\left(-\frac{5}{11} \leq Z \leq \frac{5}{11}\right) = P\left(Z \leq \frac{5}{11}\right) - P\left(Z \leq -\frac{5}{11}\right)$$

$$\text{Because } \text{pnorm}(5/11) = 0.6753 \text{ and } \text{pnorm}(-5/11) = 0.3247,$$

$$P(135 \leq X \leq 155) = 0.6753 - 0.3247 = 0.3506$$

$$(b) \text{ According to central limit theory, } \sum_{k=1}^{16} X_k \sim N(16 \times 145, \sqrt{16} \times 22).$$

$$\text{Thus, } \frac{\sum_{k=1}^{16} X_k}{16} \sim N\left(145, \frac{22}{\sqrt{16}}\right) \Rightarrow \bar{X} \sim N(145, 5.5) \text{ for a random sample of 16 plants.}$$

$$P(135 \leq \bar{X} \leq 155) = P\left(\frac{135-145}{5.5} \leq \frac{\bar{X}-145}{5.5} \leq \frac{155-145}{5.5}\right) = P\left(-\frac{20}{11} \leq Z \leq \frac{20}{11}\right) = P\left(Z \leq \frac{20}{11}\right) - P\left(Z \leq -\frac{20}{11}\right)$$

$$\text{Because } \text{pnorm}(20/11) = 0.9655 \text{ and } \text{pnorm}(-20/11) = 0.0345,$$

$$P(135 \leq \bar{X} \leq 155) = 0.9655 - 0.0345 = 0.9310$$

$$(c) \text{ According to central limit theory, } \sum_{k=1}^{32} X_k \sim N(32 \times 145, \sqrt{32} \times 22).$$

Thus, $\frac{\sum_{k=1}^{32} X_k}{32} \sim N(145, \frac{22}{\sqrt{32}}) \Rightarrow \bar{X} \sim N(145, \frac{22}{\sqrt{32}})$ for a random sample of 32 plants.

$$P(135 \leq \bar{X} \leq 155) = P\left(\frac{135-145}{22/\sqrt{32}} \leq \frac{\bar{X}-145}{22/\sqrt{32}} \leq \frac{155-145}{22/\sqrt{32}}\right) = P(Z \leq 2.5713) - P(Z \leq -2.5713)$$

$$P(135 \leq \bar{X} \leq 155) = \text{pnorm}(2.5713) - \text{pnorm}(-2.5713) = 0.9899$$

Problem 3

Let A denotes the event that individuals preferred to use right hand to write.

Let B denotes the event that individuals preferred to use right foot to kick a ball.

Then $P(B) = \frac{2012+121}{2391} = 0.8921$

$$P(B|A) = \frac{2012}{2012+142} = 0.9341$$

Since $P(B) \neq P(B|A)$, hand and foot preferences are not independent.

Exercise 2.1

Let G denotes a child is a girl.

Let B denotes a child is a boy.

Let A denotes that at least one child is boy.

Then we have $P(GG) = P(BB) = P(GB) = P(BG) = \frac{1}{4}$

(a) The question actually asks what is $P(GB|A) + P(BG|A)$.

Using Bayes rules, we have

$$P(GB|A) + P(BG|A) = 1 - P(BB|A) = 1 - \frac{P(A|BB) \times P(BB)}{P(A)} = 1 - \frac{1/4}{3/4} = \frac{2}{3}$$

(b) Let b denotes the event that I happen to see one of his children and it is a boy.

The question actually asks what is $P(GB|b) + P(BG|b)$.

Using Bayes rules, we have

$$P(GB|b) + P(BG|b) = \frac{P(b|GB) \times P(GB)}{P(b)} + \frac{P(b|BG) \times P(BG)}{P(b)} = \frac{1/2 \times 1/4}{1/2} + \frac{1/2 \times 1/4}{1/2} = \frac{1}{2}$$

Exercise 2.2

Exercise 2.4

Let A denotes the event that one has this disease.

Let B denotes the event that the test is positive.

Then we have $P(B|A) = P(\bar{B}|\bar{A}) = 0.99$ and $P(A) = 1/10000$.

We want to infer $P(A|B)$.

Using Bayes rules,

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} = \frac{P(A) \times P(B|A)}{P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})} = \frac{1/10000 \times 0.99}{1/10000 \times 0.99 + 9999/10000 \times 0.01} = 0.98\%$$

Exercise 2.5

Let A denotes the event that the prize was behind door 1.

Let B denotes the event that the prize was behind door 2.

Let C denotes the event that the prize was behind door 3.

Let c denotes the event that I choose door 1 and the host open door 3.

Then we have $P(A) = P(B) = P(C) = \frac{1}{3}$, $P(c|A) = \frac{1}{2}$, $P(c|B) = 1$ and $P(c|C) = 0$.

$$\begin{aligned} \text{Thus, } P(A|c) &= \frac{P(A) \times P(c|A)}{P(c)} = \frac{P(A) \times P(c|A)}{P(A) \times P(c|A) + P(B) \times P(c|B) + P(C) \times P(c|C)} \\ &= \frac{1/3 \times 1/2}{1/3 \times 1/2 + 1/3 \times 1 + 1/3 \times 0} = \frac{1}{3} \\ P(B|c) &= 1 - P(A|c) = \frac{2}{3} \end{aligned}$$

So, it is better to switch to door 2.

Exercise 2.12

$$\begin{aligned} I(X, Y) &= \sum_x \sum_y P(x, y) \log \frac{P(x, y)}{p(x)p(y)} = \sum_x \sum_y P(x, y) \log \frac{P(x, y)}{p(x)} - \sum_x \sum_y P(x, y) \log P(y) \\ &= \sum_x P(x) \left(\sum_y P(y|x) \log P(y|x) \right) - \sum_y \log P(y) \left(\sum_x P(x, y) \right) \\ &= -\sum_x P(x) H(Y|X=x) - \sum_y \log P(y) P(y) \\ &= -H(Y|X) + H(Y) = H(Y) - H(Y|X) \end{aligned}$$

Similarly,

$$\begin{aligned} I(X, Y) &= \sum_x \sum_y P(x, y) \log \frac{P(x, y)}{p(x)p(y)} = \sum_x \sum_y P(x, y) \log \frac{P(x, y)}{p(y)} - \sum_x \sum_y P(x, y) \log P(x) \\ &= \sum_y P(y) \left(\sum_x P(x|y) \log P(x|y) \right) - \sum_x \log P(x) \left(\sum_y P(x, y) \right) \\ &= -\sum_y P(y) H(X|Y=y) - \sum_x \log P(x) P(x) \\ &= -H(X|Y) + H(X) = H(X) - H(X|Y) \end{aligned}$$