Problem 1

(a) Choose (2) and (4)
(1) Median is tough for changing of a few very small or large values.
(3) The scatterplots can be only used to see correlation, correlation is weaker than independence.
(4) Because the sample variance converges to population variance, which means we can estimate variance by $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(b) Choose (3) and (4)
(1) It should be density histogram, when bin converge to 0
(2) It should be 0.

(c) Choose (2) and (3)
(1) Independence and Disjoint are totally different concepts!
(2) By Bayes Formula, $P(A|B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$
(3) It's just the definition of independence
(4) This formula can be used for only disjoint events.

Problem 2

(a) 1.75, since the area of each bar represents the proportion then the conclusion follows from the definition of median

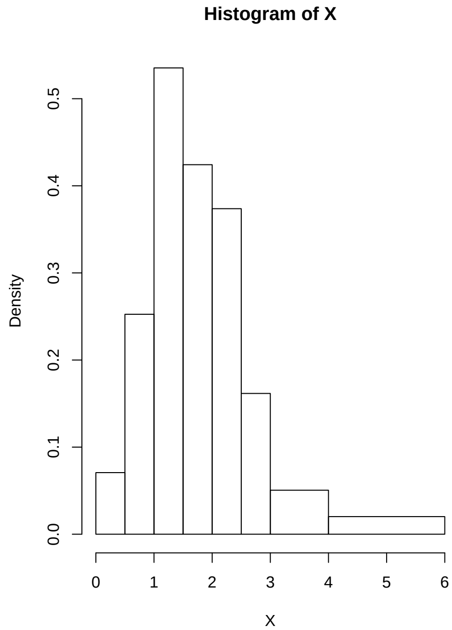
(b) 2, since the distribution is slightly right skewed

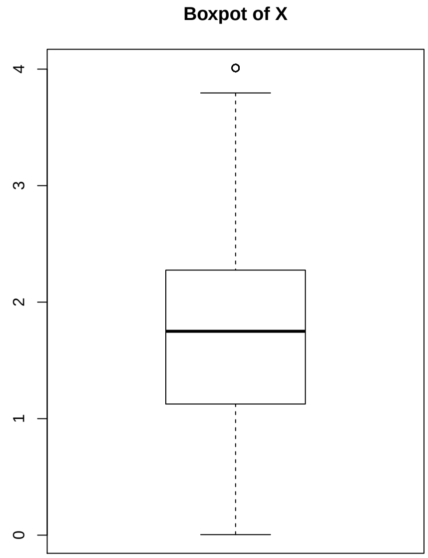
(c) 66%, since $53\% \cdot 0.5 + 42\% \cdot 0.5 + 37\% \cdot 0.5 = 66\%$

(d) any number between 2 and 2.5 can be the cut-off point, which is the third quantile Q_3, since the proportion between 1.75 and 2 is $42\% \cdot 0.25 = 10.5\% < 25\%$ and the proportion between 1.75 and 2.5 is $10.5\% + 37\% \cdot 0.5 = 29\% > 25\%$

(e) The following box plot is an example.

From part(a), we know the median is 1.75. And we get the first quantile $Q_1 \in [1, 1.5]$, since the proportion between 1.5 and 1.75 is $10.5\% < 25\%$ and the proportion between 1 and 1.75 is $10.5\% + 53\% \cdot 0.5 = 28.2\% > 25\%$. From part(d), we get the third quantile $Q_3 \in [2, 2.5]$. Hence the interquartile range $IQR = Q_3 - Q_1 \in [0.5, 1.5]$ and $Q_3 + 1.5 \cdot IQR \in [2.75, 4.75]$ and





$Q_1 - 1.5 \cdot IQR \in [-1.25, 0.75]$. Note that the whiskers draw at the last observation that is not an outlier. Hence the lower whisker is $\in [0, 0.75]$ and the upper whisker is $\in [2.75, 4)$ but not 4. And 4 is the only outlier.

`Problem 3` Let X_1, X_2, X_3 be the events that the product came from line 1, 2, or 3 respectively, and let D be the event that the product selected is defective.

(a) We are asked to find $P(D)$. By laws of total probability, we have (because X_1, X_2, X_3 are mutually exclusive and partition the whole set of outcomes)

$$\begin{aligned} P(D) &= P(D|X_1) \cdot P(X_1) + P(D|X_2) \cdot P(X_2) + P(D|X_3) \cdot P(X_3) \\ &= (0.05)(0.5) + (0.1)(0.25) + (0.08)(0.25) = .07. \end{aligned}$$

(b) We are asked now to find $P(X_1|D)$, we use Bayes' Rule:

$$\begin{aligned} P(X_1|D) &= \frac{P(X_1 \cap D)}{P(D)} = \frac{P(X_1 \cap D)}{P(D)} \frac{P(X_1)}{P(X_1)} = P(D|X_1) \frac{P(X_1)}{P(D)} \\ &= (0.05) \cdot \frac{0.50}{0.07} \approx 0.357. \end{aligned}$$

`Problem 4`

(a) We have three 0s and two 4s. If we draw two numbers without replacement, the only possible values their sum could be are $0 + 0 = 0$, $0 + 4 = 4$, or

$4+4=8$. To get a 0, we have to draw two 0s. There's $\binom{3}{2}=3$ ways to do that. To get a 4, we have to draw a 0 and a 4. There are $3\times 2=6$ ways to do that. To get an 8, we have to draw two 4s. There's only 1 way to do that.

There's a total of $\binom{5}{2}=10$ ways to draw two numbers without replacement. Hence, $f_X(0)=\frac{3}{10}, f_X(4)=\frac{6}{10}, f_X(8)=\frac{1}{10}$.

(b) $E(X)=\sum kf_X(k)=0\times\frac{3}{10}+4\times\frac{6}{10}+8\times\frac{1}{10}=\frac{32}{10}$.
var(X) $=E(X^2)-E(X)^2$.
$E(X^2)=\sum k^2f_X(k)=0\times\frac{3}{10}+16\times\frac{6}{10}+64\times\frac{1}{10}=\frac{160}{10}=16$.
Hence, var(X) $=16-3.2^2$.

(c) X+Y can take on only a few values. Denote our draw as (a,b). Then $X+Y=(a+b)+a=2a+b$. What can (a,b) be? Either (0,0), (0,4), (4,0), or (4,4). So the values X+Y can take on are 0, 4, 8, 12. Using the Multiplication Principle, we can compute the probabilities of drawing any of those pairs.
$P(0,0)=\frac{3\times 2}{20}$. $P(0,4)=\frac{3\times 2}{20}$. $P(4,0)=\frac{2\times 3}{20}$. $P(4,4)=\frac{2\times 1}{20}$.
So $f_W(0)=\frac{3}{10}, f_W(4)=\frac{3}{10}, f_W(8)=\frac{3}{10}, f_W(12)=\frac{1}{10}$, where W = X+Y.

(d) $(Y-2)^2$ can be $(0-2)^2$ or $(4-2)^2$. But those are both just 4! So $X+(Y-2)^2$ is just X+4. The values X+4 can take are just the values of X shifted up by 4; the probabilities remain the same, since $P(X+4=k)=P(X=k-4)$. So, $f_W(4)=\frac{3}{10}, f_W(8)=\frac{6}{10}, f_W(12)=\frac{1}{10}$, where W = $X+(Y-2)^2$.

`Challenge Problem`

Because then density function is symmetric, which means:

$$-\int_{-\infty}^{0}\frac{x}{\pi(1+x^2)}dx=\int_{0}^{\infty}\frac{x}{\pi(1+x^2)}dx$$

Thus,

$$EX=\int_{-\infty}^{\infty}\frac{x}{\pi(1+x^2}dx=0$$

And for variance, we have:

$$\begin{aligned}
Var(X) =& E[X]^2 \\
&= \int_{-\infty}^{\infty} \frac{x^2}{\pi(1+x^2)} dx \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{x^2}{1+x^2} dx \\
&= \frac{2}{\pi} \int_0^{1} \frac{x^2}{1+x^2} dx + \frac{2}{\pi} \int_1^{\infty} \frac{x^2}{1+x^2} dx \\
&\geq \frac{2}{\pi} \int_0^{1} \left(1 - \frac{1}{1+x^2}\right) dx + \frac{2}{\pi} \int_1^{\infty} \frac{x}{1+x^2} dx
\end{aligned}$$

Where the first term is a constant, and the second term is

$$\int_1^{\infty} \frac{2x}{1+x^2} dx = \int_1^{\infty} \frac{1}{1+x^2} d(x^2) = \ln(1+x^2)\Big|_1^{\infty} = \infty$$

Thus the variance is infinity, which mean variance doesn't exist.