2. (a) The marginal PMF of X is $p_X(0.0) = 0.388 + 0.009 + 0.003 = 0.4$, $p_X(1.0) = 0.485 + 0.010 + 0.005 = 0.5$, $p_X(2.0) = 0.090 + 0.006 + 0.004 = 0.1$.

   The marginal PMF of Y is $p_Y(0) = 0.388 + 0.485 + 0.090 = 0.963$, $p_Y(1) = 0.009 + 0.010 + 0.006 = 0.025$, $p_Y(2) = 0.003 + 0.005 + 0.004 = 0.012$.

   (b) (i) The probability that a randomly selected rat has one tumor is $P(Y = 1) = p_Y(1) = 0.025$.

   (ii) The probability that a randomly selected rat has at least one tumor is $P(Y \geq 1) = p_Y(1) + p_Y(2) = 0.037$.

   (c) (i) This is the conditional probability that $P(Y = 0|X = 1.0) = P(X = 1.0, Y = 0)/P(X = 1.0) = 0.485/0.5 = 0.97$.

   (ii) This is the conditional probability that $P(Y > 0|X = 1.0) = 1 - P(Y = 0|X = 1.0) = 0.03$.

4. (a) The table for the CDF is

			y	
	$F(x,y)$	1	2	3
	1	0.09	0.21	0.34
x	2	0.21	0.44	0.68
	3	0.34	0.67	1

   (b) In this problem $F_X(x) = F(x, \infty) = F(x, 3)$, so $F_X(1) = 0.34$, $F_X(2) = 0.68$, $F_X(3) = 1$. By the same reason, we have $F_Y(1) = 0.34$, $F_Y(2) = 0.67$, $F_Y(3) = 1$.

   (c) $P(X = 2, Y = 2) = F(2, 2) - F(2, 1) - F(1, 2) + F(1, 1) = 0.44 - 0.21 - 0.21 + 0.09 = 0.11$, as shown in the table of Exercise 1.

7. (a) $E(Y|X=1) = 1p_{Y|X=1}(1)+2p_{Y|X=1}(2) = 1\times0.66+2\times0.34 = 1.34$, $E(Y^2|X=1) = 1^2p_{Y|X=1}(1)+2^2p_{Y|X=1}(2) = 1\times0.66+2\times0.34 = 2.02$, thus $\text{Var}(Y|X=1) = E(Y^2|X=1) - E(Y|X=1)^2 = 2.02 - 1.34^2 = 0.2244$.

   (b) The table for the joint PMF is

		y	
	$P(x,y)$	1	2
	1	0.132	0.068
x	2	0.24	0.06
	3	0.33	0.17

   (c) This problem is asking for $P(Y=1) = p_Y(1) = 0.132+0.24+0.33 = 0.702$.

   (d) This is asking for $P(X=1|Y=1) = P(X=1, Y=1)/P(Y=1) = 0.132/0.702 = 0.188$.

8. (a) The regression function of Y on X is $\mu_{Y|X}(x) = E(Y|X=x) = 1p_{Y|X=x}(1) + 2p_{Y|X=x}(2)$. Thus, we have $E(Y|X=0) = 1.34$, $E(Y|X=1) = 1.2$, $E(Y|X=2) = 1.34$.

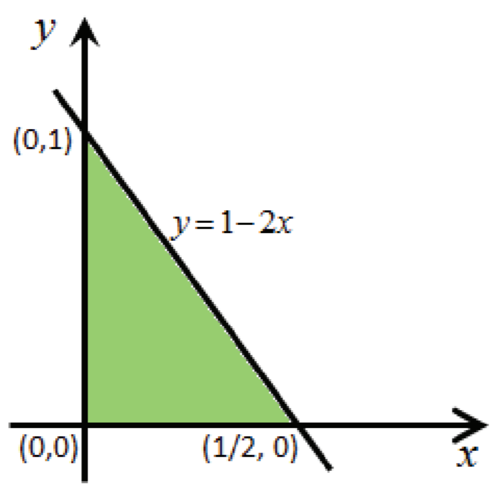
   (b) By the law of total expectation, $E(Y) = E(Y|X=0)p_X(0) + E(Y|X=1)p_X(1) + E(Y|X=2)p_X(2=) = 1.34\times2+1.2\times0.3+1.34\times0.5 = 1.298$.

12. (a) Independent

    (b) Not independent

    (c) Not independent

17. (a) The support of the joint PDF is given on the following page.



(b) The support of the joint PDF is not a rectangle, thus X and Y are not independent.

(c) We have

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{1-2x} 24xdy = 24x(1-2x), \quad \text{for} \quad 0 \le x \le 0.5,$$

and $f_X(x) = 0$ otherwise.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{(1-y)/2} 24xdx = 12x^2 \mid_0^{(1-y)/2} = 3(1-y)^2,$$

for $0 \le y \le 1$, and $f_Y(y) = 0$ otherwise. Thus,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^{0.5} 24x^2(1-2x)dx = \int_0^1 3t^2(1-t)dt \\ &= 3\frac{2!1!}{4!} = \frac{1}{4}, \end{aligned}$$

and

$$E(Y) = \int_{-\infty}^{\infty} yf_Y(y)dy = \int_0^1 3y(1-y)^2dy = 3\frac{2!1!}{4!} = \frac{1}{4}.$$

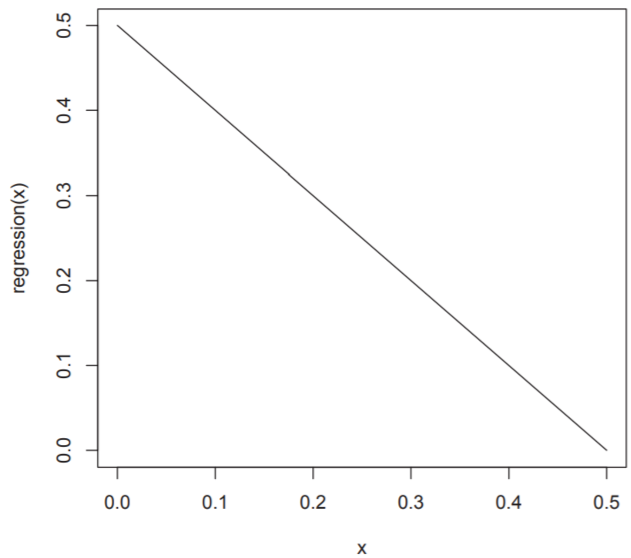
18. (a) The conditional PDF of Y given $X = x$ is

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{24x}{24x(1-2x)} = \frac{1}{1-2x},$$

for $0 \le y \le 1-2x$, and $f_{Y|X=x}(y) = 0$ otherwise. The regression function is

$$E(Y|X=x) = \int_{-\infty}^{\infty} yf_{Y|X=x}(y)dy = \int_0^{1-2x} y\frac{1}{1-2x}dy = \frac{1-2x}{2} = \frac{1}{2} - x.$$

The plot for the regression function is given as follows:



By the expression, $E(Y|X=0.3) = 0.5 - 0.3 = 0.2$.

(b) By the law of total expectation,

$$E(Y) = \int_{-\infty}^{\infty} E(Y|X=x) f_X(x)dx = \int_0^{0.5} \frac{1-2x}{2} 24x(1-2x)dx$$
$$= \int_0^{0.5} 12x(1-2x)^2 dx = \int_0^1 3t(1-t)^2 dt = 3\frac{2!1!}{4!} = \frac{1}{4}.$$

6. The number of injuries in a month is $X_1 + X_2 + \cdots + X_N$ and $\mu = E(X_i) = 1.5$. Thus the expected number to injuries in a month is $E(X_1 + X_2 + \cdots + X_N) = E(N)\mu = 7 \times 1.5 = 10.5$.

7. The total tips is $T = X_1 + \cdots + X_{N_1} + Y_1 + \cdots + Y_{N_2}$, and we know that $\mu_1 = E(X_i) = 20$, $\mu_2 = E(Y_j) = 10$, N_1 is Poisson(4), N_2 is Poisson(6). Thus, the expected value of the total amount of tips is

$$E(T) = E(X_1 + \cdots + X_{N_1}) + E(Y_1 + \cdots + Y_{N_2}) = E(N_1)\mu_1 + E(N_2)\mu_2$$
$$= 4 \times 20 + 6 \times 10 = 140.$$

11. Similar to the previous exercise,

$$\begin{aligned}E(8X+10Y) = \sum_x\sum_y (8x+10y)p(x,y) &= (8\times 0+10\times 0)\times 0.06\\
&+(8\times 0+10\times 1)\times 0.04+(8\times 0+10\times 2)\times 0.2\\
&+(8\times 1+10\times 0)\times 0.08+(8\times 1+10\times 1)\times 0.3\\
&+(8\times 1+10\times 2)\times 0.06+(8\times 2+10\times 0)\times 0.1\\
&+(8\times 2+10\times 1)\times 0.14+(8\times 2+10\times 2)\times 0.02\\
&=18.08\end{aligned}$$

and

$$\begin{aligned}E[(8X+10Y)^2] = \sum_x\sum_y (8x+10y)^2p(x,y) &= (8\times 0+10\times 0)^2\times 0.06\\
&+(8\times 0+10\times 1)^2\times 0.04+(8\times 0+10\times 2)^2\times 0.2\\
&+(8\times 1+10\times 0)^2\times 0.08+(8\times 1+10\times 1)^2\times 0.3\\
&+(8\times 1+10\times 2)^2\times 0.06+(8\times 2+10\times 0)^2\times 0.1\\
&+(8\times 2+10\times 1)^2\times 0.14+(8\times 2+10\times 2)^2\times 0.02\\
&=379.52.\end{aligned}$$

Thus,

$$\mathrm{Var}(8X+10Y) = E[(8X+10Y)^2]-E(8X+10Y)^2 = 379.52-18.08^2 = 52.6336.$$

14. First we find

$$\begin{aligned}\mathrm{Cov}(\bar{X},\bar{Y}) = \mathrm{Cov}\left(\frac{X_1+X_2+X_3}{3},\frac{Y_1+Y_2+Y_3}{3}\right) &= \frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3}\mathrm{Cov}(X_i,Y_j)\\
=\frac{1}{9}[\mathrm{Cov}(X_1,Y_1)+\mathrm{Cov}(X_2,Y_2)+\mathrm{Cov}(X_3,Y_3)] &= \frac{5}{3},\end{aligned}$$

$\mathrm{Var}(\bar{X}) = \sigma_X^2/3 = 3$ and $\mathrm{Var}(\bar{Y}) = \sigma_Y^2/3 = 4/3$. Thus,

$$\mathrm{Var}(\bar{X}+\bar{Y}) = \mathrm{Var}(\bar{X})+\mathrm{Var}(\bar{Y})+2\mathrm{Cov}(\bar{X},\bar{Y}) = 3+\frac{4}{3}+2\frac{5}{3} = \frac{23}{3}.$$

16. From the model, $X_1, X_2, \cdots, X_r$ are independent and all of them have geometric distribution with success probability p. Thus, $E(X_i) = 1/p$ and $\mathrm{Var}(X_i) = (1-p)/p^2$. Since $X = X_1+\cdots+X_r$, we have

$$E(X) = E(X_1)+\cdots+E(X_r) = \frac{r}{p},$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_r) = \frac{r(1-p)}{p^2}.$$

1. We have the joint PMF and the marginal PMF as

	$P(x,y)$	y 0	1	2	$p_X(x)$
	0	0.06	0.04	0.2	0.3
x	1	0.08	0.3	0.06	0.44
	2	0.1	0.14	0.02	0.26
	$p_Y(y)$	0.24	0.48	0.28	

Thus,

$$\begin{aligned}
E(X) &= 0 \times 0.3 + 1 \times 0.44 + 2 \times 0.26 = 0.96, \\
E(X^2) &= 0^2 \times 0.3 + 1^2 \times 0.44 + 2^2 \times 0.26 = 1.48, \\
E(Y) &= 0 \times 0.24 + 1 \times 0.48 + 2 \times 0.28 = 1.04, \\
E(Y^2) &= 0^2 \times 0.24 + 1^2 \times 0.48 + 2^2 \times 0.28 = 1.6, \\
\sigma_X^2 &= E(X^2) - E(X)^2 = 1.48 - 0.96^2 = 0.5584, \\
\sigma_Y^2 &= E(Y^2) - E(Y)^2 = 1.6 - 1.04^2 = 0.5184, \\
E(XY) &= 1 \times 0.3 + 1 \times 2 \times 0.06 + 2 \times 1 \times 0.14 + 2 \times 2 \times 0.02 = 0.78,
\end{aligned}$$

and

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0.78 - 0.96 \times 1.04 = -0.2184.$$

Hence, the linear correlation coefficient of X and Y is

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{-0.2184}{\sqrt{0.5584}\sqrt{0.5184}} = -0.4059.$$

6. (a) The marginal distribution of X is Bernoulli(0.3).

   (b) If $X = 1$, that is, the first selection is defective, then there are 2 defective and 7 non-defective products left. Thus, $Y|X = 1 \sim$ Bernoulli(2/9). By same reason, $Y|X = 0 \sim$ Bernoulli(3/9).

   (c) $p_{X,Y}(1,1) = P(Y = 1|X = 1)P(X = 1) = 2/9 \times 0.3$, $p_{X,Y}(1,0) = P(Y = 0|X = 1)P(X = 1) = 7/9 \times 0.3$, $p_{X,Y}(0,1) = P(Y = 1|X = 0)P(X = 0) = 3/9 \times 0.7$, and $p_{X,Y}(0,0) = P(Y = 0|X = 0)P(X = 0) = 6/9 \times 0.7$.

(d) $p_Y(1) = p_{X,Y}(1,1) + p_{X,Y}(0,1) = 2/9 \times 0.3 + 3/9 \times 0.7 = 0.3$, thus the marginal distribution of Y is Bernoulli(0.3) , which is the same as X.

(e) From the joint distribution of X and Y, we have

$$E(XY) = \sum_{x=0}^{1}\sum_{y=0}^{2} xy p_{X,Y}(x,y) = p_{X,Y}(1,1) = 2/9 \times 0.3.$$

Thus,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{9} \times 0.3 - 0.3 \times 0.3 = -0.02333,$$

and the linear correlation coefficient is

$$r_{X,Y}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\frac{2}{9} \times 0.3 - 0.3 \times 0.3}{\sqrt{0.3 \times (1-0.3) \times 0.3 \times (1-0.3)}} = -\frac{1}{9}.$$

3. The proof is straightforward:

$$\begin{aligned} E(\hat{\sigma^2}) = E\left[\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right] &= \frac{(n_1-1)E(S_1^2) + (n_2-1)E(S_2^2)}{n_1+n_2-2} \\ = \frac{(n_1-1)\sigma^2 + (n_2-1)\sigma^2}{n_1+n_2-2} = \sigma^2. \end{aligned}$$

4. (a) The parameter of interest is the proportion of all credit card customers who had incurred an interest charge in the previous year due to an unpaid balance. The empirical estimator is the proportion in a sample of credit card customers who had incurred an interest charge in the previous year due to an unpaid balance. Using the provided information, we can get the estimate as $\hat{p} = 136/200 = 0.68$.

(b) Yes, it is unbiased.

(c) The estimated standard error is

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.68 \times (1-0.68)}{200}} = 0.033.$$

6. (a) $E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = E(X)/m - E(Y)/n = mp_1/m - np_2/n = p_1 - p_2$, thus $\hat{p}_1 - \hat{p}_2$ is unbiased estimator of $p_1 - p_2$.

   (b) The standard error of $\hat{p}_1 - \hat{p}_2$ is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2}} = \sqrt{\frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}}.$$

   The estimated standard error is

$$S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}.$$

   (c) From the data, we have $\hat{p}_1 = X/m = 70/100 = 0.7$ and $\hat{p}_2 = Y/n = 160/200 = 0.8$, thus the estimator for $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 = -0.1$ and the estimated standard error is

$$S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{0.7 \times (1-0.7)}{100} + \frac{0.8 \times (1-0.8)}{200}} = 0.054.$$

7. (a) The model-free estimation is 0.5.

   (b) Using the commands *x=c(2.08, 2.10, 1.81, 1.98, 1.91, 2.06); 1-pnorm(2.05, mean(x),sd(x))*, we get a model based estimation of 0.2979.