Valeria Duran

Math 6358

Homework 3

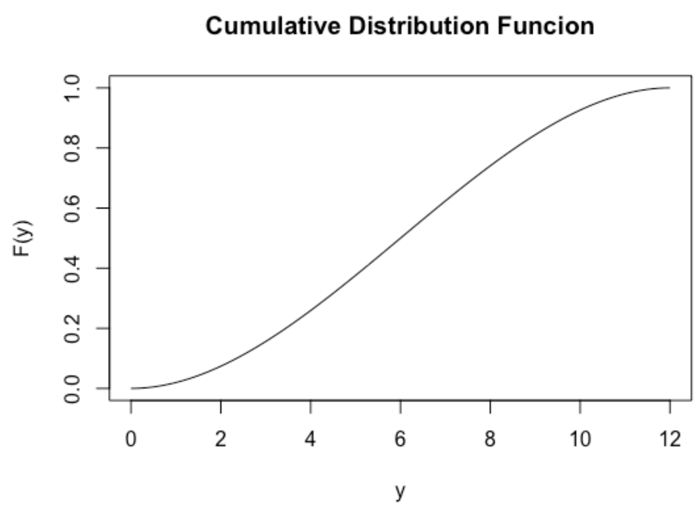
Problem 1.

A. The CDF is:

F(y) = { 0 for y <0

$\frac{1}{48}(y^2 - \frac{y^3}{18})$ for $0 \le y \le 12$

1 for y > 12}



B. P(Y ≤ 4) =

f <- function(y) {(y/24)*(1-(y/12))}

integrate(f, lower = 0, upper = 4)

0.2592593

P(Y>6) =

> integrate(f, lower = 6, upper = 12)

0.5

P(4 ≤ Y ≤ 6) =

> integrate(f, lower = 4, upper = 6)

0.2407407

C. E(Y)=

> e <- function(y) {(y/24)*(1-(y/12))*y}

> integrate(e, lower = 0, upper = 12)

6

$E(Y^2)$ =

> e <- function(y) {(y/24)*(1-(y/12))*y^2}

> integrate(e, lower = 0, upper = 12)

43.2

Var(Y) = $E(Y^2)$ - $E(Y)^2$

> 43.2 - 6^2

[1] 7.2

SD(X) = qrt(7.2)
[1] **2.683282**

D. P(|Y - μ| > 2)

= 1 - P(-2 < Y - 6 < 2)

= 1 - P(4 < Y < 8)

= > f <- function(y) {(y/24)*(1-(y/12))}

> integrate(f, lower = 4, upper = 8)

0.4814815 with absolute error < 5.3e-15

> 1 - 0.4814815

[1] **0.5185185**

E.

Y = bar length

Y - 12 = smaller bar length

Since E(Y) = 6;

0 < Y < 6 : y < 12 - Y —> y

Y = 6 : y = 12 - Y

6 < Y < 12 : y > 12 - Y —> y = 12 - Y

> d <- function(y) {(y/24)*(1-(y/12))*y}

> g <- function(y) {(y/24)*(1-(y/12))*(12-y)}

> integrate(d, lower = 0, upper = 6)

1.875 with absolute error < 2.1e-14

> integrate(g, lower = 6, upper = 12)

1.875 with absolute error < 2.1e-14

1.875 + 1.875 = **3.75**

Problem 2.

$$f(x) = \frac{1}{b-a} \quad E(x) = \frac{a+b}{2} \quad Var(x) = \frac{(b-a)^2}{12}$$

$$E(x) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2}{2} - \frac{a^2}{2} = \frac{b^2-a^2}{2(b-a)} = \boxed{\frac{b-a}{2}}$$

$$E(x^2) = \int_a^b \frac{x^2}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b^2+ba+a^2)(b-a)}{3(b-a)} = \frac{b^2+ba+a^2}{3}$$

$$Var(X) = E[X^2] - E(x)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b-a}{2}\right)^2 = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4}$$

$$= \frac{4a^2+4ab+4b^2}{12} - \frac{3a^2+6ab+3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \boxed{\frac{(b-a)^2}{12}}$$

Problem 3.

Normal distribution with mean = 137.2 and sd = 1.6. stated content is 135

A. > 1 - pnorm(135, 137.2, 1.6)

[1] **0.9154343**

B. P(X ≥ 8)

binomial distribution n=10; p = 0.9154

> 1 - pbinom(7, 10, 0.9154)

[1] **0.9537763**

C. > qnorm(.05)

[1] -1.644854

$135 = 137.2 - 1.644854\sigma$

$1.644854\sigma = 137.2 - 135$

> (137.2 - 135)/ 1.644854

[1] **1.337505**

Problem 4.

A. $E[X] = e^{\mu+\sigma^2/2}$

E[X] = 68.03

$Var(X) = e^{2\mu+\sigma^2} * (e^{\sigma^2} - 1)$

Var(X) = 14907.17

SD(X) = 122.09

B. $P(50 \leq X \leq 250)$ =

> plnorm(250, 3.5, 1.2) - plnorm(50, 3.5, 1.2)

[1] **0.319629**

C. plnorm(68.03, 3.5, 1.2)

[1] **0.7257327**

The probability is not 0.5 because the lognormal distribution is not symmetric.

Problem 5.

A. Marginal pmf of X:

For X = 20: 0.05 + 0.05 + 0.1

[1] **0.2**

For X = 25: 0.05 + 0.10 + 0.35

[1] **0.5**

For X = 30: 0 + 0.20 + 0.10

[1] **0.3**

Marginal pmf for Y:

For Y = 20: 0.05 + 0.05 + 0

[1] **0.1**

For Y = 25: 0.05 + 0.10 + 0.20

[1] **0.35**

For Y = 30: 0.10 + 0.35 + 0.10

[1] **0.55**

B. P(X ≤ 25 and Y ≤ 25) = P(X = 20 and Y = 20) + P(X = 20 and Y = 25) + P(X = 25 and Y = 20) + P(X = 25 and Y = 25) = 0.05 + 0.05 + 0.05 + 0.10

[1] **0.25**

C. For it to be independent, the probability distribution of X has to equal the conditional probability of P(X| Y=20):

> 0.05/0.1 #For X=20

[1] 0.5

> 0.05/0.1 #For X=25

[1] 0.5

> 0/0.1 #For X=30

[1] 0

Since the conditional probability and the probability distribution of X are not the same, **X and Y are not independent.**

D. E[X] =

> 20*0.2 + 25*0.5 + 30*0.3 #expected value of X

[1] 25.5

E[Y] =

```
> 20*0.1 + 25*0.35 + 30*0.55 #expected value of Y
[1] 27.25
```

E[X and Y] =

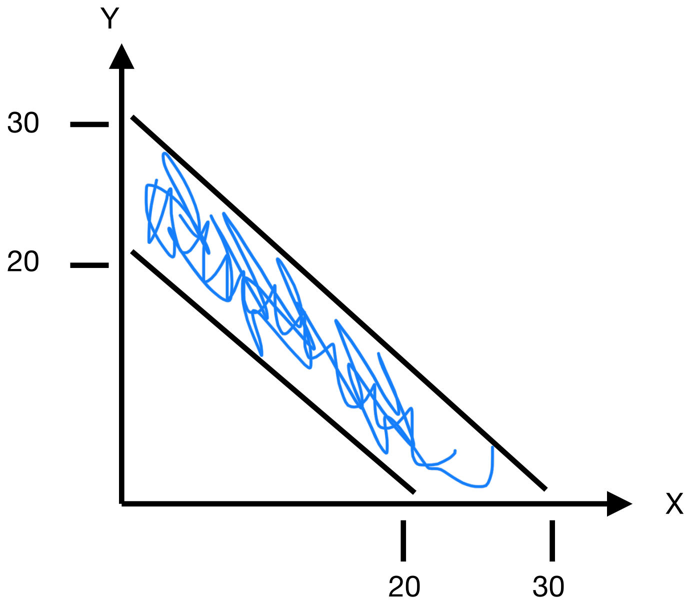
```
> 25.5 + 27.25 #expected value of X and Y
```

[1] **52.75**

E. E[Y-X] = E[Y] - E[X] = 27.25 - 25.5 = **1.75**

Problem 6.

a)



$$\int_0^{30}\int_0^{30-x} f(x,y)dydx - \int_0^{20}\int_0^{20-x} f(x,y)dydx = 1$$

$$\int_0^{30}\int_0^{30-x} kxy\,dy\,dx = \int_0^{30}\frac{k}{2}x\left[y^2\right]_0^{30-x}dx = \frac{k}{2}\int_0^{30}\left(x^3-60x^2+900x\right)dx = \frac{k}{2}\left[\frac{x^4}{4}-\frac{60x^3}{3}+\frac{900x^2}{2}\right]_0^{30}$$

$$= \frac{k}{2}\left[\frac{30^4}{4}-\frac{60(30)^3}{3}+\frac{900(30)^2}{2}\right] = 33750k$$

$$\int_0^{20}\int_0^{20-x} f(x,y)\,dy\,dx = \int_0^{20}\frac{kx}{2}\left[y^2\right]_0^{20-x}dx = \frac{k}{2}\int_0^{20}\left(x^3-40x^2+400x\right)dx = \frac{k}{2}\left[\frac{x^4}{4}-\frac{40x^3}{3}+\frac{400x^2}{2}\right]_0^{20}$$

$$= \frac{k}{2}\left[\frac{20^4}{4}-\frac{40(20)^3}{3}+\frac{400(20)^2}{2}\right] = \frac{20,000k}{3} \Rightarrow 33750k - \frac{20000k}{3} = 1 \Rightarrow \boxed{k = \frac{3}{81250}}$$

b) marginal pdf of X: $f(x) = \int_0^{30-x} f(x,y)\,dy - \int_0^{20-x} f(x,y)\,dy = \frac{kx}{2}\left[900 + x^2 - 60x - x^2 + 40x - 400\right]$

$= \frac{kx}{2}\left[500 - 20x\right] \Rightarrow k\left[250x - 10x^2\right],\ 0 \le x \le 20$

$f(x) = \int_0^{30-x} f(x,y)\,dy = \frac{kx}{2}\left[y^2\right]_0^{30-x} = \frac{kx}{2}\left[900 + x^2 - 60x\right] = \frac{k}{2}\left[450x + \frac{x^3}{2} - 30x^2\right],\ 20 < x \le 30$

$$f(x) = \begin{cases} k\left[250x - 10x^2\right], & 0 \le x \le 20 \\ \frac{k}{2}\left[450x + \frac{x^3}{2} - 30x^2\right], & 20 \le x \le 30 \\ 0, & \text{otherwise} \end{cases}$$

$$f(y) = \begin{cases} k(250y - 10y^2), & 0 \le y \le 20 \\ k\left(450y - 30y^2 + \frac{y^3}{2}\right), & 20 \le y \le 30 \\ 0, & \text{otherwise} \end{cases}$$

$\boxed{f(x,y) \neq f(x) \cdot f(y) \Rightarrow X \text{ and } Y \text{ are not independent}}$

c) $P(X+Y \le 25) = \int_0^{25}\int_0^{25-x} f(x,y)\,dy\,dx - \int_0^{20}\int_0^{20-x} f(x,y)\,dy\,dx$

$\Rightarrow P(X+Y \le 25) = \frac{19218.75k}{2} \Rightarrow \boxed{0.3548}$

d) $E(X+Y) = E(X) + E(Y)$

$= \int_0^{20} x f(x)\,dx + \int_{20}^{30} x f(x)\,dx + \int_0^{20} y f(y)\,dy + \int_{20}^{30} y f(y)\,dy$

$= \int_0^{20} xk(250x - 10x^2)\,dx + \int_{20}^{30} kx\left(450x + \frac{x^3}{2} - 30x^2\right)dx = k\left[\frac{250x^3}{3} - \frac{10x^4}{4}\right]_0^{20} + k\left[\frac{450x^3}{3} + \frac{x^5}{10} - \frac{30x^4}{4}\right]_{20}^{30}$

$$\boxed{\begin{aligned} E(X) &= 12.9846 \\ E(Y) &= 12.9846 \\ E(X+Y) &= 25.969 \end{aligned}}$$

e) $Cov(X,Y) = E(XY) - E(X)E(Y)$

$E(XY) = \int_0^{30}\int_0^{30-x} xy\, f(x,y)\,dy\,dx - \int_0^{20}\int_0^{20-x} xy\, f(x,y)\,dy\,dx = 149.53 - 13.128 = 136.402$

$Cov(X,Y) = 136.402 - (12.98)(12.98) = \boxed{-32.0784}$

$Var(X) = E[X^2] - [E(X)]^2$

$$E(X^2) = \int_0^{20} x^2 k(250x - 10x^2)\,dx + \int_{20}^{30} x^2 k\left(450x - 30x^2 + \frac{x^2}{2}\right)dx \Rightarrow 132.923 + 71.69 = 204.165$$

$Var(X) = 204.165 - (12.98)^2 = 36.135$

$SD(X) = 6.011$

$$Corr(X,Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{-32.1784}{(6.011)(6.011)} = \boxed{-0.89}$$

f) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$= 36.135 + 36.135 + 2(-32.0784)$

$= \boxed{8.1132}$

Problem 7.

n	1	2	3	4
p(n)	0.4	0.3	0.2	0.1

A. E(Y| X= x) = x*n*p(n) = x*(1*0.4 + 2*0.3 + 3*0.2 + 4*0.1)

2x

Var(Y| X = x) = E(Y^2 | X = x) - E(Y| X = x)^2 =

E(Y^2 | X = x) = x^2*n^2*p(n) = > (1^2)*0.4 + (2^2)*0.3 + (3^2)*0.2 + (4^2)*0.1

[1] 5

Var(Y| X = x) = $5x^2 - 4x^2 = x^2$

B. E[Y] = E[E(Y| X= x)] = E[2x] = 2E[X] = $2 \times 20 =$ **40**

C. Var[Y] = V[E(Y| X= x)] + E[V(Y| X= x)] =

V[E(Y| X= x)] = V[2x] = 4V[x] = 80

E[V(Y| X= x)] = E[X] = 20

Var[Y] = 80 + 20 = **100**

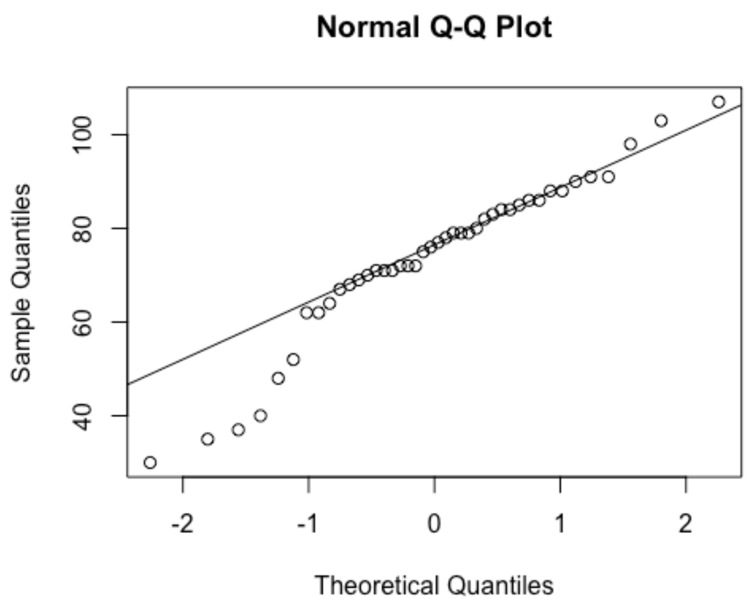
Problem 8.

> airpollution <- data.frame(airpollution)

> attach(airpollution)

A. > qqnorm(Solar)

> qqline(Solar)



B. > shapiro.test(Solar)

Shapiro-Wilk normality test

data: Solar

W = 0.93883, p-value = 0.02601

The W value is large but the p-value is less than 0.05, indicating that the solar radiation measurements are not normally distributed. The QQ plot in part a shows possible outliers in the data, indicating no normality. This shows that the Shapiro-walker test in part b gives the same results as in part a.

Problem 9.

Per day (X):

Mean = 13 oz

SD = 2 oz

14 days (W):

E[W] = 14E[X] = 14 * 13

[1] 182

SD(W) = sqrt(14) * 2

[1] 7.483315

2 6-packs of 16 oz sodas is 192 oz in total

> pnorm(192, 182, 7.483315)

[1] **0.9092754**

We should worry about the validity of the independence assumption here since the amount of bottles she currently has to possibly last her for two weeks is finite and her total consumption might actually depend on the the amount she drinks every day.

Problem 10.

E[X] = 70

Var[X] = 9

E[Y] = 170

Var[Y] = 400

$$(X, Y) \sim BVN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

$\rho = 0.9$

$$(X, Y) \sim BVN(70, 170, 3, 20, 0.9)$$

A.

$$(Y|X = x) \sim N\left(\mu_y + \rho \frac{\sigma_y}{\sigma_x}\ (x - \mu_x), \sigma_y^2(1 - \rho^2)\right)$$

$$(Y|X = x) \sim N(170 + 0.9 * \frac{20}{3}\ (x - 70), 400(1 - 0.9^2))$$

$$(Y|X = x) \sim N(170 +\ 6(x - 70), 76)$$

$$(Y|X = 68) \sim N(170 + 6(68 - 70), 76)$$

$$(Y|X = 68) \sim N(158, 76)$$

B.

$$(Y|X = x) \sim N(170 +\ 6(x - 70), 76)$$

$$(Y|X = 70) \sim N(170 +\ 6(70 - 70), 76)$$

$$(Y|X = 70) \sim N(170, 76)$$

Both distributions of Y at X = 68 and X = 70 are univariate normal. However, the univariate distribution of Y at X = 68 has a mean of 158 and a variance of 76 while the distribution of Y at X = 70 has a mean of 170 and a variance of 76.

C.

$$(Y|X = x) \sim N(170 +\ 6(x - 70), 76)$$

$$(Y|X = 72) \sim N(170 +\ 6(72 - 70), 76)$$

$$(Y|X = 72) \sim N(182, 76)$$

$$P(Y < 180\ |X = 72) = P\left[\frac{Y - \mu}{\sigma} < \frac{180 - \mu}{\sigma}\right] = P\left[Z < \frac{180 - 182}{\sqrt{76}}\right] = P(Z < -0.2294157)$$

```
> pnorm(180, 182, sqrt(76))
[1] 0.4092729
```