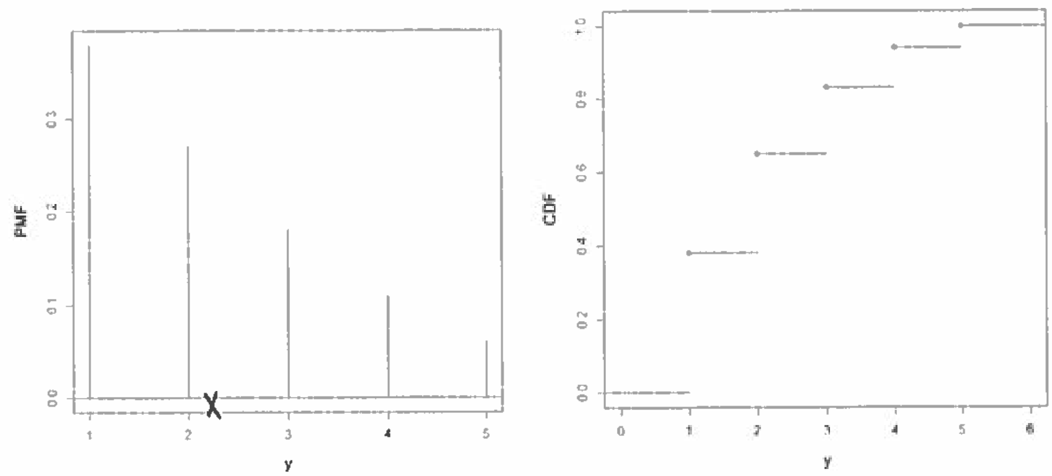
# Stat 509 Homework 2 Solution

1. (a)



R-code:

```
y=seq(1:5)
prob=c(0.38 ,0.27, 0.18, 0.11 ,0.06)
cdf=cumsum(pdf)
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(b)

$$E(Y) = \Sigma y \cdot P_Y(y) = 1\times 0.38 + 2\times 0.27 + 3\times 0.18 + 4\times 0.11 + 5\times 0.06$$
$$= 2.2$$

$$Var(Y) = E(Y^2) - [E(Y)]^2$$
$$= 1^2\times 0.38 + 2^2\times 0.27 + 3^2\times 0.18 + 4^2\times 0.11 + 5^2\times 0.06 - 2.2^2$$
$$= 6.34 - 2.2^2$$
$$= 1.5$$

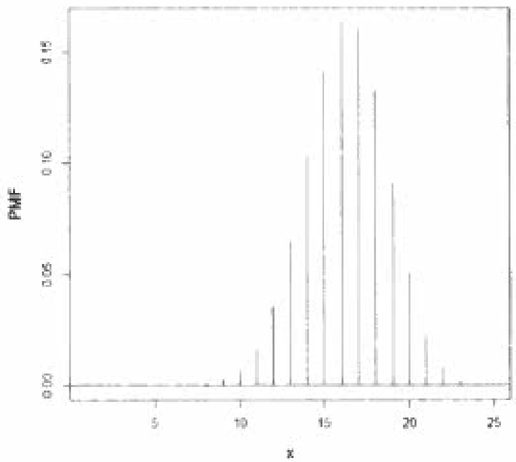
R-code:

```
E_y=sum(y*prob)
> E_y
[1] 2.2
var_y= sum(y^2*prob)- E_y^2
> var_y
[1] 1.5
```

(c)

X~b (25, 0.65), where n=25, p=0.38+0.27=0.65

E(X)=n*p=25*0.65=16.25




R-code:

```
n=25; p=0.38 +0.27
prob= dbinom(x,n,p)
x=seq(0,25,1)
plot(x,prob,type="h",xlab="x",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(x,cdf,f=0)
plot.stepfun(cdf.plot,xlab="x",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch=16,cex.lab=1.25)
```

(d)

$P(X \geq 20) = 1 - P(X \leq 19) = 0.0826247$

No, it is not unusual because the probability of this event is 0.0826, which is not close enough to 0 to conclude that it's an unusual event.

R-code:

```
n=25; p=0.38 +0.27
Prob=1-pbinom(19,n,p)
> Prob
[1] 0.0826247
```

2. (a)

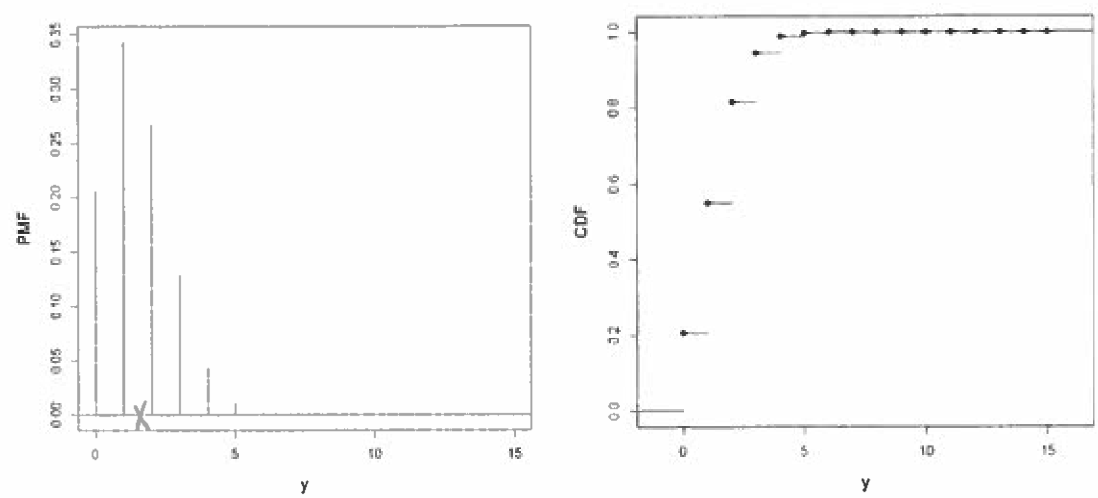
(i) Each water specimen results in containing a particular organic pollutant ("success") or not containing a particular organic pollutant ("failure").

(ii) Water specimens are independent.

(iii) The probability of containing a particular organic pollutant, p=0.1, is the same on every water specimen.

(b)

Y~b(15, 0.1)  E(Y)=15*0.1=1.5




R-code:

```
n=15; p=0.1
y=seq(0,15,1)
prob= dbinom(y,n,p)
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(c)

$$P(Y=1) = \binom{15}{1} 0.1^1 (1-0.1)^{15-1}$$
$$= 15 \times 0.1 \times 0.9^{14}$$
$$= 0.3431519$$

$$P(Y \geq 3) = 1 - P(Y=0) - P(Y=1) - P(Y=2)$$
$$= 1 - \binom{15}{0} 0.1^0 (1-0.1)^{15-0} - \binom{15}{1} 0.1^1 (1-0.1)^{15-1} - \binom{15}{2} 0.1^2 (1-0.1)^{15-2}$$
$$= 1 - 0.2058911 - 0.3431519 - 0.2668959$$
$$= 0.1840611$$

R-code:

```
> dbinom(1,15,0.1)
[1] 0.3431519
> 1-pbinom(2,15,0.1)
[1] 0.1840611
```

(d)

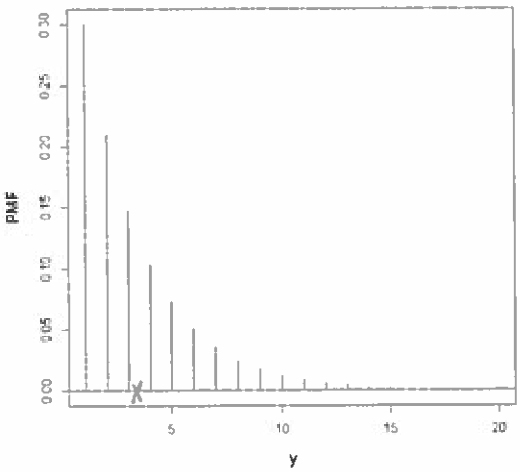
```
> 1-pbinom(9,15,0.1)
[1] 1.866202e-07
```

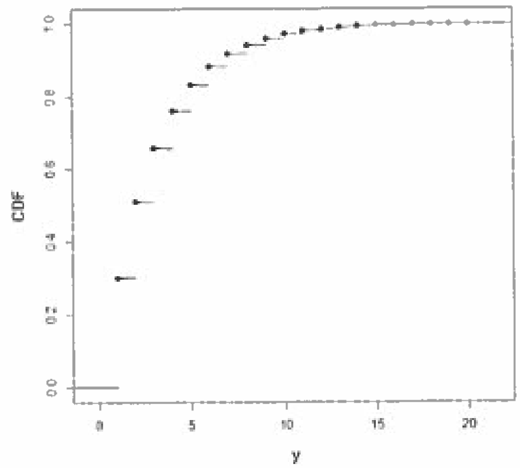
$$P(Y \geq 10) = 1 - P(Y \leq 9) = 1.866 \times 10^{-7}$$

The probability of this event is close to 0, thus this is an unusual event. The chance of containing a particular organic pollutant for each water specimen might be greater than 10 percent.

3.(a)

Y~geom(0.3)    E(Y)=1/p=1/0.3=3.33333



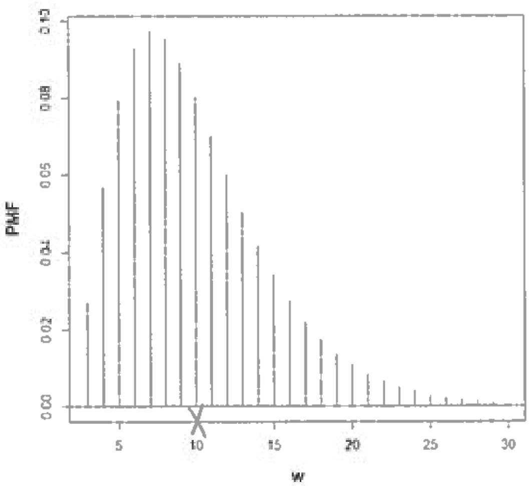



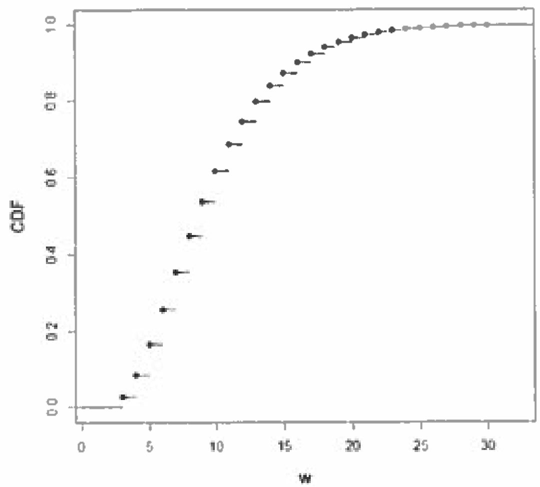

R-code:

```
y=seq(1:20)
prob=dgeom(y-1,0.3)
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(b)

W~nib(3,0.3)  E(W)=r/p=3/0.3=10




R-code:

```
w=seq(3,30,1);r=3;p=0.3
prob=dnbinom(w-r,r,p)
plot(w,prob,type="h",xlab="w",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(w,cdf,f=0)
plot.stepfun(cdf.plot,xlab="w",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(c)

(i) Each customer seeking a renewal results in driver's license is currently expired "success" or not expired "failure".

(ii) Customers are independent.

(iii) The probability of driver's license is currently expired, p=0.3, is the same on every customer seeking a renewal.

(d)

(i) $P(Y>6) = 1 - P(Y \le 6)$

$= 1 - P(Y=1) - P(Y=2) - P(Y=3) - P(Y=4) - P(Y=5) - P(Y=6)$

$= 1 - (1-0.3)^{1-1} 0.3 - (1-0.3)^{2-1} \cdot 0.3 - (1-0.3)^{3-1} \cdot 0.3$

$- (1-0.3)^{4-1} \cdot 0.3 - (1-0.3)^{5-1} \cdot 0.3 - (1-0.3)^{6-1} \cdot 0.3$

$= 1 - 0.3 - 0.21 - 0.147 - 0.1029 - 0.07203 - 0.050421$

$= 0.117649$

$P(W \ge 10) = 1 - P(W \le 9)$

$= 1 - P(W=3) - P(W=4) - P(W=5) - P(W=6) - P(W=7) - P(W=8) - P(W=9)$

$= 1 - \binom{3-1}{3-1} 0.3^3 (1-0.3)^{3-3} - \binom{4-1}{3-1} 0.3^3 (1-0.3)^{4-3} - \binom{5-1}{3-1} 0.3^3 (1-0.3)^{5-3}$

$- \binom{6-1}{3-1} 0.3^3 (1-0.3)^{6-3} - \binom{7-1}{3-1} 0.3^3 (1-0.3)^{7-3} - \binom{8-1}{3-1} 0.3^3 (1-0.3)^{8-3}$

$- \binom{9-1}{3-1} 0.3^3 (1-0.3)^{9-3}$

$= 1 - 0.027 - 0.0567 - 0.07938 - 0.09261 - 0.0972405 - 0.0952957 - 0.0889426$

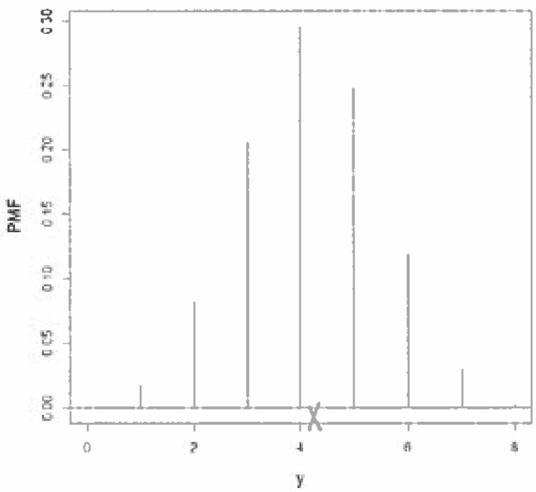
$= 0.4628312$

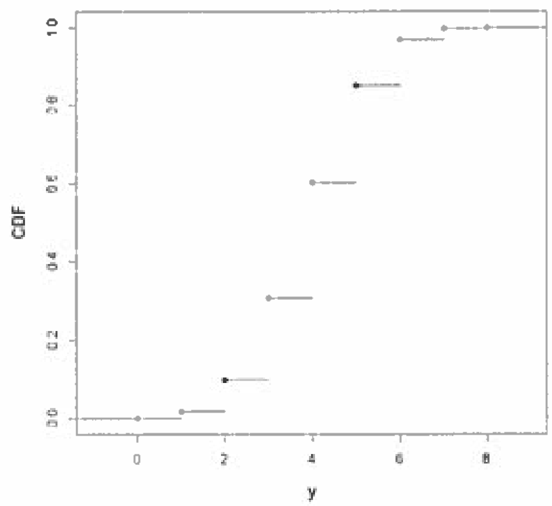
R-code:

```
prob=1-pgeom(6-1,0.3)
> prob
[1] 0.117649
prob=1-pnbinom(9-3,3,0.3)
> prob
[1] 0.4628312
```

4. (a) $Y \sim hyper(50, 8, 26)$

$$E(Y) = n \cdot \left(\frac{r}{N}\right)$$
$$= 8 \times \frac{26}{50}$$
$$= 4.16$$




R-code:

```
y = seq(0,8,1)
prob = dhyper(y,26,50-26,8)
# Plot PMF
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
# Plot CDF
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(b) (i) $P(Y=8)$

$$= \frac{\binom{26}{8}\binom{50-26}{8-8}}{\binom{50}{8}}$$

$$= 0.002909922$$

(ii) $P(Y \le 3)$

$$= P(Y=0) + P(Y=1) + P(Y=2) + P(Y=3)$$

$$= \frac{\binom{26}{0}\binom{50-26}{8-0}}{\binom{50}{8}} + \frac{\binom{26}{1}\binom{50-26}{8-1}}{\binom{50}{8}} + \frac{\binom{26}{2}\binom{50-26}{8-2}}{\binom{50}{8}} + \frac{\binom{26}{3}\binom{50-26}{8-3}}{\binom{50}{8}}$$

$$= 0.001369902 + 0.01676115 + 0.08147782 + 0.2058387$$

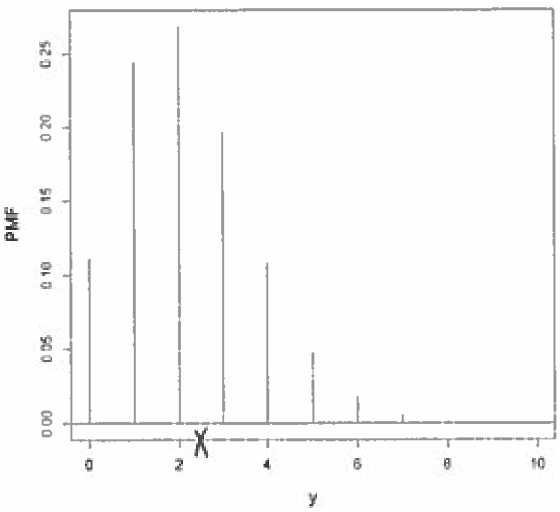
$$= 0.3054476$$

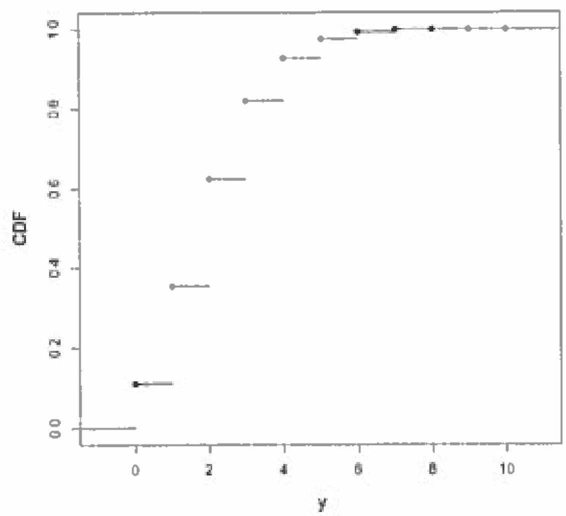
R-code:

```
> dhyper(8,26,50-26,8)
[1] 0.002909922
> phyper(3,26,50-26,8)
[1] 0.3054476
```

5. (a)

$Y \sim \text{Poisson}(2.2)$ $\qquad E(Y) = \lambda = 2.2$




R-code:

```
y = seq(0,10,1)
prob = dpois(y,2.2)
# Plot PMF
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
# Plot CDF
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(b)

(i) $P(Y=0)$

$$= \frac{2.2^0 e^{-2.2}}{0!}$$

$$= 0.1108032$$

(ii) $P(Y \geq 5)$

$$= 1 - P(Y \leq 4)$$

$$= 1 - P(Y=0) - P(Y=1) - P(Y=2) - P(Y=3) - P(Y=4)$$

$$= 1 - \frac{2.2^0 e^{-2.2}}{0!} - \frac{2.2^1 e^{-2.2}}{1!} - \frac{2.2^2 e^{-2.2}}{2!} - \frac{2.2^3 e^{-2.2}}{3!} - \frac{2.2^4 e^{-2.2}}{4!}$$

$$= 1 - 0.1108032 - 0.2437669 - 0.2681436 - 0.1966387 - 0.1081513 - 0.04758656$$

$$= 0.07249631$$

R-code:

```
> dpois(0,2.2)
[1] 0.1108032
> ppois(4,2.2)
[1] 0.9275037
> 1- ppois (4, 2.2)
[1] 0.07249631
```

(c) $C = 150 + 1000Y + 0.1Y^2$

$$E(C) = 150 + 1000\,E(Y) + 0.1\,E(Y^2)$$

$$= 150 + 1000\,E(Y) + 0.1\,(\mathrm{var}(Y) + [E(Y)]^2)$$

$$= 150 + 1000\lambda + 0.1\,(\lambda + \lambda^2)$$

$$= 150 + 1000 \times 2.2 + 0.1\,(2.2 + 2.2^2)$$

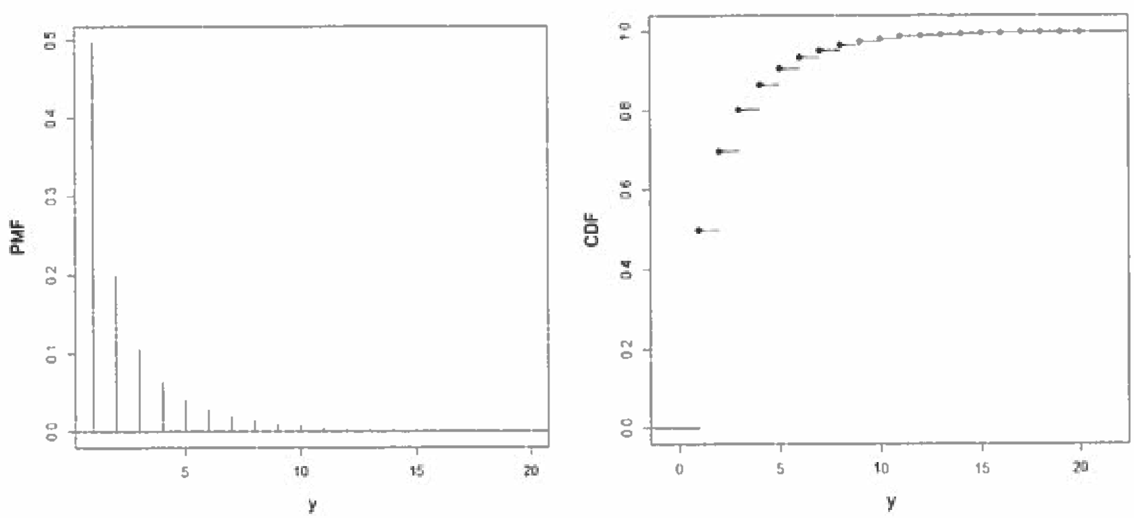
$$= 2350.704$$

6. (a)

The mode in Problem 5 is: Y=2.

It means that it's most likely that 2 cars will experience the catastrophe per year.

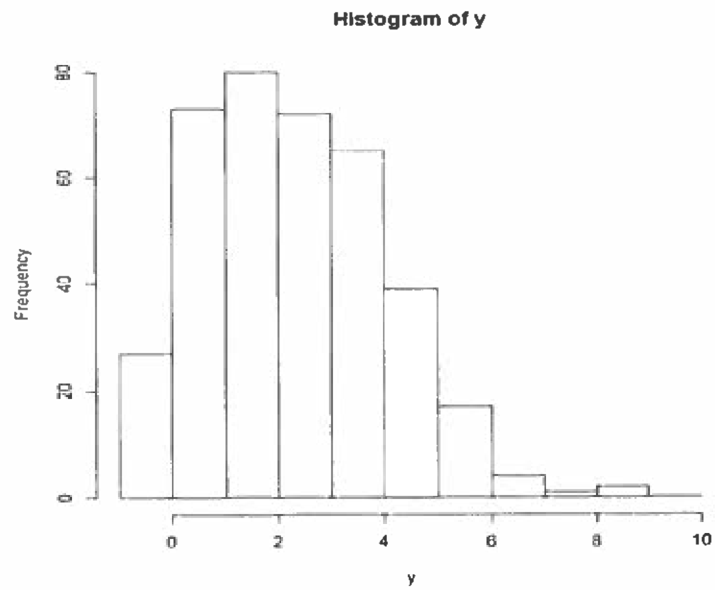
(b)




R-code:

```
y = seq(1,20,1)
p=0.2
prob = -(1-p)^y/(y*log(p))
# Plot PMF
plot(y,prob,type="h",xlab="y",ylab="PMF",ylim=c(0,max(prob)),cex.lab=1.25)
abline(h=0)
# Plot CDF
cdf = c(0,cumsum(prob))
cdf.plot = stepfun(y,cdf,f=0)
plot.stepfun(cdf.plot,xlab="y",ylab="CDF",verticals=FALSE,do.points=TRUE,main="",pch
=16,cex.lab=1.25)
```

(c) This is a histogram of the 2013-2014 English Premier League season data.




R-code:

```
y=c(rep(0,27), rep(1,73),rep(2, 80),rep(3, 72),rep(4, 65),rep(5, 39),rep(6, 17),rep(7,
4),rep(8, 1),rep(9, 2),rep(10, 0))
hist(y,xlim=c(-1,10),breaks=-1:10)
```

No, logarithmic series distribution wouldn't be a good distribution to model the number of goals scored per game. First, the observation y=0 is not allowed under logarithmic series model. Second, the distribution of the data does not resemble logarithmic series distribution at all.