Q3

(a).

Probability mass function
Probability mass function of X, P(X=k; r, p)
k represents the total trials, which is the sum of successes and failures
r represents the experiment ends at rth success (until r successes)
p represents the success rate

To end this experiment, the last trail must be a success.

$$\_\_\_\_\_ \cdots \_\_\_\_ \mathbf{S}$$

The remaining (k-1) trails has (r-1) successes, and (k-r) failures

By using the formula of probability mass function of binomial distribution,

These remaining trails has $\binom{k-1}{r-1}$ combination (r-1) successes.
Then, the probability of success (r-1) times in (k-1) trails should be:

$\binom{k-1}{r-1} \times p^{r-1} \times (1-p)^{k-1-(r-1)} = \binom{k-1}{r-1} \times p^{r-1} \times (1-p)^{k-r}$, where p is still the probability of success

Since the last trail is a success, by Multiplicative Rule,

$$\mathrm{P(X = k; r, p)} = \binom{\mathrm{k}-1}{\mathrm{r}-1} \times \mathrm{p}^{\mathrm{r}-1} \times (1-\mathrm{p})^{\mathrm{k}-\mathrm{r}} \times \mathrm{p} = \binom{\mathrm{k}-1}{\mathrm{r}-1} \times \mathrm{p}^{\mathrm{r}} \times (1-\mathrm{p})^{\mathrm{k}-\mathrm{r}}$$

Or to change the combination to factorial, the alternative can be:

$$\mathrm{P(X = k; r, p)} = \frac{(\mathrm{k}-1)!}{(\mathrm{k}-\mathrm{r})! \times (\mathrm{r}-1)!} \times \mathrm{p}^{\mathrm{r}} \times (1-\mathrm{p})^{\mathrm{k}-\mathrm{r}} \times \mathrm{p}$$

The range of k from k=r, r+1, ... to infinity

Expected Value

$$E(x) = \sum_{k=r}^{\infty} x \cdot f(x) = \sum_{k=r}^{\infty} k \cdot \binom{k-1}{r-1} \cdot p^r \cdot (1-p)^{k-r}$$

$$= \sum_{k=r}^{\infty} k \cdot \binom{k-1}{r-1} \cdot p^r \cdot (1-p)^{k-r}$$

$$= \sum_{k=r}^{\infty} k \cdot \frac{(k-1)!}{(k-r)!(r-1)!} \cdot p^r \cdot (1-p)^{k-r}$$

$= \sum_{k=r}^{\infty} r \cdot \frac{k!}{(k-r)!(r)!} \cdot p^r \cdot (1-p)^{k-r}$

$= \text{r} \cdot \sum_{k=r}^{\infty} \binom{k}{r} \cdot p^r \cdot (1-p)^{k-r} \ldots (1)$

Consider j = r+1

Because $\sum_{k=j}^{\infty} \binom{k-1}{j-1} \cdot p^j \cdot (1-p)^{k-j} = 1$

So $\sum_{k=r+1}^{\infty} \binom{k-1}{r+1-1} \cdot p^{r+1} \cdot (1-p)^{k-r-1} = 1,$

which means $\sum_{k=r}^{\infty} \binom{k-1+1}{r+1-1} \cdot p^{r+1} \cdot (1-p)^{k-r-1+1} = 1$

Then, (1) = $\text{r} \cdot \sum_{k=r}^{\infty} \binom{k}{r} \cdot p^r \cdot (1-p)^{k-r}$

$= \frac{\text{r}}{\text{p}} \cdot \sum_{k=r}^{\infty} \binom{k}{r} \cdot p^{r+1} \cdot (1-p)^{k-r}$

$= \frac{\text{r}}{\text{p}} \cdot 1$

Thus, E(x) = $\frac{r}{p}$

Variance

Var(x) = E(x^2) − (E(x))^2

$= \sum_{k=r}^{\infty} x^2 \cdot f(x) \ \text{-} \ (\frac{r}{p})^2$

Consider E(x^2):

E(x^2) $= \sum_{k=r}^{\infty} x^2 \cdot f(x) = \sum_{k=r}^{\infty} k^2 \cdot \frac{(k-1)!}{(k-r)!\,(r-1)!} \cdot p^r \cdot (1-p)^{k-r}$

$= \sum_{k=r}^{\infty} k^2 \cdot \frac{(k-1)!}{(k-r)!\,(r-1)!} \cdot p^r \cdot (1-p)^{k-r}$

$= \sum_{k=r}^{\infty} ((k+1)(k) - k) \cdot \frac{(k-1)!}{(k-r)!\,(r-1)!} \cdot p^r \cdot (1-p)^{k-r}$

$= \sum_{k=r}^{\infty} \frac{(k+1)!}{(k-r)!\,(r-1)!} \cdot p^r \cdot (1-p)^{k-r} - \sum_{k=r}^{\infty} k \cdot \frac{(k-1)!}{(k-r)!\,(r-1)!} \cdot p^r \cdot (1-p)^{k-r}$

$= (\sum_{k=r}^{\infty} r(r+1) \cdot \frac{(k+1)!}{(k-r)!\,(r+1)!} \cdot p^r \cdot (1-p)^{k-r}) - \frac{r}{p}$

$= \text{r(r+1)}(\sum_{k=r}^{\infty} \binom{k+1}{r+1} \cdot p^r \cdot (1-p)^{k-r}) - \frac{r}{p} \ \ldots (2)$

Consider i = r+2

Because $\sum_{k=i}^{\infty}\binom{k-1}{i-1}\cdot p^{i}\cdot(1-p)^{k-i}=1,$

Then $\sum_{k=r+2}^{\infty}\binom{k-1}{r+2-1}\cdot p^{r+2}\cdot(1-p)^{k-(r+2)}=1$

So $\sum_{k=r}^{\infty}\binom{k-1+2}{r+2-1}\cdot p^{r+2}\cdot(1-p)^{k-(r+2)+2}=1,$

which means $\sum_{k=r}^{\infty}\binom{k+1}{r+1}\cdot p^{r+2}\cdot(1-p)^{k-r}=1$

Then (2) = r(r+1)($\sum_{k=r}^{\infty}\binom{k+1}{r+1}\cdot p^{r}\cdot(1-p)^{k-r}) - \frac{r}{p}$

$= \frac{r(r+1)}{p^2}(\sum_{k=r}^{\infty}\binom{k+1}{r+1}\cdot p^{r-2}\cdot(1-p)^{k-r}) - \frac{r}{p}$

$= \frac{r(r+1)}{p^2} - \frac{\mathrm{r}}{\mathrm{p}}$

Thus, $\mathrm{Var(x)} = \mathrm{E(x^2)} - (\mathrm{E(x)})^2 = \frac{r(r+1)}{p^2} - \frac{\mathrm{r}}{\mathrm{p}} - (\frac{\mathrm{r}}{\mathrm{p}})^2$

$= \frac{r(r+1)-pr-r^2}{p^2}$

$= \frac{r(1-p)}{p^2}$

(c)

Consider the negative binomial distribution, r represents the time of death to completely kill the cat, and p is the probability that the cat dies per week.

Life expectancy = E(x) = $\frac{r}{p} = \frac{9}{\frac{1}{20}} = 180\ weeks$

Standard deviation = $\sqrt{Var(x)} = \sqrt{\frac{9(1-\frac{1}{20})}{(\frac{1}{20})^2}} \approx 58.48\ weeks$

Thus, the life expectancy of this cat is 180 weeks and the standard deviation is approximately equal to 58.48 weeks

(d)

Consider binomial distribution. P (X=x) = $\binom{k}{x}p^{x}(1-p)^{k-x}$, where k is the total number of trails, x is the number of successes, and p is the probability of success

In this question, let n represents his remaining lives. When he has n lives, he can die n-1 or fewer times, so x<n. Both x and n should be the non-negative number.

Let p be the probability that he is hit, which is $\frac{1}{20}$. The total trails (weeks) are 104. The event of success corresponds to the cat losing one life

(1) When he has 1 life, he can not lose any life.

P(x<1) = P(x=0) = $\binom{104}{0} \times (\frac{1}{20})^0 \times (1-\frac{1}{20})^{104} \approx 0.0048$

(2) When he has 2 lives, he can die 1 or fewer times.

P(x<2) = P(x=1) + P(x<1) = $\binom{104}{1} \times (\frac{1}{20})^1 \times (1-\frac{1}{20})^{103} + 0.0048 \approx 0.0312$

(3) When he has 3 lives, he can die 2 or fewer times.

P(x<3) = P(x=2) + P(x<2) = $\binom{104}{2} \times (\frac{1}{20})^2 \times (1-\frac{1}{20})^{102} + 0.0312 \approx 0.1027$

(4) When he has 4 lives, he can die 3 or fewer times.

P(x<4) = P(x=3) + P(x<3) = $\binom{104}{3} \times (\frac{1}{20})^3 \times (1-\frac{1}{20})^{101} + 0.1027 \approx 0.2307$

(5) When he has 5 lives, he can die 4 or fewer times.

P(x<5) = P(x=4) + P(x<4) = $\binom{104}{4} \times (\frac{1}{20})^4 \times (1-\frac{1}{20})^{100} + 0.2307 \approx 0.4008$

(6) When he has 6 lives, he can die 5 or fewer times.

P(x<6) = P(x=5) + P(x<5) = $\binom{104}{5} \times (\frac{1}{20})^5 \times (1-\frac{1}{20})^{99} + 0.4008 \approx 0.5799$

(7) When he has 7 lives, he can die 6 or fewer times.

P(x<7) = P(x=6) + P(x<6) = $\binom{104}{6} \times (\frac{1}{20})^6 \times (1-\frac{1}{20})^{98} + 0.5799 \approx 0.7355$

(8) When he has 8 lives, he can die 7 or fewer times.

P(x<8) = P(x=7) + P(x<7) = $\binom{104}{7} \times (\frac{1}{20})^7 \times (1-\frac{1}{20})^{97} + 0.7355 \approx 0.8500$

(9) When he has 9 lives, he can die 8 or fewer times.

P(x<9) = P(x=8) + P(x<8) = $\binom{104}{8} \times (\frac{1}{20})^8 \times (1-\frac{1}{20})^{96} + 0.8500 \approx 0.9232$

Thus, his survival probability is about 0.0048 if he has one life, while the probability is approximately equal to 0.9232 when he has 9 lives.

I then simulate this situation for 100000 times in R language

```
#Q3 (d)

#This is a binomial distribution situation.
#rbinom(n, size, r) simulates n times of hit times of the cat within 104 weeks
set.seed(104)
hit_time=rbinom(100000,104,0.05)
hit=data.frame(table(hit_time))

#Calculate the proportion of n times hits within 104 weeks
f=1
l_d=length(hit[,1])
d_proportion=c()
while(f<=l_d){
  d_proportion=c(d_proportion,hit[f,2]/100000)
  f=f+1
}
hit=data.frame(hit,d_proportion)

#One life, can die 0 times
one_life=hit[1,3]
#Two lives, can die 1 or fewer times
two_lives=hit[2,3]+one_life
#Three lives, can die 2 or fewer times
three_lives=hit[3,3]+two_lives
#Four lives, can die 3 or fewer times
four_lives=hit[4,3]+three_lives
#Five lives, can die 4 or fewer times
five_lives=hit[5,3]+four_lives
#Six lives, can die 5 or fewer times
six_lives=hit[6,3]+five_lives
#Severn lives, can die 6 or fewer times
seven_lives=hit[7,3]+six_lives
#Eight lives, can die 7 or fewer times
eight_lives=hit[8,3]+seven_lives
#Nine lives, can die 8 or fewer times
nine_lives=hit[9,3]+eight_lives
```

I use rbinom() to simulate this situation. Size, 104, refers to 104 weeks. The probability, 0.05, is equal to 1/20, which is the likelihood that the cat is hit. The following photo shows the hit time (successes time), frequency, and the corresponding proportion within 100000 simulations.

	hit_time	Freq	d_proportion
1	0	491	0.00491
2	1	2558	0.02558
3	2	7180	0.07180
4	3	12898	0.12898
5	4	17049	0.17049
6	5	17926	0.17926
7	6	15372	0.15372
8	7	11600	0.11600
9	8	7302	0.07302
10	9	4072	0.04072
11	10	1990	0.01990
12	11	916	0.00916
13	12	424	0.00424
14	13	153	0.00153
15	14	46	0.00046
16	15	18	0.00018
17	16	2	0.00002
18	17	3	0.00003

I then find that the simulation results are approximately same as my theoretical calculation.

```
>
> print(one_life)
[1] 0.00491
> print(two_lives)
[1] 0.03049
> print(three_lives)
[1] 0.10229
> print(four_lives)
[1] 0.23127
> print(five_lives)
[1] 0.40176
> print(six_lives)
[1] 0.58102
> print(seven_lives)
[1] 0.73474
> print(eight_lives)
[1] 0.85074
> print(nine_lives)
[1] 0.92376
>
```

The result of running code is shown above. The probability of living is 0.00491 if he has one life, and the probability is 0.92376 when he has 9 lives.

*I do not know why this problem is put in Question 3. I do not understand whether I must use negative binomial distribution to solve the problem. If I must, this is negative binomial situation.

(1) When the cat has one life. Consider he will die after 104 weeks, assume in 105th week, which means he will be hit the earliest at 105^{th} week. Let w represents weeks.

$$P(w>104) = \sum_{w=105}^{\infty} \binom{w-1}{1-1} \cdot \left(\frac{1}{20}\right)^{1} \cdot \left(\frac{19}{20}\right)^{w-1}$$

(2) When the cat has 9 lives. Consider he will lose his 9^{th} life after 104 weeks, which may begins at 105^{th} week.

$$P(w>104) = \sum_{w=105}^{\infty} \binom{w-1}{9-1} \cdot \left(\frac{1}{20}\right)^{9} \cdot \left(\frac{19}{20}\right)^{w-9}$$

It is physical impossible to calculate, but they have similar ideas with my previous binomial calculation. The following code estimate the cumulative probability, and their answers are similar to my binomial estimation.

```
> #one life.
> #The following code describes the likelihood of most 104-1=103 failures before
> #first hit, which means he will be hit at least in 104th week
> p_1 = pnbinom(103,1,0.05)
> #1-p_1 should be the answer since he needs to live at least 104 weeks
> print(1-p_1)
[1] 0.004822308
>
> #Nine lives.
> #Same idea as before. The first parameter inside pnbinom should be 104-9=95
> p_9=pnbinom(95,9,0.05)
> print(1-p_9)
[1] 0.9233176
>
```

(I made a typing error, p is for cumulative probability)