Verify the Theory of Natural Selection

Yan Haoming

November 22, 2024
November 29, 2024
December 1, 2024

1 Introduction

In this experiment, we repeated the study on mutation by Luria,S. E. and Delbrück,M. [1] The hypothesis (mutation hypothesis instead of acquired hereditary hypothesis) we tried to verified is exactly in accordance to the theory of natural selection, which inspired me to title the report in this way.

The difference between our experiment and the original work lies mainly in experimental material. Specifically, we used yeast rather than bacteria. Besides, I carried out the simulation of the process on my laptop by Python which is impossible in the year of 1943.
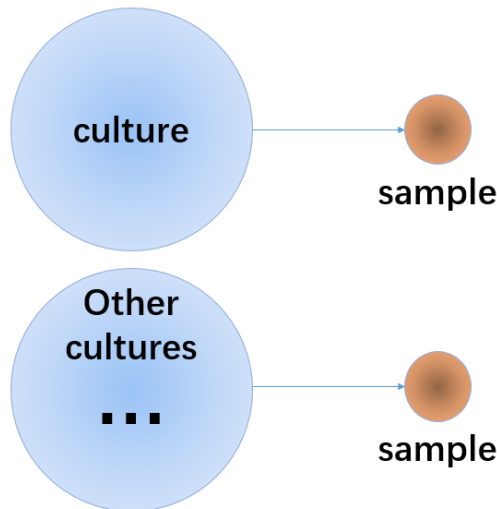
2 Theory

2.1 Key Concepts

Figure 1: Cultures and Samples

The concept of **culture** and **sample** should be distinguished clearly. Each culture was cultured independently under the same condition serving as paralleled groups. Samples were sampled out when spreading the plate representing the corresponding cultures.

Specifically speaking, cultures were grown in Multiwell Plate with 1 mL in each well. Samples were 20 µL ,which were sampled from each culture and spread onto to a solid plate.

2.2 Meaning of Symbols

To emphasis the time dependence of some variables, I add subscripts to them, which is slightly different from the original article.

- C: number of similar **cultures** in our experiment.

- t: time in units of division cycles of the yeasts.

- thus the growth rate should be exactly 1, both for sensitive yeasts and resistant yeasts.
- t_0: a certain time, prior to which mutations were not likely to occur in any experimental **cultures**. Specifically speaking the t_0 satisfies:

$$1 = \text{Ex}(Cm_{t_0})) = \text{Ex}(aC(N_{t_0} - N_0))$$

- N_t, *observable*: number of yeasts in **a growing culture** at time t.
  - $N_t - N_0$: increase in the number of yeasts.
  - $N_0 \ll N_t$, thus, $N_0 \approx 0$ in the presence of N_t.

- ρ_t: the **average** number of resistant yeasts present at time t, in one **culture**.

- r_t, *observable*: the **likely average** number of resistant yeasts in a **culture** at time t, gotten from **limited number of samples**.

- a: mutation rate, namely, the chance of mutation per yeast per time unit.

- m_t: the **average** number of mutations at time t, in **a culture**.

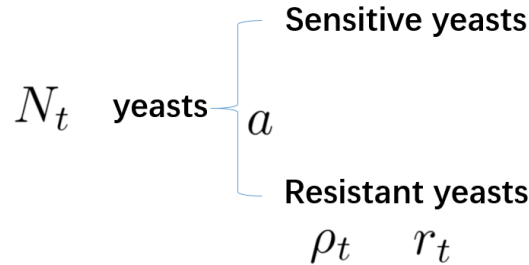- p_0: the fraction of cultures showing **no** mutation in a large series of similar cultures.

$$N_t \quad \text{yeasts} \begin{cases} \text{Sensitive yeasts} \\ a \\ \text{Resistant yeasts} \\ \rho_t \quad r_t \end{cases}$$

Figure 2: Number of Yeasts Divides into Sensitive One and Resistant One

2.3 Derivation of the Formulas

2.3.1 Acquired Hereditary Immunity Hypothesis

Assumption: a fixed small chance for each yeast to survive an attack.

$$\text{Number of resistant yeasts} \sim \text{Binomial Distribution} \tag{1}$$
$$\sim \text{Poisson Distribution} \tag{2}$$

$$\text{Var}(r_t) = \text{Ex}(r_t) = r_t \tag{3}$$
$$\tag{4}$$

2.3.2 Mutation Hypothesis

Assumption: a fixed small chance per time unit for each yeast to undergo a mutation to resistance.

$$\frac{dN_t}{dt} = N_t \tag{5}$$
$$N_t = N_0 e^t \tag{6}$$
$$dm_t = adt N_t \tag{7}$$
$$m_t = a(N_t - N_0) \tag{8}$$

$$\text{Number of yeasts mutate during } dt \sim \text{Poisson Distribution} \tag{9}$$

$$p_0 = e^{-m_t} \tag{10}$$

$$\frac{d\rho_t}{dt} = aN_t + \rho \tag{11}$$

$$\text{assume: } \rho_0 = 0 \tag{12}$$

$$\rho_t = taN_t \tag{13}$$

$$r_t = (t - t_0)aN_t \tag{14}$$

$$1 = aC(N_{t_0} - N_0) \approx aCN_{t_0} \tag{15}$$

$$N_{t_0} = N_t e^{-(t-t_0)} \tag{16}$$

$$t - t_0 = \ln(N_t Ca) \tag{17}$$

$$r_t = aN_t \ln(N_t Ca) \tag{18}$$

The following discussion is about **partial distribution**, defined as the distribution during a certain time interval from $t - \tau$ to $t - \tau + d\tau$ with t fixed while τ as parameter:

$$\text{Ex}(dm_t) = aN_\tau d\tau = aN_t e^{-\tau} d\tau \tag{19}$$

$$d\rho_t = e^\tau dm_t \tag{20}$$

$$\text{Ex}(d\rho_t) = e^\tau \text{Ex}(dm_t) = aN_t d\tau \tag{21}$$

$$\text{Var}(d\rho_t) = e^{2\tau}\text{Var}(dm_t) = aN_t e^\tau d\tau \tag{22}$$

$$\tag{23}$$

To find the total distribution from partial distribution:

$$\text{Var}(\rho_t) = \int_0^t \text{Var}(d\rho_t)d\tau \tag{24}$$

$$= aN_t(e^t - 1) \tag{25}$$

$$\text{Var}(r_t) = \int_0^{t-t_0} \text{Var}(dr_t)d\tau \tag{26}$$

$$= aN_t(e^{t-t_0} - 1) \tag{27}$$

$$\approx Ca^2 N_t^2 \tag{28}$$

$$\frac{\text{Var}(r_t)}{r_t} = \frac{CaN_t}{\ln(N_t Ca)} \gg 1 \tag{29}$$

3 Calculating Procedure in Experiment

- observe the number of yeasts in a growing culture N_t at time t.

- determine the number of resistant yeasts in each **sample**.

  - this should be carried out by observing the number of colonies several days after inoculating.

- calculate the number of resistant yeasts r_t in each **culture**:

$$r_t = \text{Ex}(r_t) = \frac{\text{volume of culture}}{\text{volume of samples}} \cdot \text{Ex}(r_{t \text{ sample}})$$

  - volume of culture = 1 mL, volume of samples = 20 μL in our experiment.

- evaluate the mutation rate a by the following formula:

$$r_t = \text{Ex}(r_t) = aN_t \ln(N_t Ca)$$

- figure out the likely variance $\text{Var}(r_t)$:

  - by mutation rate a according to the theory:

$$\text{Var}(r_t) \approx Ca^2 N_t^2$$

– by statistical methods:

$$\mathrm{Var}(r_{t\ \tilde{\mathrm{sample}}}) = \mathrm{Var}(r_{t\ \mathrm{sample}}) - \mathrm{Var}(\mathrm{sampling})$$

$$\mathrm{Var}(\mathrm{sampling}) \approx \mathrm{Ex}(\mathrm{sampling}) = r_t$$

$$\mathrm{Var}(r_t) = \left(\frac{\text{volume of culture}}{\text{volume of samples}}\right)^2 \cdot \mathrm{Var}(r_{t\ \tilde{\mathrm{sample}}})$$

Thus we calculate the modified statistical variance by:

$$\mathrm{Var}(r_t) = \left(\frac{\text{volume of culture}}{\text{volume of samples}}\right)^2 \cdot \left(\mathrm{Var}(r_{t\ \mathrm{sample}}) - \mathrm{Ex}(r_{t\ \mathrm{sample}})\right)$$

– compare the ratio of variance and average to one (Poisson distribution in Acquired hereditary immunity hypothesis):
$$\frac{\mathrm{Var}(r_t)}{\mathrm{Ex}(r_t)}$$

both by the statistical method (just based on the final result) and by the theoretical method (based on the mutation rate a).

4 Materials and Experimental Procedure

We adopted YNB, glucose, and Amino Acids to make the culture medium. The pressure of selection was introduced by canavanine, which can distinguish the resistant yeasts from sensitive yeasts.

The following solutions were prepared.

- 10x YNB 500 mL
- 10x glucose 500 mL
- 10x Amino Acids 500 mL
- 10x Amino Acids without Arginine 500 mL
- 100x canavanine 100 mL

The detailed configuration of selective culture medium is shown below:

- 10x YNB 100 mL
- 10x glucose 100 mL
- 10x Amino Acids without Arginine 100 mL
- 100x canavanine 10 mL

And pour 25 mL mixture to one plate.
As for SD culture medium:

- 10x YNB 100 mL
- 10x glucose 100 mL
- 10x Amino Acids 100 mL

And pour 25 mL mixture to one plate in the same way.

We found out the proper concentration (approximately 100 yeasts/mL) by serial dilution. And then added them to Deep-well Multiwell(96) Plate for culturing. After a week, spread the yeasts to solid culture medium. Three days later, stopping culturing and we counted the colonies in each drop in both experimental group and controlled group.
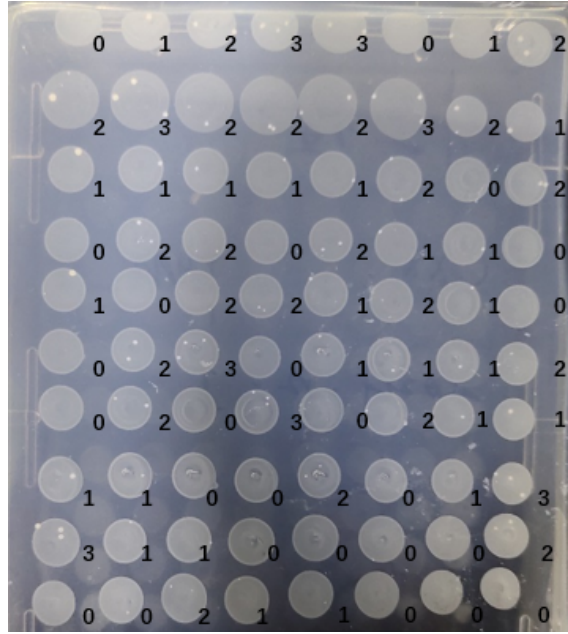
Figure 3: Controlled Group

5 Data Analysis

5.1 Calculate Sampling Variance

I firstly calculate the sampling (from culture to sample) variance by the controlled group as shown in Figure 3. In Table 1, I have shown the mean and variance of $r_{t\ \text{sample}}$. The mean value is close to variance indicating the sampling process follow Poisson distribution.

It should be noted that unbiased estimator is adopted to estimate.

Frankly speaking we could get the final result even if without controlled group. The role of controlled group is to confirm us that the randomness introduced by sampling follows Poisson Distribution.

$$0.97 = \text{Var(sampling)} \approx \text{Ex(sampling)}$$

Mean $\bar{x}$	1.1625
Variance $S^2 x$	0.9733
Capacity	80

Table 1: Statistics of Controlled Group

5.2 Number of Total Yeasts in Cultures

Culture Number	Concentration (number per mL)	Number N_t in a culture
C1	6.67×10^7	6.67×10^7
C4	1.38×10^6	1.38×10^6
C7	1.41×10^6	1.41×10^6
F12	4.57×10^7	4.57×10^7
F9	1.79×10^8	1.79×10^8
Mean	5.88×10^7	5.88×10^7

Table 2: Number of Yeasts in Cultures

To calculate mutation rate a , we should first figure out the number of yeast in a culture, namely N_t. Therefore we randomly selected five cultures from experimental group and counted the number of yeasts respectively before spreading to the plate (solid culture medium).

The results are shown in Table 2. To be noted that the volume is exactly 1 mL per culture (deep well in the plate).

Finally taking the average number to be N_t for further calculation.

$$N_t = 5.88 \times 10^7$$
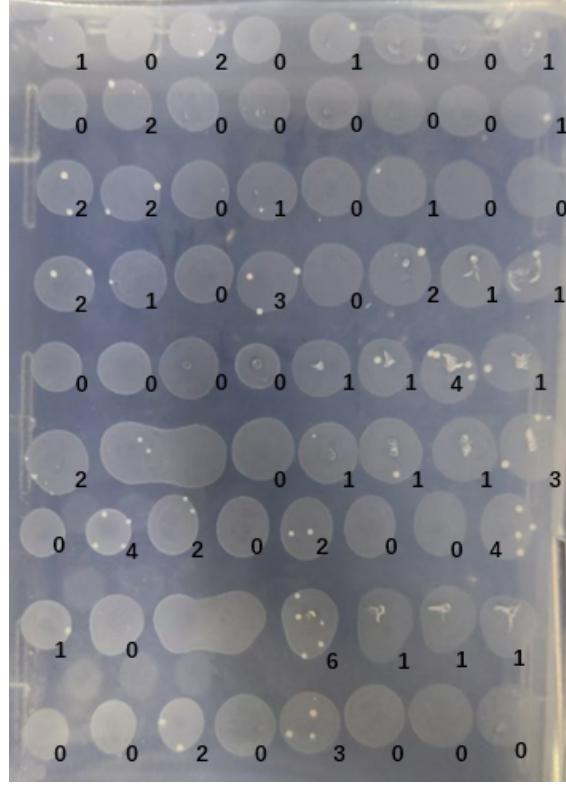
5.3 Number of Resistant Yeasts in Samples

Figure 4: Experimental Group

The critical statistics are gotten from experimental groups as shown in Figure 4 and Table 3. Again, the mean value in experimental groups (0.97) is close to the variance of controlled group (0.9733, representing sampling variance), which demonstrate the correctness of our train of thought.

Mean $\bar{x}$	0.9706
Variance $S^2 x$	1.5812
Capacity	68

Table 3: Statistics of Experimental Group

5.3.1 The Expectation Value

$$\text{Ex}(r_{t \text{ sample}}) = 0.97 \tag{30}$$
$$\text{volume of culture} = 1 \text{ mL} \tag{31}$$
$$\text{volume of samples} = 20 \text{ µL} \tag{32}$$
$$\text{Ex}(r_t) = \frac{\text{volume of culture}}{\text{volume of samples}} \cdot \text{Ex}(r_{t \text{ sample}}) \tag{33}$$
$$= \frac{1 \times 10^3}{20} \times 0.97 = 48.5 \tag{34}$$

5.3.2 The Variance: By Statistical Methods

$$\text{Var}(r_{t \text{ sample}}) = 1.5812 \tag{35}$$

$$\text{Var}(r_{t \tilde{\text{ sample}}}) = \text{Var}(r_{t \text{ sample}}) - \text{Var}(\text{sampling}) \tag{36}$$

$$\text{Var}(\text{sampling}) \approx \text{Ex}(\text{sampling}) = r_t \tag{37}$$

$$\text{Var}(r_t) = \left(\frac{\text{volume of culture}}{\text{volume of samples}} \right)^2 \cdot \text{Var}(r_{t \tilde{\text{ sample}}}) \tag{38}$$

$$\text{Var}(r_t) = \left(\frac{\text{volume of culture}}{\text{volume of samples}} \right)^2 \cdot (\text{Var}(r_{t \text{ sample}}) - \text{Ex}(r_{t \text{ sample}})) \tag{39}$$

$$= \left(\frac{1 \times 10^3}{20} \right)^2 \times (1.5812 - 0.97) = 1528 \tag{40}$$

5.4 Mutation Rate and Variance: By Theoretical Methods

We can evaluate the mutation rate a by the following formula:

$$C = 68 \tag{41}$$

$$\text{Ex}(r_t) = 48.5 \tag{42}$$

$$N_t = 5.88 \times 10^7 \tag{43}$$

$$r_t = \text{Ex}(r_t) = aN_t \ln(N_t Ca) \tag{44}$$

$$a = 1.32 \times 10^{-7} \tag{45}$$

Thus the corresponding variance is:

$$\text{Var}(r_t) \approx Ca^2 N_t^2 \tag{46}$$

$$= 4074 \tag{47}$$

5.5 Verification

5.5.1 By Statistical Methods

$$\frac{\text{Var}(r_t)}{\text{Ex}(r_t)} = \frac{1528}{48.5} = 31.5 \gg 1$$

5.5.2 By Theoretical Methods (Based on Mutation Rate)

$$\frac{\text{Var}(r_t)}{\text{Ex}(r_t)} = \frac{4074}{48.5} = 84 \gg 1$$

5.6 Conclusion

$$\frac{\text{Var}(r_t)}{\text{Ex}(r_t)} \gg 1$$

From our experiment, it has been verified that, the ratio of variance to mean value is much greater than one, both by the statistical method (just based on the final result) and by the theoretical method (based on the mutation rate a). Therefore, the acquired hereditary immunity hypothesis is wrong while the mutation hypothesis is more sensible.

6 Simulation

I wrote the computer program in the logic of each hypothesis to simulate the corresponding mechanisms. After that, I calculated the statistics including mean value and variance to confirm the validity of former theoretical analysis.
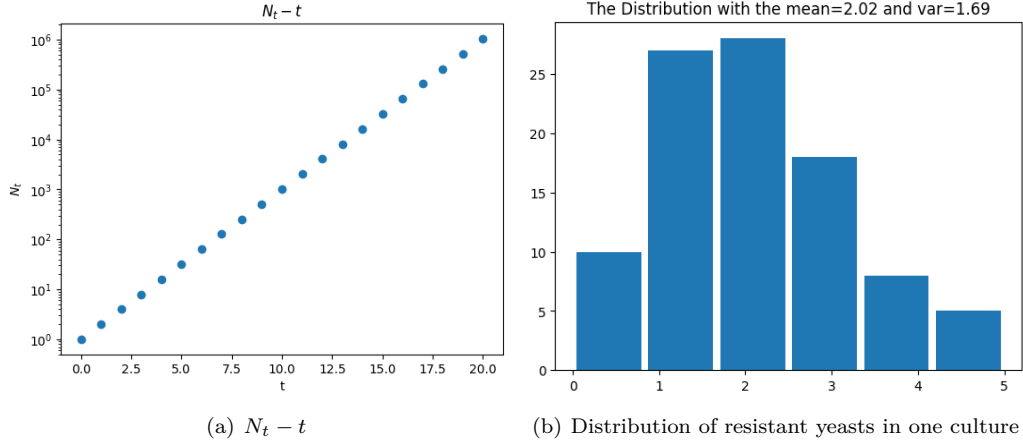
(a) $N_t - t$

(b) Distribution of resistant yeasts in one culture

Figure 5: Simulation on Acquired Hereditary Immunity Hypothesis

6.1   Acquired Hereditary Immunity Hypothesis

As we can see in Figure 5, the mean value is close to variance.

$$\text{Ex}(r_t) = 2.02 \tag{48}$$

$$\text{Var}(r_t) = 1.69 \tag{49}$$

$$\frac{\text{Var}(r_t)}{\text{Ex}(r_t)} = 1.195 \approx 1 \tag{50}$$
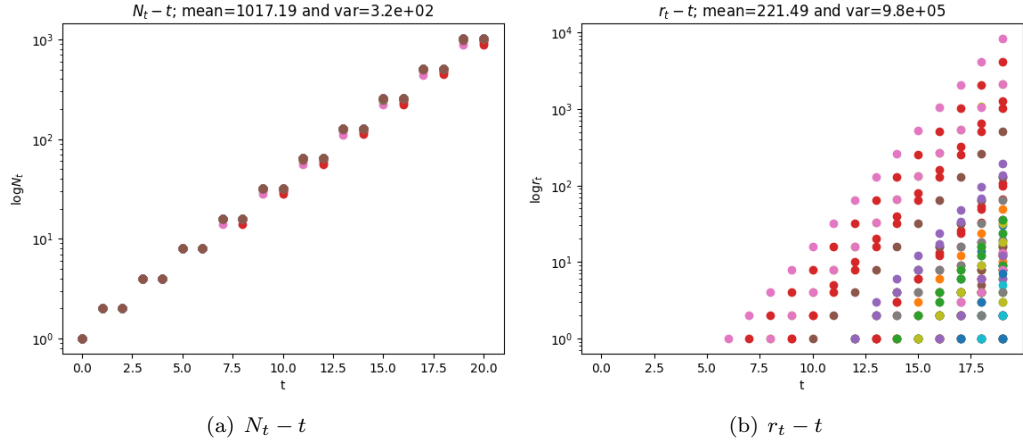
6.2   Mutation Hypothesis

(a) $N_t - t$

(b) $r_t - t$

Figure 6: Simulation on Mutation Hypothesis

As we can see in Figure 6, the variance is much greater than the mean value.

$$\text{Ex}(r_t) = 221.49 \tag{51}$$

$$\text{Var}(r_t) = 9.8 \times 10^5 \tag{52}$$

$$\frac{\text{Var}(r_t)}{\text{Ex}(r_t)} = 4.4 \times 10^3 \gg 1 \tag{53}$$

The bigger variance compared with acquired hereditary immunity hypothesis is very intuitive. On the acquired hereditary immunity hypothesis, the randomness is only introduced at the moment of selection or attack, **which happens only once**. However, under the mutation hypothesis, the source of randomness is mutation . There is mutation in each generation and therefore the **randomness will be accumulated**. That is the nature of large variance.

For the convenience of adjusting configuration of the simulation, I further analyzed the increasing trend of the ratio Var/Ex as shown in Figure 7.

It can be clearly seen that that the ration of variance to mean value increase rapidly with time.
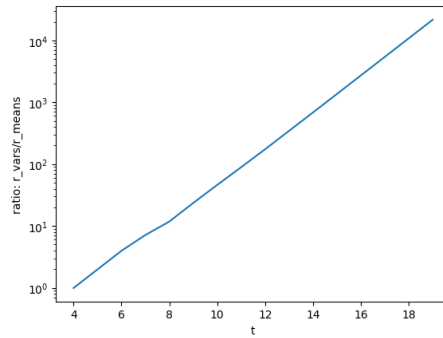
Figure 7: Increasing Ratio with Respect to Time

7 Conclusion

In this experiment we verified the mutation hypothesis, in accordance with the theory of natural selection. The simulation of the two different mechanisms helped me to grasp the nature of bigger randomness, while providing supporting evidence for our experiment. Thanks for the instruction from teaching assistant.

References

[1] S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics (Austin)*, 28(6):491–511, 1943.

```python
import numpy as np
import matplotlib.pyplot as plt
```

Acquired Hereditary Immunity Hypothesis

In [2]:
```python
t = 20 #time under consideration in the unit of cell cycle
N = [1]  #the list stores the number of yeasts at time t
for i in range(t):
    temp = 2*N[i]
    N.append(temp)
plt.scatter(np.linspace(0,t,t+1),N)
plt.gca().set_yscale('log')
plt.xlabel('t')
plt.ylabel(r'$N_t$')
plt.title(r'$N_t-t$')
```

Out[2]: Text(0.5, 1.0, 'N_t-t')
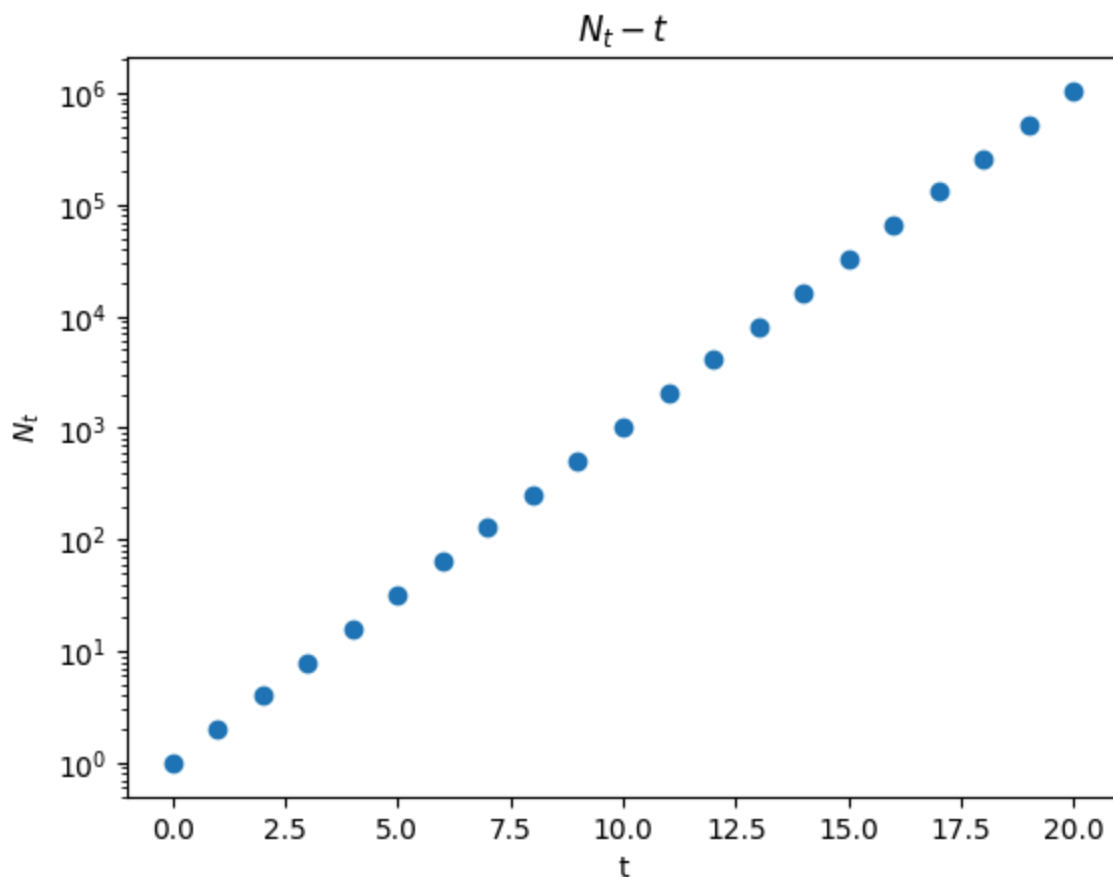
In [3]:
```python
p = 1e-1 # the chance of a yeast to survive an attack
r = []     # each element of the list is the number of resistant yeasts in one cultu
C = 96   # the number of similar cultures
for j in range(C):
    temp=0
```

```
        for i in range(len(N)):
            if np.random.rand()<p:
                temp+=1
        r.append(temp)
plt.hist(r,bins=6,rwidth=0.9)
mean,var=np.mean(r),np.var(r)
plt.title(f'The Distribution with the mean={mean:.2f} and var={var:.2f}')
```

Text(0.5, 1.0, 'The Distribution with the mean=2.41 and var=1.87')
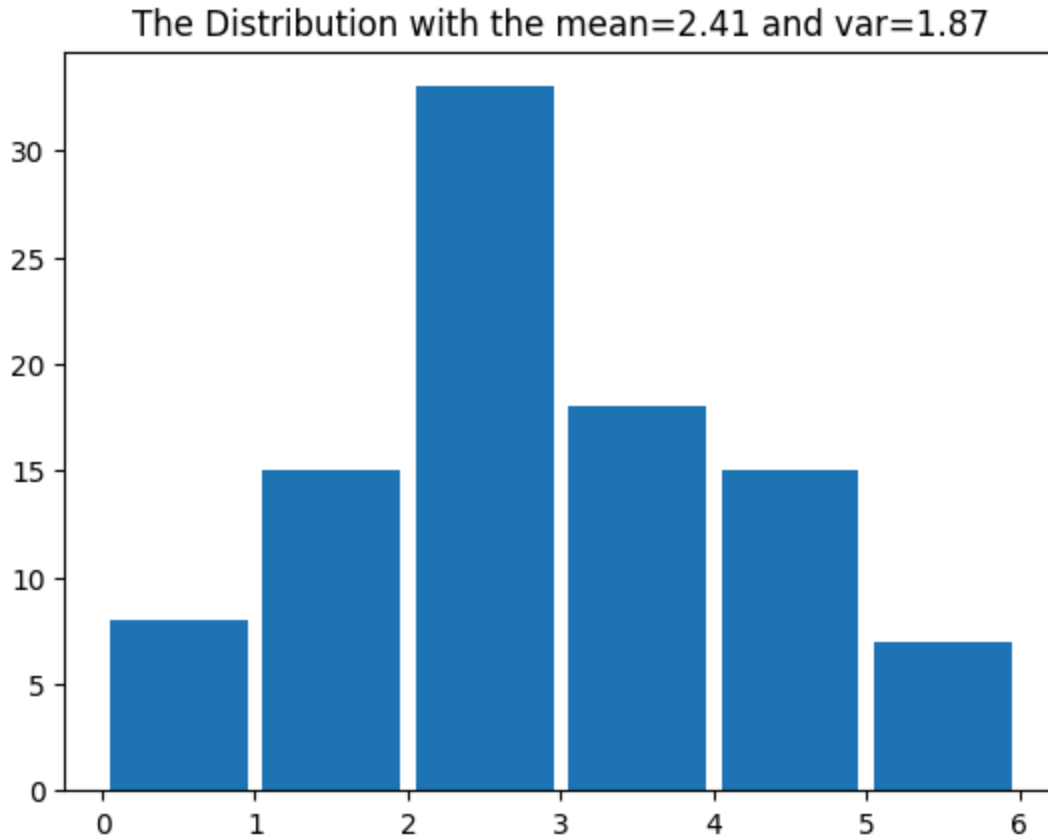
Mutation Hypothesis

```
r =[]
N = []  #the list stores the number of yeasts of several similar cultures, each ele
a = 1e-3 #p is the chance per time unit per yeast to
r_means=r_vars=[]
for j in range(C):
    n=[1] #for one culture start from one yeast
    for i in range(t):
        if i == 0: # at zero time, there was no resistant yeasts
            r_temp=0
            rr = [0]
        else:
            r_temp=0
            for k in range(n[i-1]):# there will be yeasts mutate
                if np.random.rand()<a:
                    r_temp+=1
            new_resist = 2*rr[i-1]+r_temp
```

```
                rr.append(new_resist)
            temp = 2*(n[i-1]-r_temp)
            n.append(temp)
        N.append(n)
        r.append(rr)
        #r_means.append(np.mean())
```
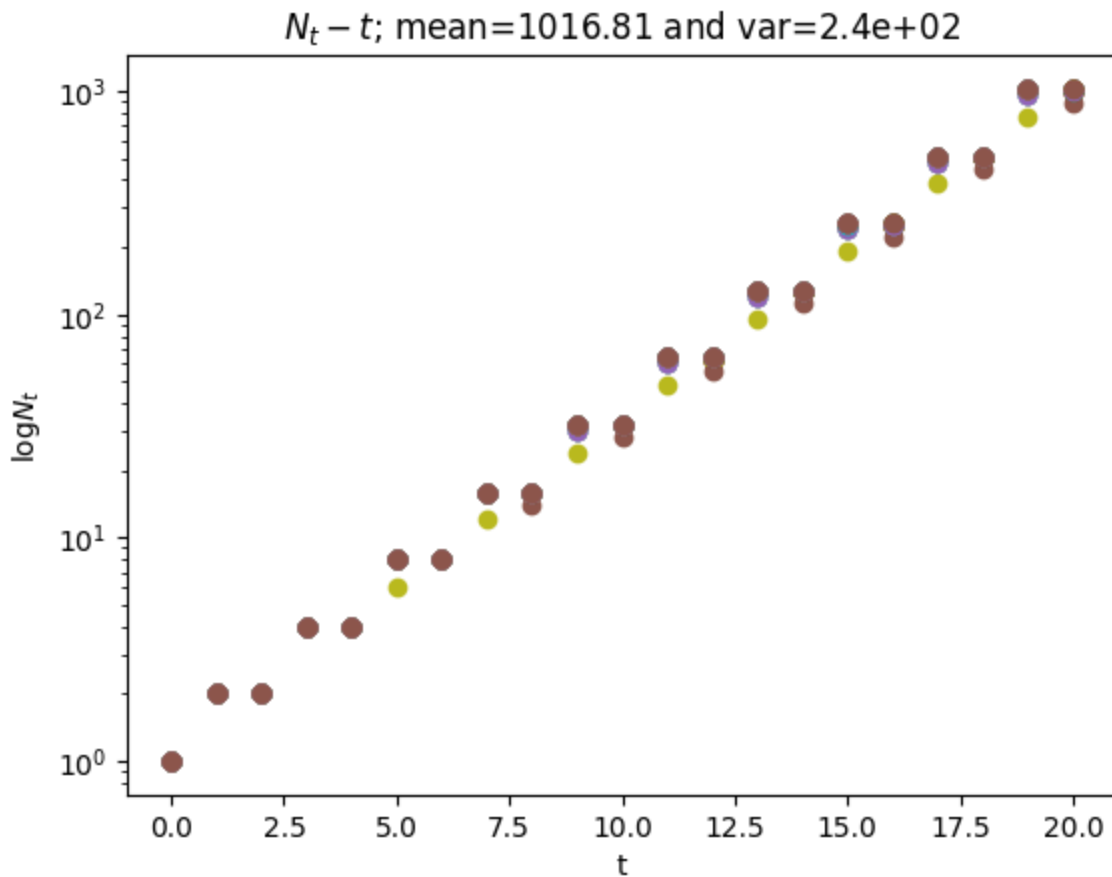
In [5]:
```
for i in range(C):
    plt.scatter(np.linspace(0,t,t+1),N[i])
plt.gca().set_yscale('log')
plt.xlabel('t')
plt.ylabel(r'log$N_t$')
N=np.array(N)
mean,var=np.mean(N.T[-1]),np.var(N.T[-1])
plt.title(r'$N_t-t$; '+f'mean={mean:.2f} and var={var:.2}')
var/mean
```
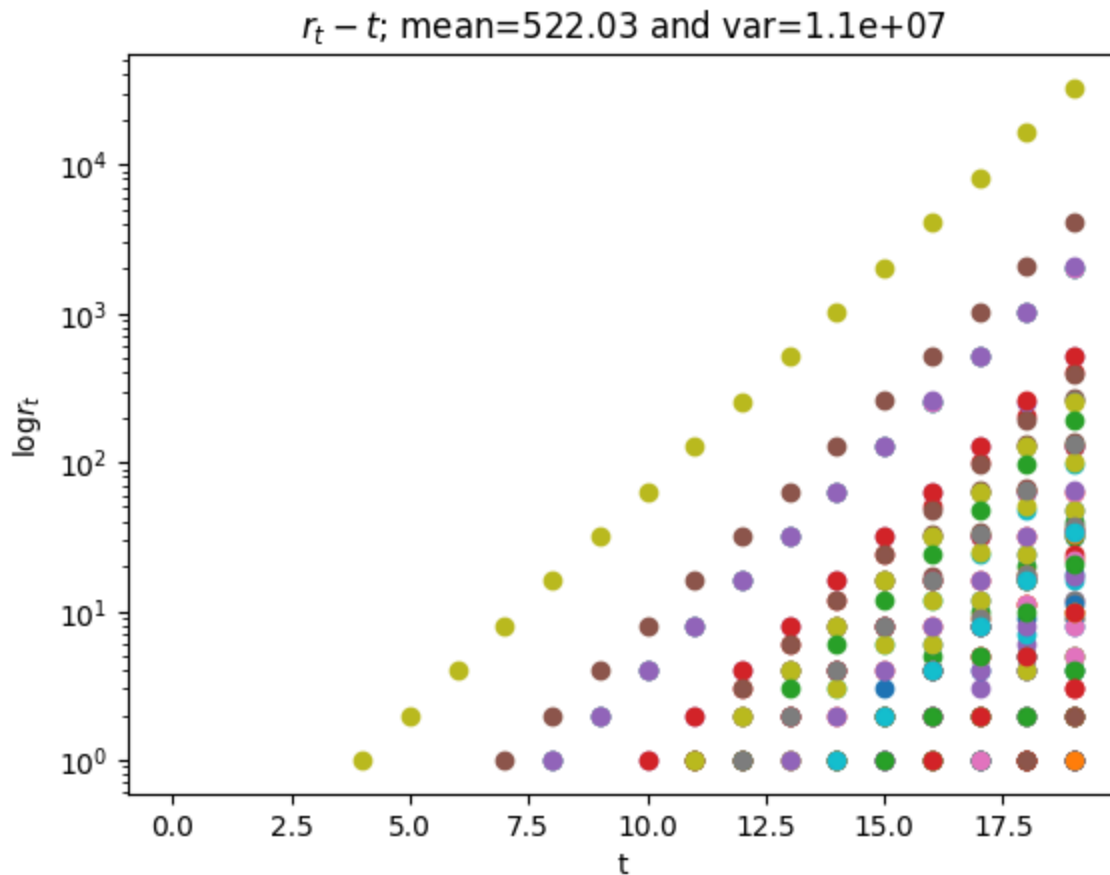
Out[5]: 0.23534149814575778

$N_t - t$; mean=1016.81 and var=2.4e+02

In [6]:
```
for i in range(C):
    plt.scatter(np.linspace(0,t-1,t),r[i])
plt.gca().set_yscale('log')
plt.xlabel('t')
plt.ylabel(r'log$r_t$')
r=np.array(r)
mean,var=np.mean(r.T[-1]),np.var(r.T[-1])
plt.title(r'$r_t-t$; '+f'mean={mean:.2f} and var={var:.2}')
var/mean
```

21607.949773645614

$r_t - t$; mean=522.03 and var=1.1e+07

In [7]:
```python
r_means=[]
r_vars=[]
for i in range(t):
    r_means.append(np.mean(r.T[i]))
    r_vars.append(np.var(r.T[i]))
r_means=np.array(r_means)
r_vars=np.array(r_vars)
plt.semilogy(np.linspace(0,t-1,t),r_vars/r_means)
plt.xlabel('t')
plt.ylabel('ratio: r_vars/r_means')
```

C:\Users\zyan\AppData\Local\Temp\ipykernel_1248\698680870.py:8: RuntimeWarning: inva
lid value encountered in divide
  plt.semilogy(np.linspace(0,t-1,t),r_vars/r_means)

Out[7]: Text(0, 0.5, 'ratio: r_vars/r_means')