Problem Analysis:

In order to improve the use efficiency under the same amount of resources, we consider using the group detection method. Question 1 requires us to make decisions about **the total number of tests, false-positive rates and false-negative rates**, establish mathematical models, and perform preliminary calculations and tests based on the number of stages and the number of groups in each stage.

First consider **the decision-making goals**. In order to perform the detection of the coronavirus more quickly, we must consider the three factors of the total number of tests, the false positive rate, and the false negative rate. Regarding the total number of detections, the smaller the total number of detections, the less resources can be saved, and more people can be tested under the premise of limited resources. Regarding the false-positive rate, that is, the percentage of negatives judged to be positives. As the grouped samples are positive, all members of the reorganization are determined to be positive, and further testing is performed until a positive individual is determined. The lower the false positive rate, the lower it can be Isolate the number of people to achieve efficient use of resources. Regarding the false-negative rate, that is, the percentage of positive being judged as negative, that is, regarding the error of the experimental instrument, reducing the number of tests can control the error of the experiment.

Second, consider two variables**-the number of stages and the group size of each stage**. The number of stages should be considered in light of actual conditions. We first look up relevant information to get the upper limit of the daily test phase as the value range of the number of phase variables. For the group size of each stage, the overall sample should be tested in groups. Since each sample is united, it is inseparable. For the convenience of calculation, we consider the factor of the test sample as the number of groups.

Third, we use statistical probability samples to detect relevant knowledge, and **use the characteristics of the overall sample to estimate part of the sample**. We use the diagnosis rate in Hong Kong to estimate the number of patients in the sample. Since the number of patients will be randomly distributed in each group sample, the number of patients will be randomly distributed among different numbers of groups to determine the total number of tests, so we use expectations to represent the average number of tests. Using the principle of random sampling, through multiple experiments, the total number of tests stabilized.

Finally, we conduct a comprehensive evaluation of the three decision-making goals. Because there are connections between the three decision-making goals, we use **gray correlation** to obtain comprehensive indicators. By adjusting the number of stages, the number of groups and the number of samples in the group to reduce the comprehensive evaluation indicators, the optimal group detection program is finally obtained.

The Construction of the Model:

**Step1 variables**

* The number of steps(steps)

Official information shows that the duration of a nucleic acid test to detect a sample is 6 hours, and the normal working time of a hospital clinic is 10 hours. Assuming that the hospital has a shift system during the epidemic, it is reasonable to consider that we believe that the detection cycle that can be performed in a day is two times, that is, the value range of the number of stages is 1-2. Since the number of stages equals to 1 means that all samples will be tested once, so we mainly consider the case where the number of stages is 2.

* Group size (N)

The group size of each stage (denoted by N) refers to the total number of samples that need to be tested at that stage. We want to group the overall sample for testing. Considering that each sample is a whole and cannot be divided, for the convenience of calculation, we only consider the case where the number of groups is the factor of the total number of samples. Where a represents the number of groups in the second stage, n represents the number of groups in the second stage, and the relationship is as follows:

**Step2 Three decision goals**

* Detection times

The calculation steps of the number of detections are as follows:

**Step1 the determination the number of infections in the total sample**

As the number of infected persons in the sample is different, the probability that the infected persons will be randomly assigned to different groups will be different, resulting in different total number of tests. We set the infection rate as , and the number of infections is:

**Step2 the consideration the total number of tests whose number of stages is 2**

Consider the case where the number of stages is 2. For the first stage, first divide the total number of samples in the first stage into groups a, where a is a factor of N. Therefore, the number of detections in the first stage is a. Since the number of infected people is randomly allocated to different groups, for each group, as long as there is one infected sample in the group, it means that the group is infected. The random distribution example diagram is as follows:

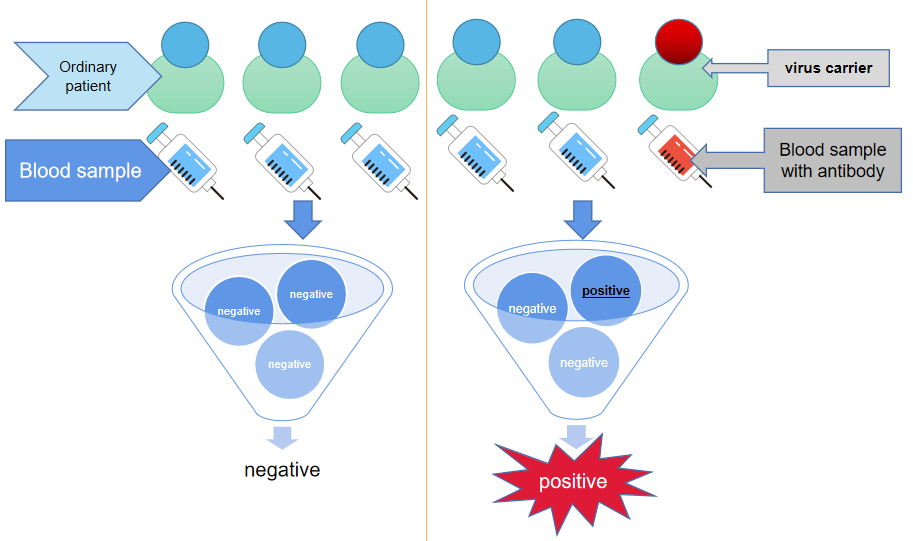


Figure :...

Then we discussed the number of groups with infected samples in the group after grouping. If the number of infected persons b is greater than the number of groups a, at most a group will be positive; if the number of infected persons b is less than the number of group a, then at most b groups will be positive.

Assuming that there are infected persons in group a and group c, the calculation formula for the total number of tests is as follows:

Because of the difference in c, the number of detections is different, and the number of infected people will be randomly assigned to different numbers of groups. From the meaning of the question, the number of infected persons in a group indicates that all members of the group are infected. The next step is to check until whether a single sample is infected. Therefore, we only need to consider the number of infected groups, so we refer to the probability of fakes. The principle. The principle of counterfeit probability is that there are some fake coins in a pile of coins. The quality of the fake coins is lower than that of the real coins. The pile of coins is randomly grouped. Among them, there are groups with fake coins whose quality is less than that of all real coins. There is a probability of the number of fake coin groups. Using the probability to find the expected number of detections indicates the average number of detections in different groups. The expected formula is as follows:

Therefore, we have to find the grouping method with the smallest expectation as the optimal group detection scheme with stage 2.

* False-positive rate

Regarding the false positive rate, that is, the percentage of negative being judged to be positive, because the grouped samples are positive, all members of the reorganization are determined to be positive, and further testing is performed until a positive individual is determined. The lower the false positive rate, the lower it can be reduced Number of people in quarantine. Suppose the false positive rate is , the calculation formula is:

* False-negative rate

Regarding the false negative rate, that is, the percentage of positives that are judged as negative, that is, the error of the experimental instrument. Let the probability of error of the experimental instrument be , and the calculation formula of the false negative rate is:

Step3 comprehensive evaluation model

From the perspective of decision-making goals, we chose the total number of tests, false positive rate, and false negative rate as indicators to evaluate the pros and cons of group detection methods. Since the various indicators are related to each other but have properties that reflect different aspects, a scientific and reasonable comprehensive evaluation model must be established. We choose the gray correlation model. The specific model is as follows:

Determine the overall sample size. Different groups will get different total number of tests, false positive rates, and false negative rates, as shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| The number of groups | The times of tests | False-positive rate | False-negative rate |
|  |  |  |  |
| ... | ... | ... | ... |
|  |  |  |  |

* Select the parent index

In step2, we find the formula about the total number of tests, false positive rate, and false negative rate. We can find that the false negative rate and false positive rate are related to the total number of tests, and the total number of tests most intuitively reflects the pros and cons of group detection , So choose the parent index of the total number of detections.

* Calculate the correlation coefficient

Define the correlation coefficient between the parent index and at the th point as:

In which can be calculated as follow:

And the formula of a and b are:

In which the range of is ,and we take .

* Calculate the degree of relevance
* Build a comprehensive evaluation model

From the relevance obtained in the previous step, calculate the weight of each indicator:

As the weight of each indicator, a comprehensive indicator is obtained.

Finally, the grouping number corresponding to the smallest comprehensive index is the grouping method with the best overall sample number N.

**Solution of the model:**

Regarding the value of the infection rate, we use the characteristics of the overall sample to estimate part of the sample, that is to say, use the confirmed rate in Hong Kong to estimate the number of patients in the sample. According to data from the National Bureau of Statistics, there are currently 70,000 infected people in Hong Kong, of which the total population of Hong Kong is 7,345,670. The infection rate is one in a thousand, so we take A=0.001.

When calculating the number of tests, since the probability of a group being positive after grouping is easy to calculate, as long as there are positive individuals in the group, it means that the group is positive. It is too complicated to calculate the probability of how many groups are positive samples, but they must exist, so we choose a large Random simulation gets the probability and expectation. Below we show the image of random times and expectation:

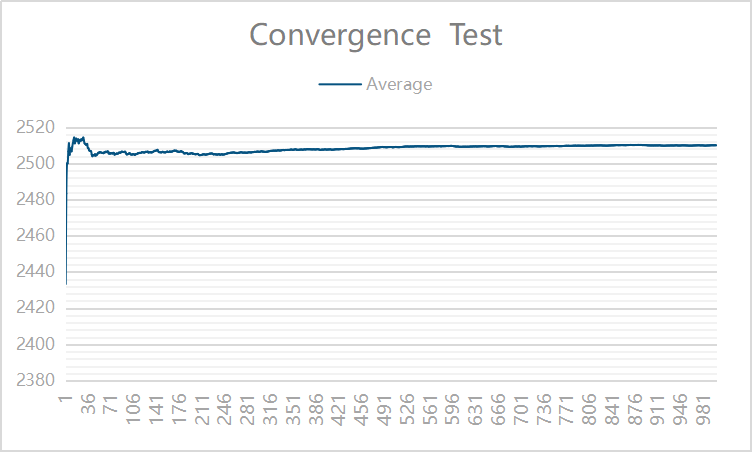


Figure : ...

From the image, when the random number reaches 350 times, the expectation of the detection times basically tends to be the same, fluctuates in a very small range. In order to reduce the calculation error, we take the random number of 10000 when calculating the expected number of detections.

Calculate the values under different groups by using the formula of false positive rate and false negative rate in the model. Here is an example where the total number of samples is 20,000, and the data values in different grouping situations are shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| The number of infected people | The times of tests | False-positive rate | False-negative rate |
| 20 | 1251.763979 | 0.031328 | 0.000285 |
| 40 | 1764.176201 | 0.042819 | 0.000396 |
| 100 | 2772.505343 | 0.065685 | 0.000624 |
| 200 | 3891.00172 | 0.08849 | 0.00087 |

According to the gray correlation model, the correlation between the three indicators and the parent indicator is shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| The number of infected people | The times of tests | False-positive rate | False-negative rate |
| correlation | 0.999999999 | 0.910954924 | 0.910954837 |

The weight of each indicator is shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| The number of infected people | The times of tests | False-positive rate | False-negative rate |
| Weight | 0.745446 | 0.139456 | 0.115098 |

The comprehensive evaluation indicators are:

The smaller the comprehensive evaluation index value, the more efficient the grouping scheme.

The comprehensive indicators calculated for each group are shown in the following table:

|  |  |  |
| --- | --- | --- |
| The number of groups | The comprehensive indicator | Rank |
| 1020 | 0.972138929 | 1 |
| 1024 | 0.97210998 | 2 |
| 1071 | 0.972096827 | 3 |

The optimal group with a population of 20,000 is obtained as 200 groups.

Repeat the above steps to get the optimal grouping scheme in the case of different sample numbers. The results are shown in the following table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Total detection | Prevalence | Number of patients | Number of groups | Time of tests | False-positive rate | False-negative rate | score |
| 4000 | 20 | 0.005 | 271.2 | 553.954888 | 0.065689 | 0.000624 | 0.054968158 |
| 4000 | 40 | 0.01 | 379.8 | 777.215848 | 0.089354 | 0.000877 | 0.070881357 |
| 10000 | 20 | 0.002 | 439.8 | 881.818882 | 0.042202 | 0.00039 | 0.037706762 |
| 10000 | 100 | 0.01 | 935.6 | 1944.595555 | 0.0909 | 0.000891 | 0.071062632 |
| 20000 | 20 | 0.001 | 605.2 | 1251.763979 | 0.031328 | 0.000285 | 0.027861071 |
| 20000 | 40 | 0.002 | 867.8 | 1764.176201 | 0.042819 | 0.000396 | 0.03767999 |
| 40000 | 40 | 0.001 | 1212.5 | 2503.446565 | 0.031328 | 0.000285 | 0.02438544 |

**sensitivity analysis**

In the above model, we use the diagnosis rate in Hong Kong as the prevalence rate in the test sample, but the prevalence rate will change for different regions. In areas with high epidemics, the prevalence rate will increase; in safe areas, the prevalence rate is relatively small. In order to cope with the detection of different areas, we adjusted the infection rate to obtain the optimal grouping strategy corresponding to different infection rates. Below we have done a sensitivity analysis for the overall sample of 4000, 10000 and 20000

The following table shows the grouping strategies corresponding to the total number of people tested for 4000 and different infection rates:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Total detection | Prevalence | Number of patients | Number of groups | Time of tests | False-positive rate | False-negative rate | score |
| 4000 | 4 | 0.001 | 122.6 | 250.35529 | 0.030939 | 0.000282 | 0.01545349 |
| 4000 | 10 | 0.0025 | 198 | 393.948366 | 0.046487 | 0.000433 | 0.031818349 |
| 4000 | 20 | 0.005 | 271.2 | 553.954888 | 0.065689 | 0.000624 | 0.054968158 |
| 4000 | 40 | 0.01 | 379.8 | 777.215848 | 0.089354 | 0.000877 | 0.070881357 |

When the total number of people tested is 10,000, the results are as follows:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Total detection | Prevalence | Number of patients | Number of groups | Time of tests | False-positive rate | False-negative rate | score |
| 10000 | 20 | 0.002 | 439.8 | 881.818882 | 0.042202 | 0.00039 | 0.037706762 |
| 10000 | 50 | 0.005 | 683.2 | 1385.714887 | 0.065251 | 0.00062 | 0.054973315 |
| 10000 | 100 | 0.01 | 935.6 | 1944.595555 | 0.0909 | 0.000891 | 0.071062632 |