Syphilis Lab Name: 70

Sometimes, there are large numbers of people that need to get screened for a particular disease or issue. For example, large numbers of American soldiers needed to get tested for syphilis during war time, and large numbers of professional athletes need to get tested for performance enhancing drugs. Since the number of individuals to be tested is quite large, we can expect that the cost of testing will also be large. How can we reduce the number of tests needed to screen everyone and thereby reduce costs?

Suppose that you have a large population (N) that you wish to test for certain characteristics in their blood or urine. If the blood or urine could be pooled by putting G samples together and then testing the pooled sample, then the number of tests required might be reduced. In this lab, you will determine the optimal group size (G) given the probability of an individual testing positive (p).

Question 1: This lab suggests that there is an optimal group size (G) of pooled samples that minimizes the total number of tests that need to be performed. Intuitively, why might the optimal value, G, not be very small? Why would it not be very big?

- it would be too expensive to test many people

- because every combined sample would test positive

2: If there are N people to be tested in sizes of group G, then what is the initial number of tests needed?

$$\frac{N}{G}$$

3: Fill in the blanks below.

The probability of an individual testing positive is p.

The probability of an individual testing negative is $1-p$.

The probability of two individuals both testing negative is $(1-p)^2$.

The probability of G individuals all testing negative is $(1-p)^G$.

The probability of at least one of the G individuals testing positive is $1-(1-p)^G$.

4: How many groups will need to be retested a second time?

$$\left(\frac{N}{G}\right)\cdot\left(1-(1-p)^G\right)$$

5: If we choose to test everyone in those second groups individually, then how many additional tests will we need?

$$\frac{N}{G} \cdot \left(1-(1-p)^G\right) \cdot G$$

6: How many tests ,T, have been administered total (counting both the first and second rounds of testing)?

$$T = \frac{N}{G} + \frac{N}{G} \cdot \left(1-(1-p)^G\right) \cdot G = N\left(\frac{1}{G} + (1-(1-p)^G\right.$$

7: You would like to minimize this equation for T with respect to G in order to minimize costs. To do this, find an expression for $\frac{dT}{dG}$.

$$T = \frac{N}{G} + N - N(1-p)^G$$

$$\frac{dT}{dG} = -\frac{N}{G^2} - N(1-p)^G \ln(1-p)$$

8: Try solving analytically for the value of G that minimizes T. Explain what goes wrong.

$$-\frac{N}{G^2} - N(1-p)^G \ln(1-p) = 0$$

$$\frac{1}{G^2} + (1-p)^G \ln(1-p) = 0$$

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

can't isolate G easily

15: We are now able to determine exactly how many tests we expect to administer on a population of 1,000,000 if we expect p=0.0001 (i.e., if we expect 100 people to be infected). Fill in all of the table info below. Some of the answers have been filled in so that you know that you are on the right track. Remember: if the infection rate p gets above a certain percentage, you will want to do the remaining tests individually.

Round	Population to be tested this round	p this round	Group Size $\frac{1}{\sqrt{p}}$	# of tests this round $\frac{N}{G}$	Population needing to be tested next round $(1-(1-p)^G)\cdot N$
1	1,000,000	0.0001	100	10,000 $1{,}000{,}000 \div \frac{1}{\sqrt{.0001}}$	9,951
2	9,951	$\frac{100}{9951}$	10	makes more sense to round up 998	956
3	956	$\frac{100}{956}$	3 ~~14~~	239/ 309	341/ 270
4	341/ 270	$\frac{100}{270}$ or $\frac{100}{341}$	do individually since $p>.29$	341/ 270	0

Question 16: How many total tests needed to be administered? Was this number significantly lower than the total population?

10,000 + 998 + 309 + 270

= 11,577

11,585

9: Instead of using calculus, this minimization problem is much better suited for data analysis. For a given value of p, you can approximate the optimal value of G by plotting your equation from #6 in Desmos or your graphing calculator and numerically finding the minimum point. Make a different graph for each of the probabilities listed below in order to fill in the table. I've given you two answers so that you know that you are on the right track.

Probability	0.15	0.1	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	0.00001
Group Size	3	4	5	7	11	15	32	45	100	317

Hint: model with power function so log-log plot

10: Now plot the data points that you have obtained in #9. Given the shape of the data, explain why a power function $y = ax^n$ is better suited than an exponential function $y = ab^x$. there are asymptotes at $x = 0$ and $y = 0$

11: Use a log-log plot to obtain the power function. Remember, you will plot ln(p versus ln(G) to obtain a linear function of the form $\ln(G) = m(\ln(p)) + b$. From there, your power function will be $G = e^b(p^m)$.

$$G = 1.027 \cdot p^{-.499} \approx \frac{1}{\sqrt{p}}$$

12: We are now ready to think about when it makes sense to group at all. Set up an inequality using your equation for T from Question #6 to determine when it would NOT make sense to group individuals into groups of size G.

$$N\left(\frac{1}{G} + 1 - (1-p)^G\right) > N$$

13: Graph the inequality in #12 to determine what values of p dictate that we should test everyone individually.

$$\frac{1}{G} + 1 - (1-p)^G > 1$$

$$\frac{1}{\left(\frac{1}{\sqrt{p}}\right)} + 1 - (1-p)^{\frac{1}{\sqrt{p}}} > 1$$

14: Explain why your answer to #13 makes sense intuitively.

$$\sqrt{p} + \cancel{1} - (1-p)^{\frac{1}{\sqrt{p}}} > \cancel{1}$$

$$\sqrt{p} > (1-p)^{1/\sqrt{p}}$$

for $p > .29$ it doesn't make sense

when the % of people infected is high it does not make sense to pool people together because you will need to retest.