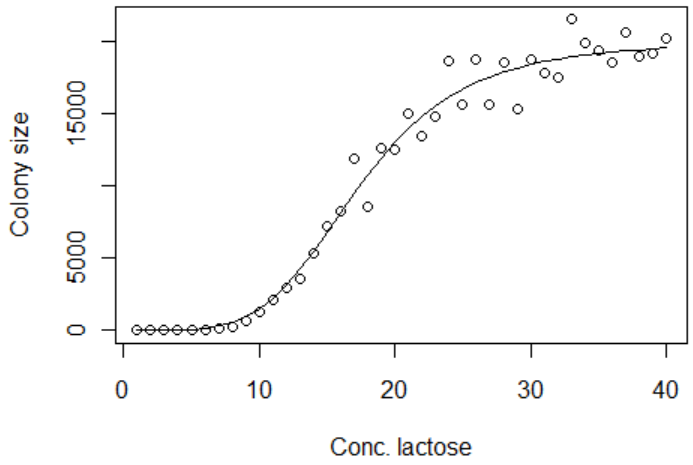
Name:

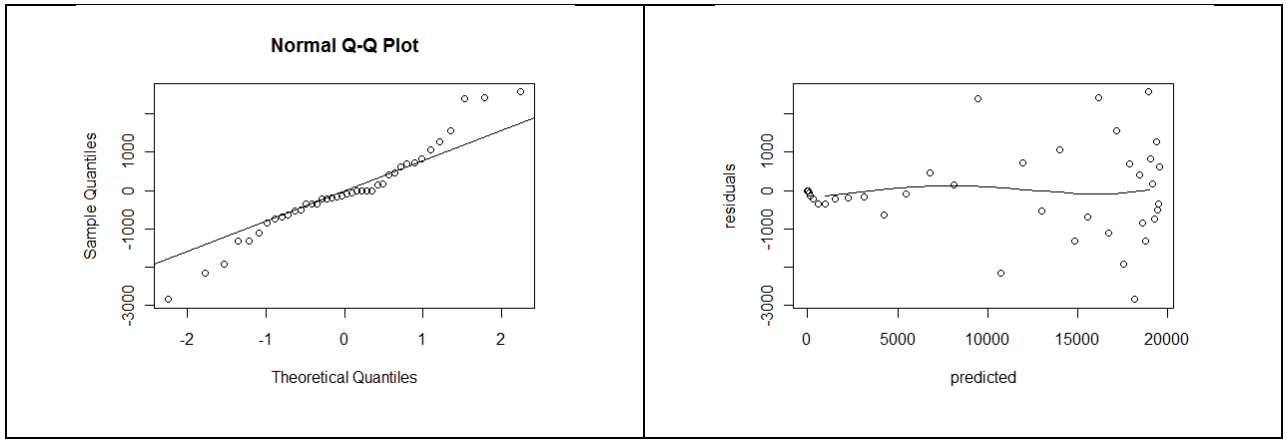
Question 1

Your lab mate has obtained the data shown in the first panel below, showing the colony size of a genetically modified yeast after twelve hours growth in a chemostat as a function of lactose concentration in the growth medium. He has fit the data to a Hill model (curve shown in the figure) using non-linear regression, and reports that the Hill model exponent is 4.5 (95% CI: 3.7 – 5.4), and that $k_{1/2}$ is 17.4 (95% CI: 16.6 – 18.4).

Having just completed qBio lab #2, you of course prepare a Q-Q plot of the residuals, and a plot of residuals vs. fitted values, as shown in the second and third plots below.



$$\binom{Colony}{Size} = \binom{Colony}{Size}_{max} \cdot \frac{[lactose]^n}{(k_{1/2})^n + [lactose]^n}$$



Comment critically on the model and the reported results. What suggestions do you have for your lab mate (other than to enroll in the qBio course next year)?

Question 2

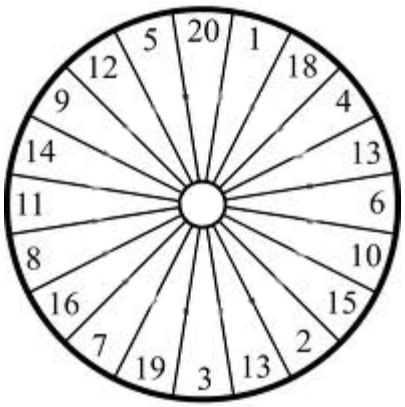
Systolic blood pressure of children in primary school has a mean of 110 mmHg and a standard deviation of 12 mmHg. You are planning to ascertain whether a proposed change to the lunch served in school has a significant effect on the students' blood pressure. You consider a change in blood pressure of 15 mmHg or more to be significant. 144 students (and their parents) have agreed to participate in your study. Half will be chosen at random to be fed the traditional lunch for four weeks, and the other half will be served the proposed new lunch. At the end of the four week study period, you will measure each student's blood pressure and analyze the results.

What size change in blood pressure do you think you have a reasonable chance of detecting? Do you have enough patients enrolled to have a reasonable chance of detecting changes in systolic blood pressure of 15 mmHg or more if the study is carried out as described above? Explain.

How might you alter the experimental plan to be able to detect smaller effects of the new lunch on systolic blood pressure? Explain why your altered plan would be an improvement.

Question 3

You are responsible for statistical analysis of data from a clinical trial. As a routine part of the trial, each patient's fasting blood glucose level is taken when they visit the study's clinic. The trial's director has received a report suggesting that one of the technicians in the clinic may be skipping the blood glucose test and making up the data. You are told that the tech was seen throwing a single dart at the board in the on-call room (shown at right) before filling in a value on each patient's chart.



Checking prior data, you determine that fasting glucose levels have a mean of 4.9 mmol/L and a standard deviation of 1.1 mmol/L. You retrieve the fasting blood glucose data collected by the technician on the day in question, as well as from another day when the data was obtained by a trusted nurse at the clinic.

```
> df.suspect
   patient.id fasting.glucose
1        3545             4.1
2        3213             5.7
3        5556             4.1
4        9544             5.9
5        5701             4.9
6        9451             4.8
7        4155             4.7
8        5487             5.0
9        2483             4.5
10       5032             5.8
11       4026             4.6
12       2365             4.1
13       2915             5.8
14       7543             4.7
15       5288             4.9
16       4487             5.3
17       6828             5.3
18       6412             5.4
19       5079             4.2
20       6723             4.2
21       5719             4.5
22       6246             5.7
23       3473             5.0
24       6778             5.2
25       5049             5.3
```

```
> df.trusted
   patient.id fasting.glucose
1        4643             6.2
2        5887             5.3
3        2724             3.6
4        4370             5.4
5        1524             5.6
6        3289             5.5
7        6044             6.3
8        8459             5.5
9        8995             8.7
10       6566             5.2
11       1441             2.8
12       6740             5.4
13       7029             4.2
14       8342             4.9
15       4100             4.5
16       3215             5.7
17       7534             3.8
18       2246             5.4
19       4963             5.2
20       1139             6.7
21       3372             5.8
22       2680             5.6
23       6771             3.7
24       3160             7.3
25       6313             5.6
```

What kind of probability distribution would you expect unadulterated fasting blood glucose levels to follow? Why?

If the technician is generating data from a dart board as accused, what kind of distribution might you expect the data to be from?

Prepare a Q-Q plot to help determine if the technician's data is being fabricated.

What can you conclude from the plot you just made?

What further analyses would you perform?

Question 4

Explain the difference between the t-test and the Mann-Whitney test (aka the two sample Wilcoxon rank sum test). What factors would you consider when deciding which test to use?

[no question on this page – just extra space]

[no question on this page – just extra space]

[no question on this page – just an extra page of graph paper]