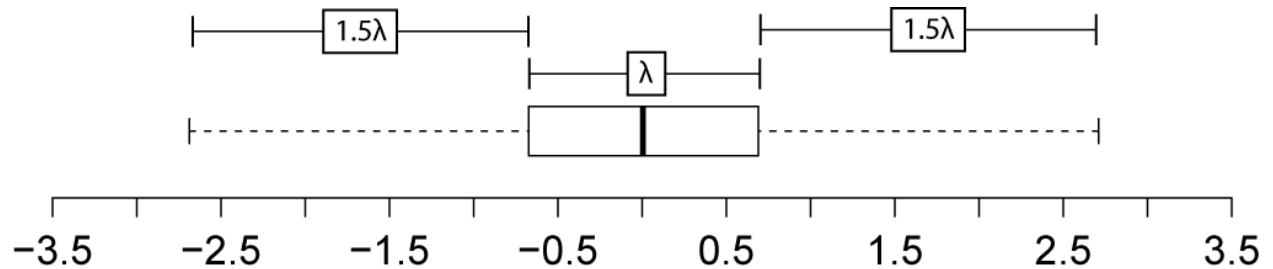


Name:

(please write your name on all pages)

Question 1

When R draws a boxplot, the default behavior is to draw the whiskers such that the distance from the central box is 1.5 times the width of the central box, as shown in the figure below.



- A) Show that for normally distributed data, this practice results in approximately 1 in 100 points lying outside of the whiskers.

(Hint: you may want to use the reference data on page 3)

Name:

- B) Why do you think the authors of R have adopted this behavior as the default, as opposed to simply drawing the whiskers to show the most extreme values in the data being plotted?

Name:

Reference Data for Question 1

x	qnorm(x)
0.500	0.0000000
0.550	0.1256613
0.600	0.2533471
0.650	0.3853205
0.700	0.5244005
0.750	0.6744898
0.800	0.8416212
0.850	1.0364334
0.900	1.2815516
0.905	1.3105791
0.910	1.3407550
0.915	1.3722038
0.920	1.4050716
0.925	1.4395315
0.930	1.4757910
0.935	1.5141019
0.940	1.5547736
0.945	1.5981931
0.950	1.6448536
0.955	1.6953977
0.960	1.7506861
0.965	1.8119107
0.970	1.8807936
0.975	1.9599640
0.980	2.0537489
0.985	2.1700904
0.990	2.3263479
0.995	2.5758293
0.996	2.6520698
0.997	2.7477814
0.998	2.8781617
0.999	3.0902323

Name:

Question 2

A recently published study demonstrated that the systolic blood pressure of graduate students increases by an average of 18 mmHg (SD=8 mmHg) due to the stress of taking a midterm exam. Your roommate is planning an experiment to test whether this stress-induced change in blood pressure is different in 2nd year students as compared to 1st year students who are enrolled in your graduate program. He has asked you to look over his experimental plan before submitting the proposed protocol to the IRB for the approval to carry out research on human subjects.

The proposed plan calls for 32 of the 39 first year students and 32 of the 37 second year students enrolled in your graduate program to participate. After receiving informed consent, the protocol calls for students to be monitored while taking their midterm exams. The blood pressure of each student will be measured 20 minutes before the exam begins, and again one hour into the exam (the exam is so unimaginably difficult that it cannot possibly be completed in less than one hour).

Because this study involved before-and -after measurements of systolic blood pressure, your roommate uses the relation...

$$N = 4 \left(\frac{SD}{precision} \right)^2$$

...to estimate how large a difference in stress response he'll be able to measure between 1st and 2nd year students (50% power; 95% CI). Substituting N = 64 and SD = 8 mmHg, he computes precision = 2 mmHg.

A) Comment critically on your roommate's analysis; be as quantitative as possible.

Name:

- B) Suggest an alternative experimental plan that would allow for more statistical power than your roommate's proposed procedure would have yielded. Can you quantify?

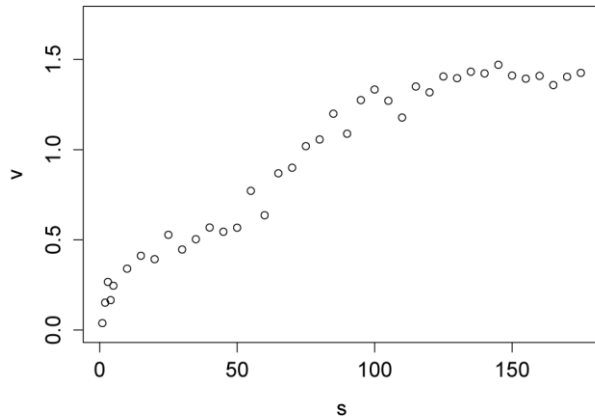
Name:

- C) State any additional advantages and/or disadvantages of your proposed plan, as compared to your roommate's.

Name:

Question 3

Data in the figure below were obtained for an enzymatic reaction, where v is the rate of consumption of the substrate and s is the substrate concentration. Three models have been proposed to explain the data, and each was fit to the data using a non-linear least squares procedure in R. A plot of the best fit curve for each model is shown.

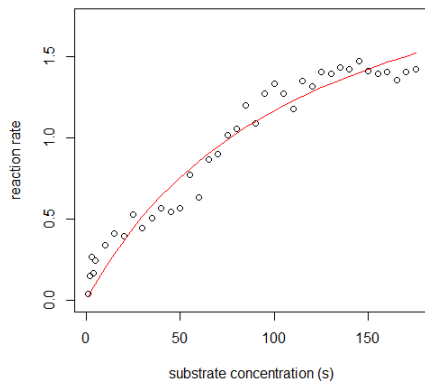


Model 1
$$v = \frac{V_{max} \cdot s}{K + s}$$

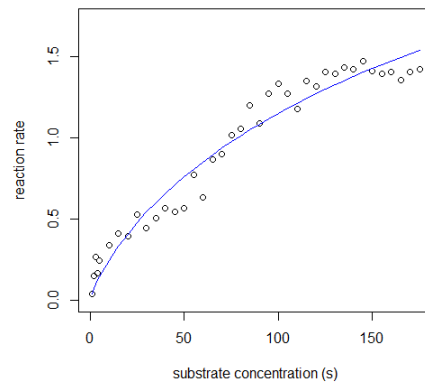
Model 2
$$v = \frac{V_{max} \cdot s^n}{K^n + s^n}$$

Model 3
$$v = \frac{V_1 \cdot s}{K_1 + s} + \frac{V_2 \cdot s^n}{K_2^n + s^n}$$

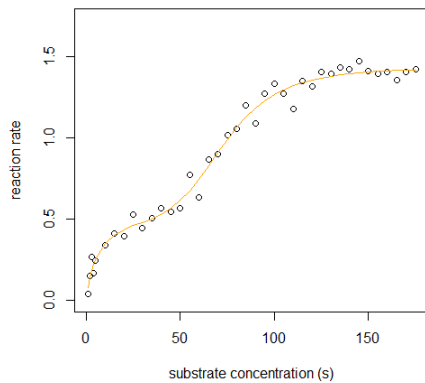
Model 1



Model 2



Model 3



Name:

A) Rank the models in order of their sum-of-squares of their residuals? Explain your reasoning.

B) Which model do you think best explains the data? Explain your reasoning.

Name:

C) What statistical tests would you perform to formally justify your response to part B? Justify your answer.

D) What does the best fit model suggest about the underlying mechanism of the consumption of substrate?

Name:

Question 4

A) Explain the difference between the SD and the SEM.

B) When planning and executing experiments, what can you do to reduce the SD of your measurements? Can you drive the SD all the way to zero? What determines how close to zero you can practically get?

Name:

- C) When planning and executing experiments, what can you do to reduce the SEM? Can you drive the SEM all the way to zero? What determines how close to zero you can practically get?

Name:

[extra space if you need it]

Name:

[extra space if you need it]

Name:

[extra space if you need it]