Question 1

An investigator wishes to determine whether sitting for a qBio midterm exam raises the blood pressure of students. Because this study involves human subjects, the experimental protocol, including data analysis, needs to be submitted to the Institutional Review Board (IRB) for approval.

The investigator plans to measure the blood pressure of each of the 15 students before (x_i) and after (y_i) the exam. The investigator claims that a one-tailed t-test comparing the x_is and y_is will be appropriate. If the p-value of this test is less than 0.05, the study will conclude that the exam significantly raises blood pressure, and will recommend that further exams be cancelled to protect the students' health.

Comment on the investigator's proposed statistical analysis (assume that no additional data can be collected for this study). Suggest any improvements to the proposed analysis and interpretation of the data, and justify those suggestions.

Question 2

Consider a simplified model of a channel where there are only two states, open and closed. At a particular voltage, the probably of an open channel being found in a closed state after some time period τ is 50%. The probability of a closed channel being found in an open state after the same time period is unknown, and will be denoted x.

The evolution of the state of this system can be represented by the matrix equation…

$$\begin{bmatrix} P(closed)_{i+1} \\ P(open)_{i+1} \end{bmatrix} = \boldsymbol{M} \begin{bmatrix} P(closed)_{i} \\ P(open)_{i} \end{bmatrix}$$

…where the time interval between state i and state i+1 is τ.

A) Write the matrix M in terms of x.

B) If closed channels never re-open at the prevailing voltage, what would the equilibrium state of the system be?

C) Can you find a value of x for which this system does not reach a stable equilibrium, but rather exhibits a sustained oscillation? Justify your reasoning.

D) If, at equilibrium, $\boldsymbol{P(open)=\frac{1}{3}}$, what is the the opening transition probability, x?

Question 3

In Dr. Elemento's work on discovering regulatory sequences from expression data, the mutual information statistic was computed to ascertain whether the presence or absence of a given motif is informative about a gene sequence being found in a particular cluster. A non-parametric test was used to ascertain if the mutual information value obtained was statistically significant.

Why was a non-parametric test used in this case? What are the advantages and disadvantages of non-parametric tests?

Question 4

The following data were obtained for the growth rate of *E. coli* as a function of glucose concentration:

Glucose Concentration (μM)	Growth Rate (hr^{-1})			
	Trial I	Trial II	Trial III	Average
0.1	0.17	0.18	0.20	0.183
0.5	0.44	0.43	0.41	0.427
1.0	0.59	0.56	0.59	0.580
5.0	0.71	0.75	0.72	0.727
10.0	0.78	0.76	0.74	0.760
20.0	0.78	0.78	0.75	0.770

Hypothesizing Michaelis-Menten like kinetics, we can write the growth rate, v, as a function of nutrient concentration, s:

$$v = v_{max} \frac{s}{K_m + s}$$

This can be rearranged so that the data are amenable to linear regression:

$$\frac{1}{v} = \frac{1}{v_{max}} + \frac{K_m}{v_{max}} \frac{1}{s}$$

Using R to regress (1/v) vs. (1/s), we get…

```
> df <- data.frame(s=c(0.1, 0.5, 1.0, 5.0, 10.0, 20.0),
                   v=c(0.183, 0.427, 0.580, 0.727, 0.760, 0.770))
> df$s.recip <- 1/df$s; df$v.recip <- 1/df$v
> df
     s     v  s.recip  v.recip
1  0.5 0.183 2.000000 5.464481
2  0.9 0.427 1.111111 2.341920
3  1.0 0.580 1.000000 1.724138
4  5.0 0.727 0.200000 1.375516
5 10.0 0.760 0.100000 1.315789
6 20.0 0.770 0.050000 1.298701

> model <- lm(v.recip ~ s.recip, data=df)
> summary(model)

Call:
lm(formula = v.recip ~ s.recip, data = df)

Residuals:
       1        2        3        4        5        6 
-0.03326  0.18238 -0.01812 -0.03293 -0.05092 -0.04715
```

(continued next page)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.32499    0.04856   27.28 1.07e-05 ***
s.recip      0.41728    0.01161   35.95 3.57e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1007 on 4 degrees of freedom
Multiple R-squared: 0.9969,     Adjusted R-squared: 0.9961
F-statistic:  1293 on 1 and 4 DF,  p-value: 3.573e-06

> confint(model)
                2.5 %    97.5 %
(Intercept) 1.1901506 1.4598219
s.recip     0.3850517 0.4494994
```

We thus conclude $V_{max} = 1.32499^{-1} = 0.755$, and that the 95% CI for this model parameter ranges from $1.4598^{-1} = 0.685$ to $1.1902^{-1} = 0.840$.

Similarly, we conclude $K_m = (0.41728)(0.755) = 0.315$, and the 95% CI ranges from $(0.385)(0.755) = 0.291$ to $(0.449)(0.755) = 0.339$.

Comment critically on the above analysis. State what you might do differently, and justify your suggestions.

(no questions on this page—it is provided for extra space)

(no questions on this page—it is provided for extra space)