BIOE 210, Spring 2022

Homework 9

Due Monday, 3/28/2021 by 5:00pm.
Upload your answers to Gradescope. If submitting a single PDF,
you must mark the location of all answers.

Part I

Consider the logistic regression model

$$\log(\text{odds}(\text{HeartAttack})) = -7.98 + 0.0731[\text{MAP}] - 0.0112[\text{Cholesterol}]$$

where MAP is the mean arterial pressure measured in mmHg and cholesterol has units of mg/dL. The model predicts the risk of having a heart attack in the next year.

1. What is the probability that a person with a MAP of 110 mmHg and cholesterol levels of 170 mg/dL will have a heart attack in the next year?
2. How would a 1 mg/dL increase in cholesterol affect a person's risk of heart attack?
3. Identify a MAP and cholesterol level that would correspond to a 50% probability of having a heart attack in the coming year.

Part II

Gene editing is a rapidly expanding branch of biomedical research. Newly-discovered enzymes allow bioengineers to modify the genetic code of an organism. One such tool is the CRISPR-Cas9 complex. CRISPR (clustered regularly interspaced short palindromic repeats) refers to a family of DNA that is stored in the genetic code of select prokaryotes. These DNA sequences are acquired from bacteriophage after a survived infection and act as an adaptive immune system for a single cell. When there is a repeat exposure to the same strain of bacteriophage, these sequences of DNA are transcribed and combine with Cas9, guiding them to the foreign genetic material for destruction. This guiding happens stochastically and is the result of complementary nucleotide interactions. Clearly, CRISPR-Cas9 could be put to great use in the medical field if given the right target. There are already clinical trials in the works for a CRISPR-Cas9 engineered cancer fighting "CAR T-cell" therapy.

To apply CRISPR-Cas9, a plasmid encoding a 20 nucleotide protospacer guide RNA (gRNA) and the Cas9 protein is introduced to the cell. Plasmids can be transfected directly into cells, or viral vectors are used to transduce the cells. Despite being one of the most efficient and reliable means of genetic modification, Cas9 activity is still promiscuous, inconsistent, and requires a protospacer adjacent motif (PAM). In other words, there are several factors that go into determining if editing will be successful or not:

- Protospacer length.
- PAM (optimized or not).
- Presence of DNA methylation.
- Number of mismatches between the guide and target sequences.

Your goal is to use logistic regression to predict the probability of successful cleavage by Cas9. Download and load the file `cas9.mat`. The file contains a MATLAB table `cas9` with 100 observations of the above predictors and the success of cleavage.

1. Fit a logistic regression model to predict the probability of cleavage using all of the predictors in the table. Show the output of the model.
2. Which of the predictor variables have a significant effect on the probability of Cas9 cleavage?
3. For each of the significant effects, calculate the odds ratio for a unit change in the input variable.