# Mathematical and Statistical Foundations for Data Science (CMPINF 2105)
# "Pen and Paper" Homework 4: Probability Theory, Regression, and Inference (Modules 7 & 8)

1. In a certain population, 2% of people have a particular disease. A test for the disease is 98% accurate (true positive rate) and has a false positive rate of 5%. If a randomly selected person tests positive, what is the probability that they actually have the disease?
2. A sample of 15 students has an average test score of 76. Test the claim that the average test score of this population is 80 with a standard deviation of 8 using a 0.05 significance level. Make sure to:
   - State the null and alternative hypotheses.
   - Calculate the test statistic.
   - Determine the critical value and make a decision.

   Include these steps in your answer. Use a t-test reference table.
3. A traffic control office records the number of cars passing through an intersection. On average, 5 cars pass through per minute. What is the probability that exactly 7 cars pass through in a given minute? What is the probability that at most 2 cars pass through in a given minute?
4. Explain why the statement "correlation implies causation" is incorrect. Provide an example where correlation does NOT imply causation in your explanation.
5. Consider two models for a dataset: Model A with $R^2 = 0.75$ and 5 predictors, and Model B with $R^2 = 0.80$ and 10 predictors. With a sample size of 100, calculate the adjusted $R^2$ for both models and determine which model is better based on adjusted $R^2$.
6. In what situation might a high $R^2$ value not necessarily indicate a good model fit?
7. Given a dataset with binary outcomes for a response variable, $y$, and a numeric predictor, $x$, we fit a logistic regression model $\beta_0 + \beta_1 \cdot x$ where $\beta_1 = 0.5$. How does $y$ change with a unit increase in $x$?