MATH 166 Statistics - Spring 2023
Tufts University, Department of Mathematics
Instructor: James M. Murphy
Practice Exam 2

Instructions: This exam has three questions and is out of a total of 30 points. Each question is worth 10 points. No graphing calculators, books, or notes are allowed. Be sure to show all work for all problems. No credit will be given for answers without work shown. If you do not have enough room in the space provided you may use additional paper. Be sure to clearly label each problem and attach them to the exam. You have 75 minutes. Good luck! :-)

Your Printed Name: ________________________________

Problem	Score
1	
2	
3	
Total	

Academic Honesty Certification:

I certify that I have taken this exam without the aid of unauthorized people or objects.

Signature: ________________________________ Date: ____________________

Question 1

Consider the random variables X=“graduated from high school” and Y=“has an income above \$25,000”. We interview 1000 people and observe the following breakdown, written out in a contingency table:

	Not a High School Graduate	High School Graduate
Income below 25,000	210	79
Income above 25,000	153	558

(a) Set up a test to evaluate whether X and Y are independent. Compute all relevant statistics and explain carefully, but do not evaluate.

(b) Discuss how experimental design (e.g. who is interviewed and where/when) may impact the results of such a test.

Question 2

Suppose we are given the following (x_i, y_i) pairs in $\mathbb{R}^2$: $(1,1), (2,3), (5,3), (7,6), (9,5)$.

(a) Derive the equation for the linear least squares line through the above data by computing the coefficients (β_0, β_1) that minimize $\sum_{i=1}^{5}(\beta_1 x_i + \beta_0 - y_i)^2$. This must be proved by calculus, not recalled from memory.

(b) Suppose the data point $(9,5)$ is replaced with $(9,50)$. At a descriptive level, how will this change the solution in (a)? Explain.

Question 3

Suppose we observe the following data points sampled i.i.d. from an unknown random variable taking values in $(0, 1)$: $\{.1, .15, .2, .22, .24, .45, .8\}$.

(a) Compute and plot a histogram estimator with uniform bin size $h = 1$.

(b) Compute and plot a histogram estimator with uniform bin size $h = .25$.

(c) Compute and plot a histogram estimator with uniform bin size $h = .01$.

(d) In light of (a)-(c), discuss the role of bin size in histogram estimation. Which of the above makes the most sense to you? Why?