

## Assignment 7. Distribution of sample means

Your name

March 2, 2024

Solve the following problems – each problem on a separate page – and submit to Gradescope for grading.

## 1 Two problems [3 points]

Solve two problems below and pay special attention to how each problem is stated. Though they both use a table of the standard normal distribution, the solutions are different.

Average fetal heart rate is normally distributed with a mean of 140 bpm and with standard deviation of 12 bpm.

**Question 1.** What is the probability that a randomly chosen heart rate differs from the mean by 25 bpm?

**Question 2.** What is the probability that the average heart rate of 50 randomly selected fetuses differs from the mean by 25 bpm?

Show steps of your computation and clearly explain how the two solutions are different.

## 2 Confidence intervals [3 points]

78 students surveyed said they study on average 15 hours per week with a standard deviation of 2.3 hours.

What is the 90% confidence interval of the average study time for the entire student population?

Show all the steps of your calculation.

### 3 Naive Bayes classifier and its success rate [9 points]

You are given the following labeled dataset used for building a Naive Bayes classifier.

**Build a classifier on the training set.** To self-check that you mastered Naive Bayes, build it using these data. Building a Naive Bayes classifier implies computing all conditional probabilities for all combinations of different values for each attribute given one of each class. After you finished counting, add Laplace correction.

#	A	B	C	Class
1	0	0	0	Yes
2	0	0	1	Yes
3	0	1	0	Yes
4	0	1	1	No
5	1	0	0	Yes
6	1	0	0	Yes
7	1	1	0	No
8	1	0	1	Yes
9	1	1	0	No
10	1	1	0	No

Training set.

**Measure success rate.** Now use the classifier to classify the following labeled records and write the results of your classification in the last column of the table below. You do not need to compute the probabilities of ‘Yes’ and ‘No’ - just which class is more probable.

#	A	B	C	Class	Predicted
11	0	0	0	Yes	
12	0	1	1	Yes	
13	1	1	0	Yes	
14	1	0	1	No	
15	1	0	0	Yes	

Test set.

Next, compute the success rate of your classifier.  
The success rate is:

**Infer confidence interval of success rate.** We cannot do any inference for a sample of size 5. So let’s imagine that we tested the classifier on 100 records and obtained the same success rate. Compute confidence interval for a real success rate of your classifier at confidence level 85. Show steps of your computation.