Problem Set 2

Applied Data Science 2

Semester 1 2022/23

We expect this problem set to take around an hour to complete. But professors are sometimes wrong![citation missing] If this or future problem sets are too long, please let us know, so we can adjust and plan accordingly.

Looking at student grades

In an introductory biomedical informatics class at the University of Awesome, the mean course mark is 86, with a standard deviation of 5.0

- Use this information to generate a distribution of marks that might be expected for a class with 100 students.
- Plot that distribution as a histogram with informative axis labels and in a pretty colour
- In this distribution, how many students have a mark higher than 91 or lower than 81? How many students have a mark higher than 96 or lower than 76?
- Discuss one potential problem with creating a predicted grade distribution in this way

Getting good grades

Oh no! There is an exam that you have not prepared for! The exam consists of 20 multiple-choice questions. You know none of the answers! Each question has four answer choices, of which exactly one is correct. You need to answer 10 or more questions correctly in order to pass. The examiners have used a random number generator to determine which of the answer choices (A, B, C, or D) should have the correct answer.

- Which of the following strategies gives you a higher chance of passing?
  - Randomly selecting one of the four answers for each question
  - Selecting "A" for every answer
  - They are the same
- Often, exams are set in such a way that the choice of the correct answer (A, B, C, or D) is random, but with the additional constraint that each of the four choices should be used equally often for the correct answer. If the exam is set in such a way, does this change your answer? Why or why not?

Last update by DJ MacGregor in 2022