# Homework 1

P8130 Fall 2022

Due: September 23, 2022 at midnight EST

**P8130 Guidelines for Submitting Homework**

- Your homework must be submitted through Courseworks. No email submissions!
- Only one PDF file should be submitted, including all derivations, graphs, output, and interpretations. When handwriting is allowed (this will be specified), scan the derivations and merge ALL PDF files (http: //www.pdfmerge.com/).
- You are encouraged to use R for calculations, but you must show all mathematical formulas and derivations. Please include the important parts of your R code in the PDF file but also submit your full, commented code as a separate R/RMD file.
- To best follow these guidelines, we suggest using Word (built in equation editor), R Markdown, Latex, or embedding a screenshot or scanned picture to compile your work.

DO NOT FORGET: You are encouraged to collaborate on homeworks, explain things to each other, and test each other's knowledge. But **do NOT hand out answers to someone who has not done any work**. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. Write your own solutions!

## Problem 1 (5 points)

Please classify each of the following variables as qualitative (specify if binary, nominal, or ordinal) or quantitative (specify if discrete or continuous):

a) homework feedback, labeled as "poor", "fair", "good", "very good"

b) homework feedback, labeled as "fail", "pass"

c) country of birth

d) the quantity of grapes (in lbs) to make 3 liters of wine

e) number of TAs in the P8130 course

## Problem 2 (15 points)

In a study of 133 individuals with a recent bike crash history, depression scores were measured using a standardized test. The depression scores for 14 of these individuals are as follows:

45, 39, 25, 47, 49, 5, 70, 99, 74, 37, 99, 35, 8, 59

a) Compute the following descriptive summaries of these data: mean, median, range, SD.

b) Describe the box plot and the underlying distribution of the data. Use some of the following terms: left-skewed, right-skewed, symmetric, bimodal, unimodal distribution.

Additionally, 140 individuals with a recent car crash history also participated in the study. The depression scores for 13 of these individuals are given below:

67, 50, 85, 43, 64, 35, 47, 97, 58, 58, 10, 56, 50

a) Using R, make a side-by-side box plot of the depression scores stratified by type of accident. Make sure you label your figure appropriately.

b) Describe each of the box plots and the underlying distribution of the data. Use some of the following terms: left-skewed, right-skewed, symmetric, bimodal, unimodal distribution.

c) Comparing the 2 box plots, which group appears to have a lower typical depression score?

## Problem 3 (10 points)

Suppose we toss one fair 12-sided die:


a) Let's define the event A as "an even number appears". What is the probability of the event A?

b) Let's define the event B as "number 10 appears". What is the probability of the event B?

c) Compute P(B U A).

d) Are events A and B independent? Why? Prove your answer.

## Problem 4 (10 points)

5% of women above age of 75 have dementia. Among women (75+ years old) with dementia, 80% have positive findings on their CT scan. Among women (75+ years old) who don't have dementia, 10% will have a positive CT scan findings. A randomly-selected woman (75+ years old) had a positive CT scan findings.

What is the probability that she actually has dementia? Compute by hand and show the key steps. The answer can be hand written.