

# ps2 basic statistics (due in 3 weeks+presentations)

[version: Wednesday 4<sup>th</sup> September, 2024 14:57]

1. Use your own data, and describe it briefly what it is and why you use it. Either give me a link to it or upload to canvas, if it's big, subset/take random sample. If you dont have it, find it asap, dont hesitate to ask for help. The data should be of substantive interest to you and ideally serve you for the capstone [2]
2. Calculate mean, median, and mode; and produce a histogram [4].
3. Do a scatterplot and crosstabs [4].
4. Can and should use software for calculations. Ideally Python, R, Stata, etc; worse case scenario Excel. Unless it's like  $< 10$  numbers, then can do it by hand. If data are too big, eg 500k obs in brfss, find some report, paper, book, that has a small table summarizing it, say avg for NJ counties and use that.

---

## general directions (always the same):

- ps is due in Canvass by the beginning of the class
- keep it short; max: 5 single spaced pages; typically way less, say 1 or 2
- if you are stuck, email me early! also email if you want some feedback and make sure you are on the right track, etc
- show your work, a "naked" number won't do! unless indicated otherwise, always do calculations by hand
- likewise, numbers should be interpreted—we are not only interested in calculating values of interest, but we are interested in their meaning! whenever you calculate your final quantity if interest, interpret it!
- if your handwriting is bad, please type
- i may want to discuss your assignment in class, which should be beneficial to you and give you more feedback; if however, you'd like to keep it private, let me know!
- numbers in brackets are relative importance of each item for grading; adds up to 10