basic statistics; due in 2 weeks

[version: Wednesday 30th September, 2020 14:48]

1. you can either use your own dataset (just retain only several observations) or use the following data:

name	years of education	income in millions
donald	12	12
ted	20	1
bernie	18	.5
hilary	17	3

2. compute the mean, median, and mode of income–which measure describes the data best? explain

   A: computation is very straightforward–just plug into the formulas; explaining which describes data best, is little more complicated, and there is not one right answer, but the answer must always include discussion of distribution! Distributions are very important! In this case, as in any case involving dollar values, income is right skewed due to Donald. When there is skewness, it is usually a good idea to report the median, and note explicitly the outliers. In short, median would be a typical value, describing well a typical person except the outlier, the Donald.

3. crosstabulate mean and median income by gender–what do you conclude?

   A: crosstabulation in this case simply means calculating statistics by groups. For females, there is just one person, and hence you can either say that statistics does not exist or is just the obs for Hilary. Conclusions here may be different, but it will involve comparison between the 2 groups, and perhaps discussion of distribution. Again, distributions are very important!

4. do a scatterplot (can do it by hand or using software), correlate income and education: what do you conclude?

   A: Again, discussion of distributions is the key–you have to say that there is an outlier, Donald, very different from everybody else. And some discussion of corr: sign and strength. You should be briefly creative and discuss general observation: expected, unexpected, causality, what else would correlate, etc etc.

---

general directions (always the same):

- ps is due in Sakai's dropbox
- keep it short; max: 5 single spaced pages; typically way less, say 1 or 2
- if you are stuck, email listserv early! also can email listserv 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!
- ideally you want to use this class to write a research part of your capstone
- think if you can engage civically–see section 'civic engagement component' in syllabus