STARFISH SCHOOL

A Virtual Bootcamp for Astronomy Graduate Students

WEEK 3 EXERCISES (GROUP)

Version 1.0

Exercise

1. Download the data set `xvalues.csv` from the website
2. Generate a histogram for these values using bin widths of 2, from -8 to 4. *Before going to part 3,* what do you notice about this distribution? Would you hypothesize what distribution the data came from?
3. Generate a new histogram for these values using bin widths of 2, starting instead from -7.
4. Make a boxplot of these data and find the summary statistics
5. Make a kernel density estimate plot of the distribution. How does this compare to the other options?
6. Based on your figures, comment on the pros and cons of each estimate of the distribution (histogram, boxplot, KDE)
7. Standardize the data from question 1 and make a new histogram and boxplot. Compare these to your histogram and boxplot in question 1.
8. What are the mean and standard deviation of the standardized data?
9. Check the 68-95-99 rule using the standardized data. Is the empirical rule applicable here? Why or why not?

Stretch Goals:

1. Make an empirical CDFs of the data and compare to the CDF of a normal.
2. Make a Q-Q plot (look up what this is)!