



Practicum in Statistical Computing

2021 Fall / APSTA-GE.2352

Lab week 4

Kwan Bo Shim

Sep 28, 2021

P A R T 0 1

Week 4

Week 4

- 1. Poll**
- 2. Attendance**
- 3. Homework**
- 4. Slack**

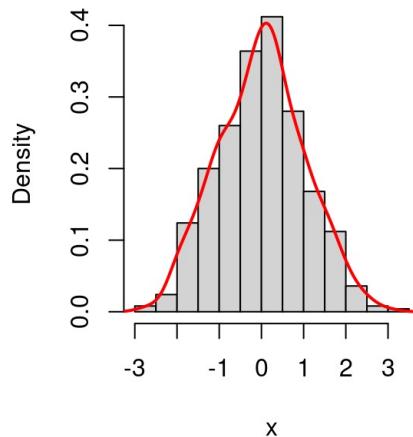
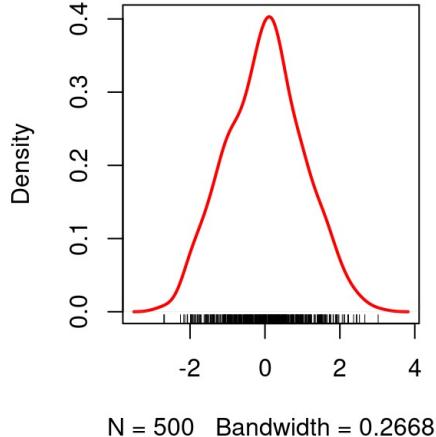
Week 4

1. Density plot with histogram

2. Bandwidth

Density estimation

1. Nonparametric **density estimation** is an approach to estimating the distribution of the population from which the sample was drawn. The density estimation is nonparametric in the sense that the sample data suggest the shape, rather than imposing the shape of a known population distribution with particular values of its parameters.
2. Nonparametric **kernel density estimation** can be thought of as a method of averaging and smoothing the density estimate that would be provided by a histogram.
3. Changing either the **kernel function or the smoothing parameter** would affect the overall density that is estimated. The former by changing the relative weights and the latter by changing the range of data used in the estimate.

Histogram and density**Only density plot**

Definitions:

If the *frequency* of the i th bar is f_i , then its *relative frequency* is $r_i = f_i / n$, where n is the sample size. Its *density* is $d_i = r_i / w_i$, where w_i is its width.

```
density(vlss$Age, bw = "SJ", adjust = 0.5,  
kernel="gaussian")
```

1. **bw:** smoothing bandwidth; The kernels are scaled such that this is the standard deviation of the smoothing kernel

2. **adjust:** Used with $bw * adjust \rightarrow$ convenience

3. **kernel:** giving the smoothing kernel to be used. Default="gaussian".



"Finding my NYC apartment!"

Searched several units in following area..

1. Greenwich_village
2. East_village
3. Chelsea
4. Kips Bay



Searched several units in following area..

1. Rent (\$)
2. Area
3. Utility & wifi (\$)
4. Room type (studio or 1-bed)
5. Washer (yes or no)
6. Floor
7. Grocery (minute)
8. Café (minute)

PART 01

Data will be provided on Slack! (`find_my_apartment.csv`)
You have to find two units (One for my friend)



shutterstock.com · 1406108012

1. Create density plots of overall rent!
2. Using **for loop** and **conditionals** to find..
 - first, create a column computing actual rent (actual rent = rent + utility + wifi)
 - There is a 50% reduction in utility in Greenwich Village
 - 50% reduction in wifi in Kips Bay if rent > 2500\$
 - (a) Rent < 2400(\$)
 - (b) Room type is either fine
 - (c) I prefer to have washer
 - (d) No higher than 5th floor
 - (e) Grocery in 10 minutes
 - (f) Café near 10 minutes
3. My friend's condition is..
 - (a) I must have a dish washer!
 - (b) Café near 7 min
 - (c) Never in Chelsea area

