edX   **MITx: 14.310x Data Analysis for Social Scientists**   <u>**Help**</u>

Module 3: Gathering and Collecting Data, Ethics, and Kernel Density Estimates > Summarizing and Describing Data > Kernel Density Estimation, Part I - Quiz

Kernel Density Estimation, Part I - Quiz

🔖 **Bookmark this page**

Question 1

1 point possible (graded)
The choice of the boundaries for a kernel density plot is dictated by a trade off between _____ caused by too small of a bandwidth and _____ as a result of too large of a bandwidth.

- ○ a. variance, bias
- ○ b. bias, variance
- ○ c. bias, smoothness
- ○ d. variance, smoothness

Explanation

If your bandwidth is smaller than optimal, your function will look very jagged. As Professor Duflo mentions, the smaller the bandwidth, the bigger the variance. When you pick a bandwidth that is too large for your data, the function will try to find data that does not exist. This will cause the

result kernel density function to be biased, meaning that there is an important characteristics of the data that you are missing as a result of selecting a non-optimal bandwidth.

<div style="border:1px solid #ccc; padding:10px; display:inline-block;">Submit</div>   You have used 0 of 2 attempts

Question 2

1 point possible (graded)

For which of the following scenarios would it be better to use a histogram over a kernel density plot?

- ○ a. To show a continuous distribution function of student's math grades

- ○ b. To demonstrate the impact of an intervention on school attendance rates

- ○ c. To show the difference in the distribution of male and female height

- ○ d. To display the proportion of the U.S. households that fall within certain income brackets (e.g. between 0 and $15,000; between $15,000 and $30,000; etc.)

Explanation

As discussed in this segment, a kernel density plot provides a continuous probability density function of a random variable. In the last segment, we learned that a histogram is also an estimate of a probability distribution for a continuous variable, but, unlike the kernel density function, it is not a continuous function. Instead, a histogram demonstrates the number of observations or the proportion of observations that fit into a certain "bin." From this, a histogram would be preferable to a kernel density function for cases in which we want to demonstrate the proportion of observations that fall into certain bins, such as when we want to show the proportion of households that fall within specified income brackets.

Submit   You have used 0 of 2 attempts

Discussion
Topic: Module 3 / Kernel Density Estimation, Part I - Quiz

Show Discussion

Instrumental Variables, and Experimental Design

▸ Exit Survey

▸ Final Exam

POWERED BY
OPENedX