

Updating KDE Plot to Show Cumulative Distribution Function

Learn / Courses / Exploratory Data Analysis In Pgt...

← Course Outline →

Daily XP 585

Exercise

Exploring with KDE plots

Kernel Density Estimate (KDE) plots are a great alternative to histograms when you want to show multiple distributions in the same visual.

Suppose you are interested in the relationship between marriage duration and the number of kids that a couple has. Since values in the `num_kids` column range only from one to five, you can plot the KDE for each value on the same plot.

The `divorce` DataFrame has been loaded for you. `pandas` has been loaded as `pd`, `matplotlib.pyplot` has been loaded as `plt`, and `Seaborn` has been loaded as `sns`. Recall that the `num_kids` column in `divorce` lists only N/A values for couples with no children, so you'll only be looking at distributions for divorced couples with at least one child.

Instructions 3/3 30 XP

- Update the code for the KDE plot from the previous step to show a cumulative distribution function for each number of children a couple has.

Take Hint (-9 XP)

script.py

Light Mode

```
1 # Update the KDE plot
  to show a cumulative
  distribution function
2 sns.kdeplot
  (data=divorce,
   x="marriage_duration",
   hue="num_kids", cut=0,
   ----)
3 plt.show()
```

Run Code Submit Answer

Plots

Marriage Duration Distributions by Number of Kids (Usi

← Previous Plot 3/3 Next Plot →

IPython Shell

Slides

In [1]:

Question

Update the code for the KDE plot from the previous step to show a cumulative distribution function for each number of children a couple has.

Explanation of the Question

This task involves modifying the KDE plot to represent cumulative distribution functions (CDFs) instead of density estimates. A CDF shows the probability that a variable takes a value less than or equal to a specific value.

Answer

```
# Update the KDE plot to show cumulative distribution function
import seaborn as sns
import matplotlib.pyplot as plt

sns.kdeplot(
    data=divorce,
    x="marriage_duration",
    hue="num_kids",
    cumulative=True, # Enable cumulative distribution
    cut=0, # Restrict KDE smoothing to the data range
    palette="coolwarm"
)
plt.title("Cumulative Distribution of Marriage Duration by Number of Kids")
plt.xlabel("Marriage Duration (years)")
plt.ylabel("Cumulative Probability")
plt.show()
```

Explanation of the Answer

The code enables the cumulative distribution function (CDF) in the KDE plot by setting `cumulative=True`. This converts the density plot into a cumulative plot, showing the proportion of data points below each value. Other enhancements like `cut` and labels improve clarity and visualization.