Z-Score

```
import numpy.random as npran
import pandas as pd
import matplotlib.pyplot as plt
```

Question

> Suppose that the mid-term exam of the course EDF5400 has a **mean of 82** and **SD of 6** and follows a roughly **normal distribution**. The instructor decides ? that he wants to open help sessions for the **lowest performing 20%**[1] of the students on the basis of the mid-term exam. Where should the instructor set the **cut score** [2] to put the lowest performing **20% of the students** [3] into help sessions? (Round your answer to the nearest whole number.)

Answer

We *want to know* the cutoff score (which would be x in the formula above). We *do know* or can easily find the z-score for 20% because the z-score tells us the percentage of points beyond a certain point on the distribution. In the graph below I've just plopped a darker line down on the plot. This is just an estimate of what 20% of the left side might look like. Meaning, all the area of histogram to the left of the dark blue line makes up exactly 20% of the data. So, how do we figure out what the exact values of the cutoff are?

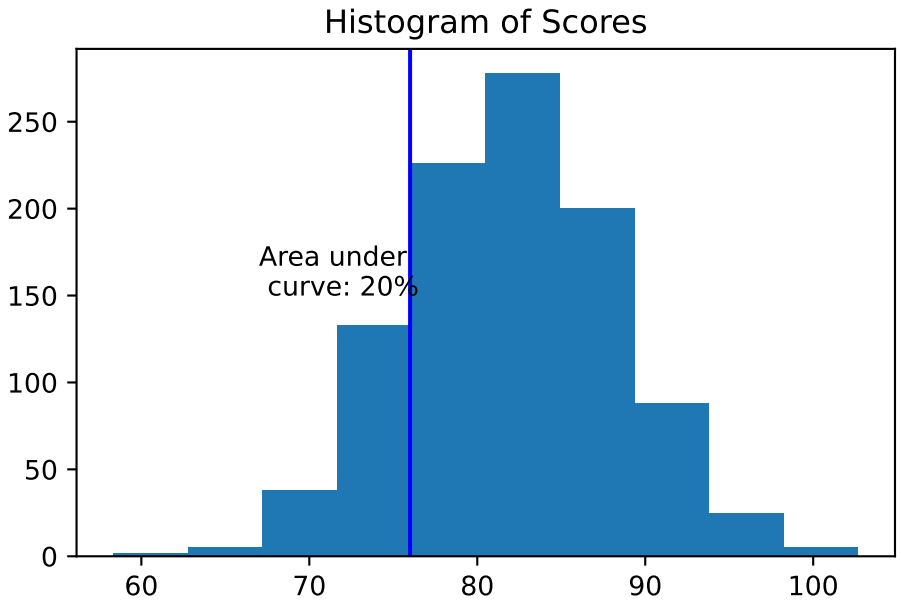
[1] Because we're focused on the "lowest", then we know we're going to be looking at the values in a single (the left/lower) tail of the distribution of scores.

[2] Okay, so the outcome that we're looking for is the cut score, which means it'll be an unstandardized number, right Because we need to be able to easily identify below what score must students have to be included in the help session.

[3] Since we already know we're interested in the lower tail (left side), the 20% is all we needed to know exactly at what point on the distribution (being standardized) to place the cut off. I.e., we want the lower 20% tail.

```
data = npran.normal(loc=82, scale=6, size=1000)
data = pd.DataFrame(data = data, columns = ['scores'])
plt.hist(data['scores'])
plt.title("Histogram of Scores")
plt.axvline(76, color='b')
plt.text(67, 150, "Area under \n curve: 20%")
```

```
Text(67, 150, 'Area under \n curve: 20%')
```



Certain calculators can tell you, but using the table is fairly easy here. We know we want 20%, so find .20 or close to it on the table. When you do you'll see the z-score for 20%:

$$z_{.20} = -.8416$$

We know the formula for z-score is $z = \frac{x-\mu}{\sigma}$. If we use the values we know in this formula...

$$-.8416 = \frac{x - 82}{6}$$

then we're just plugging the values we know into the equation and solving for x, which is our raw score.

$$6 * z_{.20} + 82 = 76.95$$