Frequency Distributions: Takeaways ↗

Syntax

- Generating a frequency distribution table for a `Series` :

```
frequency_table = Series.value_counts()
```

- Sorting the values of frequency distribution table:

```
### In an ascending order (default) ###
freq_table_asc = Series.value_counts().sort_index()
```

```
### In a descending order ###
freq_table_asc = Series.value_counts().sort_index(ascending = False)
```

- Finding proportions and percentages in a frequency distribution table:

```
### Proportions ###
proportions = Series.value_counts(normalize = True)
```

```
### Percentages ###
percentages = Series.value_counts(normalize = True) * 100
```

- Finding the percentile rank of a value (score) in some array:

```
from scipy.stats import percentileofscore
percentile_rank = percentileofscore(a = some_array, score = some_score,
                                     kind = 'weak')
```

- Finding percentiles:

```
### Only the quartiles ###
quartiles = Series.describe()
```

```
### Any percentile we want ###
percentiles = Series.describe(percentiles = [.1, .15, .33, .5, .592, .9])
```

- Generating a grouped frequency table:

```
### With 5 class intervals ###
gr_freq_table_5 = Series.value_counts(bins = 5)


### With 10 class intervals ###
gr_freq_table_10 = Series.value_counts(bins = 10)
```

Concepts

- A table that shows the frequency for each unique value in a distribution is called a **frequency distribution table**.
- The frequencies can be expressed as:
  - Absolute counts (**absolute frequencies**).
  - Proportions or percentages (**relative frequencies**).
- The percentage of values that are equal or less than a value x is called the **percentile rank** of x. For instance, if the percentile rank of a value of 32 is 57%, 57% of the values are equal to or less than 32.
- If a value x has a percentile rank of $p\%$, we say that x is the p_{th} **percentile**. For instance, if 32 has a percentile rank of 57%, we say that 32 is the 57th percentile.
- Frequency distribution tables can be grouped in **class intervals** to form **grouped frequency distribution tables**. As a rule of thumb, 10 is a good number of class intervals to choose because it offers a good balance between information and comprehensibility.

Resources

- An intuitive introduction to frequency distribution tables.
- An intuitive introduction to grouped frequency distribution tables.
- The Wikipedia entry on frequency distributions.