Frequency Distributions: Takeaways

Syntax

- Generating a frequency distribution table for a series:

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

- Sorting the values of frequency distribution table:

```
freq_table_asc = Series.value_counts().sort_index()
freq_table_desc = Series.value_counts().sort_index(ascending=False)
```

- Finding proportions in a frequency distribution table:

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

- 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:

```
quartiles = Series.describe()
percentiles = Series.describe(percentiles=[0.1, 0.15, 0.33, 0.5, 0.592, 0.9])
```

- Generating a grouped frequency table:

```
gr_freq_table_5 = Series.value_counts(bins=5)
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.