Exploring categorical variables

Row and column proportions (conditional proportions)

Import data

In [1]:

```
import pandas as pd

ROOT = "https://raw.githubusercontent.com/kirenz/modern-statistics/main/data/"
DATA = "loans.csv"

df = pd.read_csv(ROOT + DATA)
```

Contingency table with proportions

Normalize by dividing all values by the sum of values:

- If passed ‘all’ or True, will normalize over all values.
- If passed ‘index’ will normalize over each row.
- If passed ‘columns’ will normalize over each column.
- If margins is True, will also normalize margin values.

Row proportions

In [3]:

```
# A contingency table with row proportions for the application type and homeownership variables.
pd.crosstab(df.application_type , df.homeownership, normalize='index').round(3)
```

Out[3]:

homeownership	mortgage	own	rent
application_type			
individual	0.451	0.138	0.411
joint	0.635	0.122	0.242

- 0.411 corresponds to the proportion of individual applicants who rent.

Column proportions

In [4]:

```
# A contingency table with row proportions for the application type and homeownership variables.
pd.crosstab(df.application_type , df.homeownership, normalize='columns').round(3)
```

Out[4]:

homeownership	mortgage	own	rent
application_type			
individual	0.802	0.865	0.906
joint	0.198	0.135	0.094

- The value 0.906 indicates that 90.6% of renters applied as individuals for the loan.
- This rate is higher compared to loans from people with mortgages (80.2%) or who own their home (86.5%)