



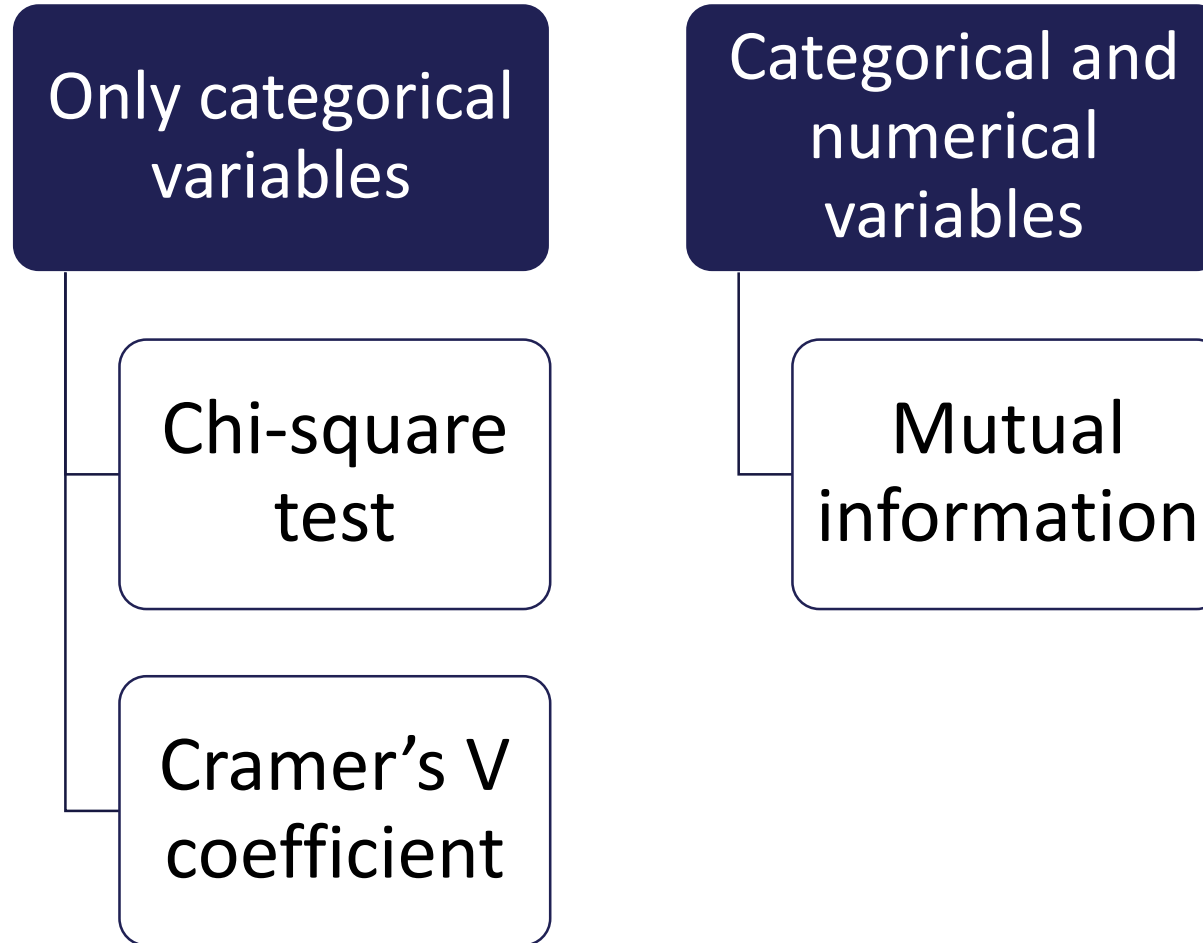
Correlation with categorical variables



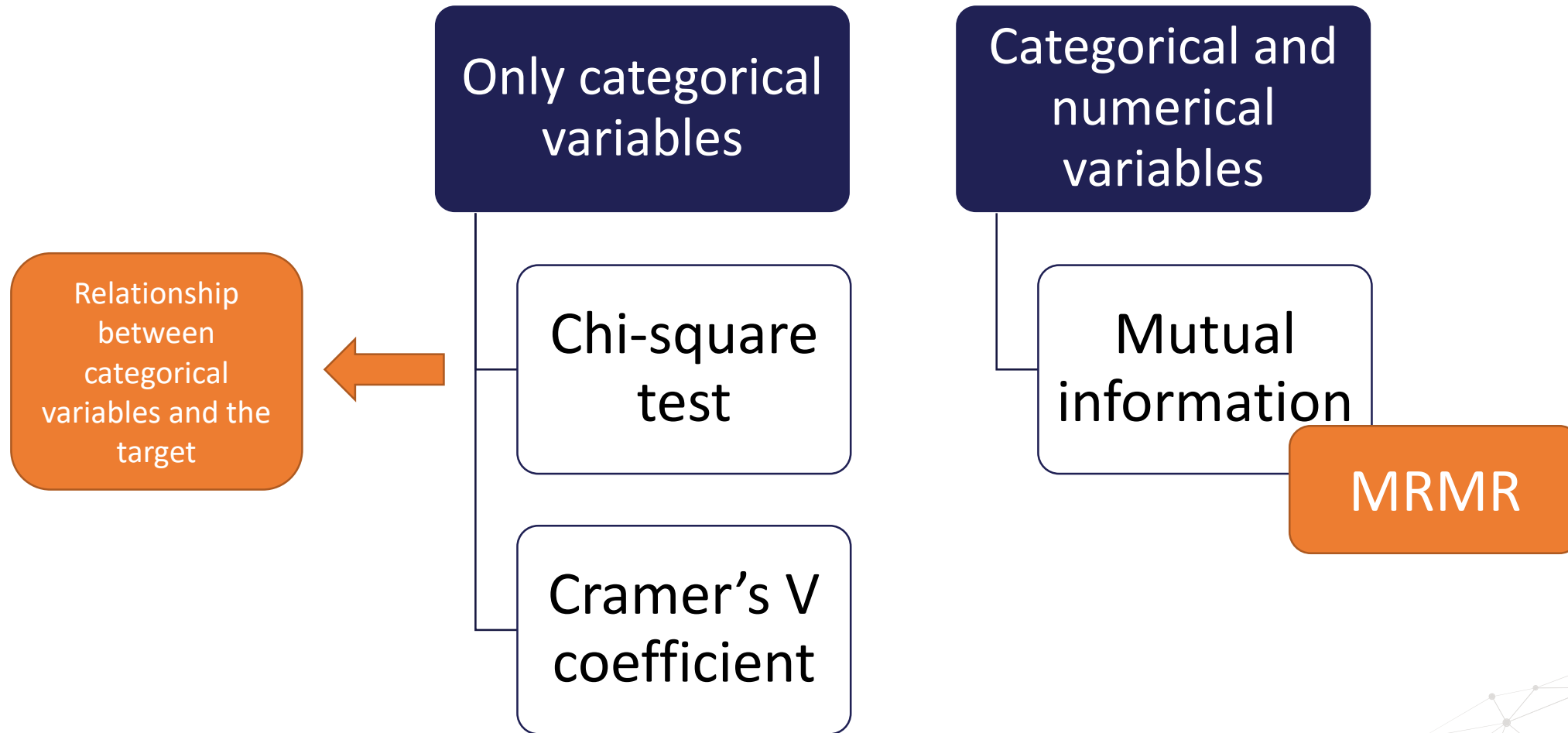
Correlation

- Correlation refers to a statistical relationship or association between 2 random variables.
- Pearson and Spearman are methods to determine association between **continuous variables**.
- For **discrete variables** → Kendall's tau
- What if we have **categorical variables**?

Correlation with categorical variables



Correlation with categorical variables





What if...

- Encode the variable and evaluate correlation afterwards?
 - Coefficients for continuous variables may not be accurate on discrete variables.





Important

- The categorical encoding methods will affect the variable's relationship with the target and also among themselves.



THANK YOU

www.trainindata.com