Statistical methods for feature selection

Statistical tests used in feature selection

ANOVA

Correlation

Chi-square

Mutual information

Statistical tests used in feature selection

	Categorical features	Numerical features
Regression (continuous target)		Correlation
Classification (categorical target)	Chi-square	Anova

Statistical tests used in feature selection

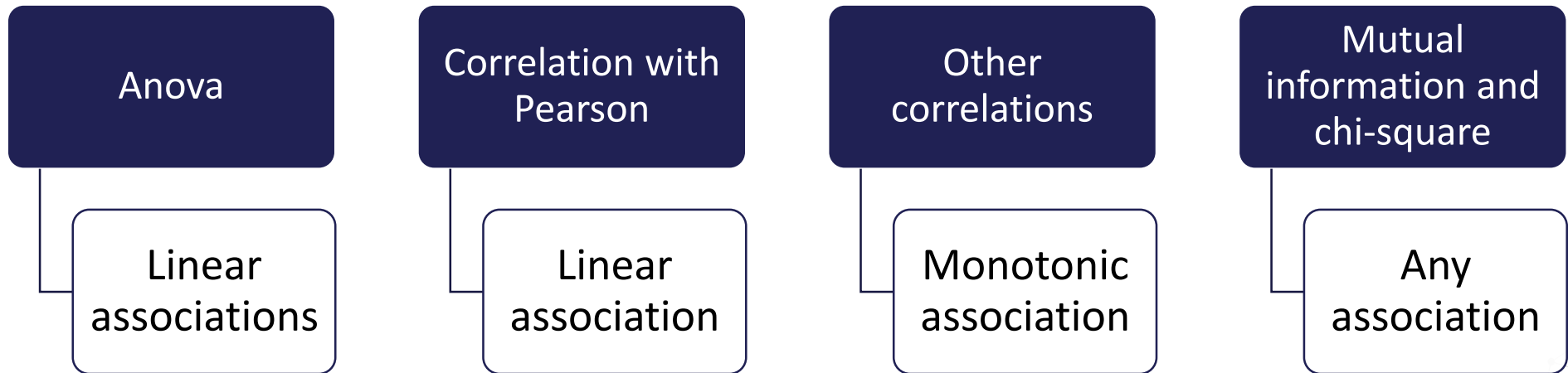|	Categorical features	Numerical features
Regression (continuous target)	Mutual information	Correlation /Mutual information
Classification (categorical target)	Chi-square /Mutual information	Anova /Mutual information

Mixed predictiors

Datasets with both numerical and categorical features ➔ mutual information.

Attention: Python implementations will work, it doesn't mean that we can trust the results!!!

Statistical tests used in feature selection

Anova	Correlation with Pearson	Other correlations	Mutual information and chi-square
Linear associations	Linear association	Monotonic association	Any association

THANK YOU

www.trainindata.com