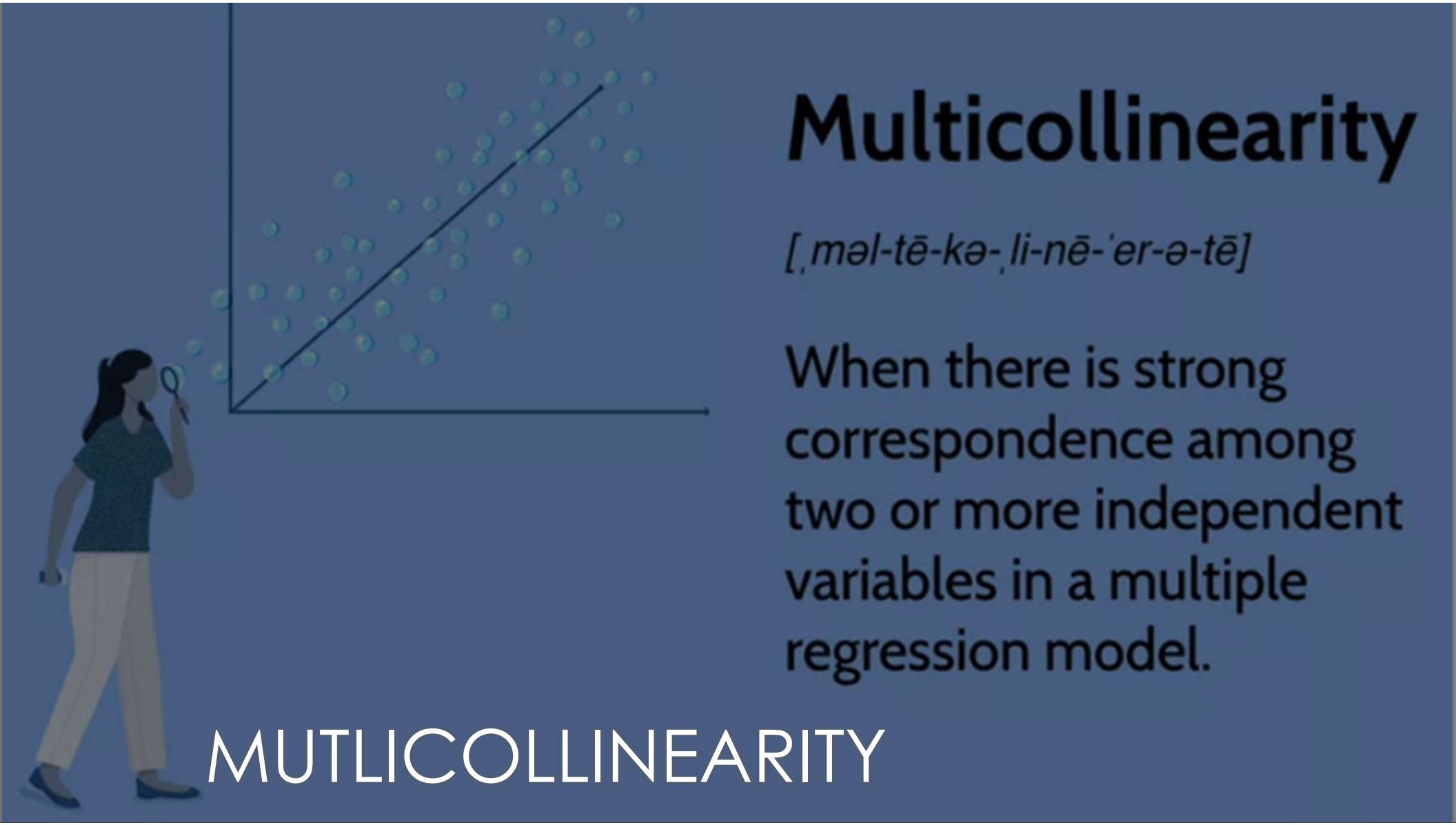

MUTLICOLLINEARITY

WHAT IS MULTICOLLINEARITY?

- MULTICOLLINEARITY IS A SITUATION IN STATISTICS WHERE TWO OR MORE INDEPENDENT VARIABLES ARE HIGHLY CORRELATED WITH EACH OTHER. THIS CAN MAKE IT DIFFICULT TO DETERMINE THE INDIVIDUAL EFFECT OF EACH PREDICTOR ON THE DEPENDENT VARIABLE, LEADING TO UNSTABLE AND UNRELIABLE REGRESSION COEFFICIENTS AND INFLATED ERRORS.

How to handle multicollinearity?

- Dropping one of the correlated features will help in bringing down the multicollinearity between variables.
- Combining variables into a single feature which can reduce multicollinearity.
- Regularizing Techniques like Ridge or Lasso regression can be used to reduce multicollinearity by penalizing the coefficients of correlated variables. These methods shrink the coefficients. Ridger tends to shrink the coefficient while Lasso can set some coefficients to zero.

- Principal Component Analysis(PCA): PCA is a dimensionality reduction technique that can be used to transform the correlated variables into a smaller set of uncorrelated variables.