

FILTRADO DE OUTLIERS

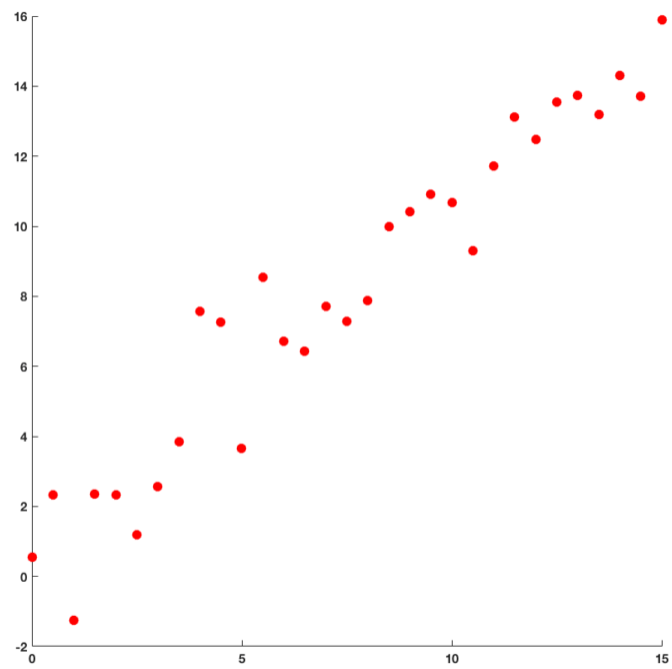
- Professor: Rubén Alvarez

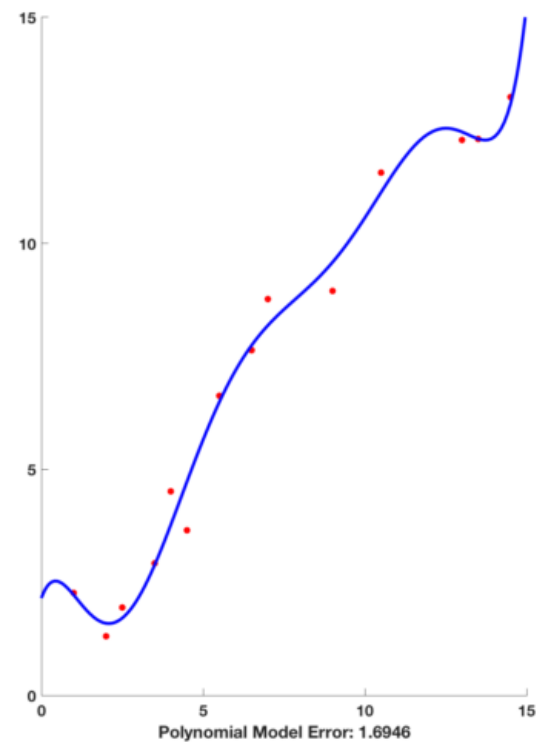
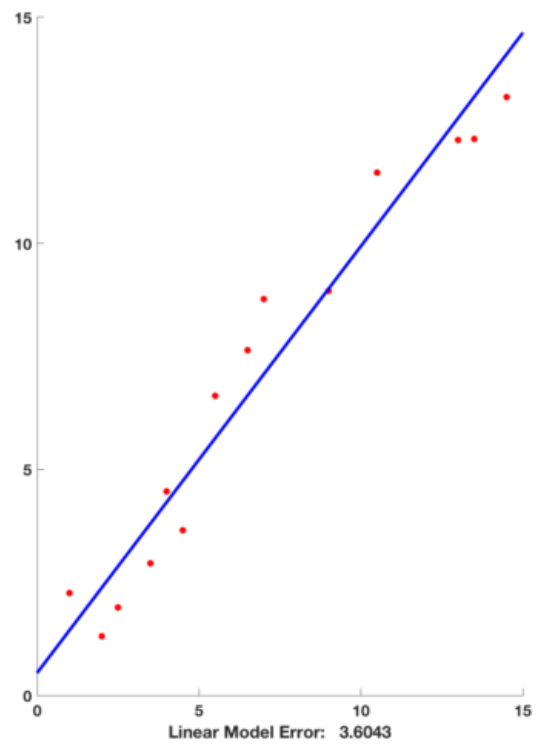
¿QUÉ ES UN OUTLIER?

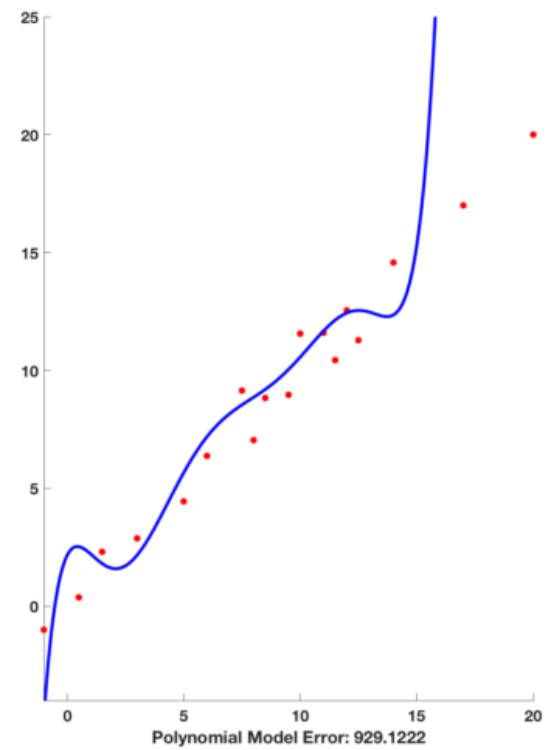
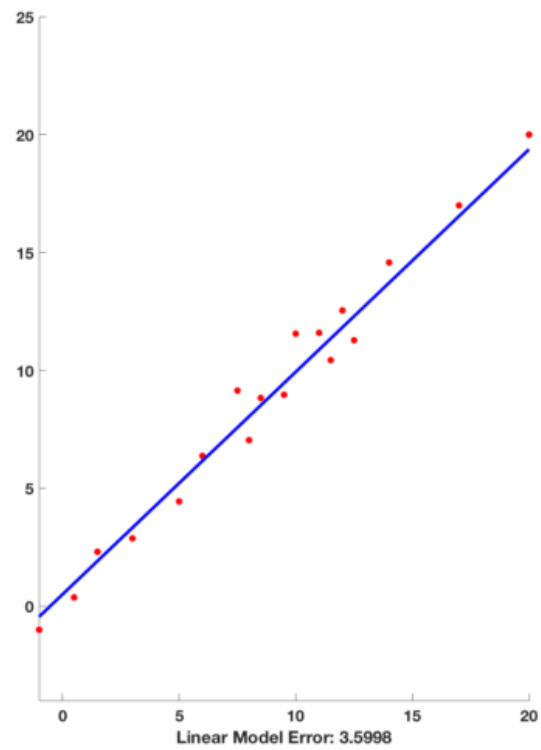
CONTENIDO

- ¿Qué es un filtro?
- ¿Qué es un outlier?
- ¿Filtros dinámicos vs filtros estáticos?
- Métricas de ruido

RUIDO EN PREDICCIÓN





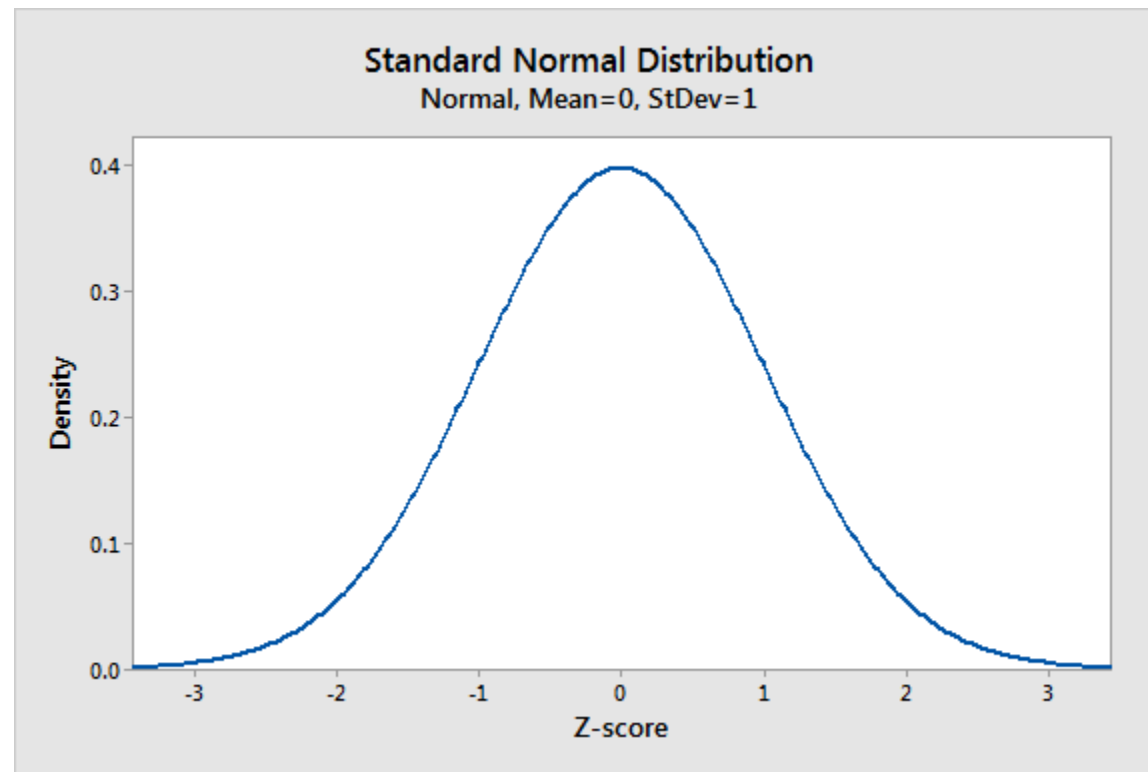


ROOT-MEAN- SQUARE DEVIATION - RMS

$$RMSD = \sqrt{\frac{(\sum_{i=1}^N x_i - \bar{x})^2}{N}}$$

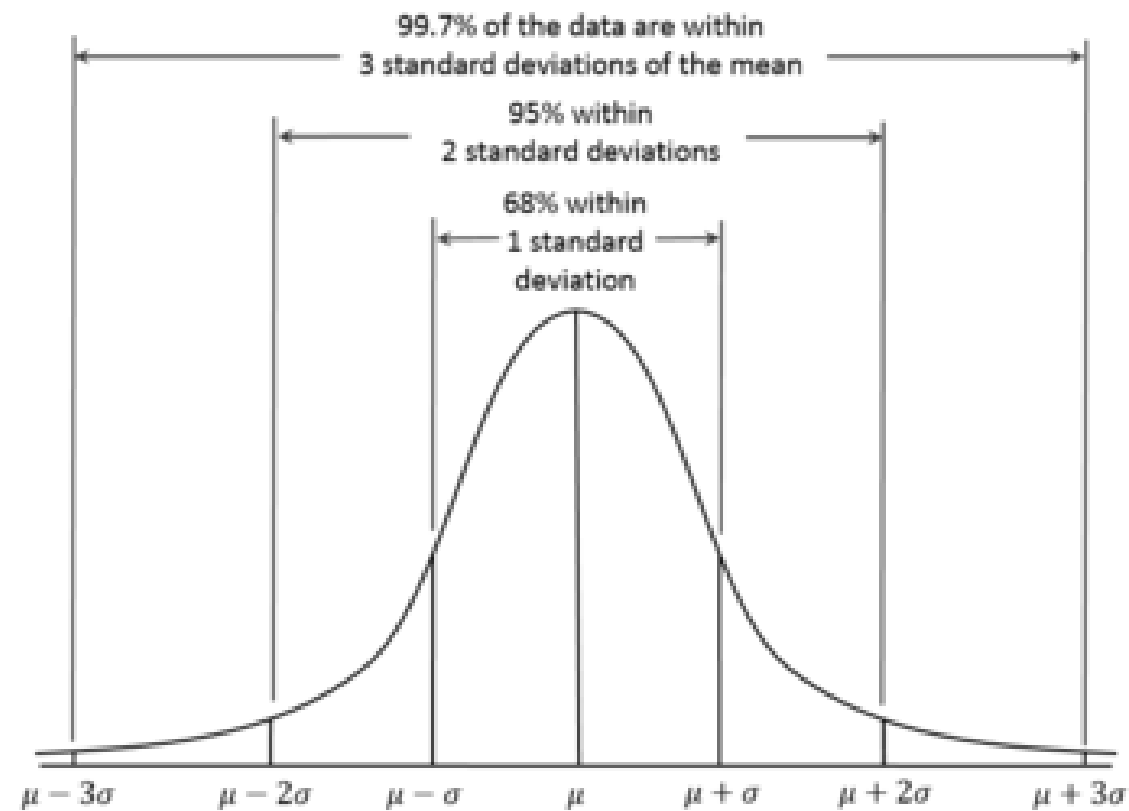
REMEMBER... GAUSSIAN DISTRIBUTION

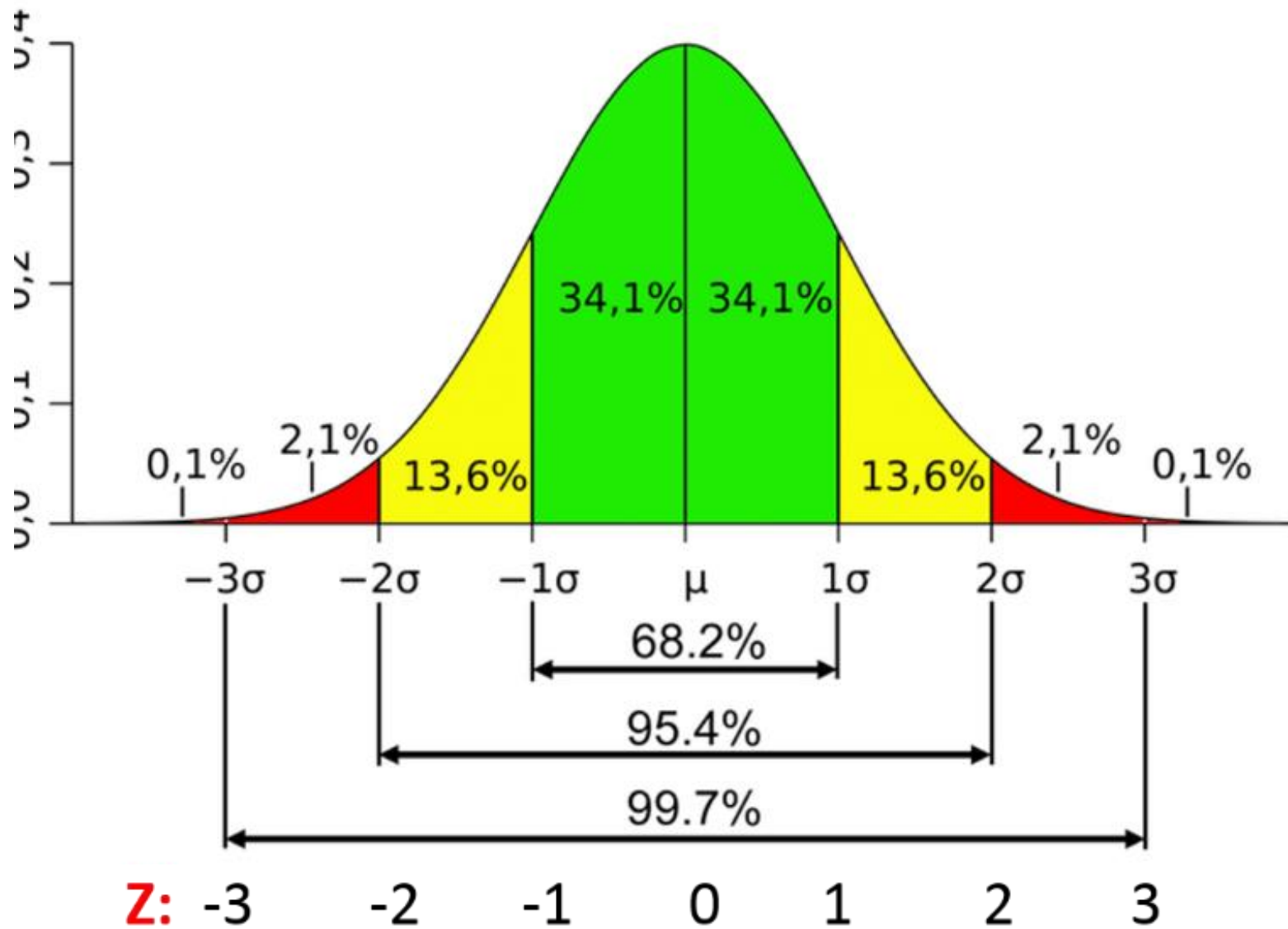
NORMAL DISTRIBUTION



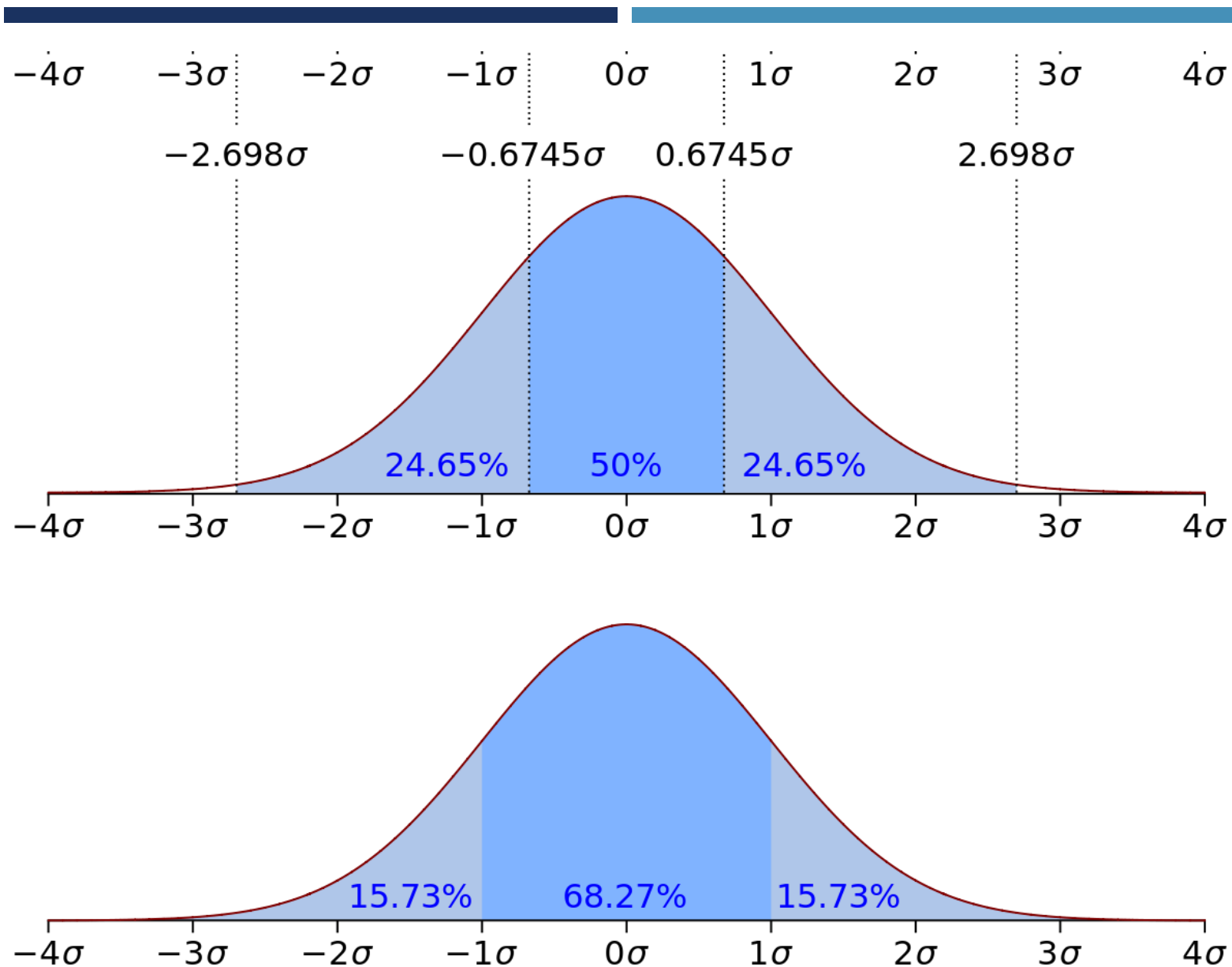
$$Z = \frac{X - \mu}{\sigma}$$

NORMAL DISTRIBUTION

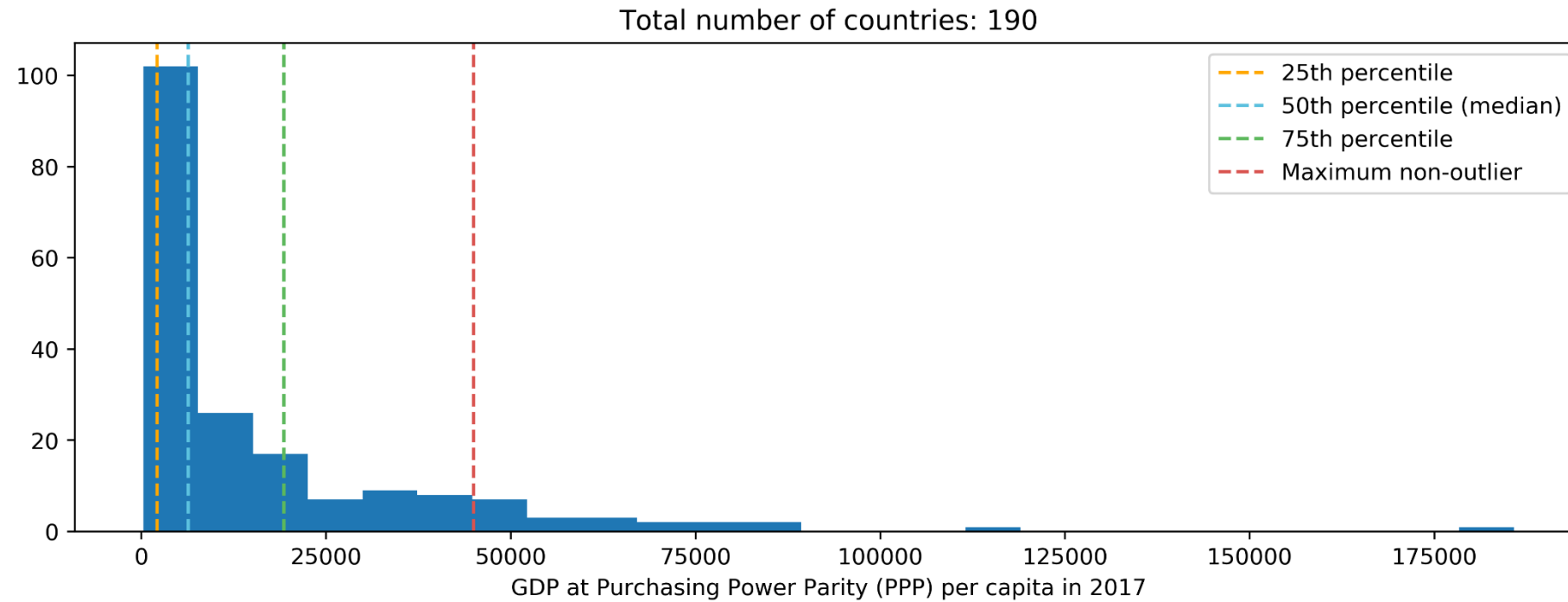




NORMAL DISTRIBUTIONS



INTER-QUARTILE RANGE (IQR)



INTER-QUARTILE RANGE (IQR)

PROCESO DE FILTRADO

- Por cada una de las variables.
- Obtener su media y su desviación estándar.
- Recorrer todo los valores por cada una de las variables
 - Obtener su valor Z
 - De acuerdo al parámetro de desviaciones estándar, filtrar los puntos de ruido.

EJERCICIO

- Obtenga los histogramas de cada una de las variables de su problema de clasificación de proyecto final.
- Aplique un método de filtrado de outliers a todas las variables de su dataset.
- Vuelva a obtener los histogramas de todas sus variables y compare con la versión antes del filtrado, además de mencionar qué porcentajes de datos fueron eliminados por el proceso de filtrado.

¿Preguntas?