



UNIVERSITY OF EDINBURGH  
Business School

# Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

**Dr Xuefei Lu**

The University of Edinburgh Business School

A photograph of the University of Edinburgh Business School building, a modern structure with a glass and metal facade. The building is shown from a low angle, emphasizing its height. The sky is blue with some clouds. A large blue rectangle is overlaid on the right side of the image, containing the text 'Week 1 Quiz'.

# Week 1 Quiz



UNIVERSITY OF EDINBURGH  
Business School

29 Buccleuch Place

# Quiz: Predictive modelling

## Q1. What component is not part of the predictive analytics process?

- A. Formulating the business questions.
- B. Building the predictive model.
- C. Interpreting the result.
- D. They are all part of the process.

## Q3. Decision rules outputted by a predictive model need to be...

- A. Understandable or accurate
- B. At least understandable
- C. At least accurate
- D. None of the above.

## Q2. What is the ideal outcome of a predictive process?

- A. A predictive model.
- B. Decision rules.
- C. An interpretation the results.
- D. A predictive model and its decision rules
- E. None of the above.

should consider many aspects e.g.: accurate ... it's more than these.

Because trade-off, depend on many situation.

# Quiz: Defining Your Problem and Variables

**Q1. What is the most appropriate technique for identifying the level of product stock?**

change every time

- A. Binary classification depend on time
- B. Regression
- C. Time series analysis
- D. Multiclass classification

**Q3. What is the most appropriate technique for identifying the GDP of a country?**

Time

- A. Binary classification
- B. Regression
- C. Time series analysis
- D. Multiclass classification

**Q2. What is the most appropriate technique for identifying someone's province/county of living?**

- A. Binary classification
- B. Regression
- C. Time series analysis
- D. Multiclass classification

**Q4. What is the most appropriate technique for identifying the attendance of an event?**

- A. Binary classification
- B. Regression
- C. Time series analysis
- D. Multiclass classification

# Quiz: Summary Statistics

**Q1. What summary statistic works best for getting a grasp on the central tendency of skewed distributions?**

- ☒ A. Median
- ☐ B. Mean
- ☐ C. Range
- ☐ D. Interquartile range

# Quiz: Summary Statistics

Q2.

Series 1 = [9, 2, 1, 10, 5, 2, 2, 8, 6, 2]

Series 2 = [8, 8, 1, 1, 2, 2, 2, 1, 1, 8]

s1 - mean 4.7 median 3.50

s2 - mean 3.4 median 2

**Based on these two series, which of the following statements is true?**

- A. 1 has a higher median, but lower mean than 2.
- B. 1 has a higher median, and higher mean than 2.
- C. 1 has a lower median, but higher mean than 2.
- D. 1 has a lower median, and lower mean than 2.
- E. None of these statements are true.

# Quiz: Summary Statistics

Q2.

**Series 1 = [9, 2, 1, 10, 5, 2, 2, 8, 6, 2]**

**Series 2 = [8, 8, 1, 1, 2, 2, 2, 1, 1, 8]**

3.37 iqr DIFF  $8-2=6$

3.21 IQR  $8-1=7$

**Based on these two series, which of the following statements is true?**

- A. 1 has a higher standard deviation, but lower interquartile range than 2.
- B. 1 has a higher standard deviation, and higher interquartile range than 2.
- C. 1 has a lower standard deviation, but higher interquartile range than 2.
- D. 1 has a lower standard deviation, and lower interquartile range than 2.
- E. None of these statements are true.

# Quiz: Visual VAR and Coding

**How can the shape of the variables be important for building predictive modelling?**

- A. You need to know the shape of the dependent variable's distribution to build a model.
- B. You need to know the shape of the independent variable(s) distribution to build a model.
- C. You need to know the shape of both the independent and dependent variables' distribution to build a model.
- D. The need to understand the shape of the variables' distribution will depend on the algorithm.

LR require a normal distribution

Neural Network and Decision Tree don't care about the distribution