The where and why of cost variation in the National Health Service

Understanding cost variation with machine learning

Henry Wilde
Supervised by: Dr Jonathan Gillard and Dr Vincent Knight

October 9, 2018

Henry Wilde
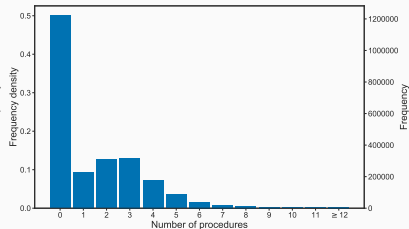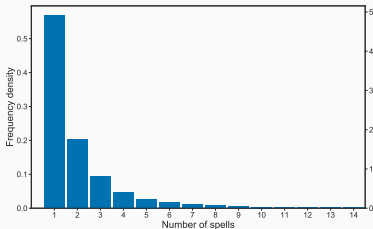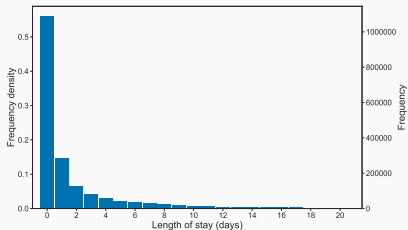*Supervised by:* Dr Jonathan Gillard and Dr Vincent Knight

The data

- The Cwm Taf University Health Board
- April 2012 through April 2017
- 2.4 million patient-episode records with 260 attributes

PATIENT ID	SPELL ID	EPISODE ID	Net Cost	Age (years)	HRG	Admission Date	Discharge Date	Length of Stay (days)
ID_123456	M1001	M1001-1	858.14	74	EA05Z	2015-05-06	2015-05-06	0.0
	M1211	M1211-1	333.95	74	FZ38F	2015-07-15	2015-08-01	17.0
		M1211-2	706.09	74	FZ38F	2015-07-15	2015-08-01	17.0
		M1211-3	8671.31	74	RC16Z	2015-07-15	2015-08-01	17.0

Where does variation come from?

Key attributes and what they look like

Why does variation (not) happen?

Clustering and partitioning

Traditionally:

- Age groupings
- Condition-based populations

A more modern way:

- Clustering of spell-level data
- Healthcare utilisation
- Pathway mining and similarity

Thanks for listening!

@daffidwilde