Engineering Variables of Mixed Type

Mixed Variables

Mixed Variables

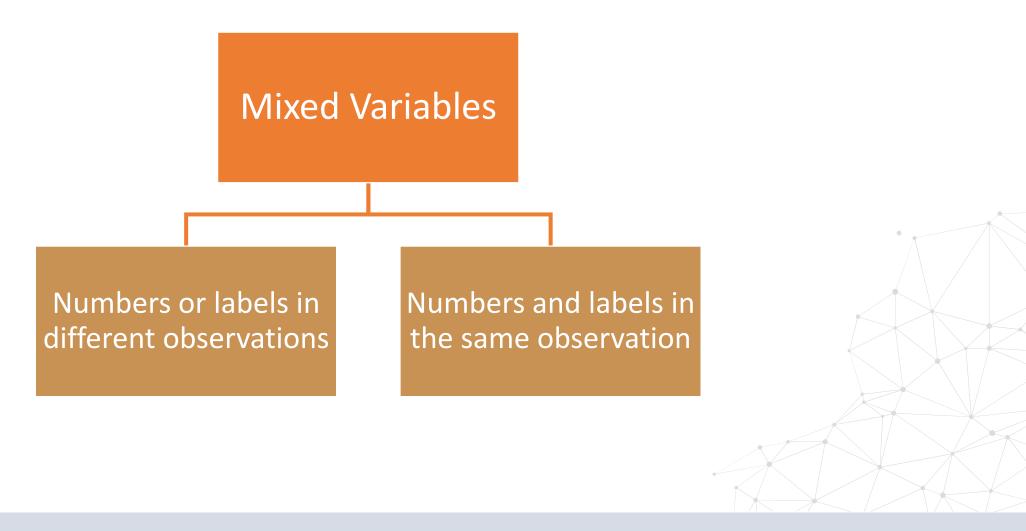Numbers or labels in different observations

Numbers and labels in the same observation

Mixed Variables: case 1

Observations show either numbers or categories among their values

- Number of credit accounts (1-100, U, T, M)
  - U = unknown, T = unverified, M = unmatched)
- Number of missed payments (1-3, D, A)
  - D = defaulted, A = arrangements

Mixed variables: case 1

Payments
0
1
0
A
3
D
0

Cat	Num
Nan	0
Nan	1
Nan	0
A	Nan
Nan	3
D	Nan
Nan	0

Mixed Variables: case 2

Observations show both numbers and categories in their values

- Cabin (Titanic) (A15, B18, …)
- Ticket (Titanic) (A103349)
- Vehicle registration (AB500)
- Postcode (SE18)

Mixed variables: case 2

Cabin
A01
B22
C33
A21
C02
D15
E07

Cat	Num
A	1
B	22
C	33
A	21
C	2
D	15
E	7

Accompanying Jupyter Notebook

- Read the accompanying Jupyter Notebook

- Download the sample dataset attached to this lecture.

THANK YOU

www.trainindata.com