Outline

Outline

Learning objectives

- Create dummy variables
- Construct a regression model with dummy variables
- Forecast using regression models with dummy variables

Outline

How to forecast impact of special events

- Among factors that may impact patient visits, special events such as public holidays, festive days, strikes, sport events, etc may play an important role:

  1. Use dummy variables
  2. Use experience and expertise
  3. Transform time series

Outline

Dummy variables

If a categorical variable takes only two values (e.g., 'Yes' or 'No'), then an equivalent numerical variable can be constructed taking value 1 if yes and 0 if no. This is called a **dummy variable**.

	A	B
1	Yes	1
2	Yes	1
3	No	0
4	Yes	1
5	No	0
6	No	0
7	Yes	1
8	Yes	1
9	No	0
10	No	0
11	No	0
12	No	0
13	Yes	1
14	No	0

Dummy variables

If there are more than two categories, then the variable can be coded using several dummy variables (one fewer than the total number of categories).

	A	B	C	D	E
1	Monday	1	0	0	0
2	Tuesday	0	1	0	0
3	Wednesday	0	0	1	0
4	Thursday	0	0	0	1
5	Friday	0	0	0	0
6	Monday	1	0	0	0
7	Tuesday	0	1	0	0
8	Wednesday	0	0	1	0
9	Thursday	0	0	0	1
10	Friday	0	0	0	0
11	Monday	1	0	0	0
12	Tuesday	0	1	0	0
13	Wednesday	0	0	1	0
14	Thursday	0	0	0	1
15	Friday	0	0	0	0

Uses of dummy variables

Seasonal dummies

- For quarterly data: use 3 dummies
- For monthly data: use 11 dummies
- For daily data: use 6 dummies

Uses of dummy variables

Seasonal dummies

- For quarterly data: use 3 dummies
- For monthly data: use 11 dummies
- For daily data: use 6 dummies

Outliers

- If there is an outlier, you can use a dummy variable to remove its effect.

Uses of dummy variables

Seasonal dummies

- For quarterly data: use 3 dummies
- For monthly data: use 11 dummies
- For daily data: use 6 dummies

Outliers

- If there is an outlier, you can use a dummy variable to remove its effect.

Public holidays

- For daily data: if it is a public holiday, dummy=1, otherwise dummy=0.

Intervention variables

Spikes

- Equivalent to a dummy variable for handling an outlier.

Intervention variables

Spikes

- Equivalent to a dummy variable for handling an outlier.

Steps

- Variable takes value 0 before the intervention and 1 afterwards.

Intervention variables

Spikes

- Equivalent to a dummy variable for handling an outlier.

Steps

- Variable takes value 0 before the intervention and 1 afterwards.

Change of slope

- Variables take values 0 before the intervention and values $\{1, 2, 3, \dots\}$ afterwards.
- this could be also handled using trend()

Include any special event using dummies

- Christmas Eve: if Christmas Eve, $v_t = 1$, $v_t = 0$ otherwise
- New year's Day: if New year's Day, $v_t = 1$, $v_t = 0$ otherwise.
- and more: Ramadan and Chinese new year, school holiday, etc

lag and lead variables

- Lagged values of a predictor:
  - ▸ Create new variables by shifting the existing variable backwards
- Lead values of a predictor:
  - ▸ Create new variables by shifting the existing variable forwards

Example: x is advertising which has a delayed effect

x_1 = advertising for previous month;

x_2 = advertising for two months previously;

⋮

Interactions

For example, sometimes the effect of a partiucluar event might be different if it is on a weekend or a week day or its efefct might be different in each shift:

- you need to introduce an interaction variable
- you can use a new dummy as : `v1*v2`

Outline

Following lab session 8, we would like to develop a model that includes three more variable as dummies:

1. Import the `se.csv` file from the project directory
2. Join them to the data you created with temperature for Lab 8
3. Split the data into train and test
4. Train data using two regression models 4.1. using temperature and seasonality 4.2. using seasonality , temperature and `Black Friday`, `Christmas Day` and `Halloween Day`
5. Produce forecast
6. Calculate forecast accuracy