Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

A parcel delivery service operates in two different regions, **region "A"** and **region "B"**. Delivery trucks leave the central warehouse and travel to region A and deliver parcels in that region. Similarly delivery trucks also leave the central warehouse and travel to region B and deliver parcels in that region.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Region

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Region
→ **A**
→ **B**

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Region → **A**
Region → **B**

One '*dummy variable*' needed

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Region ⟨ **A**
           B One '*dummy variable*' needed (*REGA*)

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

**...the Dummy Variable regression technique**

Region ⟨ A / B

One '*dummy variable*' needed (*REGA*)

	A	B
REGA	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

$Region$ ⟨ A / B   One '*dummy variable*' needed (*REGA*)

	A	B
REGA	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

Region → A
Region → B

One '*dummy variable*' needed (*REGA*)

	A	B
REGA	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

$Region$ ↗ A ↘ B   One '*dummy variable*' needed (*REGA*)

	A	B
REGA	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 Parcels + \beta_3 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

...the Dummy Variable regression technique

$Region \begin{cases} A \\ B \end{cases}$ One '*dummy variable*' needed (*REGA*)

	A	B
REGA	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 Parcels + \beta_3 TruckAge$$

$$Minutes = -33.13 + 106.84 REGA + 10.02 Parcels + 3.21 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Example (deliveries2.xlsx)

The parcel delivery service obtains some additional data, this time for a third region it delivers to, region C. So now, the data consists of deliveries across three regions, region A, region B and region C.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Example (deliveries2.xlsx)

The parcel delivery service obtains some additional data, this time for a third region it delivers to, region C. So now, the data consists of deliveries across three regions, region A, region B and region C.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Example (deliveries2.xlsx)

The parcel delivery service obtains some additional data, this time for a third region it delivers to, region C. So now, the data consists of deliveries across three regions, **region A**, **region B** and **region C**.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Example (deliveries2.xlsx)

The parcel delivery service obtains some additional data, this time for a third region it delivers to, region C. So now, the data consists of deliveries across three regions, **region A**, **region B** and **region C**.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes

Y variable

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   **Region**   **Parcels**   **TruckAge**

Y variable   ←-------- X variables --------→

Linear Regression for Business Statistics
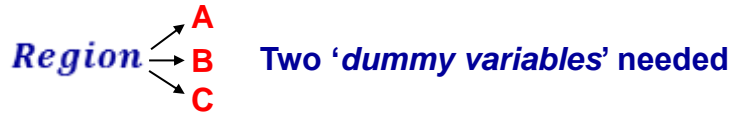
Incorporating 'Categorical Variables' in a regression model.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   *Region*   *Parcels*   *TruckAge*
↑   ↑   ↑   ↑
Y variable   ←-------- X variables --------→

Region
 → **A**
 → **B**   Two '*dummy variables*' needed
 → **C**

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   *Region*   *Parcels*   *TruckAge*

Y variable   ←-------- X variables --------→

Region → A
         → B   Two '*dummy variables*' needed
         → C

REGA
REGB

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
→ B
→ C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	Region	Parcels	TruckAge
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A, B, C   Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	Region	Parcels	TruckAge
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
Region → B
Region → C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
          → B
          → C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

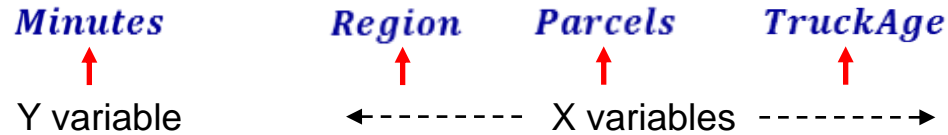| Minutes	Region	Parcels	TruckAge
↑	↑	↑	↑
Y variable	←-------- X variables --------→		

Region → A
B
C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

$Minutes$ $Region$ $Parcels$ $TruckAge$

Y variable ←-------- X variables --------→

$Region$ → A, B, C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.
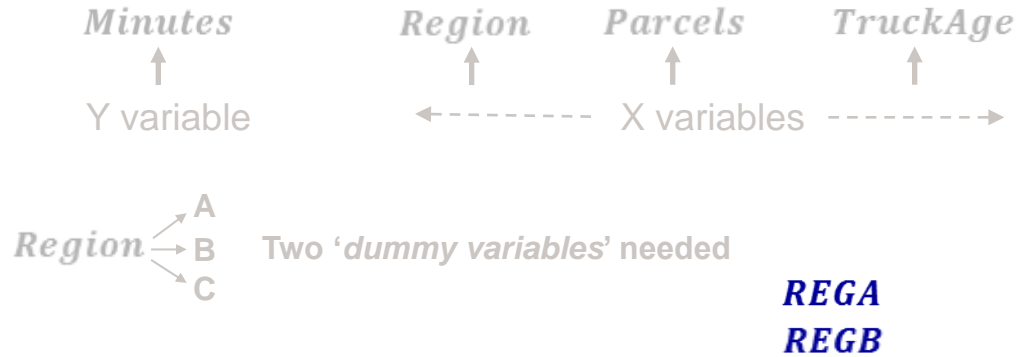
$$Minutes \qquad Region \quad Parcels \qquad TruckAge$$

$$\uparrow \qquad\qquad \uparrow \qquad\quad \uparrow \qquad\qquad \uparrow$$

Y variable ←-------- X variables --------→

$$Region \begin{array}{l} \nearrow A \\ \to B \\ \searrow C \end{array}$$ Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \; \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   Region   Parcels   TruckAge

↑   ↑   ↑   ↑

Y variable   ← - - - - - - - -   X variables   - - - - - - - - →

Region → A, B, C   Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	Region	Parcels	TruckAge
↑	↑	↑	↑
Y variable	← - - - - - - - -	X variables	- - - - - - - - →

Region → A
Region → B
Region → C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←-------- X variables --------→		

Region → A
→ B
→ C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$
↑

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	← - - - - - - - -	X variables	- - - - - - - →

Region → A
Region → B   Two '*dummy variables*' needed
Region → C

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
→ B
→ C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   Region   Parcels   TruckAge
↑   ↑   ↑   ↑
Y variable   ←-------- X variables --------→

Region → A
         → B   Two '*dummy variables*' needed
         → C

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

$Minutes$ $Region$ $Parcels$ $TruckAge$

↑   ↑   ↑   ↑

Y variable   ← - - - - - - - X variables - - - - - - - →

$Region$
- A
- B
- C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

$Minutes$ $Region$ $Parcels$ $TruckAge$

↑   ↑   ↑   ↑

Y variable   ← – – – – – – – X variables – – – – – – – →

$Region$
- A
- B
- C

Two '*dummy variables*' needed

	A	B	C
$REGA$	1	0	0
$REGB$	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A, B, C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   Region   Parcels   TruckAge
↑   ↑   ↑   ↑
Y variable   ←-------- X variables --------→

Region → A
       → B   Two '*dummy variables*' needed
       → C

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

$$Minutes \qquad Region \qquad Parcels \qquad TruckAge$$

Y variable   ← - - - - - - - - X variables - - - - - - - - →

$$Region \nearrow A \qquad \rightarrow B \qquad \searrow C$$

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

$Minutes$ $Region$ $Parcels$ $TruckAge$

↑   ↑   ↑   ↑

Y variable   ◄------- X variables -------►

$Region$ → A / B / C   Two '*dummy variables*' needed

	A	B	C
$REGA$	1	0	0
$REGB$	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Difference in overall delivery time it takes to make deliveries to Region A as compared to Region C, all other variables kept at the same level.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   Region   Parcels   TruckAge

↑   ↑   ↑   ↑

Y variable   ←-------- X variables --------→

Region
- A
- B
- C

Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

+

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

↑

Difference in overall delivery time it takes to make deliveries to Region A as compared to Region C, all other variables kept at the same level.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes   *Region*   *Parcels*   *TruckAge*

↑   ↑   ↑   ↑

Y variable   ←------- X variables -------→

Region → A / B / C   Two '*dummy variables*' needed

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Difference in overall delivery time it takes to make deliveries to Region A as compared to Region C, all other variables kept at the same level.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
         → B   Two '*dummy variables*' needed
         → C

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

↑

Is the difference in the fixed time to make deliveries across the two regions.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	Region	Parcels	TruckAge
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

$Region$ → A, B, C   Two '*dummy variables*' needed

	A	B	C
$REGA$	1	0	0
$REGB$	0	1	0

p-value > α

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

↑

Is the difference in the fixed time to make deliveries across the two regions.

Linear Regression for Business Statistics

Incorporating 'Categorical Variables' in a regression model.

Minutes	*Region*	*Parcels*	*TruckAge*
↑	↑	↑	↑
Y variable	←--------	X variables	--------→

Region → A
           B   Two '*dummy variables*' needed
           C

	A	B	C
REGA	1	0	0
REGB	0	1	0

$$Minutes = \beta_0 + \beta_1 REGA + \beta_2 REGB + \beta_3 Parcels + \beta_4 TruckAge$$

Difference in overall delivery time it takes to make deliveries to Region B as compared to Region C, all other variables kept at the same level.