

# *L**A**T**E**X* annual update

If you installed *L**A**T**E**X* with the `tinytex` package, you may want to update it to the latest version. Around the same time each year, TeX Live releases a new version of the distribution. If you have an older version of TeX Live, you may encounter errors when needing to install new *L**A**T**E**X* packages.

You can do that by running the following code in the console.

```
1 tinytex::uninstall_tinytex()  
2  
3 tinytex::install_tinytex()
```

# Linear regression

The theoretical regression model:

$$\underbrace{y_i}_{\text{Dependent Variable}} = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}}_{\text{Independent Variables}} + \underbrace{\epsilon_i}_{\text{Random Error}}$$

An estimated regression equation:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_j x_{ij}$$

The loss function minimized with the OLS procedure:

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_j x_{ij})^2$$

# Trend surface

General form:

$$y_i = f(\text{lat}_i, \text{long}_i) + \epsilon_i$$

First degree:

$$y_i = \beta_0 + \beta_1 \text{lat}_i + \beta_2 \text{long}_i + \epsilon_i$$

Second degree:

$$y_i = \beta_0 + \beta_1 \text{lat}_i^2 + \beta_2 \text{lat}_i + \beta_3 \text{lat}_i \cdot \text{long}_i + \beta_4 \text{long}_i + \beta_5 \text{long}_i^2 + \epsilon_i$$

# Spatially-varying Coefficients

## Expansion Method:

Use the trend surface function in place of a constant coefficient:

$$y_i = f_{(lat_i, long_i)}^1 + f_{(lat_i, long_i)}^2 x_{i1} + \epsilon_i$$

$$f_{(lat_i, long_i)}^1 = \beta_{01} + \beta_{02} lat_i + \beta_{03} long_i$$

$$f_{(lat_i, long_i)}^2 = \beta_{11} + \beta_{12} lat_i + \beta_{13} long_i$$

## Geographically Weighted Regression:

Instead of fitting a regression model to all the observations, it fits to a subset of them each time based on the bandwidth defined.

It is like a spatial moving average; instead of calculating an average, we are fitting a regression model to each window.

# Spatial error model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} + \underbrace{\epsilon_i}_{\text{Random Error}}$$

The error term is no longer independent of each other but follows an autoregressive form:

$$\epsilon_i = \lambda \sum_k w_{ik} \epsilon_k + u_i$$

Noting that the  $u_i$  is the pure random component now.

# Capstone activity expectation

For lab activity submission, you need to at least conduct **descriptive statistics**.

Load the data and use the `summary` method or provide some descriptive plots like histograms.

```
1 summary(df)
```

mpg	cyl	disp	hp
Min. :10.40	Min. :4.000	Min. : 71.1	Min. : 52.0
1st Qu.:15.43	1st Qu.:4.000	1st Qu.:120.8	1st Qu.: 96.5
Median :19.20	Median :6.000	Median :196.3	Median :123.0
Mean :20.09	Mean :6.188	Mean :230.7	Mean :146.7
3rd Qu.:22.80	3rd Qu.:8.000	3rd Qu.:326.0	3rd Qu.:180.0
Max. :33.90	Max. :8.000	Max. :472.0	Max. :335.0

# Activities for today

- We will work on the following chapter from the textbook:
  - Chapter 28: Activity 13: Area Data V
  - Chapter 30: Activity 14: Area Data VI
- The hard deadline is **Tuesday, March 10.**