

# Regression Analysis

## Analysis of Variance

**Nicoleta Serban, Ph.D.**

*Professor*

School of Industrial and Systems Engineering

Basics Concepts

# About This Lesson



# ANOVA: Analysis of Variance

Population 1:  $(\mu_1, \sigma_1^2) \rightarrow \text{Sample 1: } (Y_{1,1}, \dots, Y_{1,n_1}) \rightarrow (\bar{Y}_1, s_1^2)$

Population 2:  $(\mu_2, \sigma_2^2) \rightarrow \text{Sample 2: } (Y_{2,1}, \dots, Y_{2,n_2}) \rightarrow (\bar{Y}_2, s_2^2)$

.....

Population k:  $(\mu_k, \sigma_k^2) \rightarrow \text{Sample k: } (Y_{k,1}, \dots, Y_{k,n_k}) \rightarrow (\bar{Y}_k, s_k^2)$

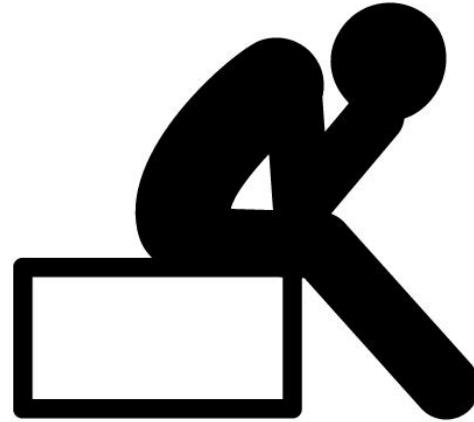
ANOVA: Comparing the means of multiple samples

# ANOVA Example 1: Global Suicide

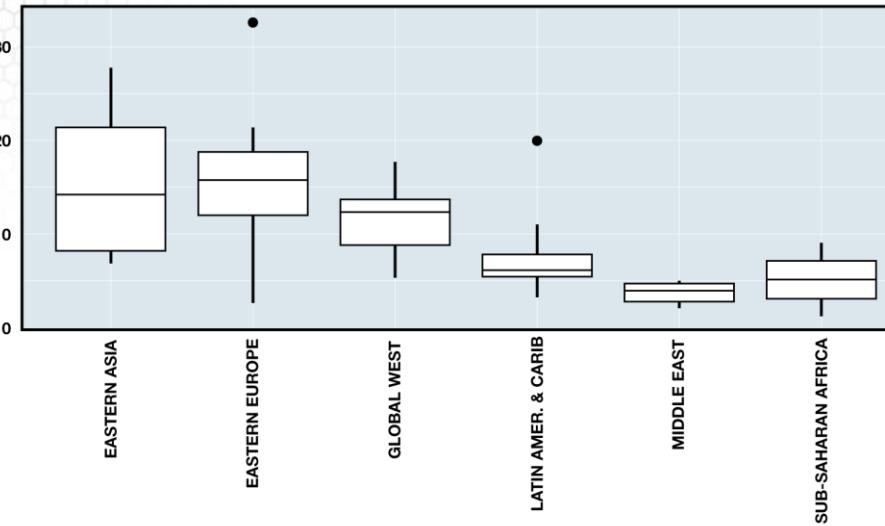
**Data Source:**

Suicide Rate: Kaggle

[https://www.kaggle.com/russellyates88/  
suicide-rates-overview-1985-to-2016](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016)



# ANOVA Example 1: Suicide Rate & Region



1. Is there a difference in the suicide rate by region?
2. Which region has higher suicide rate?



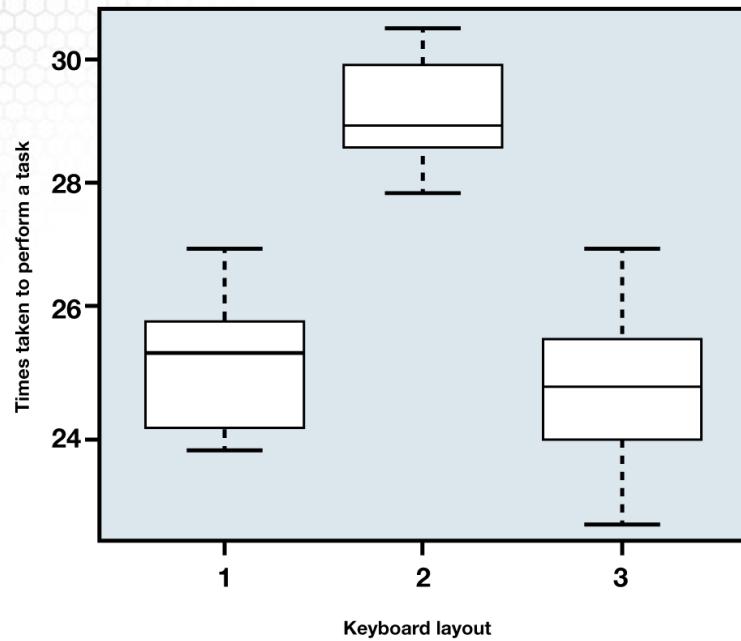
# ANOVA Example 2: Keyboard Layout

Three different keyboard layouts are being compared in terms of typing speed.



Layout 1	Layout 2	Layout 3
23.8	30.2	27.0
25.6	29.9	25.4
24.0	29.1	25.6
25.1	28.8	24.2
25.5	29.1	24.8
26.1	28.6	24.0
23.8	28.3	25.5
25.7	28.7	23.9
24.3	27.9	22.6
26.0	30.5	26.0
24.6	*	23.4
27.0	*	*

# Operation Time by Keyboard Layout



1. Is there a difference in the time taken to perform a task?
2. Which layout is more effective?

# ANOVA: Objectives

Primary objectives in ANOVA:

1. Analysis of the variability in the data – the ANOVA table
2. Testing for equal means

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

3. Estimation of simultaneous confidence intervals for the mean differences

$$\mu_i - \mu_j \text{ for } i \text{ and } j = 1, \dots, k$$

# Summary

