**Universidade de São Paulo**

**Faculdade de Filosofia, Letras e Ciências Humanas**

**Departamento de Ciência Política**

**FLS-6183 & FLP-468**

**Métodos Quantitativos de Pesquisa II**

**2º semestre / 2019**

**Lorena G. Barberia**

**Lab 4 // Class 5**

**Multicollinearity**

In this assignment, we will continue to work with simulated data. In the last class, we created and used matrix operations. Today, we will create data with a simulation using the matrix command to generate the correlations between the explanatory variables.

1. **Please fill in the table with the results you obtained after running the do file commands.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case 1. Correlation (x1, x2)=0** | | | | **Case 2. Correlation (x1, x2)=0.4** | | |
| **Sample size** | **N=10** | **N=30** | **N=100** | **N=10** | **N=30** | **N=100** |
| **Coefficient x1** |  |  |  |  |  |  |
| **Std Error x1** |  |  |  |  |  |  |
| **Vif x1 (1/Vif)** |  |  |  |  |  |  |
| **Coefficient x2** |  |  |  |  |  |  |
| **Std Error x2** |  |  |  |  |  |  |
| **Vif x2 (1/Vif)** |  |  |  |  |  |  |
| **Pw Corr x1 x2** |  |  |  |  |  |  |
| **Case 3. Correlation (x1, x2)=0.7** | | | | **Case 4. Correlation (x1, x2)=0.9** | | |
| **Sample size** | **N=10** | **N=30** | **N=100** | **N=10** | **N=30** | **N=100** |
| **Coefficient x1** |  |  |  |  |  |  |
| **Std Error x1** |  |  |  |  |  |  |
| **Vif x1 (1/Vif)** |  |  |  |  |  |  |
| **Coefficient x2** |  |  |  |  |  |  |
| **Std Error x2** |  |  |  |  |  |  |
| **Vif x2 (1/Vif)** |  |  |  |  |  |  |
| **Pw Corr x1 x2** |  |  |  |  |  |  |

1. **Let us interpret the results in the table above.**
   1. **In case 1, what did you observe between the samples as the sample size increases?**
   2. **Case 2 is a case of a positive, but weak correlation between both explanatory variables. When N = 10 what did you observe? How did the standard errors change between N=10 and N= 30? Did you obtain better results when you increased your N to 100 observations? Is there a difference in magnitude of the effects?**
   3. **Case 3 is a case of a positive and slightly stronger correlation between both explanatory variables. When N = 10 what did you observe? How did the standard errors change between N=10 and N= 30? Did you obtain better results when you increased your N to 100 observations? Is there a difference in magnitude of the effects?**
   4. **Case 4 is a case of a positive and strong correlation between both explanatory variables. When N = 10 what did you observe? How did the standard errors change between N=10 and N= 30? Did you obtain better results when you increased your N to 100 observations? Is there a difference in magnitude of the effects?**

1. **Compare the case 2 scatter plots with the case 4 plots, describe what them represent, what do you observe increasing the sample. Please, think in terms of correlation.**
2. **What does the VIF tell us about the correlation? Compare the simulations for case 2 and case 4 to illustrate your explanation.**