



# SAS<sup>®</sup> Challenge II

## Organizador: André Felipe B. Menezes

### Objetivos

Aprimorar conhecimentos de manipulação em bancos de dados por meio do software SAS<sup>®</sup>.

### Requisitos

É essencial ter conhecimento básico em SAS<sup>®</sup>, tempo e o software instalado.

### Bancos de Dados

Os banco de dados *df's* são estimações de parâmetros de diferentes processos espaciais simulados no ambiente estatístico R. Por exemplo, o banco *df1* é um processo espacial no qual foi estimado os parâmetros *sill* e *range* por diferentes métodos (os sufixos indicam os métodos).

As variáveis com prefixo *sill* são estimações de diferentes estimadores do parâmetro denominado patamar ( $\sigma^2$ ), no qual seu verdadeiro valor é 60. Enquanto que as variáveis com prefixo *range* são estimações de diferentes estimadores do parâmetro denominado alcance ( $\phi$ ), em que seu verdadeiro valor é 30. Os sufixos no nomes das variáveis, por exemplo *\_\_mat* denota o estimador.

### Etapas

1. Importe todos os bancos em **distintos** *SAS-data-set*.
2. Verifique se os nomes das variáveis estão corretos. Caso contrário organize-os.
3. Existem valores outliers em todas as variáveis dos bancos. Para as variáveis com prefixo *sill* considere outlier valores acima de 100, enquanto que para as variáveis com prefixo *range* considere outlier os valores acima de 80. Assim substitua-os pela média de cada variável, guarde em um novo SAS-data-set. (**Observação:** a média deve ser calculada sem os valores outliers; isso deve ser feito para todos os bancos).
4. Ajustado os bancos, calcule para o banco *df5*  $\mathcal{B}(\hat{\sigma}^2)$ ,  $\mathcal{B}(\hat{\phi})$ ,  $REQM(\hat{\sigma}^2)$  e  $REQM(\hat{\phi})$ , guarde em dois SAS-data-set, denominados *vies* e *reqm*.  
Lembre-se que, dado um estimador  $\hat{\theta}$  do parâmetro  $\theta$ ,  $\mathcal{B}(\hat{\theta})$  é o viés de  $\hat{\theta}$  e  $REQM(\hat{\theta})$  é a raiz quadrada do erro quadrático médio de  $\hat{\theta}$ .
5. Construa o histograma para a variável *sill\_\_mat* do banco *df3* e ajuste uma distribuição. Para isso existem diferentes procedimentos alguns exemplos são PROC UNIVARIATE, PROC SEVERITY, etc.

## Observações

- Os nomes das variáveis são: "sill\_mat", "range\_mat", "sill\_100\_100", "range\_100\_100", "sill\_100\_300", "range\_100\_300", "sill\_300\_100" , "range\_300\_100".
- O termo “ajuste uma distribuição” na etapa 4, diz respeito a encontrar uma distribuição de probabilidade para uma série de dados relativos à medição de um fenômeno ou variável. Mais detalhes em [https://en.wikipedia.org/wiki/Distribution\\_fitting](https://en.wikipedia.org/wiki/Distribution_fitting).