

UTILIZAÇÃO DE BOOTSTRAP COMO MÉTODO PARA ESTIMATIVA DO INTERVALO DE CONFIANÇA EM PEQUENAS AMOSTRAS

Trabalho de Avaliação

RICARDO MIGUEL E SOUZA DE SOUZA

João Pessoa - PB
2025

REALIZAÇÃO:



IBAPE
INSTITUTO BRASILEIRO
DE AVALIAÇÕES E PERÍCIAS DE ENGENHARIA



IBAPE-PB
INSTITUTO BRASILEIRO
DE AVALIAÇÕES E PERÍCIAS DE ENGENHARIA

PATROCÍNIO:

CONFEA
Conselho Federal de Engenharia
e Agronomia



CREA-PB
Conselho Regional de Engenharia
e Agronomia da Paraíba



mútua
Caixa de Assistência dos Profissionais do Crea

UTILIZAÇÃO DE BOOTSTRAP COMO MÉTODO PARA ESTIMATIVA DO INTERVALO DE CONFIANÇA EM PEQUENAS AMOSTRAS

ACADÊMICO

Resumo:

A engenharia/arquitetura da CAIXA é constantemente demandada na avaliação e/ou reavaliação de empreendimentos. A técnica mainstream é, naturalmente, a regressão linear usada na maioria dos casos. Contudo as grandes empresas, notadamente quando se trata de reavaliações, alegam que os números retornados pela RL não condizem com os valores que estão sendo praticados no mercado. Pelo alegado juntam vários contratos de compra e venda no empreendimento. Desnecessário dizer que estes contratos agregados as observações de mercado para análise por RL, não tem impacto significativo nos coeficientes, portanto no valor previamente estabelecido. Por outro lado, a CAIXA informa em seus sistemas os desligamentos das unidades para cada empreendimento, na medida em que os contratos se vinculam aos adquirentes pessoa física. Estas informações tem caráter serial, históricas e mostram o desempenho e perfil das vendas. Claramente estes dados seriais podem compor uma média móvel ou mesmo uma regressão simples, para unidades com mesma tipologia, com provável viés de especificação (auto-regressão). Alternativamente à RL, a análise exclusiva destes dados pode ser feita por bootstrap, método de reamostragem de dados com repetição. Este método será apresentado neste trabalho.

Palavras chave: Bootstrap, Reamostragem, Intervalo de Confiança, Algoritmos.

Abstract:

CAIXA's engineering/architecture department is constantly required to evaluate and/or reevaluate projects. The mainstream is, naturally, linear regression, used in most cases. However, large companies, especially when it comes to reevaluations, claim that the numbers returned by linear regression do not match the values being practiced in the market. According to this claim, they combine several purchase and sale contracts in the project. These contracts, aggregated with market observations for analysis by RL, do not have a significant impact on the coefficients, and therefore on the previously established value. On the other hand, CAIXA reports in its systems the unit shutdowns for each project, as the contracts are linked to individual buyers. This information is historical and shows the performance and profile of sales. Clearly, this serial data can compose a moving average or even a simple regression, for units with the same typology, with probable specification bias (autoregression). As an alternative to RL, the exclusive analysis of these data can be done by bootstrapping, a method of resampling data with repetition. This method will be presented in this work.

Key words: Bootstrap, Resampling, Confidence Interval, Algorithms.

INDICE

1. INTRODUÇÃO	4
2. O MÉTODO BOOTSTRAP.....	4
3. O EXCEL NA REAMOSTRAGEM DE DADOS	4
4. O CÓDIGO VBA VISUAL BASIC FOR APPLICATION DO EXCEL	5
5. O CÓDIGO VISUAL BASIC COMENTADO	6
6. OUTRAS FUNCIONALIDADES DA REAMOSTRAGEM	8
7. CONCLUSÕES	8
7.1 REFERÊNCIAS	9

REALIZAÇÃO:



PATROCÍNIO:



1. INTRODUÇÃO

O uso da metodologia científica nas avaliações refletiu no crescimento e na diversidade de pacotes estatísticos voltados para este mercado. Dentre estes pacotes há vários produzidos no mercado nacional e de uso corrente pelos profissionais avaliadores. Outros oriundos do mercado internacional, embora de reconhecida qualidade no meio acadêmico e das grandes corporações, não possuem tanta representatividade fora destes.

Uma das prováveis causas é a sua complexidade, além do custo do investimento na aquisição e treinamento.¹ Contudo, pouco importando a origem destes produtos, poucos disponibilizam a reamostragem como técnica para análise dos dados, notadamente em pequenas amostras.

2. O MÉTODO BOOTSTRAP

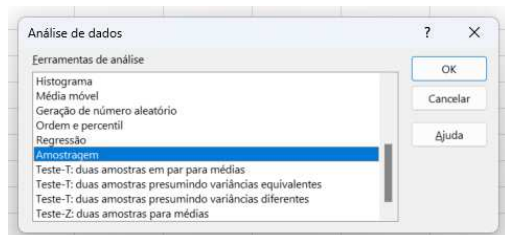
Método não paramétrico e que se infiltra na econometria aplicada, o *bootstrapping*, em estatística, é uma técnica de reamostragem que consiste em criar múltiplas amostras a partir de uma amostra original, com reposição. É utilizado para estimar a distribuição amostral de uma estatística, como a média ou a mediana, ou para construir intervalos de confiança.

Como se seleciona aleatoriamente elementos da amostra original, com reposição, significa que o mesmo elemento pode ser selecionado mais de uma vez em uma amostra bootstrap.²

3. O EXCEL NA REAMOSTRAGEM DE DADOS

Embora o *bootstrapping* em si, seja geralmente um processo simples, ele frequentemente requer um grau de repetição tão grande que só pode ser realizado com o auxílio de um computador. Em comparação com os métodos tradicionais de teste de hipóteses, o *bootstrapping* normalmente produz resultados mais fáceis de interpretar e válidos sob uma gama mais ampla de condições possíveis.³

O Excel na versão 365 (nas anteriores a partir do 2007), tem na Guia Desenvolvedor – Suplementos do Excel, as Ferramentas Análise de Dados e Ferramentas de Análise VBA que devem estar carregadas.



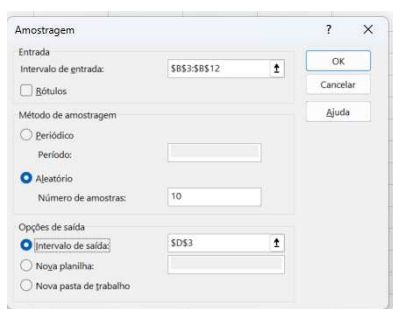
Chamando a Análise de Dados, temos a Amostragem como opção.

¹ O Excel® pertencente ao pacote do Office, embora tenha módulos de estatística na análise de dados e seja programado por Visual Basic para operações desta natureza, não é um pacote eminentemente estatístico.

² Exemplificando com uma amostra de 5 elementos 91, 92, 93, 94, 95 uma amostra bootstrap é 91, 91, 92, 93, 93.

³ <https://www.ebsco.com/research-starters/science/bootstrapping-statistics>

Que retorna o diálogo com “Intervalo de Entrada”, pergunta se os dados têm Rótulos ou não, e escolha do método de amostragem, escolhido “Aleatório” e o intervalo de saída.



Como esta operação deve ser repetida centenas de vezes, devemos operacionalizar automatizando o procedimento pelo VBA, de modo a evitar o formulário de Amostragem.

4. O CÓDIGO VBA VISUAL BASIC FOR APPLICATION DO EXCEL⁴

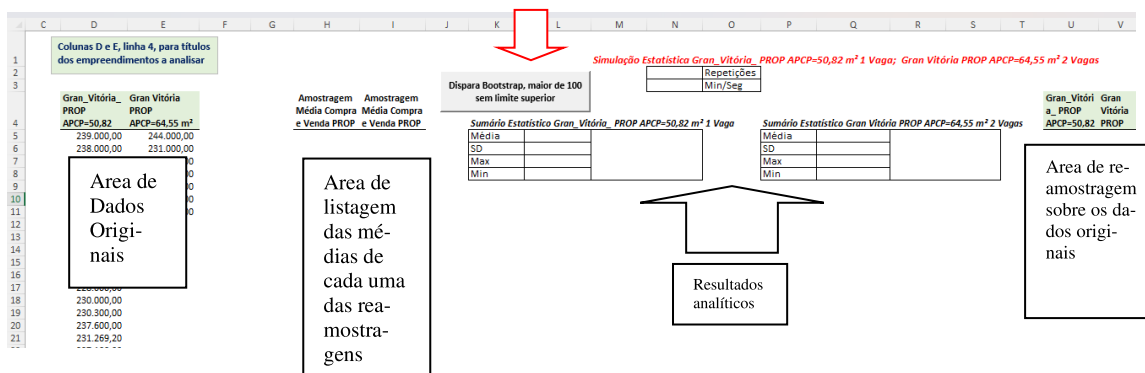
A lista de dados que será objeto de reamostragem, está num workbook add-on, arquivo Excel com código incorporado, resultando numa extensão *.xlsm.

Portanto, o arquivo é duplicável, o código é extremamente simples, posto que é acionado por apenas um botão flutuante que dispara o procedimento, razão pela qual não foi produzido um menu padrão ou mesmo uma barra flutuante.

Tanto o arquivo, quanto as planilhas podem ter qualquer nome, ou mesmo renomeadas.

O lay-out da planilha onde será desenvolvida a reamostragem, é como segue:

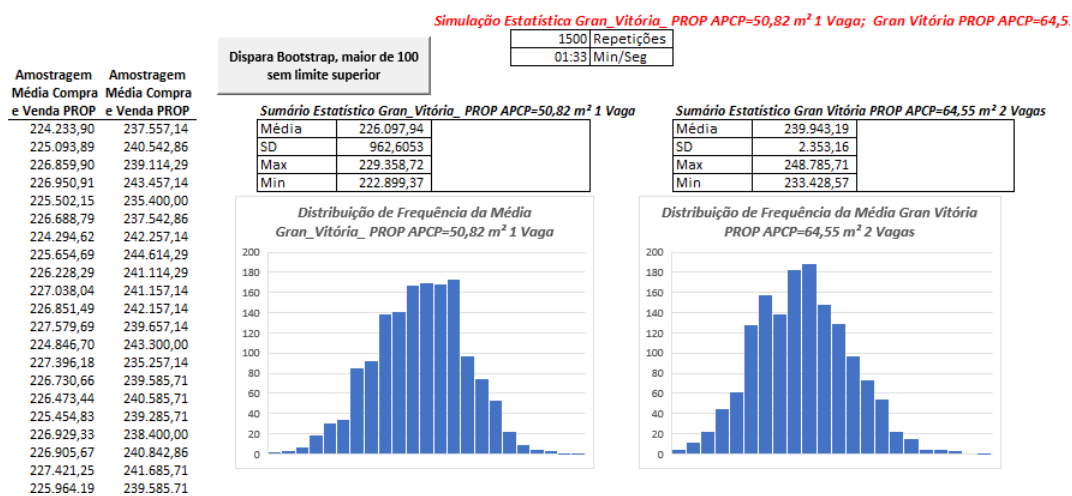
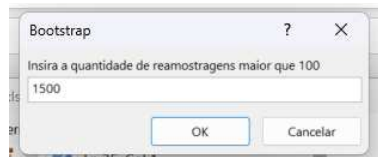
Nas colunas D e E, linha 4, sejam células D4 e E4, coloca-se o rótulo resumo da tipologia, pode-se analisar duas tipologias simultaneamente. A partir das células D5 e E5, em forma de listagem, todos os dados, pouco importando sua quantidade.



⁴ Como breve histórico, o uso das planilhas eletrônicas tem seus primórdios no Visicalc em 1979. Em 1983 surgiu o Lotus 1-2-3 que dominou o mercado de planilhas até o início da década de 90. Aplicativos com menor penetração como Quattro Pro, Multiplan e outros, não concorriam com o Lotus que era padrão nos departamentos das empresas. Em 1990 foi lançado o Excel. Em 1993 a Microsoft introduziu o conceito de pasta de trabalho composta de múltiplas planilhas e o suporte a macros do VBA. A partir do Excel 97 a combinação com o VBA se tornou vitoriosa. Hoje a dobradinha Excel+VBA é certamente a ferramenta mais poderosa para executar tarefas repetitivas, como a modelagem estatística, em questões de segundos.

A seta vermelha indica o botão de disparo da macro que pressionado, pede para que o usuário informe o número de reamostragens que deseja, evidentemente, o tempo dispendido pelo código para retornar o resultado, é função deste número.

Ilustrando o resultado, temos para 1500 reamostragens:



O tempo demandado para 1500 reamostragens das duas tipologias, foi de 1 minuto e 33 segundos, velocidade como função do processamento da máquina.

Como efeito prático, o sumário mostra os resultados para as médias, desvios padrão, mínimos e máximos das reamostragens para as duas tipologias, além das respectivas curvas de distribuição.

5. O CÓDIGO VISUAL BASIC COMENTADO

Dentro do Visual Basic, Módulo “Principal”, estão os procedimentos.

```
Sub Amostragem()
' desenvolvido em 03 de abril de 2025, para monitorar os valores propostos dos desligamentos,
' e constantes do SIOPI dos empreendimentos. Por Ricardo Miguel e Souza de Souza.

Application.Calculation = xlCalculationAutomatic
On Error Resume Next

If bChange = False Then
  AtivAnalisisTool
  bChange = True
End If

Tempo = Now()

Graph_Delete

' Limpa os resultados para lançar novos resultados
Range("N2:N3").ClearContents 'numero de loopings e tempo transcorrido
Range("L5:L8").ClearContents 'resumo da estatística da média
Range("Q5:Q8").ClearContents 'resumo da estatística da mediana/data

Application.ScreenUpdating = False

iLoop = 0

iLoop = Application.InputBox(prompt:="Insira a quantidade de reamostragens maior que 100", Title:="Bootstrap",
  If iLoop <= 100 Or IsNumeric(iLoop) = False Then Exit Sub
```


Inicialmente o código verifica se as ferramentas de Análise estão instaladas. Após abre o cronômetro do windows e limpa os resultados que estavam expostos, assim como os seus gráficos. Em seguida pede o número de reamostragens para o usuário.

```
Set ORange = Range("D7").CurrentRegion 'base original com o cabeçalho
Set NewRange = ORange.CurrentRegion.Offset(1, 0).Resize(ORange.Rows.Count - 1, ORange.Columns.Count) 'região da
Set ORange = NewRange.Columns(2) 'coluna da primeira tipologia, local em que o usuario coloca os dados
Set NewRange = NewRange.Columns(1) 'idem para a segunda tipologia

i = WorksheetFunction.CountA(NewRange)
z = WorksheetFunction.CountA(ORange)
Set DRange = NewRange.Offset(, 17) 'reamostra da primeira tipologia 17 colunas a direita da inicial
Set DataRange = NewRange.Offset(, 18) 'reamostra da segunda tipologia 18 colunas a direita da inicial

DRange.ClearContents

Range("H5:I10000").ClearContents 'limpa os resultados para inserir novos pelo looping
```

Acima, como trabalhamos com áreas de dados, definimos cada uma delas como variáveis de objeto. Abaixo, o *core* do código, chamando a macro nativa do Excel, ATPVBAEN.XLAM e o local da média de cada uma das amostras, colunas 8 e 9, a partir da quinta linha.

```
For J = 1 To iLoop
    Application.Run "ATPVBAEN.XLAM!Sample", NewRange, _
        DRange, "R", i

    Application.Run "ATPVBAEN.XLAM!Sample", ORange, _
        DataRange, "R", z

    Cells(4 + J, 8) = Application.WorksheetFunction.Average(DRange) 'media da primeira tipologia
    Cells(4 + J, 9) = Application.WorksheetFunction.Average(DataRange) 'media da segunda tipologia
    DRange.ClearContents
    DataRange.ClearContents
Next J
```

Os resultados são expostos no segmento do código abaixo, colocando os números no sumário, gerando os gráficos por uma macro de apoio, encerrando o cronômetro e finalizando o procedimento.

```
'Resultados Sumário da Média da primeira tipologia
Range("L5") = WorksheetFunction.Average(Range("H5:H" & iLoop + 4)) 'ok resultado
Range("L6") = WorksheetFunction.StDev(Range("H5:H" & iLoop + 4))
Range("L7") = WorksheetFunction.Max(Range("H5:H" & iLoop + 4))
Range("L8") = WorksheetFunction.Min(Range("H5:H" & iLoop + 4))

'Resultados Sumário da Média da segunda tipologia
Range("Q5") = WorksheetFunction.Average(Range("I5:I" & iLoop + 4))
Range("Q6") = WorksheetFunction.StDev(Range("I5:I" & iLoop + 4))
Range("Q7") = WorksheetFunction.Max(Range("I5:I" & iLoop + 4))
Range("Q8") = WorksheetFunction.Min(Range("I5:I" & iLoop + 4))

'Gera os gráficos com a nova amostragem
Graph_Developed

'Resultados da simulação
Range("N2") = iLoop
Range("N3") = Now() - Tempo
Application.ScreenUpdating = True

End Sub
```

6. OUTROS USOS/FUNCIONALIDADES DA REAMOSTRAGEM

Embora não seja um pacote estatístico específico, o Microsoft Excel do Office possui módulos de programação por VBA além de funções estatísticas que permitem obter os principais parâmetros da regressão linear.

Para o caso de pequena base de dados $Y=f(x_1, x_2, x_3...x_n)$ e quando o intervalo de confiança gerado pela regressão linear tiver grande amplitude, como suporte a decisão pode-se estimar a dispersão da média do avaliando com o auxílio da reamostragem. Para tanto é necessária a combinação dos códigos acima com a matriz LINEST do Excel.

A estimativa pontual então é listada após cada reamostragem, e tal como no caso dos desligamentos, seus parâmetros estatísticos são calculados e resumidos.

7. CONCLUSÕES

Os engenheiros/arquitetos da CAIXA detêm uma demanda gigantesca de reavaliações de empreendimentos em marcha. Daí a atualização da base de dados e, conseqüentemente, dos modelos econométricos, é um gargalo que exige superação dos profissionais.

Também há casos em que o estoque de 10% de unidades de um empreendimento de, digamos, 300 unidades residenciais, tem informações de compra e venda nos últimos meses que estão aptas a análise por bootstrap, sendo um método mais eficaz, rápido e produz resultados confiáveis.

Finalmente, é claro que a nossa amostra é uma representante válida da população.⁵

⁵ O método bootstrap coleta amostras com substituição da amostra. Cada subamostragem é uma distribuição independente idêntica, pressupondo que as subamostras vem da mesma distribuição da população e cada amostra é extraída independentemente das outras amostras.

7.1. REFERÊNCIAS

- Walkenbach, John. (2004). *Excel 2003 Power Programming With VBA*. John Wiley Consumer.
- Green, John; Bullen, Stephen. (2004). *Excel 2003 VBA Programmer's Reference*. John Wiley Consumer.
- Walkenbach, John. (2004). *Microsoft Excel 2002 Formulas*. John Wiley Consumer.
- Jacobson, Reed. (1997). *Microsoft Excel 97 Visual Basic Passo a Passo*. Microsoft Press.
- Soucie, Ralph et al. (1994). *Excel 5 For Windows Técnicas Profissionais*. Berkeley Brasil Editora.
- Jellen, Bill; Syrstad, Tracy. (2004). *Macros e VBA para Microsoft Excel*. Editora Campus.
- Roman, Steve. (2000). *Desenvolvendo Macros no Excel*. Editora Ciência Moderna.
- Levine, David; Berenson, Mark. (1998) *Estatística Teoria e Aplicações Usando Microsoft® Excel em Português*. Livros Técnicos e Científicos Editora.
- Madala, G.S. (2001). *Introdução à Econometria*. Livros Técnicos e Científicos Editora.
- Matos, Orlando Carneiro. (2000). *Econometria Básica*. Editora Atlas.
- Hill, Carter; Griffiths, William; Judge, George. (2003). *Econometria*. Editora Saraiva.
- Gujarati, Damodar. (2000). *Econometria Básica*. Makron Books.

SITES RECOMENDADOS:

- | | |
|---|---|
| http://www.add-ins.com | http://unistat.com |
| http://pucrs.br/famat/statewb | http://xlstat.com |
| http://mrexcel.com | http://msdn.microsoft.com/vba/ |
| http://vbnumericalmethods.com | http://excelbestpractices.com/excel/ |
| http://xcelfiles.com | http://www.ozgrid.com/Services/neural-network-software.htm |
| http://jjoao.cjb.net | http://www.vbnumericalmethods.com/VBNumericalMethods/ |
| | http://www.mindspring.com/~tflynn/vbalinks.html |