



14CBA
CONGRESSO BRASILEIRO DE ATUÁRIA

Bootstrap – Contribuições no cálculo do Risco Ajustado (IFRS17)

■ Bootstrap

Solução

- O *Bootstrap* é uma técnica estatística não paramétrica computacionalmente intensiva de reamostragem, BLADLEY EFRON (1979)
- “A ideia de que é possível obter propriedades de grandes amostras a partir de poucas observações”
- Permite estimar a distribuição de uma estatística por meio de repetidas amostras extraídas com reposição da amostra original → amostras são independentes do valor que veio antes delas

■ Bootstrap

Solução

- realizado n sorteios, sendo n o número de observações disponíveis na amostra, com reposição da amostra inicial, o que origina uma amostra *bootstrap*. Este procedimento deve ser repetido B vezes para que se obtenham B amostras *bootstrap*. Em posse destas B amostras, é possível construir uma distribuição *bootstrap* da variável de interesse
- A distribuição *bootstrap* converge para a distribuição verdadeira quando B tende ao infinito
- Amplamente utilizado em inferência estatística para obter estimativas e intervalos de confiança quando as suposições tradicionais são difíceis de atender ou quando se lida com amostras pequenas

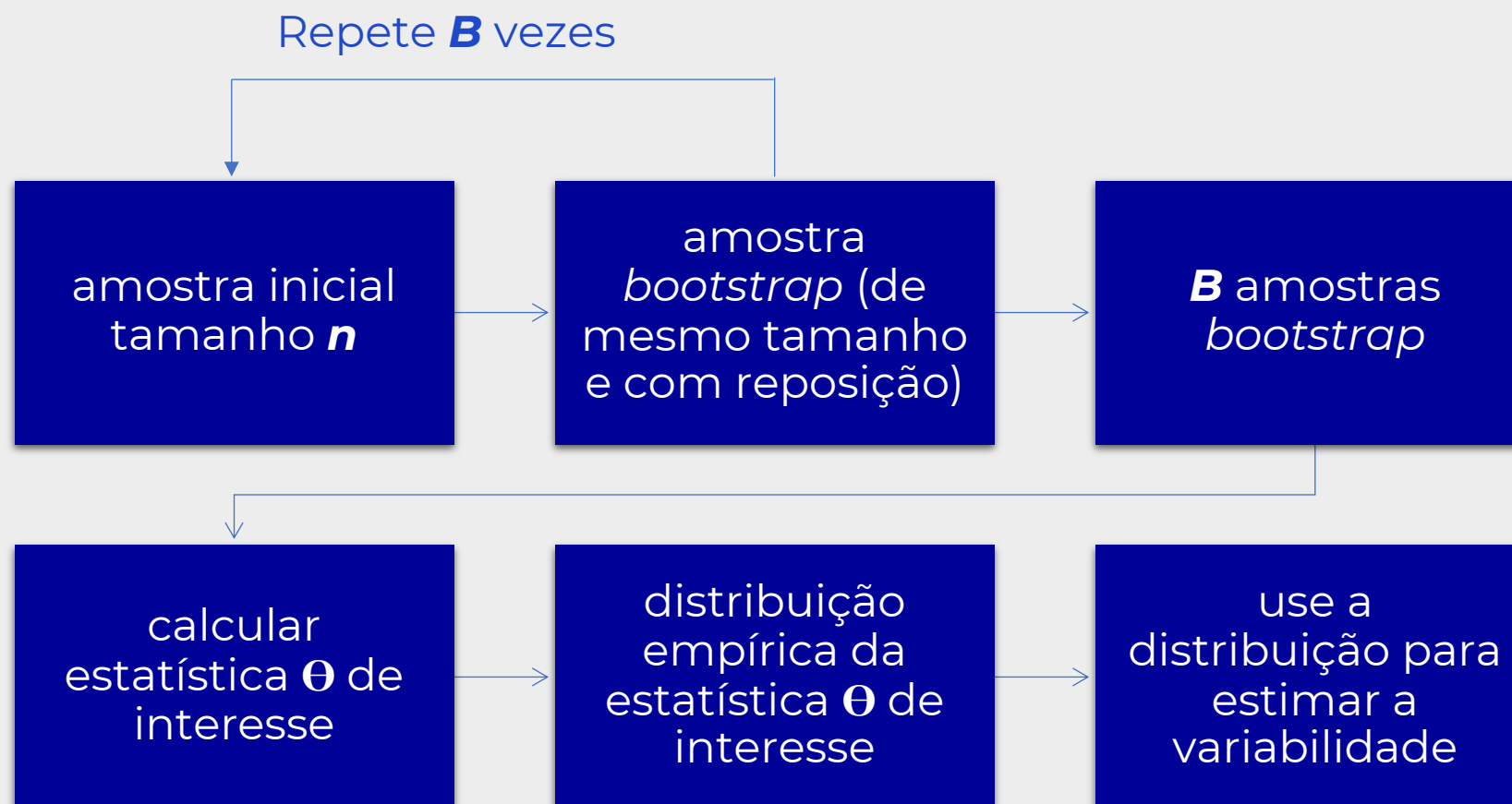
■ Bootstrap

Solução

- Aplicação em *Machine Learning* para:
 - Validação Cruzada: criar subconjuntos de dados para validar o desempenho do modelo
 - Métodos Ensemble (Bagging): criar múltiplas amostras de treino (*Random Forest*)
 - Estimativa de Erro: ajuda a entender a variabilidade do erro do modelo já treinado
 - Intervalos de Confiança: métricas de desempenho (como acurácia ou precisão), ajudando a quantificar a incerteza nas previsões
 - Correção de Overfitting: ajuda a avaliar o modelo em múltiplas amostras para garantir que o desempenho não seja apenas resultado do sobreajuste
- Para *Deep Learning*: *Hiperparametrização, dropout e data augmentation*

■ Bootstrap

Solução - Processo macro



■ Bootstrap

Solução - Vantagens do método

- Não requer suposição paramétrica: Flexível para diferentes distribuições de dados
- Versátil: Adaptável e aplicável a diversas estatísticas e situações
- Facilidade de implementação: Simples de usar com softwares estatísticos ou mesmo as planilhas

■ Bootstrap

Solução - Limitações

- Custo computacional: Pode ser intensivo em termos de processamento, mesmo assim menos se comparado com uma simulação paramétrica baseada em processos.
- Dependência da amostra original: Qualidade das estimativas depende da qualidade e variabilidade da amostra original

■ Contexto

Problema

- Necessidade: Calcular o Risco Ajustado (RA) do fluxo referente a uma carteira de seguro, com objetivo de atender a norma IFRS17
- O RA reflete a compensação que a companhia requer para convivência com a incerteza sobre o valor dos fluxos de caixa que surgem do risco não financeiro
- Dada alguma situação específica que impossibilite a realização de simulações paramétricas dos fluxos baseadas nos processos e/ou características do risco ou, ainda, no caso do fluxo ser gerado por fórmulas fechadas ou por valores médios de forma que somente haverá como resultado um único fluxo amostral representativo do fluxo de caixa futuro estimado
- Tais situações podem gerar dificuldades para estimar o RA

■ Simulação

Parâmetros

- Simulações controladas do fluxo referentes aos valores mensais (1ª simulação com 120 meses e a 2ª simulação com 24 meses) utilizando uma distribuição conhecida, no caso a distribuição *Gamma* ($shape = 2, scale = 3$)
- A média da distribuição *Gamma* = $shape * scale$
- O desvio padrão da distr. *Gamma* = $raiz(shape * scale^2)$
- Foi utilizado teste KS (*Kolmogorov-Smirnov*) para aderência → H_0 : A amostra é retirada da *Gamma*

■ Simulação

Parâmetros

- RA será calculado utilizando a distribuição das somas dos fluxos (soma de *Gammas*) geradas via *bootstrap* e pela abordagem do VaR →
$$Z_{(\text{percentil}\%)} - \text{média} = Z_{(\text{percentil}\%)} - (\text{shape} * n * \text{scale})$$
- Exemplo para a simulação de $n = 120$ meses e com percentil 80 → $758,80 - (2 * 120 * 3) = 758,80 - 720 = 38,80$
- A simulação utilizou o Python com as bibliotecas *numpy*, *scipy.stats* e *matplotlib*

■ Simulação

Minuto Teoria: Soma de *Gammas* vs Teorema Central do Limite (TCL)

Se você tem uma amostra X_1, X_2, \dots, X_n de uma distribuição Gamma com parâmetros α (forma) e β (escala), ou seja, $X_i \sim \text{Gamma}(\alpha, \beta)$ para $i = 1, 2, \dots, n$, então:

- **Soma das Amostras:** $S = X_1 + X_2 + \dots + X_n$

Portanto, se você soma n variáveis independentes X_i , onde cada $X_i \sim \text{Gamma}(\alpha, \beta)$, então a soma S segue uma distribuição Gamma com:

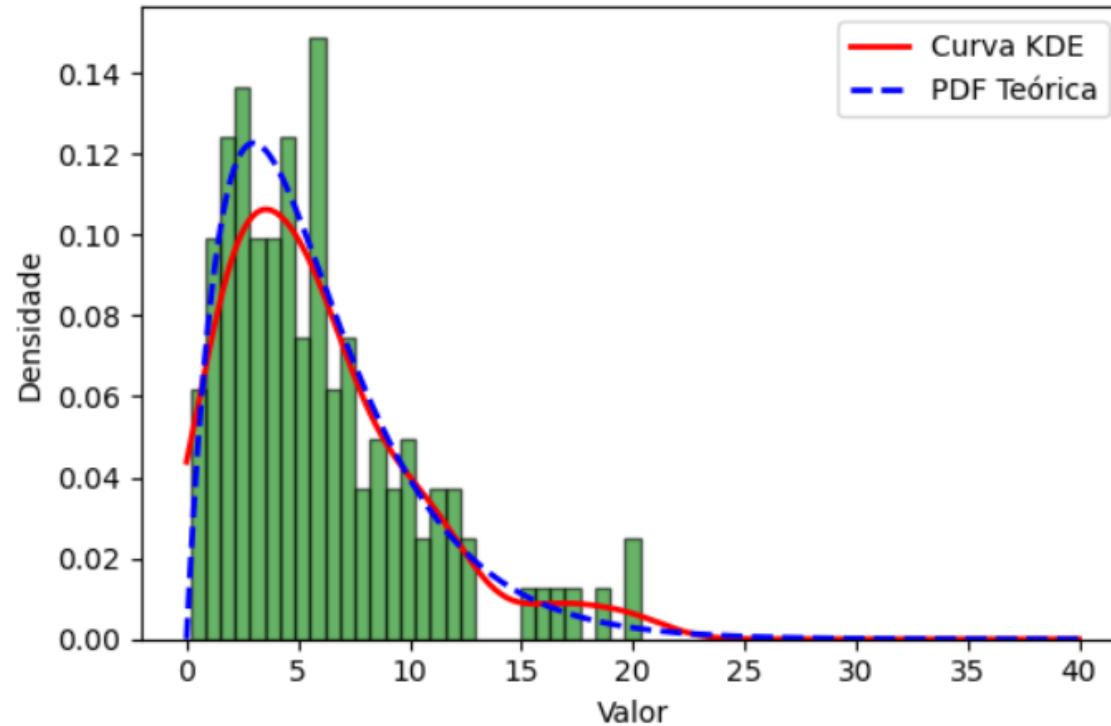
- **Forma:** $\alpha_{\text{total}} = n \cdot \alpha$
- **Escala:** $\beta_{\text{total}} = \beta$

Aplicação à Distribuição Gamma: Se a forma da distribuição Gamma é grande, a distribuição da soma ou média de várias amostras dessa distribuição Gamma se aproximará de uma distribuição Normal devido ao TCL.

■ Simulação 1

120 meses – amostra gerada da *Gamma* (2, 3)

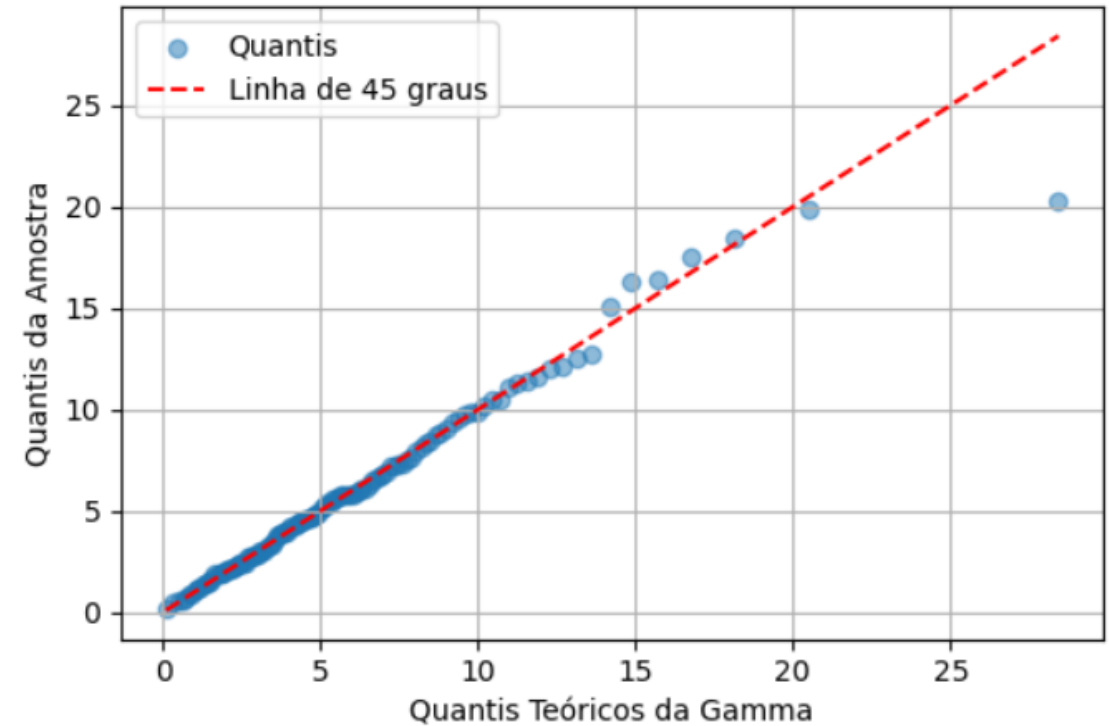
Histograma das Amostras com Curva KDE e PDF Teórica



Média: 5.8551 Esperada: 6.0000
 Desvio padrão: 4.2922 Esperada: 4.2426
 Máximo: 20.3179
 Mínimo: 0.1588

gamma test - shape= 2.0000 , scale = 3.0000
 Estatística KS: 0.0534
 Valor p: 0.8654

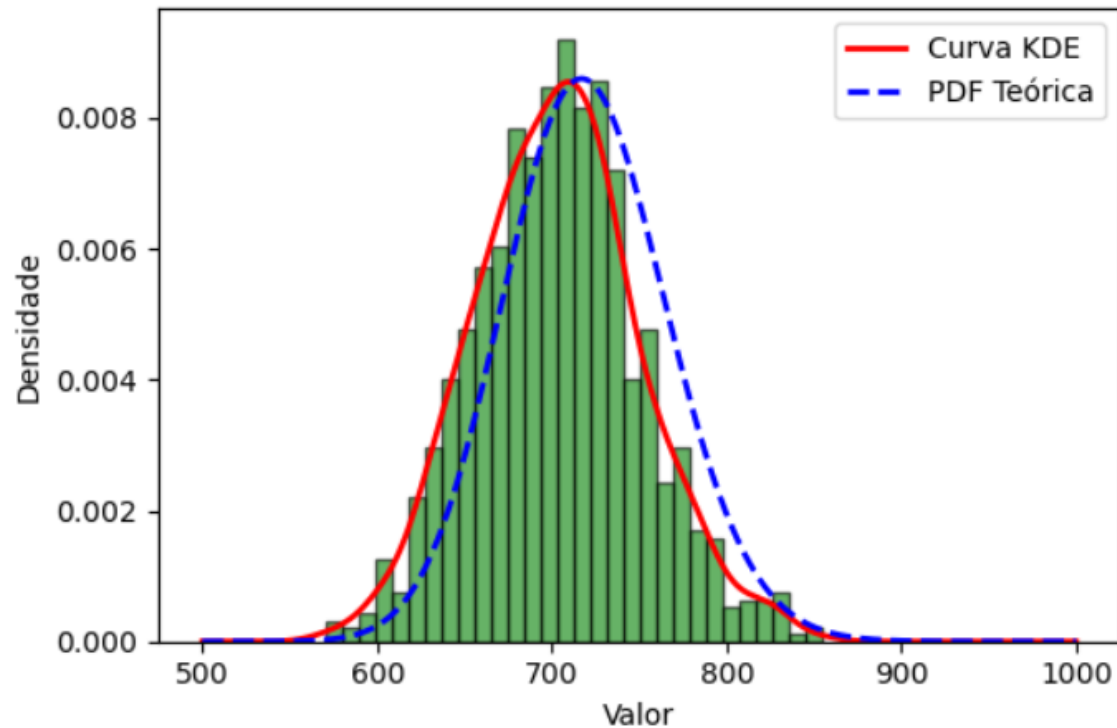
Q-Q Plot para a Distribuição Gamma



■ Simulação 1

Aplicação Bootstrap amostra tamanho 1000 – 120 meses

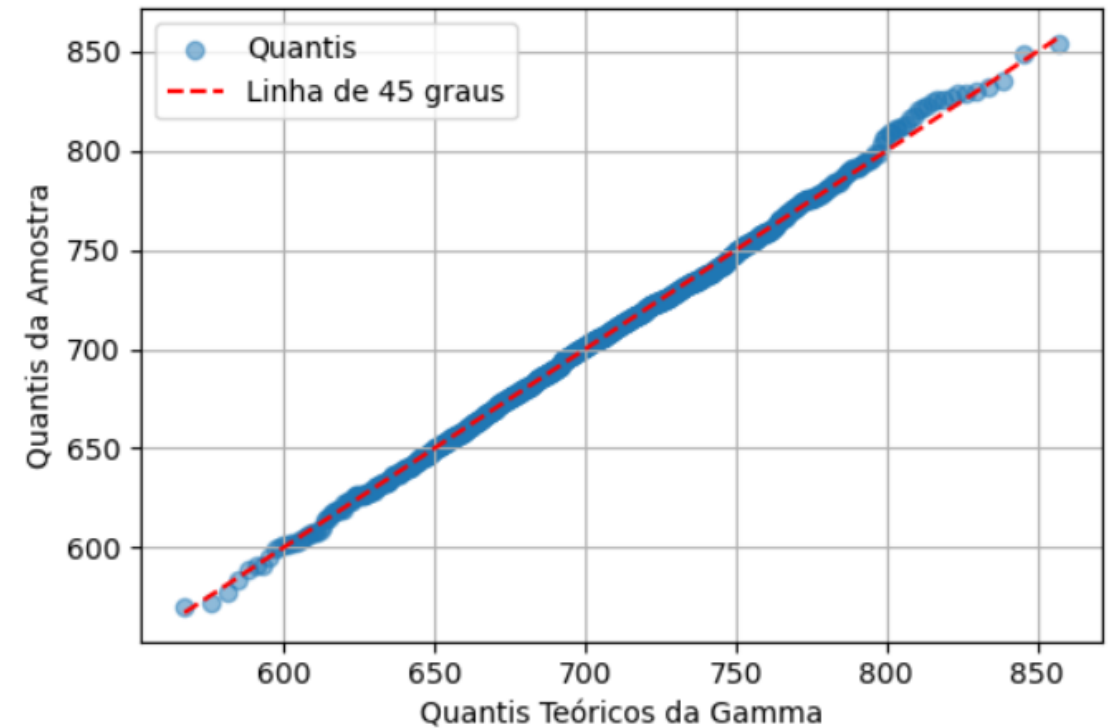
Histograma das Amostras com Curva KDE e PDF Teórica



Média: 703.1598 Esperada: 720.0000
 Desvio padrão: 46.9101 Esperada: 46.4758
 Máximo: 854.1746
 Mínimo: 570.5720

gamma test - shape= 240.0000 , scale = 3.0000
 Estatística KS: 0.1460
 Valor p: 0.0000

Q-Q Plot para a Distribuição Gamma



■ Simulação 1

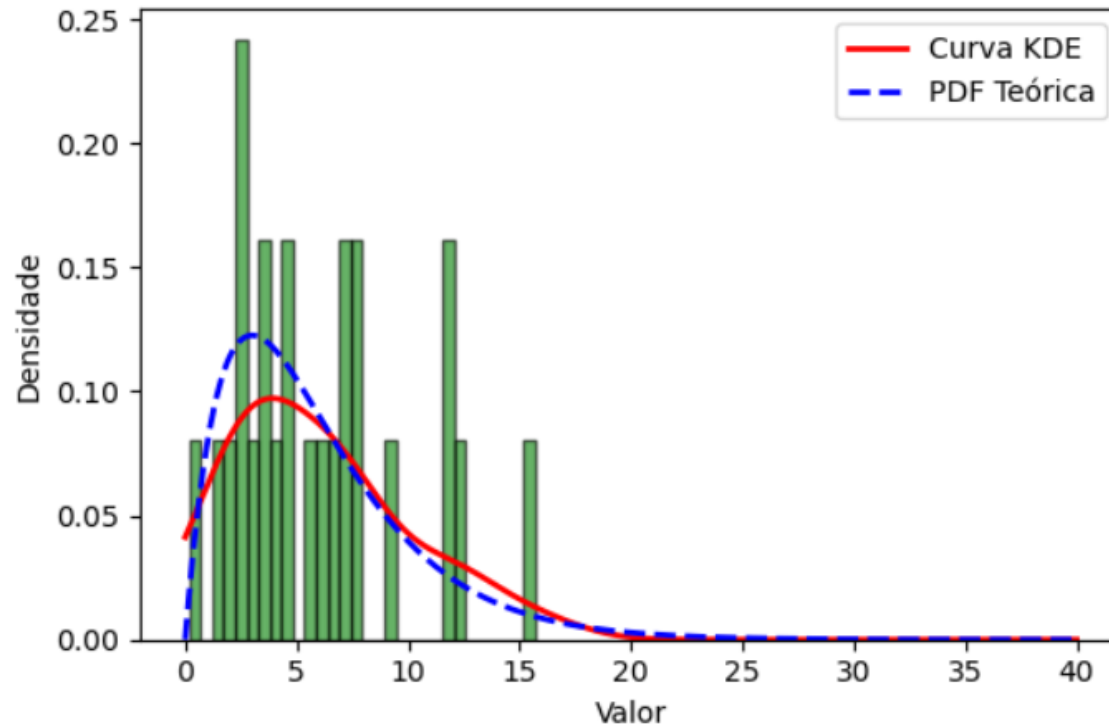
Cálculo RA – 120 meses

----- percentil 65.0%					
real	->	percentil 65.0%: 737.0426	, média: 720.0000	, RA: 17.0426	, 0.024
aprox normal	->	percentil 65.0%: 737.9081	, média: 720.0000	, RA: 17.9081	, 0.025
simulação	->	percentil 65.0%: 720.5476	, média: 703.1598	, RA: 17.3877	, 0.024
----- percentil 70.0%					
real	->	percentil 70.0%: 743.6282	, média: 720.0000	, RA: 23.6282	, 0.033
aprox normal	->	percentil 70.0%: 744.3719	, média: 720.0000	, RA: 24.3719	, 0.034
simulação	->	percentil 70.0%: 725.5482	, média: 703.1598	, RA: 22.3884	, 0.031
----- percentil 80.0%					
real	->	percentil 80.0%: 758.7950	, média: 720.0000	, RA: 38.7950	, 0.054
aprox normal	->	percentil 80.0%: 759.1150	, média: 720.0000	, RA: 39.1150	, 0.054
simulação	->	percentil 80.0%: 738.6733	, média: 703.1598	, RA: 35.5134	, 0.049
----- percentil 90.0%					
real	->	percentil 90.0%: 780.1665	, média: 720.0000	, RA: 60.1665	, 0.084
aprox normal	->	percentil 90.0%: 779.5611	, média: 720.0000	, RA: 59.5611	, 0.083
simulação	->	percentil 90.0%: 763.4338	, média: 703.1598	, RA: 60.2740	, 0.084
----- percentil 95.0%					
real	->	percentil 95.0%: 798.1131	, média: 720.0000	, RA: 78.1131	, 0.108
aprox normal	->	percentil 95.0%: 796.4459	, média: 720.0000	, RA: 76.4459	, 0.106
simulação	->	percentil 95.0%: 782.4505	, média: 703.1598	, RA: 79.2907	, 0.110

■ Simulação 2

24 meses

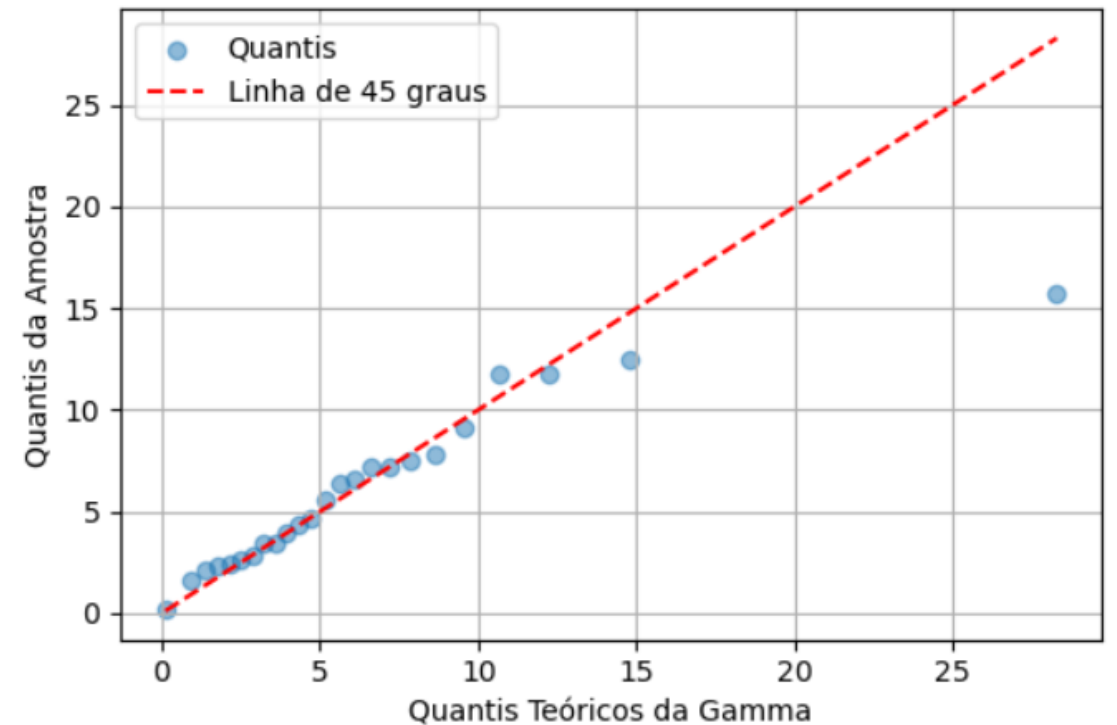
Histograma das Amostras com Curva KDE e PDF Teórica



Média: 5.9516 Esperada: 6.0000
 Desvio padrão: 3.8646 Esperada: 4.2426
 Máximo: 15.6798
 Mínimo: 0.1784

gamma test - shape= 2.0000 , scale = 3.0000
 Estatística KS: 0.0820
 Valor p: 0.9926

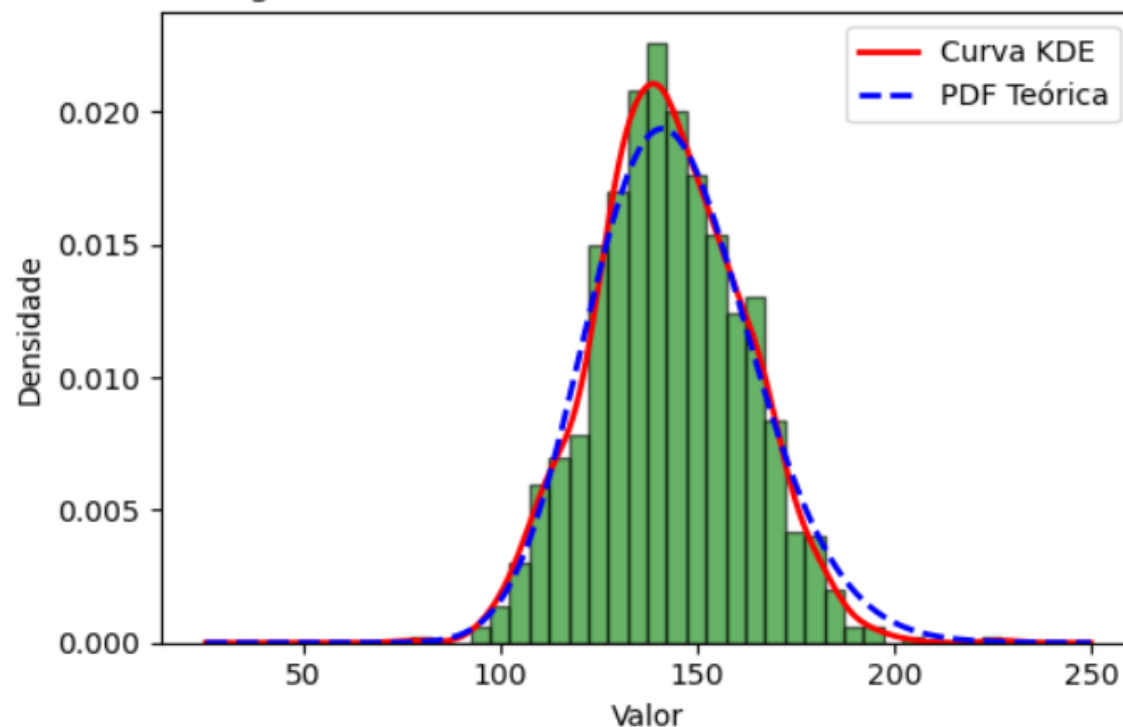
Q-Q Plot para a Distribuição Gamma



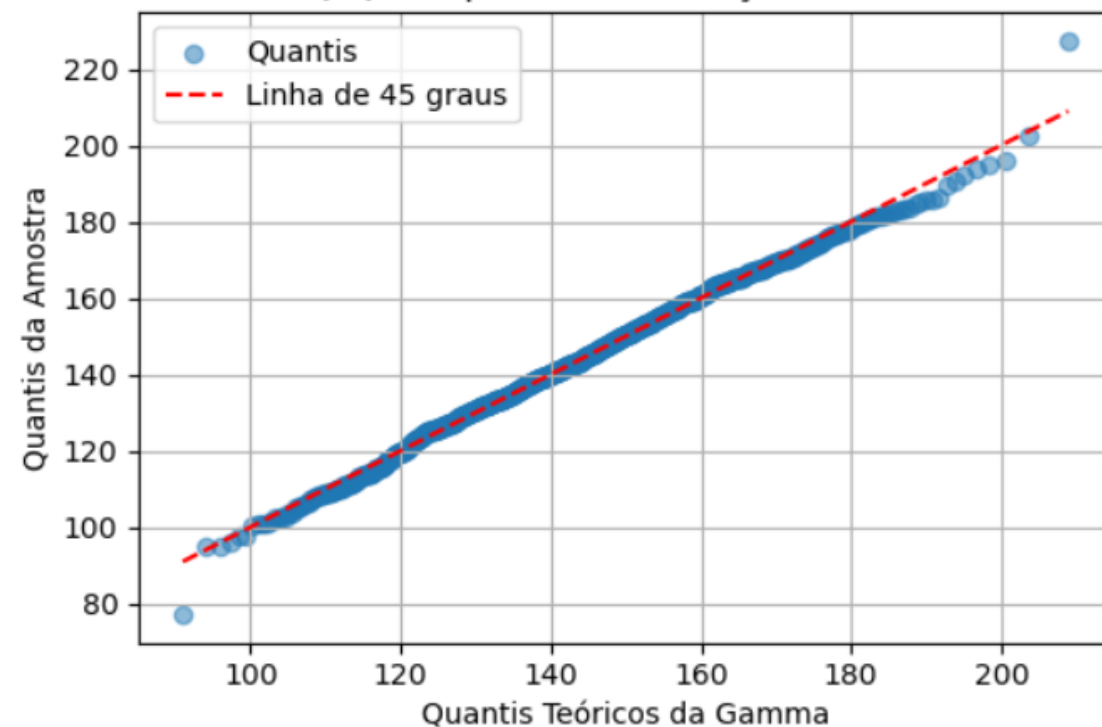
■ Simulação 2

Aplicação Bootstrap amostra tamanho 1000 – 24 meses

Histograma das Amostras com Curva KDE e PDF Teórica



Q-Q Plot para a Distribuição Gamma



Média: 142.7209 Esperada: 144.0000
Desvio padrão: 18.9407 Esperada: 20.7846
Máximo: 227.2089
Mínimo: 77.2072

gamma test - shape= 48.0000 , scale = 3.0000
Estatística KS: 0.0377
Valor p: 0.1135

■ Simulação 2

Cálculo RA – 24 meses

```
----- percentil 65.0%
real      -> percentil 65.0%: 151.1266 , média: 144.0000 , RA: 7.1266 , 0.049
aprox normal -> percentil 65.0%: 152.0087 , média: 144.0000 , RA: 8.0087 , 0.056
simulação -> percentil 65.0%: 149.7511 , média: 142.7209 , RA: 7.0302 , 0.049
----- percentil 70.0%
real      -> percentil 70.0%: 154.1331 , média: 144.0000 , RA: 10.1331 , 0.070
aprox normal -> percentil 70.0%: 154.8995 , média: 144.0000 , RA: 10.8995 , 0.076
simulação -> percentil 70.0%: 152.4480 , média: 142.7209 , RA: 9.7271 , 0.068
----- percentil 80.0%
real      -> percentil 80.0%: 161.1381 , média: 144.0000 , RA: 17.1381 , 0.119
aprox normal -> percentil 80.0%: 161.4928 , média: 144.0000 , RA: 17.4928 , 0.121
simulação -> percentil 80.0%: 159.0918 , média: 142.7209 , RA: 16.3709 , 0.114
----- percentil 90.0%
real      -> percentil 90.0%: 171.1961 , média: 144.0000 , RA: 27.1961 , 0.189
aprox normal -> percentil 90.0%: 170.6365 , média: 144.0000 , RA: 26.6365 , 0.185
simulação -> percentil 90.0%: 167.3856 , média: 142.7209 , RA: 24.6647 , 0.171
----- percentil 95.0%
real      -> percentil 95.0%: 179.8064 , média: 144.0000 , RA: 35.8064 , 0.249
aprox normal -> percentil 95.0%: 178.1876 , média: 144.0000 , RA: 34.1876 , 0.237
simulação -> percentil 95.0%: 173.7597 , média: 142.7209 , RA: 31.0387 , 0.216
```

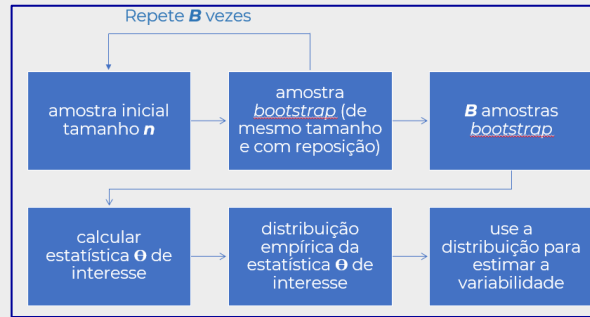
■ Simulação

Provocação

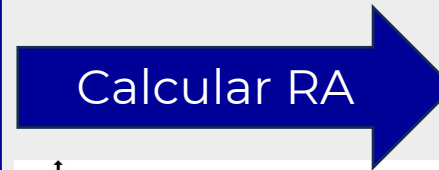
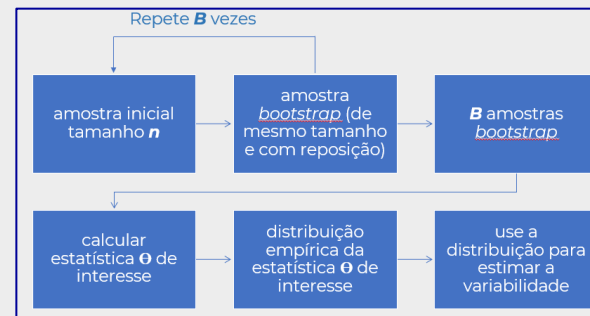
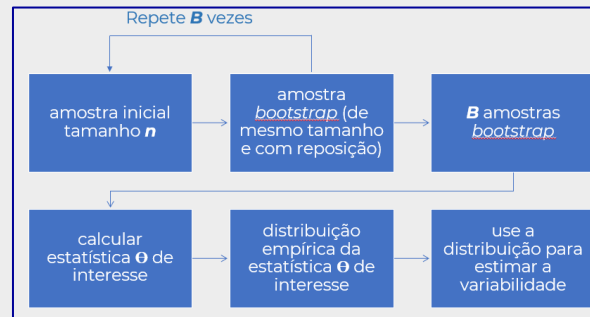
- Qual a variabilidade ou o risco embutido na simulação?
- Se fosse realizada mais uma rodada da amostragem bootstrap o resultado do RA seria o mesmo? Houve “sorte” ou “azar”? ou ainda, o resultado foi uma escolha conforme a conveniência?
- Não seria interessante tratar o RA como uma variável aleatória? (neste caso o TCL não ajuda)
- O próximo passo é gerar uma distribuição para o RA repetindo o processo de cálculo do RA, visto até aqui, várias vezes, (vide esquema a seguir)

■ Simulação

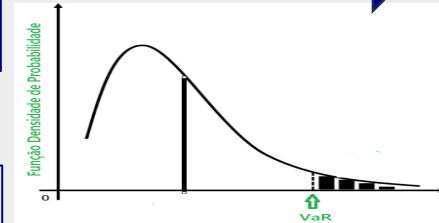
Processo ampliado



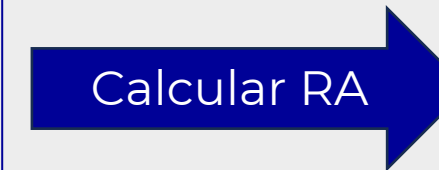
...



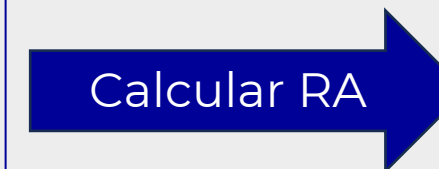
RA_1



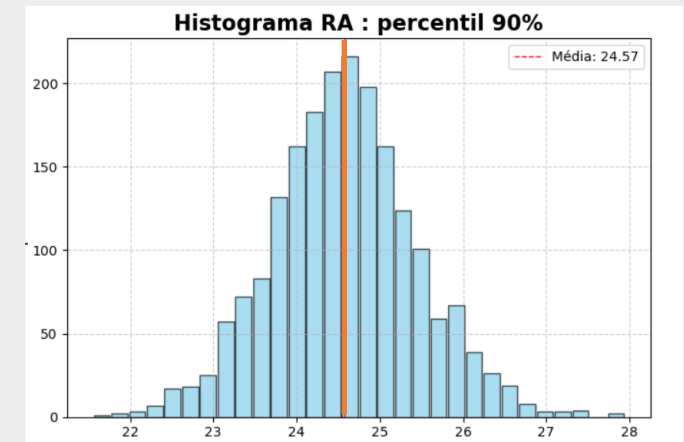
...



RA_{k-1}

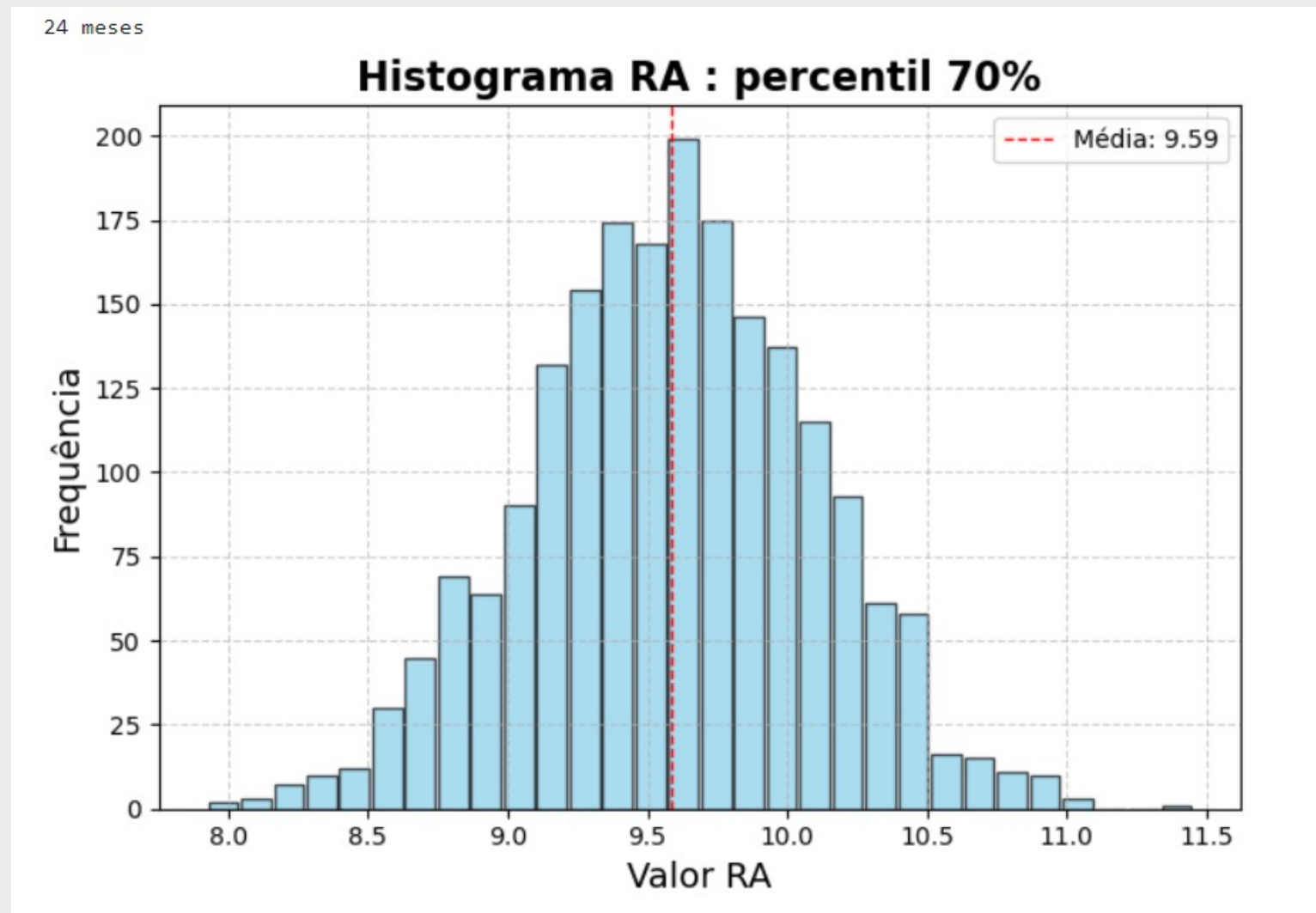


RA_k



■ Simulação 1

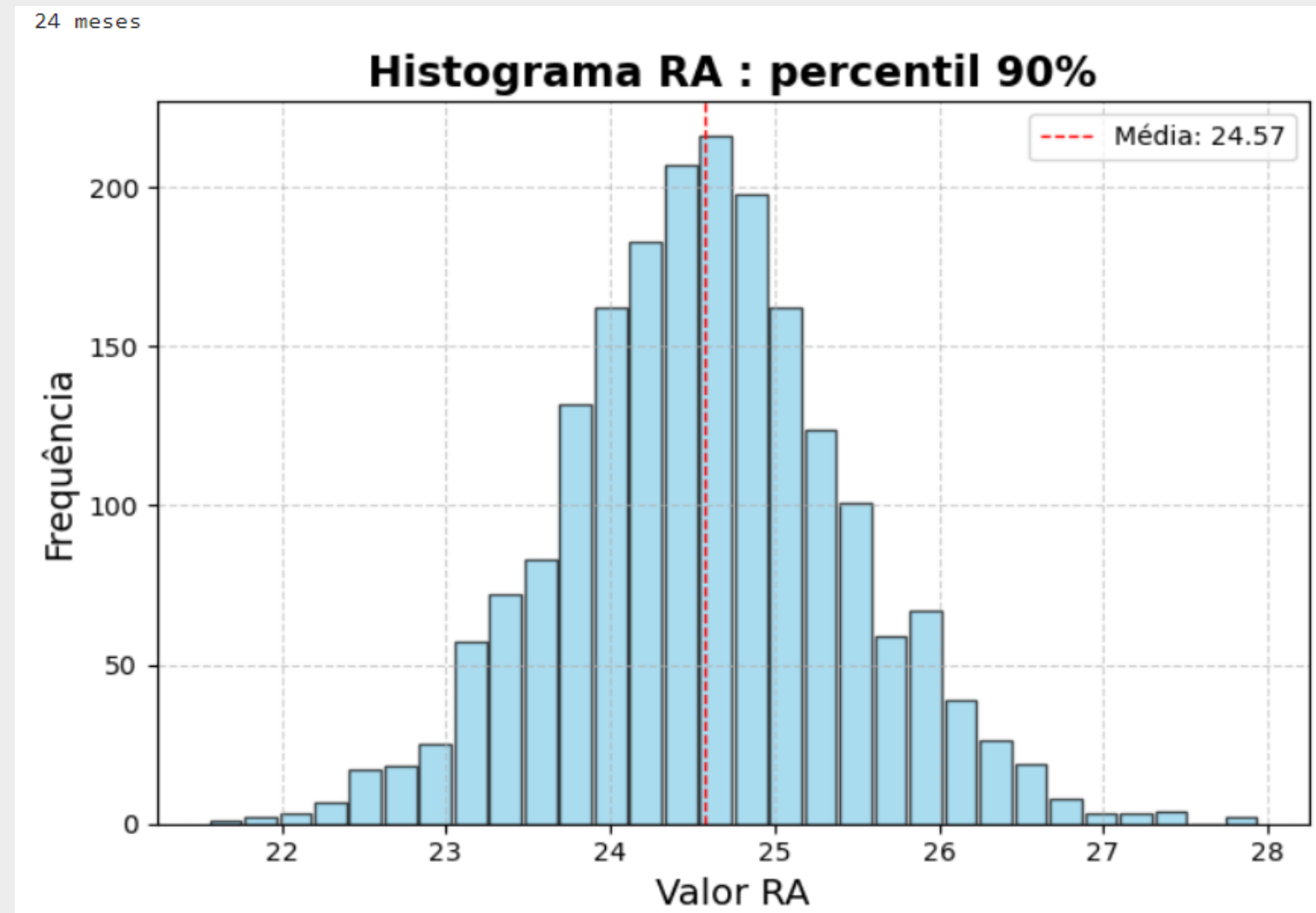
Distribuição RA – 24 meses – Dá-Ihe Monte Carlo!!! 2k x 1k



, RA: 10.1331
 , RA: 10.8995
 , RA: 9.7271

■ Simulação 2

Distribuição RA – 24 meses



RA: 27.1961
 RA: 26.6365
 RA: 24.6647

Contexto - caso real

01/04/23	9,986,859
01/05/23	1,610,111
01/06/23	491,773
01/07/23	268,747
01/08/23	117,764
01/09/23	33,294
01/10/23	13,446
01/11/23	9,507
01/12/23	4,483
01/01/24	4,230
01/02/24	2,832
01/03/24	2,034
01/04/24	1,359
01/05/24	1,259
01/06/24	1,263
01/07/24	1198
01/08/24	878
01/09/24	597
01/10/24	590
01/11/24	1,793
01/12/24	1,555
01/01/25	2,109

- Única amostra, amostra pequena e a distribuição de origem desconhecida

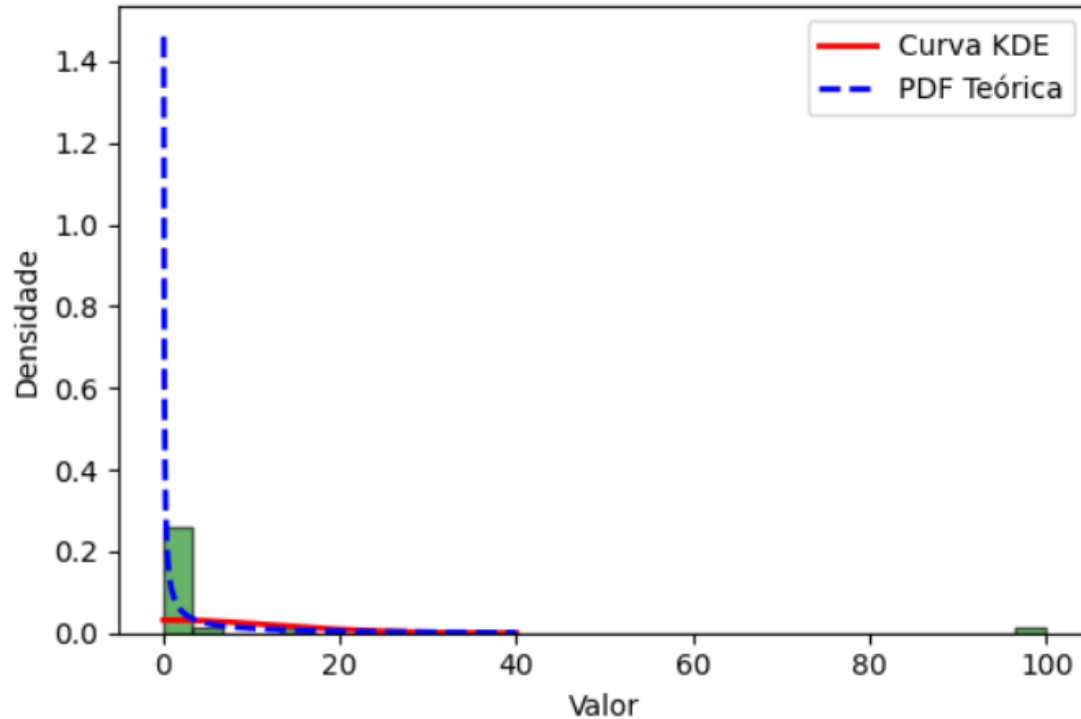
Resultado Bootstrap

Bloco	Frequência	% cumulativo	% relative
22,78	0	0,00%	0,00%
43,15	5	0,50%	0,50%
63,52	51	5,60%	5,10%
83,88	141	19,70%	14,10%
104,25	179	37,60%	17,90%
124,62	220	59,60%	22,00%
144,98	169	76,50%	16,90%
165,35	131	89,60%	13,10%
185,72	61	95,70%	6,10%
206,08	32	98,90%	3,20%
226,45	10	99,90%	1,00%
246,82	0	99,90%	0,00%
267,18	1	100,00%	0,10%
Mais	0	100,00%	0,00%
total	1000		100,00%
média	118,09		
V@R	131,32	65%	
RA	13,24	1.323.523	

■ Simulação - caso real

22 meses

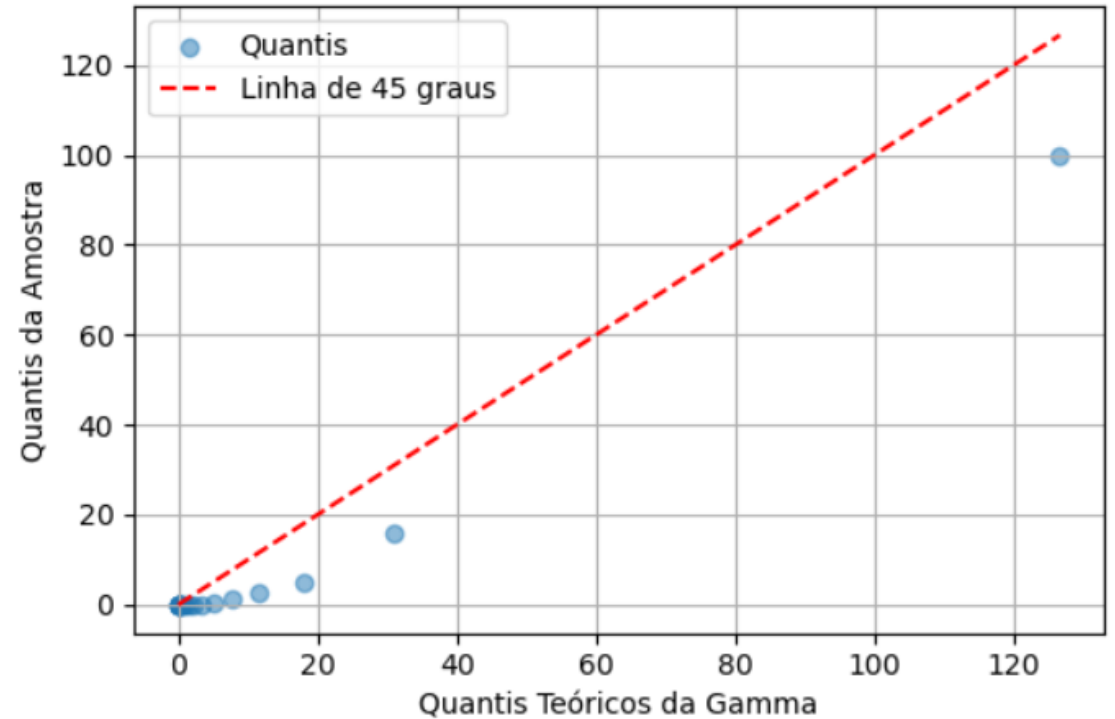
Histograma das Amostras com Curva KDE e PDF Teórica



Média: 5.7080 Esperada: 5.7420
Desvio padrão: 20.8339 Esperada: 13.5340
Máximo: 99.8686
Mínimo: 0.0059

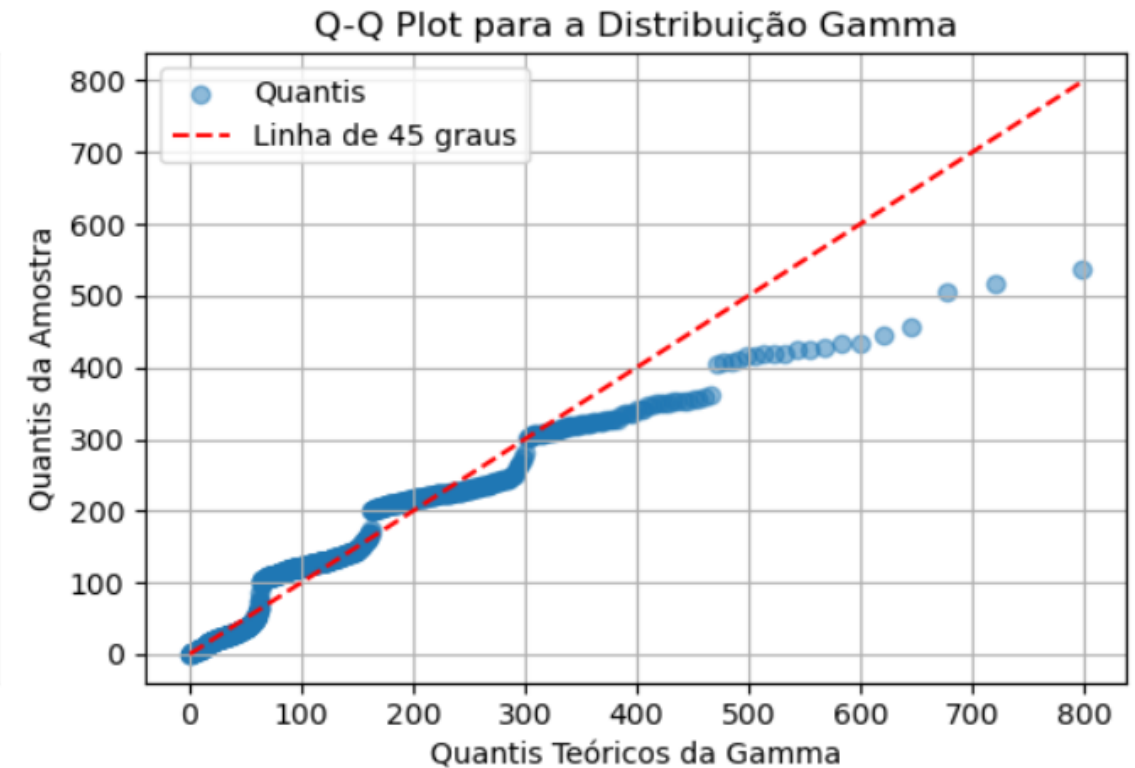
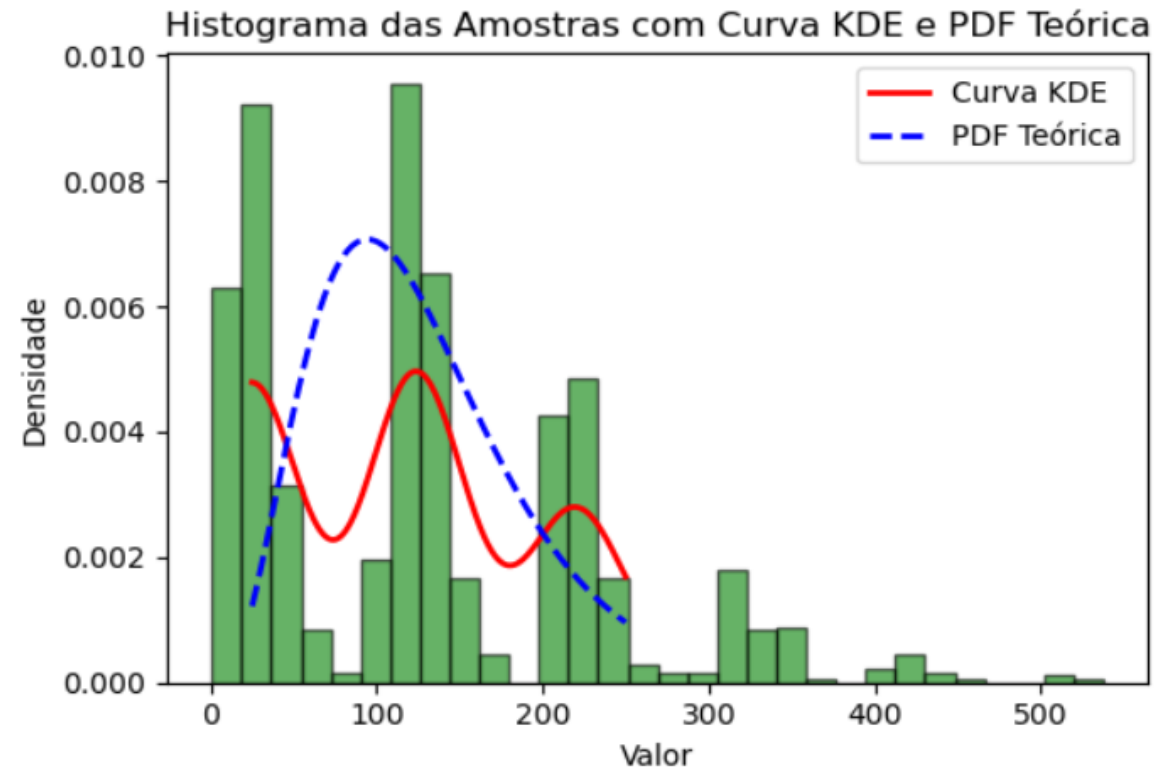
gamma test - shape= 0.1800 , scale = 31.9000
Estatística KS: 0.3230
Valor p: 0.0153

Q-Q Plot para a Distribuição Gamma



■ Simulação - caso real

Aplicação Bootstrap amostra tamanho 1000 – 22 meses



Média: 129.2657 Esperada: 126.3240
Desvio padrão: 100.8451 Esperada: 63.4802
Máximo: 537.7270
Mínimo: 0.7561

gamma test - shape= 3.9600 , scale = 31.9000
Estatística KS: 0.2630
Valor p: 0.0000

■ Simulação - caso real

Cálculo RA – 22 meses

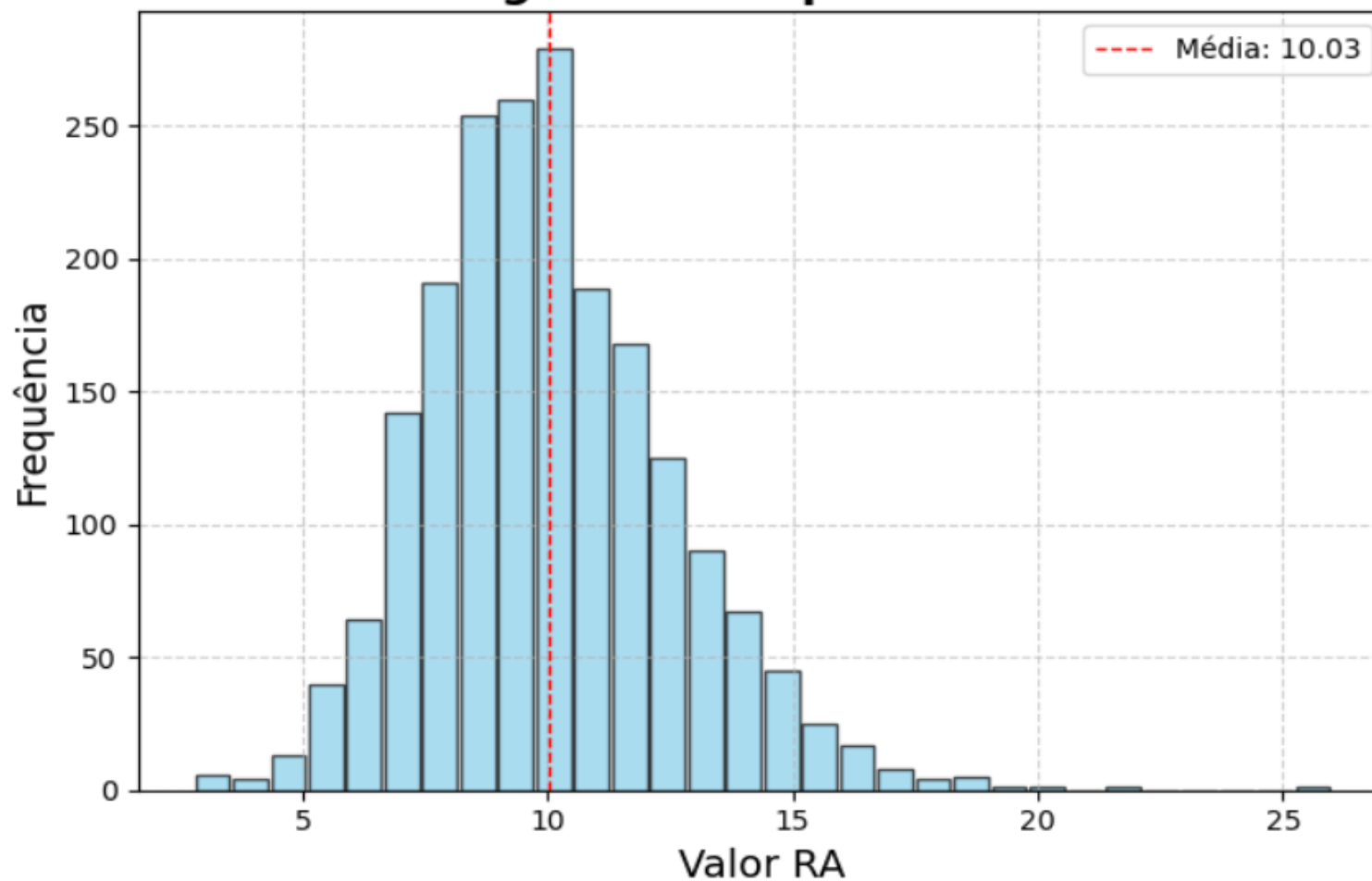
				Calculado
----- percentil 65.0%				
aprox normal (dados)	->	percentil 65.0%: 168.1234 , média: 129.2657 , RA: 38.8577		
simulação	->	percentil 65.0%: 139.1306 , média: 129.2657 , RA: 9.8649	→	13.24
----- percentil 70.0%				
aprox normal (dados)	->	percentil 70.0%: 182.1489 , média: 129.2657 , RA: 52.8832		
simulação	->	percentil 70.0%: 157.3285 , média: 129.2657 , RA: 28.0628	→	18.94
----- percentil 80.0%				
aprox normal (dados)	->	percentil 80.0%: 214.1391 , média: 129.2657 , RA: 84.8734		
simulação	->	percentil 80.0%: 217.9118 , média: 129.2657 , RA: 88.6462	→	32.22
----- percentil 90.0%				
aprox normal (dados)	->	percentil 90.0%: 258.5039 , média: 129.2657 , RA: 129.2382		
simulação	->	percentil 90.0%: 245.0963 , média: 129.2657 , RA: 115.8306	→	48.16
----- percentil 95.0%				
aprox normal (dados)	->	percentil 95.0%: 295.1411 , média: 129.2657 , RA: 165.8754		
simulação	->	percentil 95.0%: 323.6076 , média: 129.2657 , RA: 194.3419	→	64.84

■ Simulação - caso real

Distribuição RA – 22 meses

22 meses

Histograma RA : percentil 65%

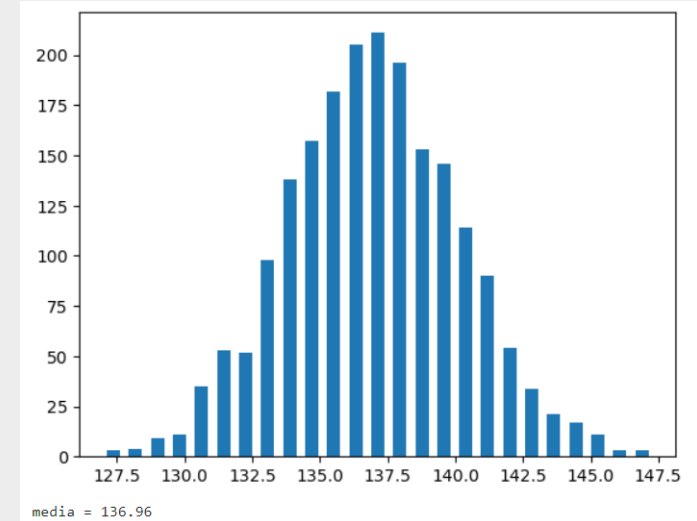
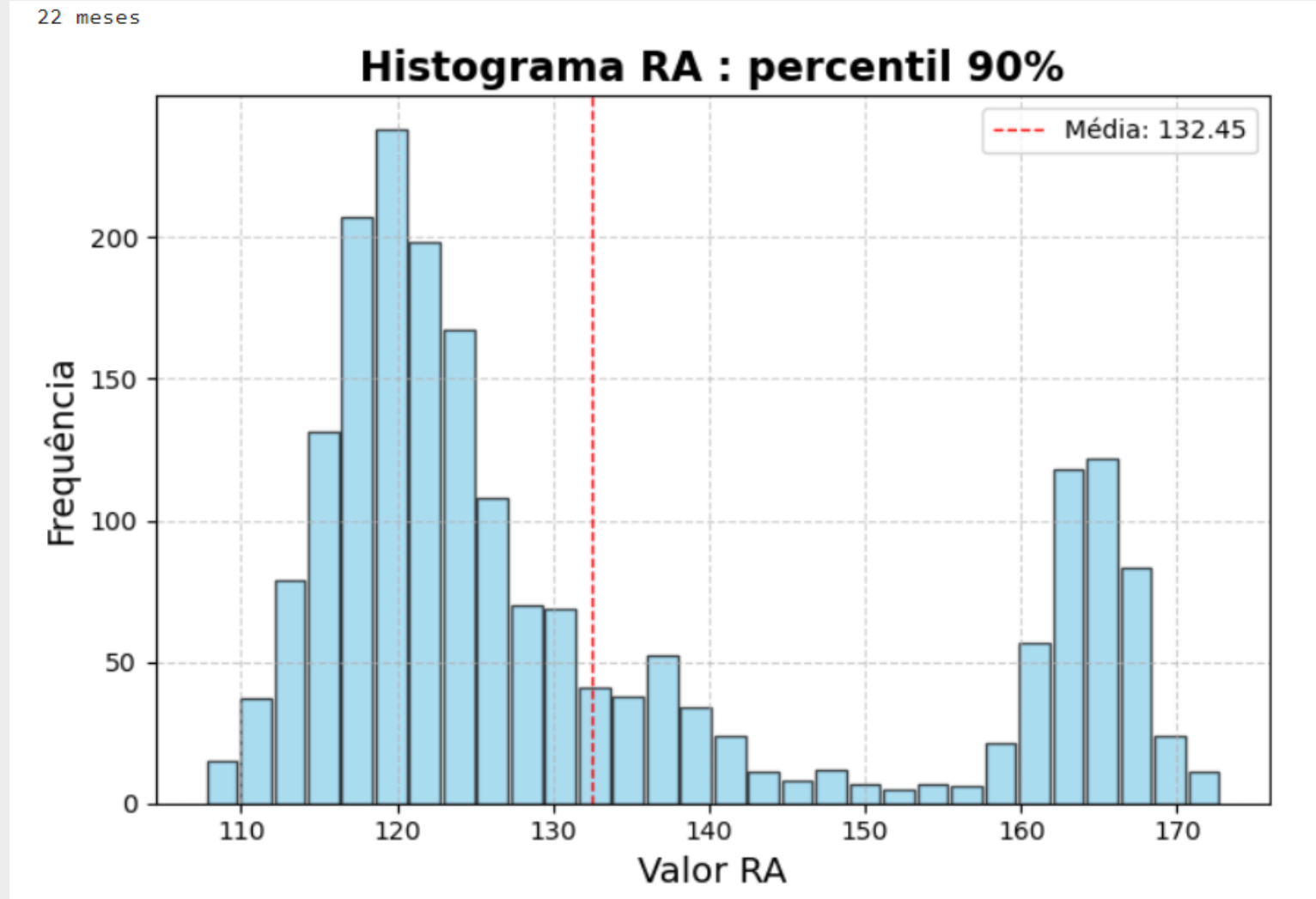


RA: 38.8577
RA: 9.8649

13.24

■ Simulação - caso real

Distribuição RA – 22 meses



RA: 129.2382

RA: 115.8306

48.16

■ Resultados

- O método provou ser satisfatório principalmente quando aplicado em situações “mais” comportadas (como na simulação controlada)
- Contudo, como era esperado, o método é suscetível a variabilidade da amostra
- Em relação aos percentis, o método se comportou melhor para os valores reduzidos (65% ou 70%), independente da situação (*trade-off* entre significância x custo x robustez)
- A abordagem da RA como variável aleatória proporciona uma opção de ferramenta para validar a performance (acurácia e precisão) do método e, neste sentido, pode ser um complemento nas análises de auditoria
- A aproximação pela Normal é eficaz, como esperado, dado o conhecimento/identificação da distribuição origem. Na ausência destes, fica dependente da aplicação do método para ser utilizado (para usar os dados amostrais), bem como, da qualidade da amostra
- Para fluxos com variabilidade alta, verificar outras estratégias para amostragem e/ou uso da mediana ao invés da média e/ou uma outra abordagem diferente do VaR
- Pela simplicidade e versatilidade, a metodologia *bootsrap* é uma opção a ser explorada dentro dos trabalhos atuariais, independentemente de ter uma ou mais amostras disponíveis

■ Referências

- Efron, B. Computers and the theory of statistics: Thinking the unthinkable, *SJAM Rev.*, 21, 460-480
- Efron, B. Bootstrap methods: Another look at the jackknife, *Annals of Statistics*, 7 (1979), n.1, to appear



14CBA
CONGRESSO BRASILEIRO DE ATUÁRIA

Obrigado!