

Distribuição de Probabilidade e Simulação

Adriana Backx Noronha (FEA/USP)

Material organizado dentro do contexto de ensino de Simulação Aplicada a Administração. Como utilizar o Excel para gerar número aleatório com base em uma distribuição de probabilidade contínua uniforme ou distribuição de probabilidade discreta não-uniforme.

Simulação é um conjunto de ferramentas que nos permite gerar cenários e trazer informações que auxiliem no processo de decisão. A base para gerar simulação é o conceito de distribuição de probabilidade; mais ainda, as ferramentas que permitam gerar um conjunto de números dentro de uma distribuição de probabilidade determinada. Esse texto apresenta como gerar as distribuições de probabilidade, permitindo aplicar o processo de simulação.

Basicamente são três passos que nos permitem selecionar, organizar e gerar números de acordo com determinada distribuição de probabilidade; esses números serão utilizados na simulação. O quadro 1 apresenta uma relação entre o tipo de distribuição e o tipo de variável.

	DISTRIBUIÇÃO UNIFORME Todos os valores com igual probabilidade	DISTRIBUIÇÃO NÃO UNIFORME Algun valor pode ter distribuição diferente
VARIÁVEL CATEGÓRICA/DISCRETA Assume valores inteiros	<p>Todos os valores têm a mesma probabilidade de serem sorteados</p> <p>Ex: jogo de dados; gênero</p> <p>Aleatórioentre() ou PROCV com probabilidades iguais</p>	<p>Ex: nº de defeitos; nº pessoas em uma fila</p> <p>PROCV(aleatório(); tabela) Binomial, Poisson</p>
VARIÁVEL CONTÍNUA Assume qualquer valor entre dois inteiros	<p>Variável quantitativa contínua; probabilidades iguais</p> <p>Ex: Custo de um produto variando entre x e y</p> <p>Aleatório() $Mín + (Máx - Min) * aleatório()$</p>	<p>Ex: Tempo entre duas chamadas</p> <p>Normal, Triangular, Exponencial, Gama</p>

Quadro 1 – Relação entre tipo de distribuição e tipo de variável

Fonte: da autora (desenhado por Anna Célia Affonso dos Santos)

Em resumo, os três passos são:

- 1) Identificar no enunciado do problema quais são os dados a serem utilizados para gerar a simulação; qual o tipo de variável e qual o tipo de distribuição (uniforme ou não uniforme)
- 2) Identificar em que quadrante está a distribuição de probabilidade que se pretende utilizar para gerar os números (cruzamento entre tipo de variável e tipo de distribuição)
- 3) Escolher as funções do Excel a serem utilizadas.

Observando o Quadro 1, basicamente temos quatro possibilidades. Neste texto, duas possibilidades são descritas em maiores detalhes: 1) Distribuição contínua uniforme; 2) Distribuição Discreta não-uniforme. Os textos foram organizados com base em material desenvolvido pelo Prof. Ronaldo Zwicker.

1. Distribuição Contínua uniforme

Números aleatórios com distribuição contínua uniforme constituem o recurso básico em simulação. Em EXCEL a função ALEATÓRIO gera números reais aleatórios uniformemente distribuídos entre 0 e 1 (Figura 1). É importante observar que o intervalo é fechado à esquerda e aberto à direita, portanto, o valor 0 pode ser gerado enquanto o valor 1 com certeza não será gerado. A precisão é de 15 casas decimais, por exemplo, 0,528978839355029.



Figura 1 - Função no Excel para gerar números reais aleatórios uniformemente distribuídos

Como a cada alteração da planilha são recalculadas todas as fórmulas, a cada recálculo é computado um novo valor aleatório. Para inibir este recálculo automático deve ser ligada a opção correspondente em **Ferramentas / Opções / Cálculo / Manual** e desligada a opção **Recalcular antes de Salvar**. Com o recálculo automático desligado é necessário acionar a tecla **F9** para computar novamente as fórmulas.

Para gerar números reais aleatórios uniformemente distribuídos num intervalo qualquer $[a, b)$ fechado à esquerda e aberto à direita deve ser usada a fórmula: $a + \text{ALEATÓRIO}() * (b - a)$ (Figura 2), por exemplo, para o intervalo 12,5 a 20,0 usamos $12,5 + \text{ALEATÓRIO}() * (20,0 - 12,5)$. O valor é gerado com 15 algarismos significativos, por exemplo, 12,6883416653930.



Figura 2 - Fórmula no Excel para gerar números reais aleatórios uniformemente distribuídos em um intervalo definido por $[a,b]$

Se o objetivo for gerar números inteiros no intervalo $[k1, k2]$, portanto o intervalo inclui $k1$ e $k2$, deve-se usar a fórmula $k1 + \text{INT}(\text{ALEATÓRIO}() * (k2 - k1 + 1))$ (Figura 3) onde INT é a função que retorna a parte inteira de um número qualquer sem arredondamento. Isso gera uma distribuição discreta uniforme.

$$k1 \quad | \quad \overbrace{\quad k1 + \text{INT}(\text{ALEATÓRIO}() * (k2 - k1 + 1)) \quad} \quad | \quad k2$$

Figura 3 - Fórmula no Excel para gerar números inteiros aleatórios uniformemente distribuídos em um intervalo definido por $[k1, k2]$

1.1 Exercício

O objetivo do exercício é verificar que os números gerados pela função ALEATÓRIO realmente se distribuem ao acaso no intervalo $[0, 1)$ e que a sua distribuição é, ou pelo menos parece ser, contínua (qualquer valor do intervalo pode ser gerado) e uniforme (cada valor tem igual probabilidade de ser gerado).

Para observar estas propriedades deverá ser gerada uma tabela com 1001 números aleatórios a partir da qual serão construídos os gráficos indicados. Sugere-se usar 1001 números em lugar de 1000 pois esse valor fornece divisões de escala mais uniformes para os eixos das abcissas nos gráficos. As tabelas e gráficos que devem ser gerados são os seguintes:

- **1) Gráfico da posição vertical dos números gerados** - Para observar que os números se distribuem de forma aleatória e contínua sugere-se usar um gráfico personalizado de tipo “linhas em 2 eixos” usando o assistente de gráfico. Após gerar o gráfico acionar via botão direito do mouse a opção formatar sequência de dados e definir: padrões / linha / nenhuma. Esse gráfico mostra a posição de cada valor gerado ao longo do eixo Y.
- **2) Gráfico da convergência ou estabilização da média** - A média dos números gerados deve tender à média teórica que vale 0,5. Para observar a convergência da média deve-se calcular a média dos primeiros 1, 2, 3, ... números gerados e em seguida traçar o gráfico destas médias ao longo do tempo. Tipicamente o gráfico mostra que a média oscila significativamente nos momentos iniciais e tende a se estabilizar em torno do valor 0,5 com oscilações cada vez mais suaves.
- **3) Tabela e gráfico de freqüências** - Para observar que os números se distribuem uniformemente no intervalo $[0, 1)$ é necessário dividir esse intervalo em faixas de mesmo tamanho e contar quantos números caem em cada faixa. O total de números que caem em cada faixa deverá ser aproximadamente o mesmo. A contagem é feita construindo uma tabela de freqüências com o uso da função FREQUÊNCIA. Em seguida constrói-se um gráfico de barras para visualizar os resultados da tabela de freqüências. Este gráfico mostra que as barras oscilam em torno de um valor médio que é igual ao número de pontos dividido pelo número de faixas considerado. O número de faixas deve ser coerente com o objetivo da tabela/gráfico.

- **4) Gerar outros conjuntos de números aleatórios** - O objetivo é verificar como se definem outros intervalos de variação e de que forma se geram números aleatórios inteiros, sempre com distribuição uniforme. Construir uma tabela com 5 colunas conforme indicado:

- coluna 1: números reais aleatórios entre 0,00 e 1,00
- coluna 2: números reais aleatórios entre 0,00 e 30,00
- coluna 3: números reais aleatórios entre 10,00 e 40,00
- coluna 4: números inteiros aleatórios entre 1 e 6 (incluir os extremos)
- coluna 5: números inteiros aleatórios entre 10 e 40 (incluir os extremos)

Cada coluna deve conter 1000 valores. Nas seis linhas finais de cada coluna computar:

- 1) os valores mínimo e máximo observados da coluna;
- 2) a média observada e a média teórica;
- 3) o desvio padrão observado e o desvio padrão teórico (exceto para as colunas 4 e 5 de números inteiros).

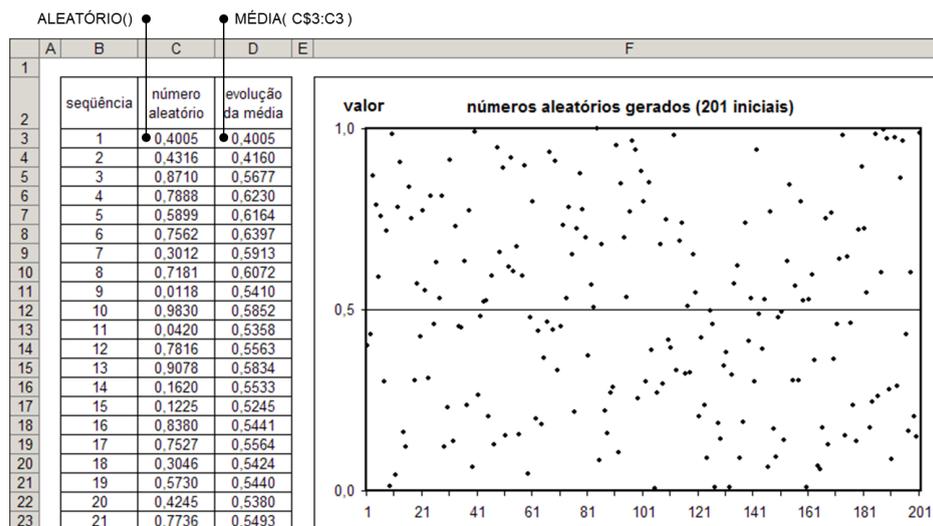
Lembrar que o desvio padrão de uma distribuição contínua uniforme entre a e b vale $(b - a) / \text{raiz}(12)$.

1.2 Solução

Apresenta-se a seguir a solução para esse exercício proposto.

1) Gráfico da posição vertical dos números gerados

A Figura 4 apresenta o gráfico com os números aleatórios gerados. O gráfico sugere que qualquer valor pode ser gerado (portanto a distribuição é contínua) e não se identifica nenhum padrão na distribuição dos pontos (portanto os números são aleatórios).



As colunas C e D foram formatadas para mostrar os valores arredondados para 4 dígitos significativos. Internamente a precisão é de 15 dígitos.

Figura 4 - Dados e Gráfico com números aleatórios gerados no Excel

2) Gráfico da convergência ou estabilização da média

Conforme verificado na Figura 5, como era de se esperar a média dos números aleatórios gerados tende a 0,5. Na medida que aumenta a quantidade de números a média torna-se cada vez mais próxima de 0,5. Na realidade a probabilidade de ocorrer um desvio significativo da média em relação a 0,5 tende a zero, mas sempre continua a existir uma chance mínima do desvio se tornar expressivo.

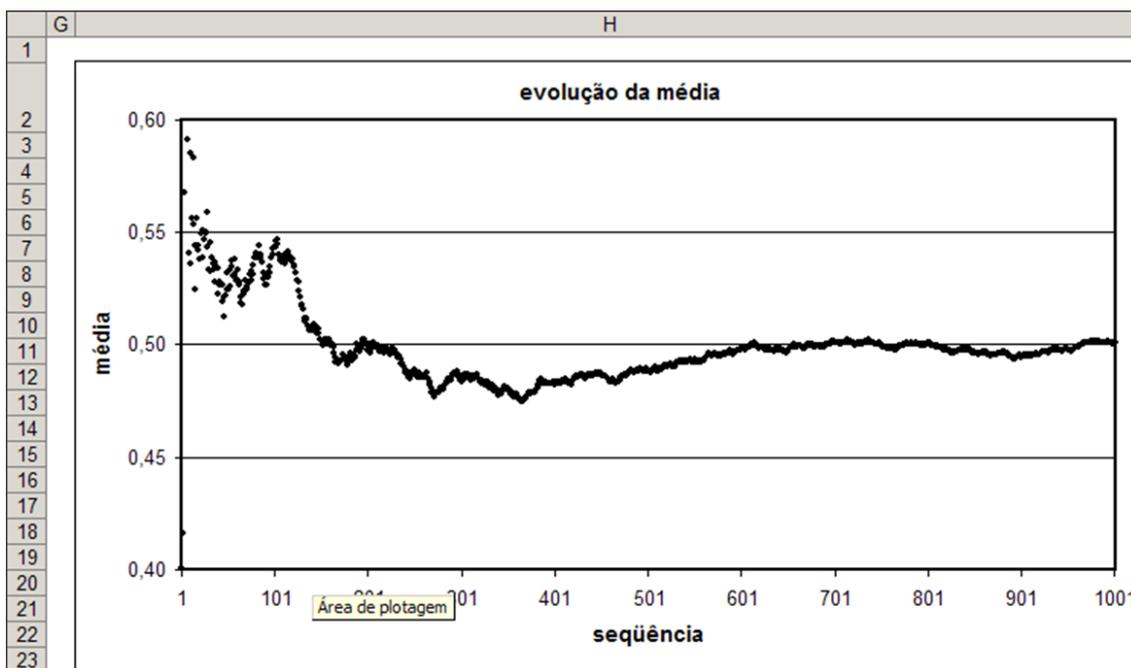


Figura 5 - Gráfico de tendência da média dos números aleatórios entre [0,1) gerados.

3) Tabela e gráfico de frequências

A Figura 6 apresenta o intervalo de frequência, considerando o tamanho de 0,05. Observe que o início da tabela de frequência é em zero; porém, a quantidade de números entre 0 e 0,05 é apresentada na linha do 0,05, que é o valor de 47. O gráfico mostra que os números gerados se distribuem uniformemente no intervalo [0, 1). Acionamentos sucessivos da tecla F9 não permitem identificar faixas privilegiadas com mais ou menos números.

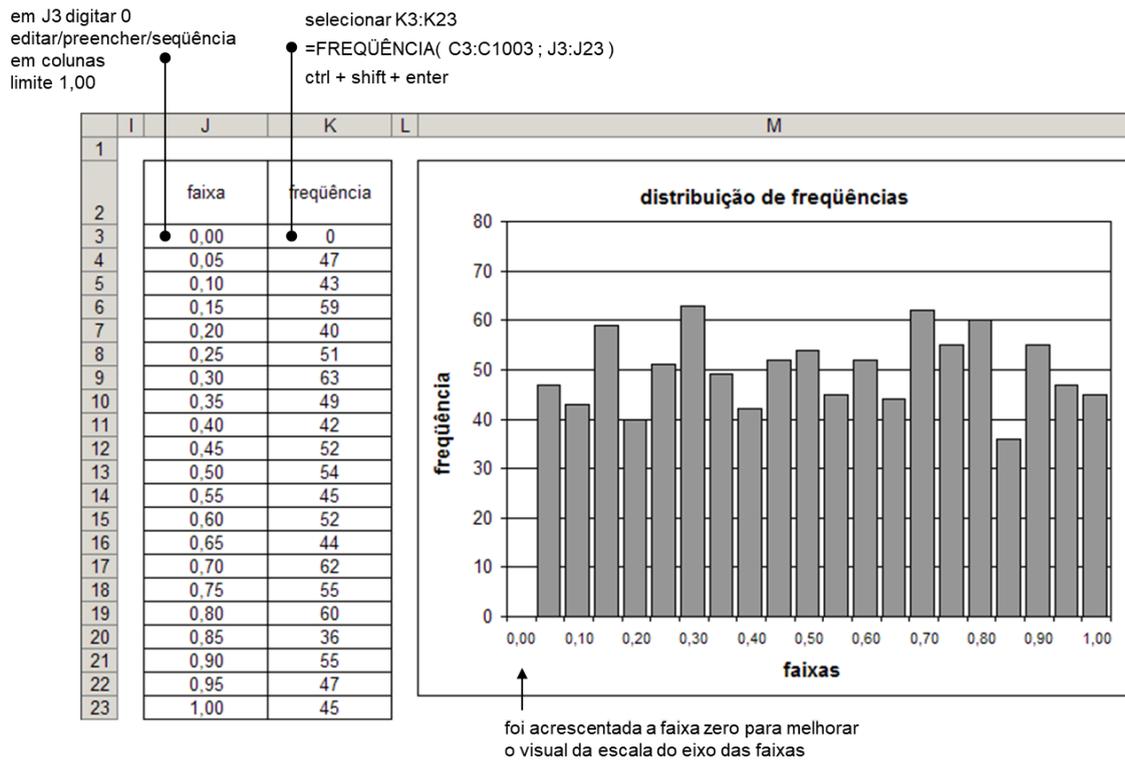


Figura 6 - Intervalos de freqüências e gráfico dos intervalos desenvolvidos no Excel

4) Gerar outros conjuntos de números aleatórios

A Figura 7 apresenta os exemplos de geração de números aleatórios conforme apontado no item 4 do enunciado do Exercício 1, ou seja:

- coluna 1: números reais aleatórios entre 0,00 e 1,00
- coluna 2: números reais aleatórios entre 0,00 e 30,00
- coluna 3: números reais aleatórios entre 10,00 e 40,00
- coluna 4: números inteiros aleatórios entre 1 e 6 (incluir os extremos)
- coluna 5: números inteiros aleatórios entre 10 e 40 (incluir os extremos)

As células, com fundo cinza, são células de entrada de dados. A partir destas células constrói-se o restante da planilha. Deve-se poder alterar o valor destas células sem necessidade de alterar qualquer fórmula da planilha.

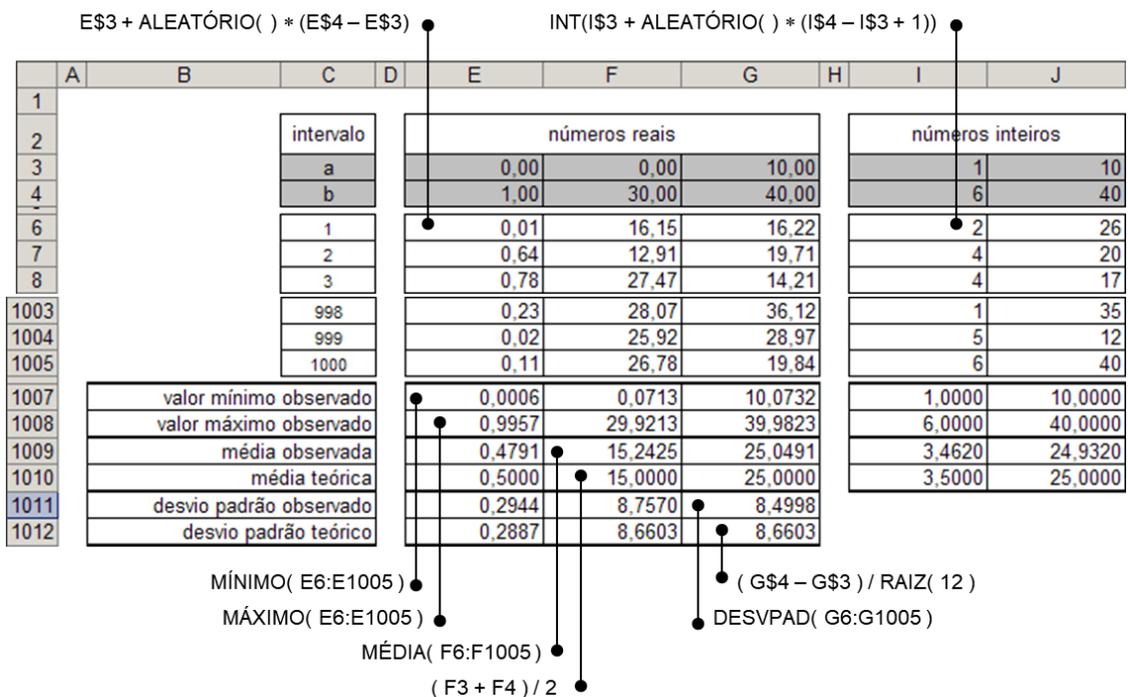


Figura 7 - Exemplo de geração de números aleatórios como o item 4, do Exercício 1.

2. Distribuição Discreta não-uniforme

Em simulação pode ser necessária a geração de valores específicos aleatórios que se distribuem de acordo com probabilidades pré-determinadas. Por exemplo, gerar 1500 números aleatórios dentre os indicados na tabela de forma que os números gerados se distribuam de acordo com as probabilidades também indicadas na Tabela 1, a seguir:

valor	probabilidade
11	0,06
12	0,14
13	0,20
14	0,50
15	0,10

Tabela 1 - Exemplo de distribuição de probabilidade não uniforme de variável discreta.

Neste caso estamos diante de uma distribuição discreta de probabilidades pois podem ocorrer apenas valores dentre um conjunto finito de valores. Cada valor possui uma probabilidade específica de ocorrer.

Distribuições discretas muitas vezes podem ser geradas utilizando uma distribuição contínua e uniforme de probabilidades, mas isso pode ser um tanto complexa e trabalhosa. Em EXCEL recomenda-se que distribuições deste tipo sejam geradas através do uso da função PROCV associada com a função ALEATÓRIO. Esta forma é preferível pois torna as planilhas de simulação muito mais flexíveis apesar de exigir um esforço inicial de entendimento maior.

No gráfico da Figura 8, a seguir, está representada uma função escada que poderia servir para gerar os números desejados. É importante observar que esta função “acumula” as probabilidades (dadas na tabela) ao longo do eixo x de forma a associar a probabilidade do número com um intervalo correspondente nesse eixo. Dessa forma a probabilidade 0,5 (da quarta linha da tabela) corresponde ao intervalo que vai de 0,4 a 0,9 no eixo x.

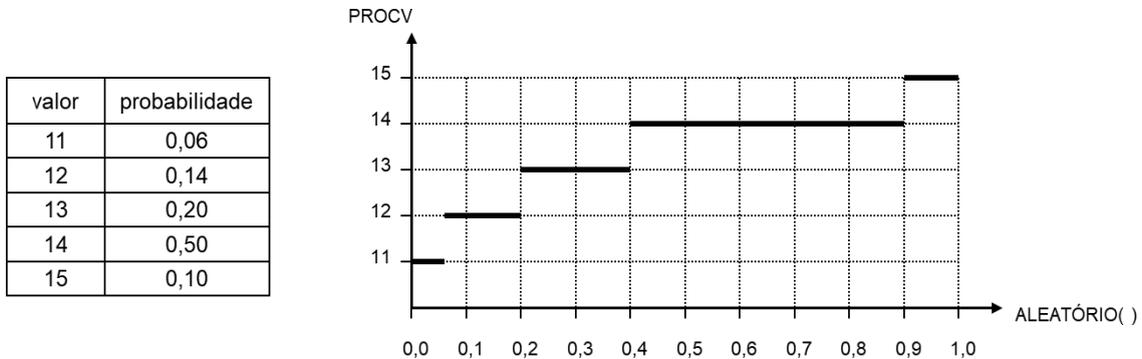


Figura 8 - Distribuição de probabilidade e gráfico com representação

Com este gráfico podemos gerar os valores com a distribuição de probabilidades definida na tabela. Conforme solicitado, basta gerar 1500 abcissas aleatórias com distribuição contínua uniforme no intervalo [0, 1) através da função ALEATÓRIO e por meio do gráfico determinar as suas correspondentes ordenadas. Por exemplo, se foi gerada a abcissa 0,528978839355029 então a ordenada retornada será 14.

É exatamente isto que a função PROCV realiza quando usada para modelar problemas de simulação. Através dela implementamos a função escada apresentada e convertemos abcissas em ordenadas de modo a preservar a distribuição discreta de probabilidades originalmente especificada para estas ordenadas.

Um dos argumentos da função PROCV é a tabela correspondente ao gráfico em escada. Esta tabela possui normalmente duas colunas:

- **Coluna 1:** deve conter obrigatoriamente o início de cada intervalo de valores no eixo das abcissas (em ordem crescente);
- **Coluna 2:** deve conter os correspondentes valores das ordenadas, isto é, os valores que devem ser gerados.

Como pode ser observado na Figura 9, a tabela PROCV possui formato específico o que normalmente exige uma conversão da tabela original de probabilidades. Sugere-se manter nas planilhas a tabela original (ela é mais fácil de entender e facilita a entrada de dados) e através de fórmulas do EXCEL construir a tabela PROCV correspondente.

tabela PROCV	
início do intervalo	valor
0,00	11
0,06	12
0,20	13
0,40	14
0,90	15

Figura 9 - Construção da Tabela PROCV

A forma final de uso da função PROCV envolve 3 argumentos. Existe um quarto, opcional, que não tem interesse no contexto em discussão. Os 3 argumentos obrigatórios são os seguintes:

- **Argumento 1:** valor da abscissa a partir da qual a função PROCV irá determinar a ordenada correspondente. Em geral é um número aleatório com distribuição uniforme no intervalo [0, 1) ou a própria chamada da função ALEATÓRIO.
- **Argumento 2:** intervalo de células onde se localiza a tabela PROCV (não esquecer de usar o \$ para tornar absoluta a referência ao intervalo).
- **Argumento 3:** número da coluna relativa à tabela PROCV onde constam os valores das ordenadas. Em geral esse argumento vale 2 pois as ordenadas estão na coluna 2 da tabela PROCV (não confundir coluna da tabela PROCV com coluna de planilha).

A Figura 10 resume essas informações.

<p>PROCV (abscissa aleatória ; tabela PROCV ; coluna das ordenadas)</p> <p>por exemplo:</p> <p>PROCV (ALEATÓRIO() ; E\$4:F\$11 ; 2)</p> <p>PROCV (0,528978839355029 ; E\$4:F\$11 ; 2)</p> <p>PROCV (H5 ; \$E\$4:\$F\$11 ; 2)</p>

Figura 10 - Função PROCV e seus argumentos

Observar que os valores iniciais dos intervalos precisam constar obrigatoriamente em ordem crescente na coluna 1 da tabela PROCV. Como o primeiro intervalo começa em zero fica evidente que o primeiro valor desta coluna será sempre zero. Observar também que não consta da tabela o valor 1 que fecha o último intervalo.

2.1. Funcionamento da Função PROCV

A função PROCV é de uso muito mais amplo do que aquele aqui discutido. Internamente o que ela realiza é uma busca de um valor (argumento 1) ao longo da coluna 1 de uma tabela

(argumento 2) retornando o valor correspondente na coluna indicada (argumento 3). Disso advém o seu nome PROCURA VERTICAL em tabelas.

Notar que a procura é realizada por definição ao longo da coluna 1 da tabela. Nessa procura ela não tenta localizar um valor igual pois se assim fosse ela não conseguiria funcionar como uma função escada. Efetivamente a procura termina ao localizar o valor na tabela **mais próximo que seja menor ou igual ao valor procurado**. Localizado este valor então ela retorna o valor correspondente da coluna resultado na mesma linha. A Figura 11 apresenta alguns exemplos

tabela PROCV		exemplos de procura		
início do intervalo	valor	valor procurado	mais próximo menor ou igual	valor retornado
0,00	11	0,39	0,20	13
0,06	12	0,04	0,00	11
0,20	13	0,41	0,40	14
0,40	14	0,40	0,40	14
0,90	15	0,06	0,06	12
		0,95	0,90	15
		0,00	0,00	11
		11,80	0,90	15
		-0,05	-	erro (#N/D)

Figura 11 - Exemplo de construção da Tabela PROCV e uso dos valores gerados

2.2 Exercício

Gerar os 1500 números aleatórios de acordo com a distribuição especificada a seguir. Em seguida contar quantos valores foram gerados para comprovar o funcionamento do procedimento de geração via PROCV.

valor	probabilidade
11	0,06
12	0,14
13	0,20
14	0,50
15	0,10

Tabela 1 - Tabela com os valores de probabilidade, já apresentada anteriormente

2.3 Solução

A Figura 12 apresenta o exemplo construído para gerar os números aleatórios, conforme solicitado. A Figura 13 apresenta as fórmulas.

Observar que as únicas informações dadas são os valores indicados nas células de fundo cinza da tabela de probabilidades. Todo o resto da planilha deve ser gerado através de fórmulas. Além disso, a planilha deve ser flexível de modo que se possa alterar a tabela de probabilidades, incluindo ou excluindo valores, sem necessidade de qualquer alteração nas fórmulas.

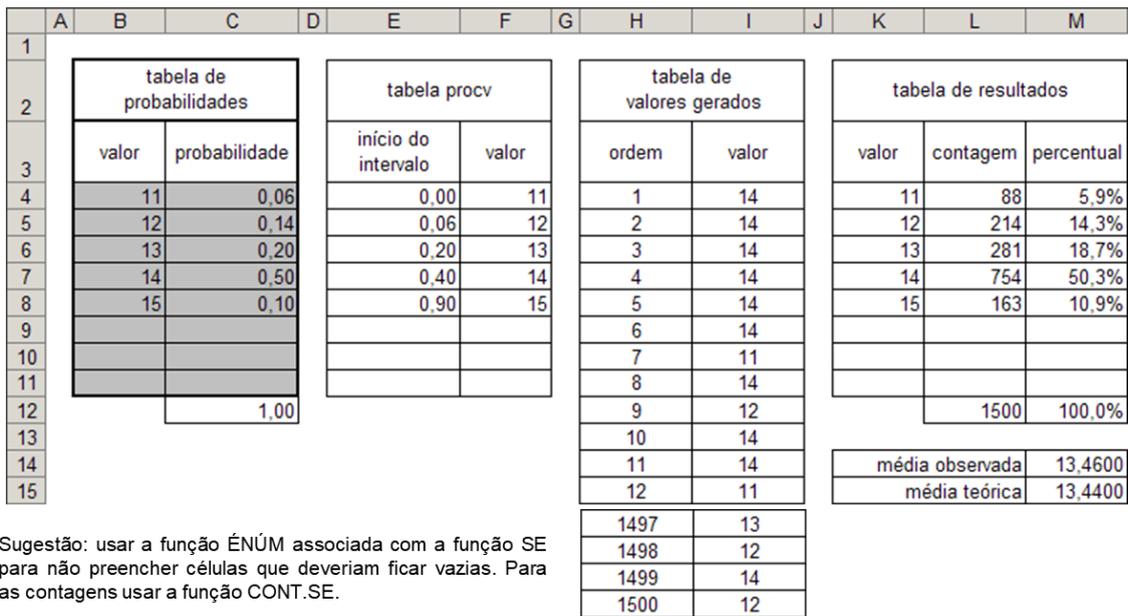


Figura 12 - Exemplo de organização dos dados na planilha e utilização da distribuição PROCV

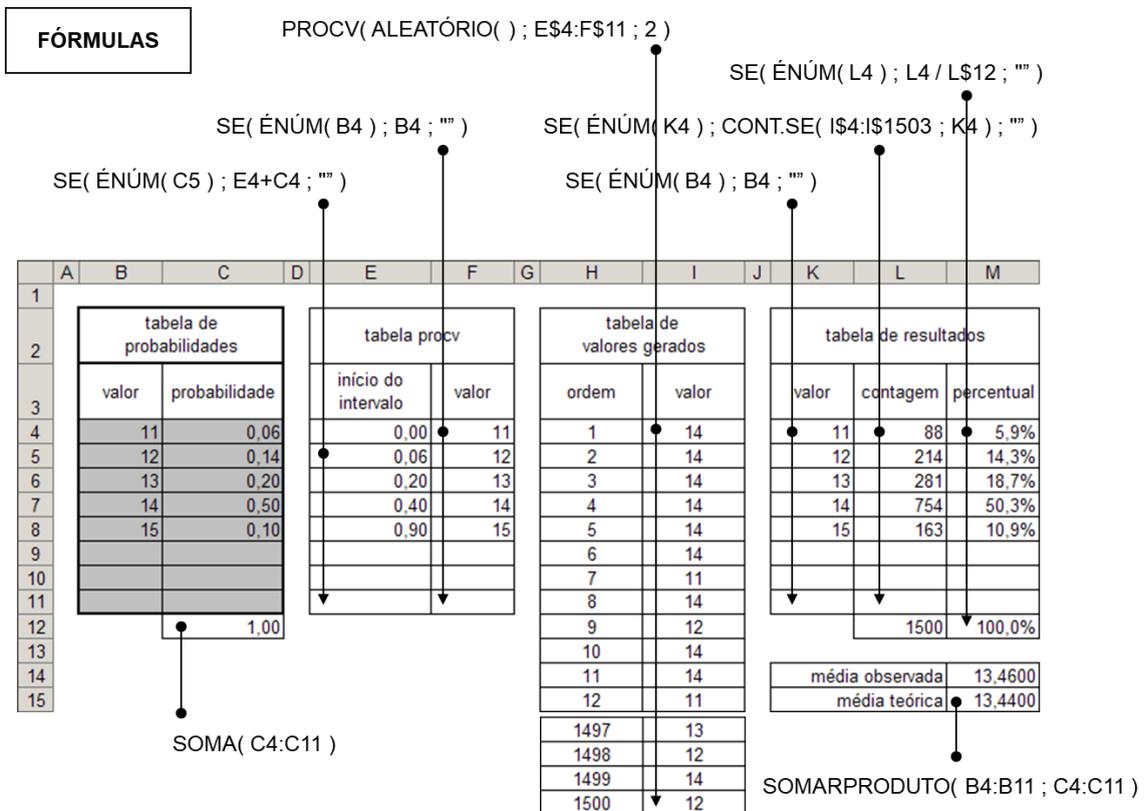


Figura 13 - Fórmulas que foram utilizadas para organização dos dados na planilha e utilização da distribuição PROCV.