Ilustração do Teorema do Limite Central através de simulação de Monte Carlo

Adriano Nascimento da Paixão, 20180015150
Carlos Alberto Alves de Meneses, 201800032
Joana D'arc Nunes da Silva, 20180078535

2022-11-20

1. Objetivos

- Demonstrar o Teorema do Limite Central por meio de metódos computacionais;
- Gerar números pseudo-aleatórios para amostras de diferentes tamanhos;
- Utilizar a simulação de Monte Carlo para demostrar a convergência em distribuição.

2. Objetivo principal

- Iliustrar o Teorema do Limite Central através de simulação de Monte Carlo, considerando diferentes tamanhos de amostras e diferentes distribuições de probabilidade.

3. Introdução

A ideia de convergência em distribuição ou do que chamamos a distribuição limte é de fundamental importância em Estatística. A motivação para tal estudo é a de que de que muitas vezes a distribuição de uma determinada variável aleatória de interesse pode ser dificíl de ser encontrada ou de se trabalhar. Porém, supondo que a informação de que dispomos é suficientemente grande, a distribuição de interesse pode muitas vezes ser aproximada por uma distribuição limite conhecida; por exemplo, uma distribuição normal-padrão. Essa informação disponível é na maioria das vezes traduzida pelo número total de observações (amostras) sobre o experimento. Uma fração significativa dos resultados é consequência do chamado *Teorema Central do Limite*. O teorema central do limite pode ser visto como uma informação adicional àquela que é dada pela lei dos grandes números, na seguinte forma. Para uma sequência (X_n) de variáveis aleatórias i.i.d. e integráveis de esperança μ, a lei forte dos grandes números diz que a média amostral $\bar{X_n}$ converge quase certamente para μ. Sendo assim, a distribuição de $\bar{X_n}$ deve convergir para uma distribuição concentrada em μ. O teorema central do limite traz uma informação adicional. Ele afirma que, se a variância σ^2 da distribuição comum for finita, a convergência dá-se de tal forma que a distribuição de $\sqrt{n}(\bar{X_n} - \mu)$ pode ser cada vez mais bem aproximada pela distribuição $N(0, \sigma^2)$. Equivalentemente, a distribuição de $\bar{X_n}$ torna-se cada vez mais próxima da distribuição $N(\mu, \frac{\sigma^2}{n})$. Dessa forma, o teorema central do limite dá uma descrição sobre como a distribuição de $\bar{X_n}$ converge para a distribuição concentrada em μ,(VASCONCELLOS, 2021).

3. Resultados Fundamentais

Definição 3.1. Seja (X_n) uma sequência de variáveis aleatórias reais e seja F_n a sua função de distribuição de (X_n), para cada n. Suponha que exista função de distribuição F tal que $F_n(x) -> F_(x)$ para todo x $\in \Re$ que é ponto de continuidade de F. Neste caso, dizemos que a seqência (X_n) converge em distribuição para F.

Notação: "$X_n \to^D F$"

Teorema 3.1. [*Teorema Central do Limite clássico*] Seja (X_n) uma sequência de variáveis aleatórias i.i.d. com $E[X_1^2] < \infty$. Suponha $\sigma^2 = Var(X_1) \neq 0$ e seja $\mu = E(X_1)$. Seja $S_n = X_1 + X_2 + ... + X_n$. Tem-se

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \to^D N(0,1)$$

.

Considere a média amostral $\bar{X_n} = \frac{S_n}{n}$. Observe que uma forma equivalente de se enunciar o Teorema 3.1 é

$$\sqrt{n}(\bar{X}_n - \mu) \to^D N(0, \sigma^2)$$

, que é muito usada na Inferência Estatística, pois descreve diretamente o comportamento da média amostral como estimador de μ. Em outras palavras, podemos trabalhar diretamente com a distribuição normal ao invés de trabalharmos com uma distribuição mais complicada ao termos conhecimento que a mesma converge para uma $N(\mu, \frac{\sigma^2}{n})$.

Exemplo 3.1. [*Aproximação da binomial pela normal*] Seja (x_n) uma seqência de variáveis aleatórias i.i.d. com distribuição comum Bernoulli de parâmetro p. Nesse caso, sabemos que $S_n = X_1 + ... + X_n$ tem distribuição binomial com parâmetros n e p. Ainda pelo teorema 3.1, $\frac{(S_n - np)}{\sqrt{(np(1-p))}} \to^D N(0,1)$. De outra forma, isso significa que, para valores grandes de n, a distribuiçÃo $N(np, np(1-p))$ fornece uma boa aproximação para a distribuição *Binomial(n,p)*. Suponha, por exemplo, que queremos a probabilidade de que, em n = 400 lançamentos independentes de uma moeda honesta, o número total de caras não ultrapasse 220. O valor procurado é

$$\frac{1}{2^{440}} \sum_{k=0}^{220} \binom{400}{k}$$

,

cuja obtenção númerica é bastante trabalhosa. O resultado arredondado para três casas decimais é 0,980. Uma aproximação pode ser obtida facilmente usando aqui a distribuição N(200,100). Nesse caso calculamos $\phi(220 - 200/10) = \phi(2)$, onde ϕ é a função de distribuição para a distribuição N(0,1). Essa última aproximação arredondada para três casas decimais é 0,977, que representa uma redução de apenas 0,3% do valor verdadeiro.

```
## [1] 0.9772499
```

4. Simulações

4.1. Apresentando Aleatoriedade

Uma parte crítica da modelagem de simulações é o uso de processos aleatórios. Um *processo aleatório* é aquele que gera um resultado diferente de acordo com algumas regras cada vez que é executado. Nesse trabalho, iremos gerar números pseudo-aleatórios para gerarem amostras que serão utilizadas nas simulações para demonstrar o Teorema Central do Limite (TCL), usando o método de Monte Carlo.

4.2 Reproduzindo Aleatoriedade

Como queremos obter os mesmos números aleatórios exatos, iremos inicialmente definir uma *semente aleatória*.

4.3. Replicação

Para usar os métodos de Monte Carlo, precisaremos replicar algum processo aleatório muitas vezes. Existem duas maneiras de fazer isso: com *replicate ()* ou com *for ()*, loops. Nesse trabalho utilizaremos loop‘s. A simulação de Monte Carlo inicia-se com a geração de números pseudo-aleatórios. Números pseudo-aleatórios constituem uma sequência de valores que, apesar de serem gerados de maneira determinística, possuem características semelhantes a variáveis aleatórias independentes U(0,1).

Geradores de números aleatórios são usados em simulação estocástica, amostragem, análise numérica, teste de programas, jogos de azar, tomada de decisões, etc.

A distribuição U(0,1) fornece a representação probalística básica da aleatoriedade em um computador e os geradores de todas as outras distribuições requerem que uma sequência de variáveis aleatórias U(O,1) seja simulada.

4.3.1 Exemplo: Geração de números pseudo-aleatórios para uma distribuição exponencial.

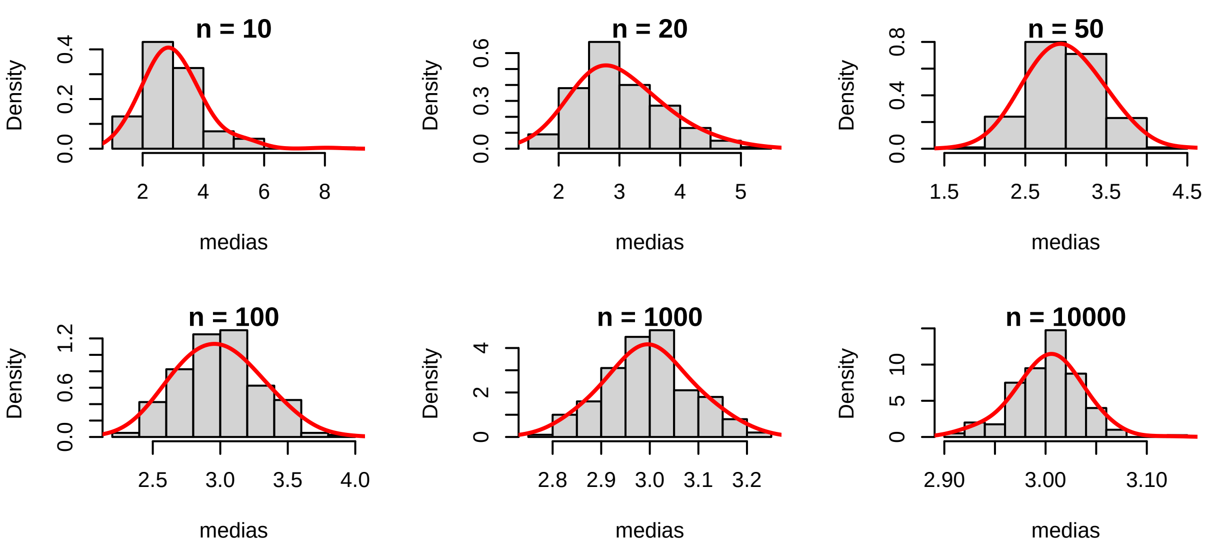
```
#Programa R demonstrando a geração de
#n realizações de variáveis aleatórias
#exponenciais com parâmetro lamb,
#utilizamos a função *runif()*
#para gerarmos números aleatórios
#uniformes

rexpon <- function(n, lamb=1/3)
{
  u <- runif(n, 0, 1) # gera vetor u
  #(U(0,1))
  x <- -log(u) / lamb # gera vetor x
  #com distrib. exp.
  return(x) # retorna o vetor x
}

# exemplo

rexpon(20, 1/3)
```

```
##  [1] 0.32711113 3.97832343 2.96558525 1.67137652 0.28884625 4.80319026
##  [7] 0.32145407 0.17074212 1.24292219 1.39032819 8.35222230 4.74000762
## [13] 5.20233873 1.12616319 2.87052799 0.78471220 2.09327795 0.99545153
## [19] 0.02438052 2.90247436
```

4.3.2 TCL:Exemplo utilizando a distribuição exponencial.

A densidade de uma exponencial com parâmetros λ é dada pela expresão: $f_(x) = \lambda\ e^{-\lambda x}$, x$\geq 0$.

Gerando dados por simulação a partir de uma exponencial com $\lambda = \frac{1}{3}$, para cada um dos seguintes tamanhos n de amostras: 10,20,50,100,1000 e 10000.



Resultados

- Obtivemos 200 valores da média amostral;

- Utilizamos esses 200 valores para construir um histograma;

Os oitos histogramas nos mostram que, à medida que o tamanho n da amostra cresce, a forma do histograma se aproxima cada vez mais de uma curva normal, comprovando o *Teorema Central do Limite.*

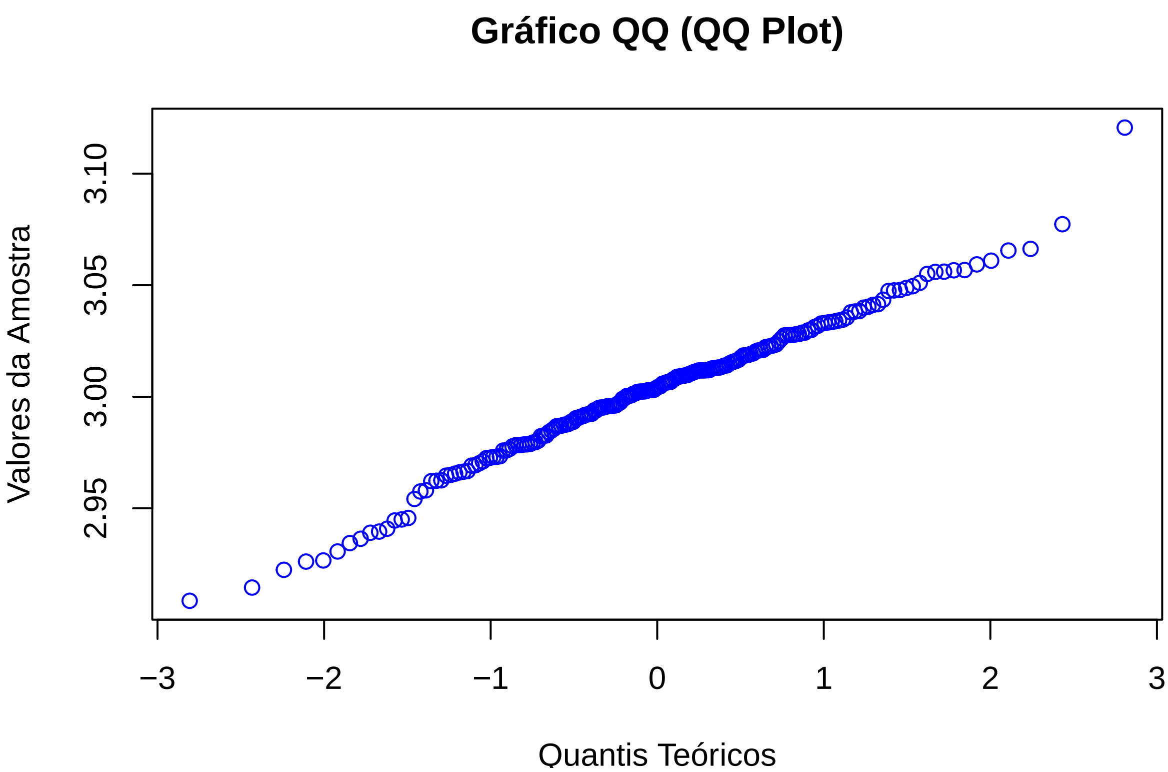
5. Análise Exploratória

5.1 Teste da Normalidade

Utilizaremos o teste de Shapiro-Wilk para verificarmos a normalidade dos dados.

```
##
##  Shapiro-Wilk normality test
##
## data:  medias
## W = 0.98723, p-value = 0.06896
```

5.2 Plotando o Gráfico



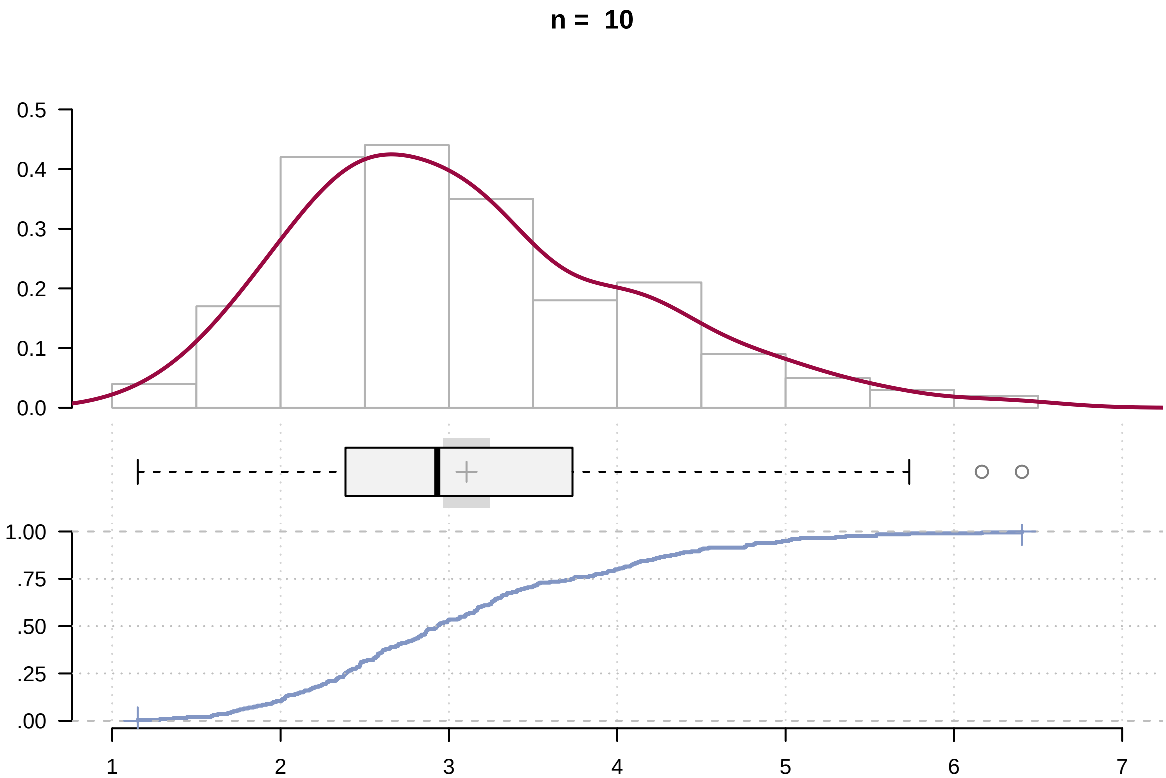
Análise: p-value = 0,2379, indica uma probabilidade de que essa variável veio de uma distribuição normal o que é comprovada pela análise gráfica.

5.3 Estatística descritiva com o pacote *DescTools*

O pacote *DescTools* foi desenvolvido com o objetivo de fornecer uma análise descritiva de forma muito rápida e completa, permitindo ao analista ganho de tempo nessa tarefa. A principal função no pacote é *Desc* que foi concebida para descrever as variáveis de acordo com sua natureza, produzindo medidas descritivas e uma representação gráfica adequada.

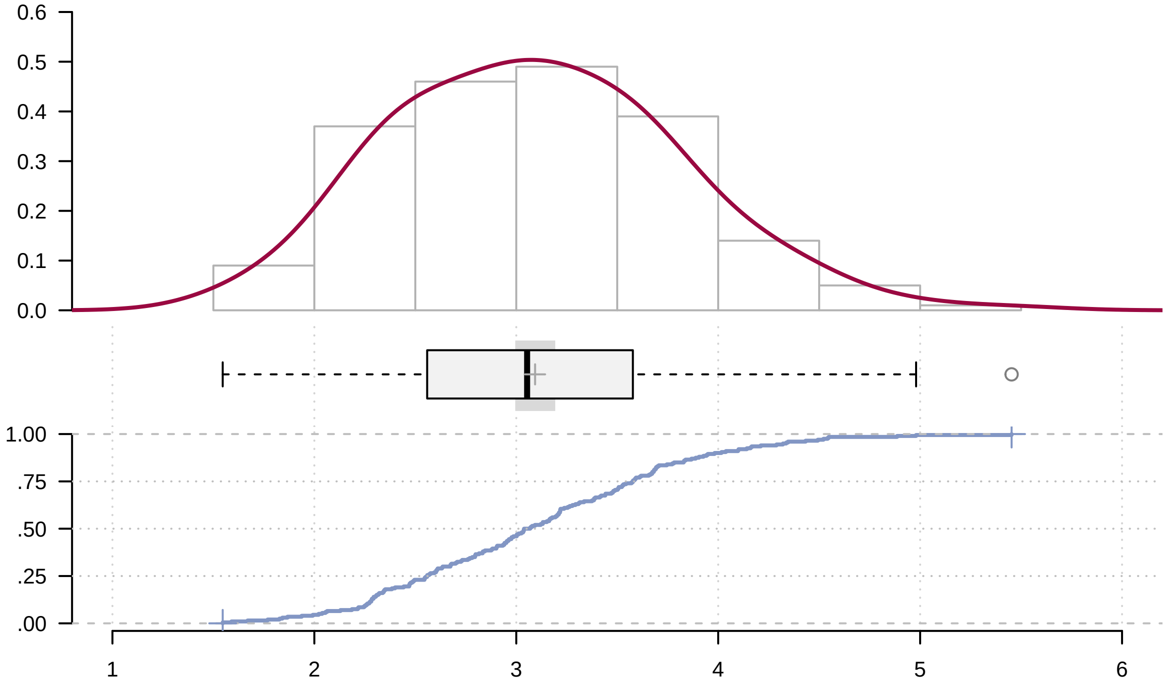
```
## ------------------------------------------------------------------------------
## n =  10
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        200       200         0       = n         0  3.104841   2.965081
##               100.0%      0.0%                0.0%             3.244602
```

```
##
##        .05       .10       .25    median       .75       .90       .95
##   1.739444  1.973759  2.387068  2.931093  3.729969  4.488185  4.981897
##
##      range        sd     vcoef       mad       IQR      skew      kurt
##   5.252649  1.002312  0.322822  0.881572  1.342901  0.709943  0.221763
##
## lowest : 1.151241, 1.283738, 1.364432, 1.444553, 1.586859
## highest: 5.538531, 5.539843, 5.735027, 6.165951, 6.403891
##
## ' 95%-CI (classic)
```

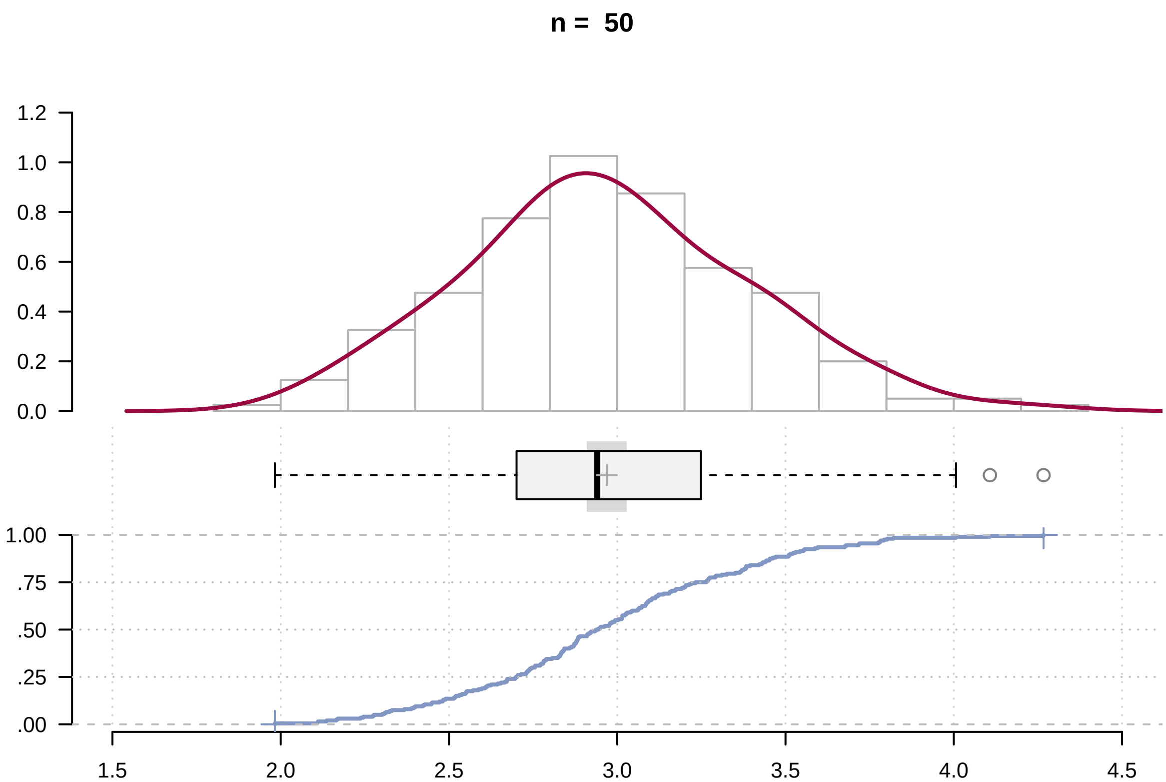



```
## ------------------------------------------------------------------------------
## n =  20
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        200       200         0       = n         0  3.093052   2.994691
##               100.0%      0.0%                 0.0%            3.191413
##
##        .05       .10       .25    median       .75       .90        .95
##   2.037961  2.262536  2.558814  3.053585  3.577027  3.985371   4.321258
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   3.906783  0.705410  0.228063  0.749570  1.018214  0.327175  -0.049957
##
## lowest : 1.546148, 1.589663, 1.669446, 1.768528, 1.827099
## highest: 4.542226, 4.547678, 4.885943, 4.979931, 5.452931
##
## ' 95%-CI (classic)
```

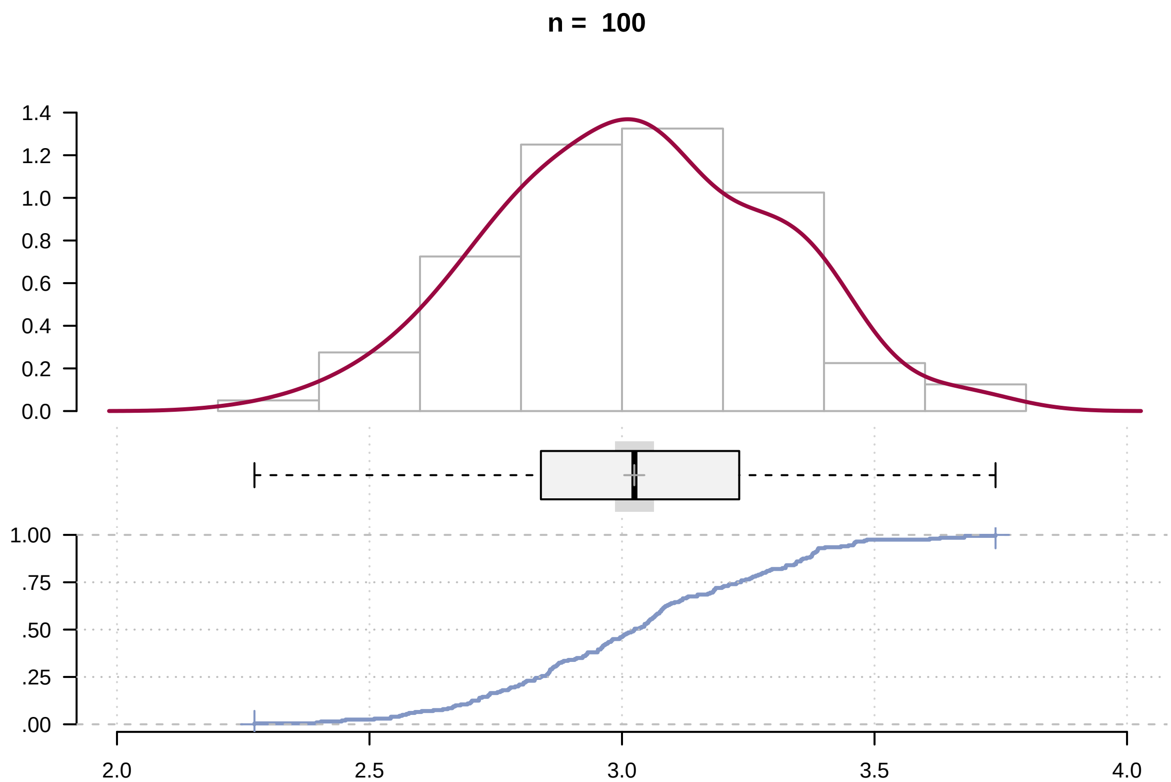
n = 20



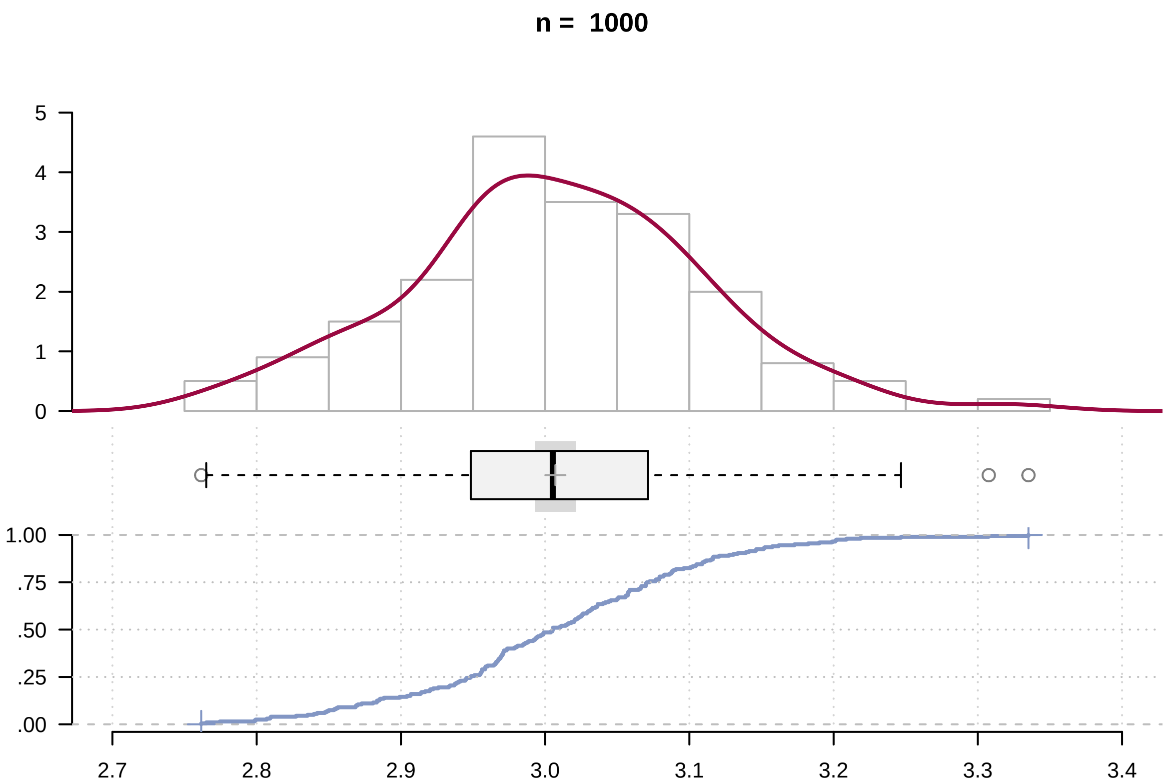

```
## ------------------------------------------------------------------------------
## n =  50
##
##     length          n        NAs     unique         0s       mean     meanCI'
##        200        200          0        = n          0   2.968914   2.910151
##                100.0%       0.0%                  0.0%              3.027678
##
##        .05        .10        .25     median        .75        .90        .95
##   2.300790   2.426785   2.701796   2.940700   3.240780   3.515142   3.717048
##
##      range         sd      vcoef        mad        IQR       skew       kurt
##   2.284106   0.421433   0.141949   0.397262   0.538984   0.242500  -0.094424
##
## lowest : 1.982594, 2.109037, 2.110487, 2.136096, 2.164974
## highest: 3.802139, 3.821412, 4.006606, 4.107342, 4.2667
##
## ' 95%-CI (classic)
```



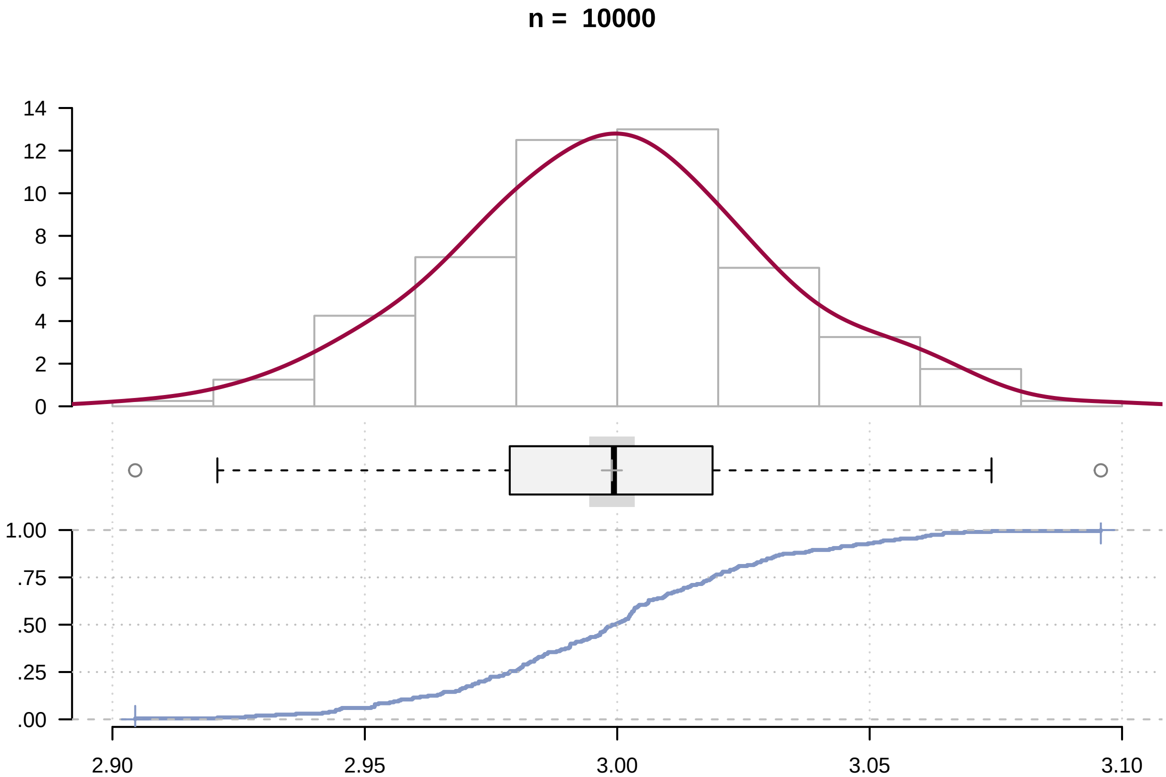

```
## ------------------------------------------------------------------------------
## n =  100
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        200       200         0       = n         0  3.024341   2.985948
##               100.0%      0.0%                0.0%             3.062734
##
##        .05       .10       .25    median       .75       .90        .95
##   2.572692  2.679513  2.840114  3.024602  3.230157  3.377249   3.458068
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   1.467421  0.275338  0.091041  0.291609  0.390043  0.003101  -0.302847
##
## lowest : 2.272232, 2.394857, 2.403702, 2.445071, 2.45261
## highest: 3.609173, 3.629923, 3.677729, 3.6784, 3.739654
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  1000
##
##     length          n        NAs     unique         0s       mean    meanCI'
##        200        200          0        = n          0   3.007109   2.992902
##                100.0%       0.0%                  0.0%              3.021317
##
##        .05        .10        .25     median        .75        .90        .95
##   2.839520   2.870014   2.948589   3.005270   3.070973   3.130994   3.173253
##
##      range         sd      vcoef        mad        IQR       skew       kurt
##   0.573530   0.101891   0.033883   0.093158   0.122385   0.109665   0.217103
##
## lowest : 2.761571, 2.76506, 2.774486, 2.798427, 2.799289
## highest: 3.208918, 3.218866, 3.246766, 3.307578, 3.335101
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  10000
##
##     length          n        NAs     unique         0s       mean    meanCI'
##        200        200          0        = n          0   2.998938   2.994461
##                100.0%       0.0%                  0.0%              3.003414
##
##        .05        .10        .25     median        .75        .90        .95
##   2.945007   2.957105   2.978736   2.999341   3.018811   3.042107   3.055005
##
##      range         sd      vcoef        mad        IQR       skew       kurt
##   0.191285   0.032106   0.010706   0.029401   0.040075   0.035908   0.127774
##
## lowest : 2.904512, 2.920787, 2.92631, 2.92839, 2.932365
## highest: 3.064545, 3.064674, 3.068741, 3.074135, 3.095797
##
## ' 95%-CI (classic)
```



Listamos a seguir as saídas que essa função fornece para o caso de variáveis numéricas

- *length.*, o comprimento do vetor;
- *n.*, o número de observações;
- NAs., o número de observações faltantes;
- *unique.* , o número de observações distintas entre si;
- 0s. , o número total de valores nulos;
- *mean.* , a média aritmética do vetor;
- *meanSE.* , fornece um intervalo de 95% de confiança para a média, com base no erro padrão da média,;
- .05, .., .95 percentil de x, iniciando com 5%, 10%, 1_0. quartil, mediana, etc.;
- *range.* , a amplitude do vetor x;
- *sd.* ,o desvio-padrão do vetor x;
- *vcoef.*, o coeficiente de variação de x, definido como ;
- *mad.*, o desvio médio absoluto;
- *IQR.* , a amplitude interquartil, definida por 3ˆ0. quartil - 1ˆ0. quartil;
- *skew.*, o coeficiente de assimetria do vetor x;
- *kurt.*, o coeficiente de curtose do vetor x;
- *lowest.*, os cinco menores valores do vetor;
- *highest.*, os cinco maiores valores do vetor.

A tabela a seguir expõem as principais métricas estatísticas da distribuição Exponencial geradas:

	Médias	sd	Mediana	Curtose
10	3.104841	1.002312	2.931093	0.2217635
20	3.093052	0.705410	3.053585	-0.0499570
50	2.968914	0.421433	2.940700	-0.0944240
100	3.024341	0.275338	3.024602	-0.3028470
1000	3.007109	0.101891	3.005270	0.2171030
10000	2.998938	0.032106	2.999341	0.1277740

6. TCL:Exemplo utilizando a distribuição Uniforme.

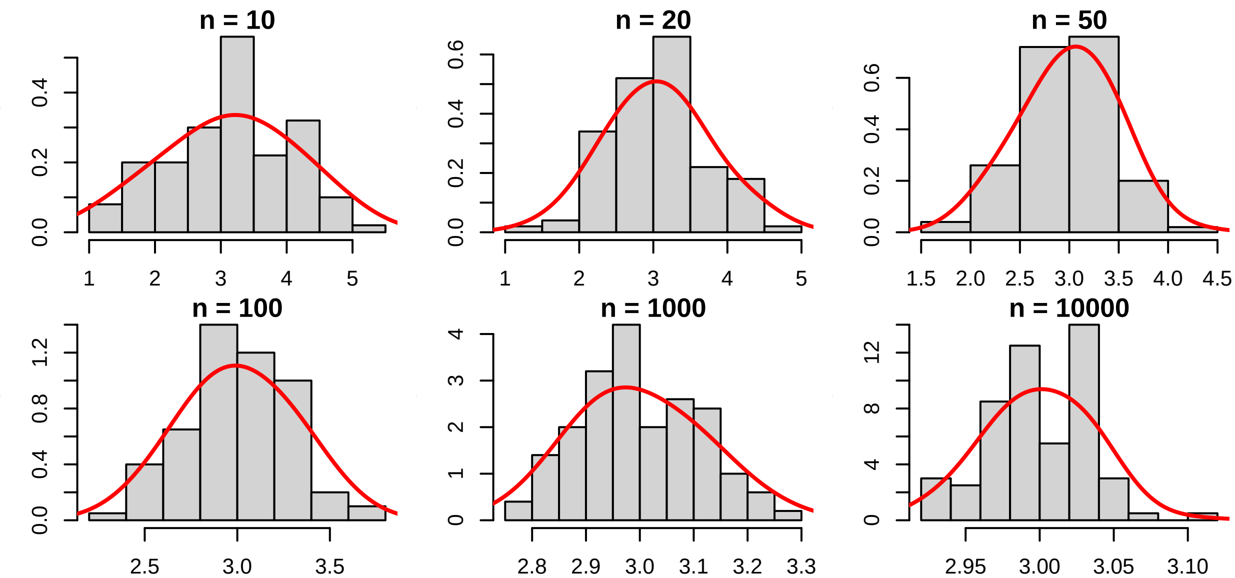
Definição 6.1 Uma variável aleatória X tomando um número finito de valores com iguais probabilidades diz-se que tem uma *distribuição uniforme*.

A **distribuição uniforme** é a mais simples para variáveis aleatórias continuas, sendo utilizada para modelar a ocorrência de eventos cuja probabilidade é constante em intervalos de mesma amplitude. Uma variável aleatória X tem distribuição uniforme no intervalo $[a, b]$, denotada por $X\ U[a, b]$, se sua função densidade de probabildade for dada por:

$$f(x) = \begin{cases} \frac{1}{(b-a)}, \text{ se a} \leq \text{x} \geq \text{b} \end{cases}$$

0, caso contrário.

6.2 TCL: Códigos no R para a elaboração dos gráficos com as simulações - Uniforme



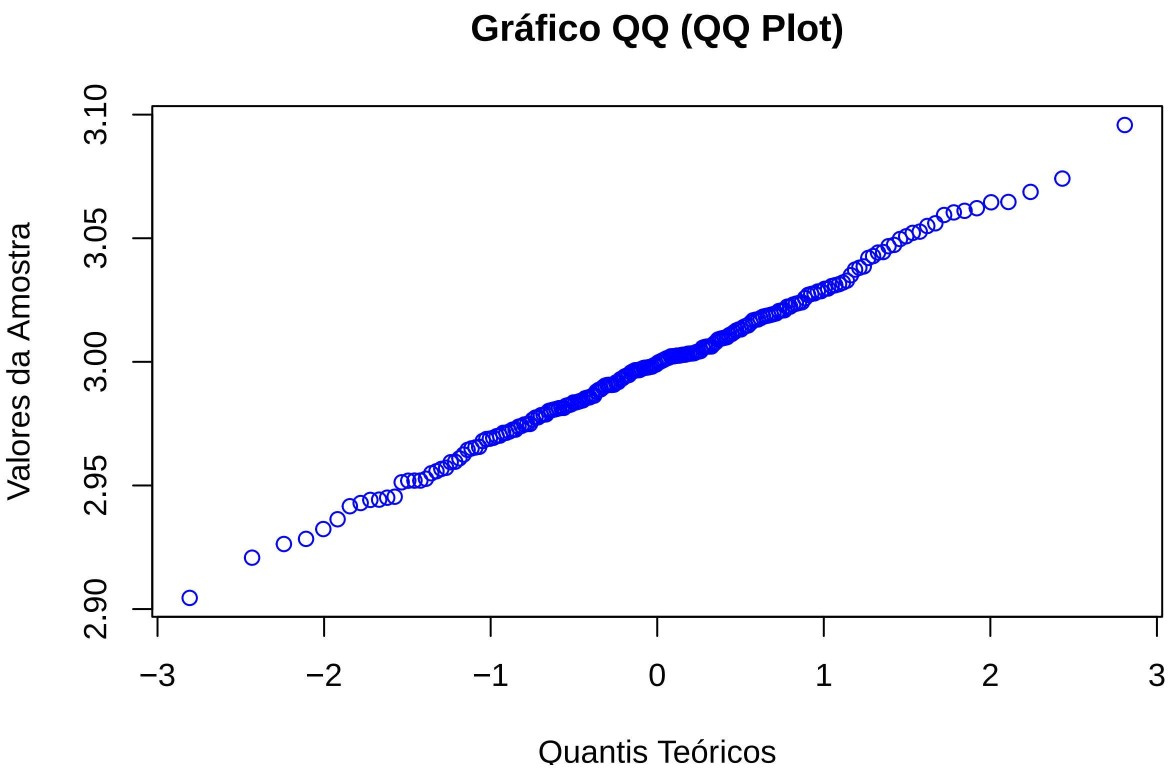
Os seis histogramas nos mostram que, à medida que o tamanho n da amostra cresce, a forma do histograma se aproxima cada vez mais de uma curva normal, comprovando o *Teorema Central do Limite.*

6.3 Teste da Normalidade

Utilizaremos o teste de Shapiro-Wilk para verificarmos a normalidade dos dados.

```
##
##  Shapiro-Wilk normality test
##
## data:  medias.unif
## W = 0.98074, p-value = 0.1518
```

6.4 Plotando o Gráfico

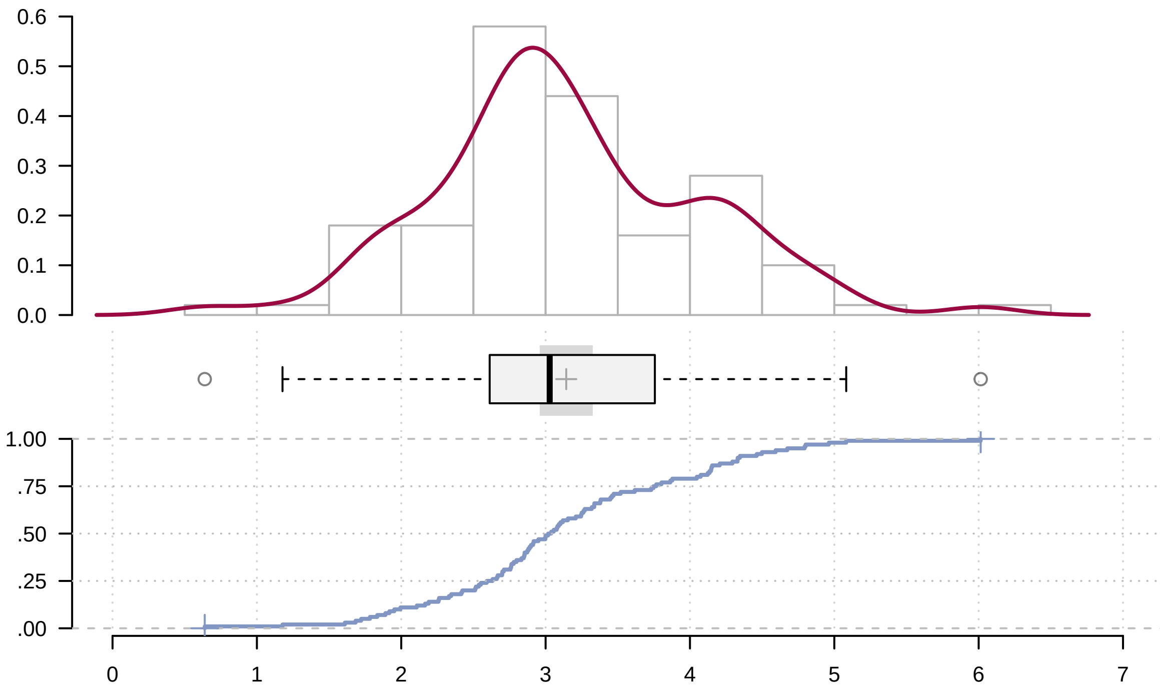


Análise: p-value = 0,06896, indica uma probabilidade de que essa variável veio de uma distribuição normal o que é comprovada pela análise gráfica.

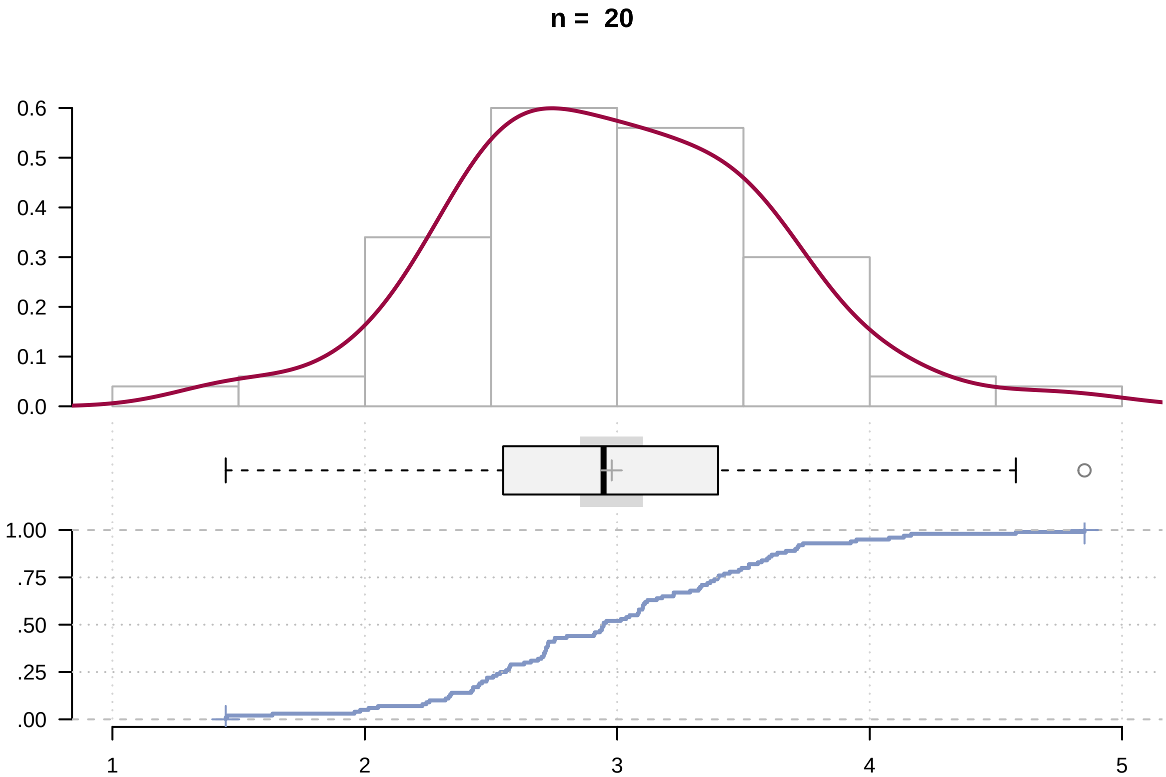
6.5 Estatística descritiva com o pacote *DescTools*

```
## ------------------------------------------------------------------------------
## n =  10
##
##      length         n        NAs     unique         0s       mean     meanCI'
##         100       100          0        = n          0  3.1431018  2.9611720
##                100.0%       0.0%                  0.0%             3.3250317
##
##         .05       .10        .25     median        .75        .90         .95
##   1.7802127 1.9902772  2.6225485  3.0282379  3.7521932  4.3317776  4.6803440
##
##       range        sd      vcoef        mad        IQR       skew        kurt
##   5.3753948 0.9168851  0.2917134  0.7434725  1.1296447  0.2712963  0.2673881
##
## lowest : 0.6387322, 1.1774455, 1.6093421, 1.6838461, 1.7224465
## highest: 4.7940637, 4.8008903, 4.9621393, 5.0824105, 6.0141269
##
## ' 95%-CI (classic)
```

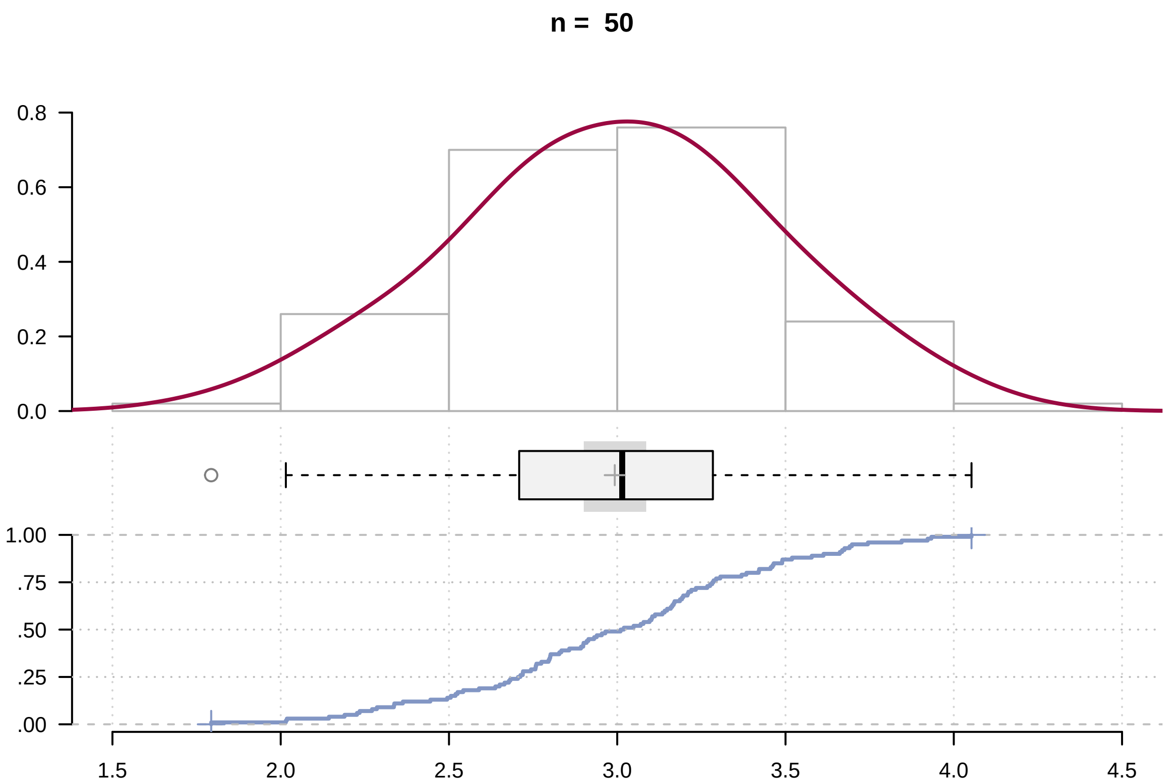
n = 10



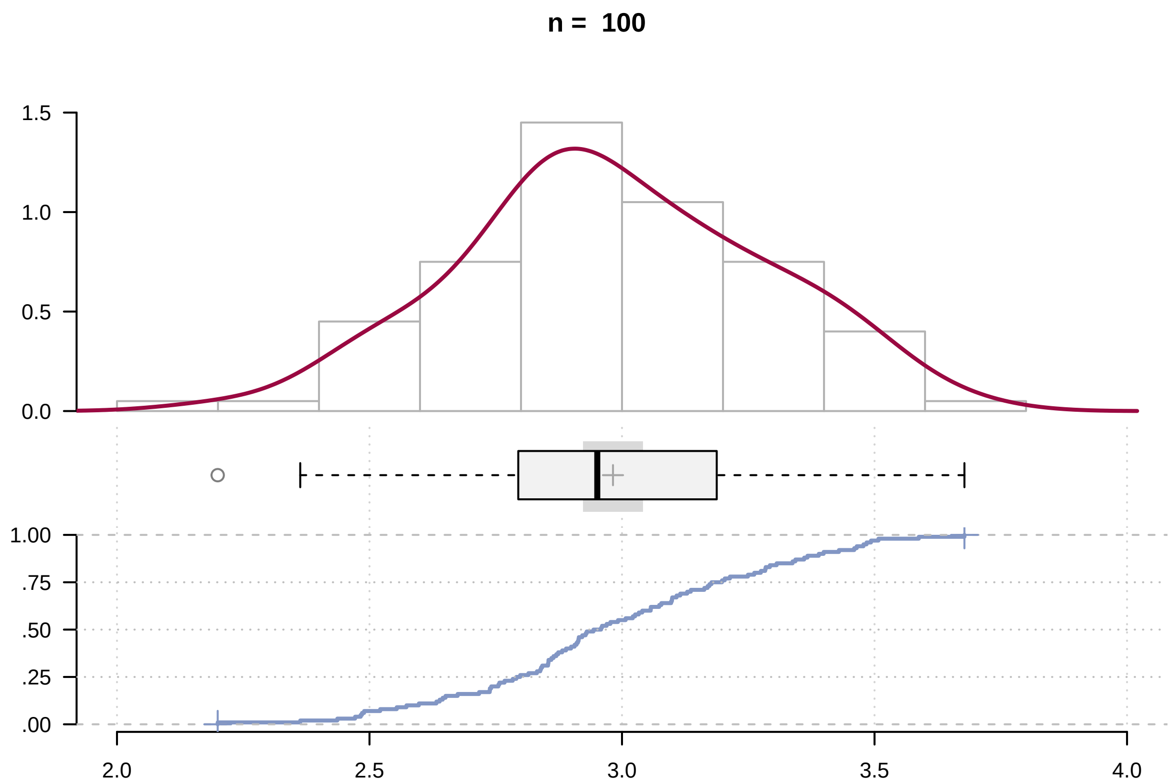

```
## -----------------------------------------------------------------------------
## n =  20
##
##      length         n       NAs    unique        0s      mean    meanCI'
##         100       100         0       = n         0  2.977908   2.854809
##                100.0%      0.0%                 0.0%            3.101007
##
##         .05       .10       .25    median       .75       .90       .95
##    2.013011  2.313356  2.554122  2.945761  3.398583  3.705801  3.953854
##
##       range        sd     vcoef       mad       IQR      skew      kurt
##    3.402436  0.620392  0.208331  0.637923  0.844461  0.172731  0.308399
##
## lowest : 1.448831, 1.454257, 1.633929, 1.959063, 1.98167
## highest: 4.076656, 4.135136, 4.164316, 4.579352, 4.851267
##
## ' 95%-CI (classic)
```



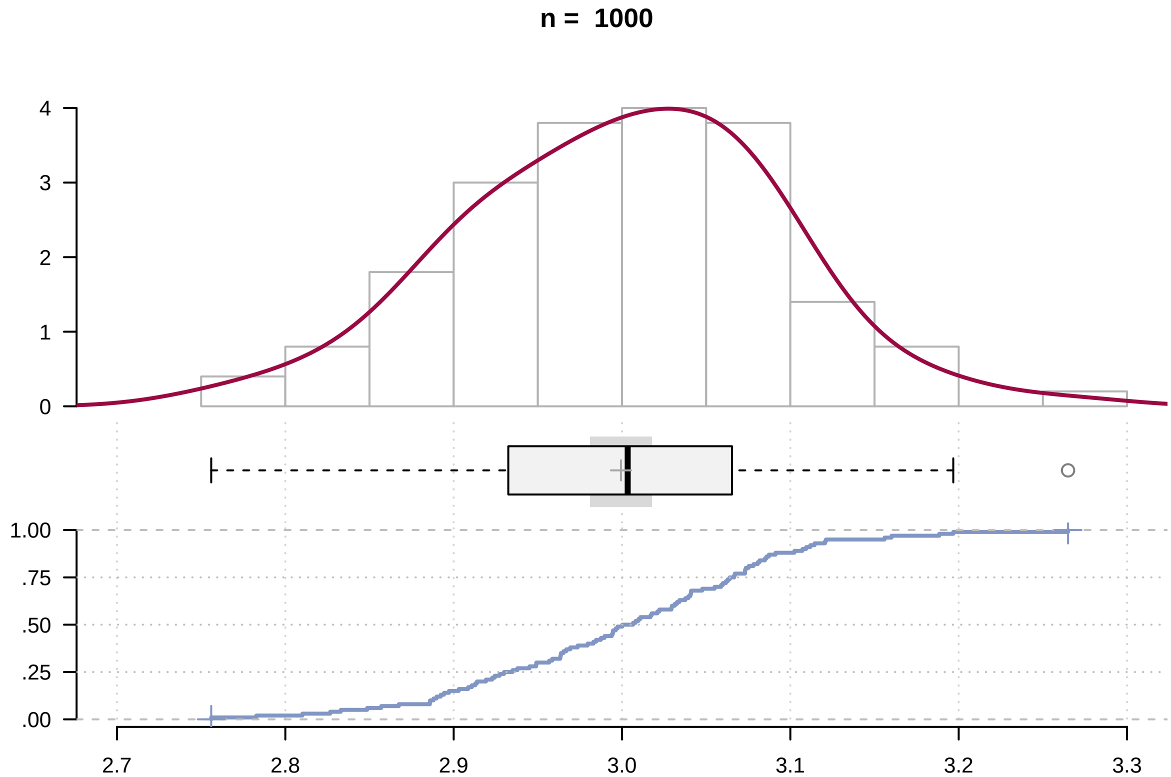

```
## ------------------------------------------------------------------------------
## n =  50
##
##      length           n         NAs     unique          0s        mean     meanCI'
##         100         100           0        = n           0    2.992624    2.900013
##                   100.0%        0.0%                   0.0%                3.085236
##
##         .05         .10         .25     median         .75         .90         .95
##    2.224880    2.337376    2.710427   3.014753    3.283267    3.617801    3.699617
##
##       range          sd       vcoef        mad         IQR        skew        kurt
##    2.259181    0.466740    0.155964   0.434623    0.572841   -0.110976   -0.354584
##
## lowest : 1.793445, 2.015336, 2.018268, 2.143294, 2.189758
## highest: 3.744792, 3.845132, 3.922184, 3.933358, 4.052626
##
## ' 95%-CI (classic)
```



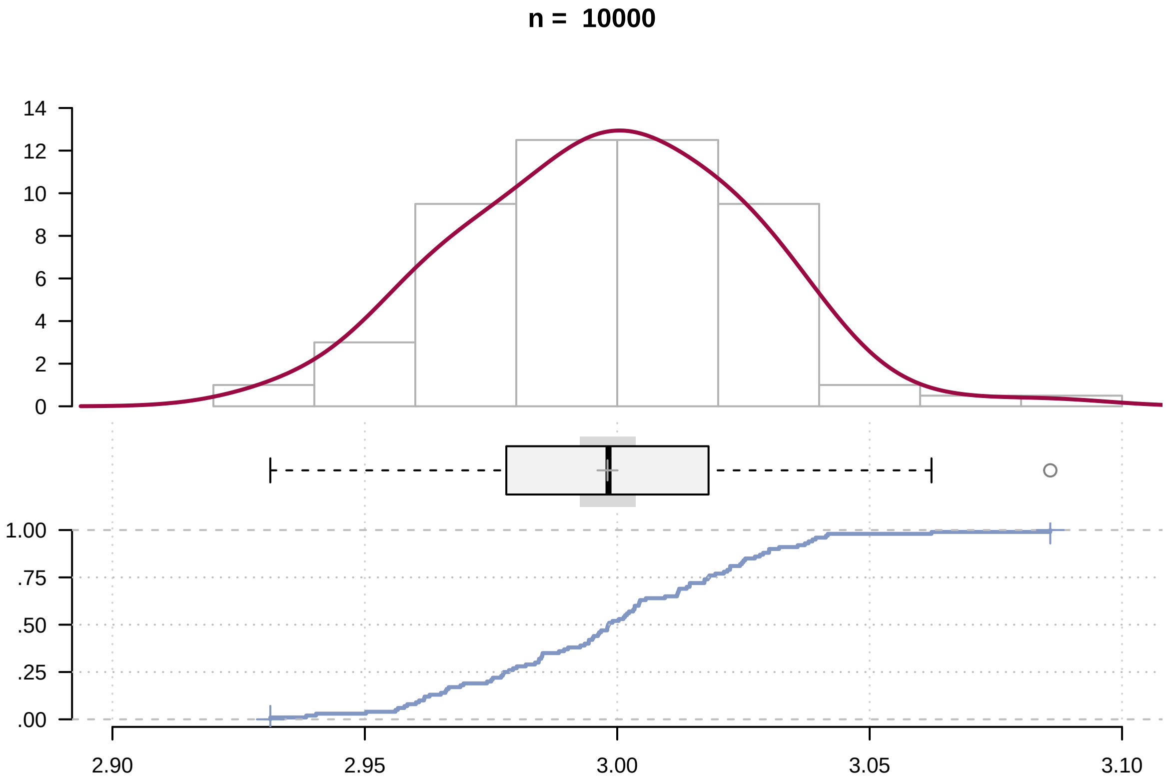

```
## ------------------------------------------------------------------------------
## n =  100
##
##     length         n        NAs    unique         0s       mean     meanCI'
##        100       100          0       = n          0   2.982156   2.922829
##               100.0%       0.0%                  0.0%              3.041482
##
##        .05       .10        .25    median        .75        .90        .95
##   2.485097  2.595342   2.796479  2.951018   3.182368   3.390628   3.478418
##
##      range        sd      vcoef       mad        IQR       skew       kurt
##   1.478560  0.298991   0.100260  0.288832   0.385889  -0.005024  -0.407928
##
## lowest : 2.199439, 2.362889, 2.436311, 2.471503, 2.48265
## highest: 3.484247, 3.49299, 3.507794, 3.587643, 3.677999
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  1000
##
##      length          n        NAs     unique         0s       mean    meanCI'
##         100        100          0        = n          0   2.999369   2.981022
##                 100.0%       0.0%                   0.0%              3.017717
##
##         .05        .10        .25     median        .75        .90        .95
##    2.847828   2.887817   2.933710   3.003387   3.064653   3.107393   3.122854
##
##       range         sd      vcoef        mad        IQR       skew       kurt
##    0.508969   0.092470   0.030830   0.095758   0.130943  -0.074060   0.045698
##
## lowest : 2.755988, 2.782777, 2.810167, 2.826658, 2.832947
## highest: 3.155937, 3.16011, 3.188469, 3.196786, 3.264957
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  10000
##
##      length         n       NAs    unique        0s      mean    meanCI'
##         100       100         0       = n         0  2.998075  2.992542
##                100.0%      0.0%                0.0%            3.003609
##
##         .05       .10       .25    median       .75       .90       .95
##    2.956533  2.961629  2.978281  2.998272  3.018020  3.030278  3.038699
##
##       range        sd     vcoef       mad       IQR      skew      kurt
##    0.154504  0.027888  0.009302  0.029424  0.039739  0.084003  0.026326
##
## lowest : 2.931278, 2.938387, 2.940335, 2.950248, 2.95607
## highest: 3.03931, 3.04133, 3.041709, 3.062253, 3.085781
##
## ' 95%-CI (classic)
```



A tabela a seguir expõem as principais métricas estatísticas da distribuição Uniforme geradas:

	Médias	sd	Mediana	crtose
10	3.028238	0.9168851	3.028238	0.2673881
20	2.977908	0.6203920	2.945761	0.3083990
50	2.992624	0.4667400	3.014753	-0.3545840
100	2.982156	0.2989910	2.951018	-0.4079280
1000	2.999369	0.0924700	3.003287	0.0456980
10000	2.998075	0.0278880	2.998272	0.0263260

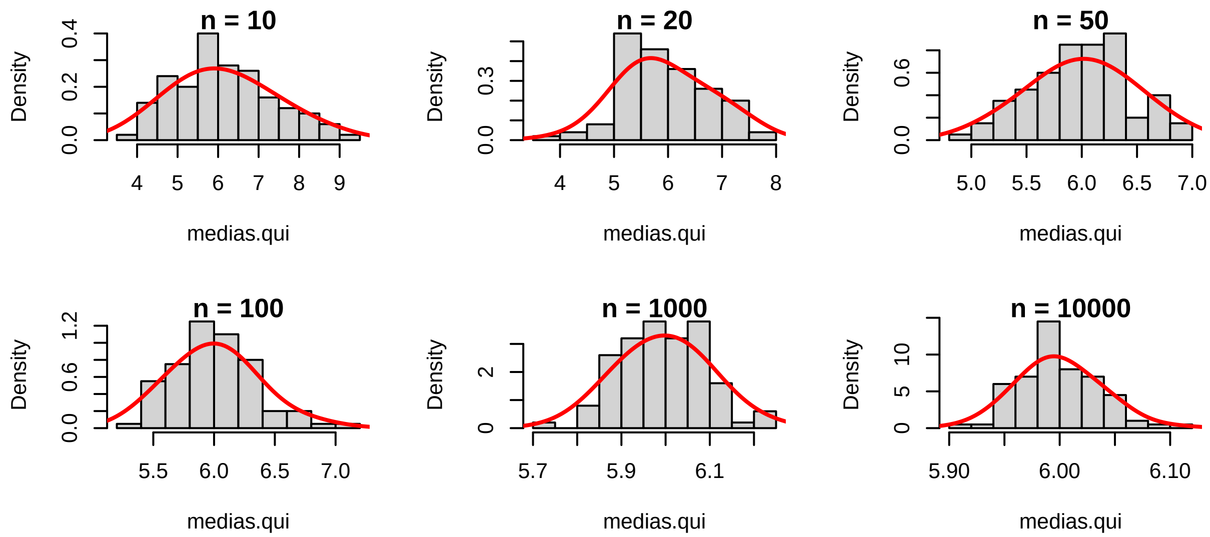
7. TCL:Exemplo utilizando a distribuição qui-quadrado.

Definição 7.1 Uma variável aleatória contínua X tem *distribuição qui-quadrado* com v graus de liberdade, denotada por X $\sim X_v^2$, se sua função densidade de probabilidade for dada por:

$$f(x) = \left\{ \frac{1}{2^(v/2)\Gamma_(v/2)} \; x^{(v/2)-1} \; e^{-x/2} \right.$$

0, se x $<$ 0

7.1 TCL: Códigos no R para a elaboração dos gráficos com as simulações - qui-quadrado



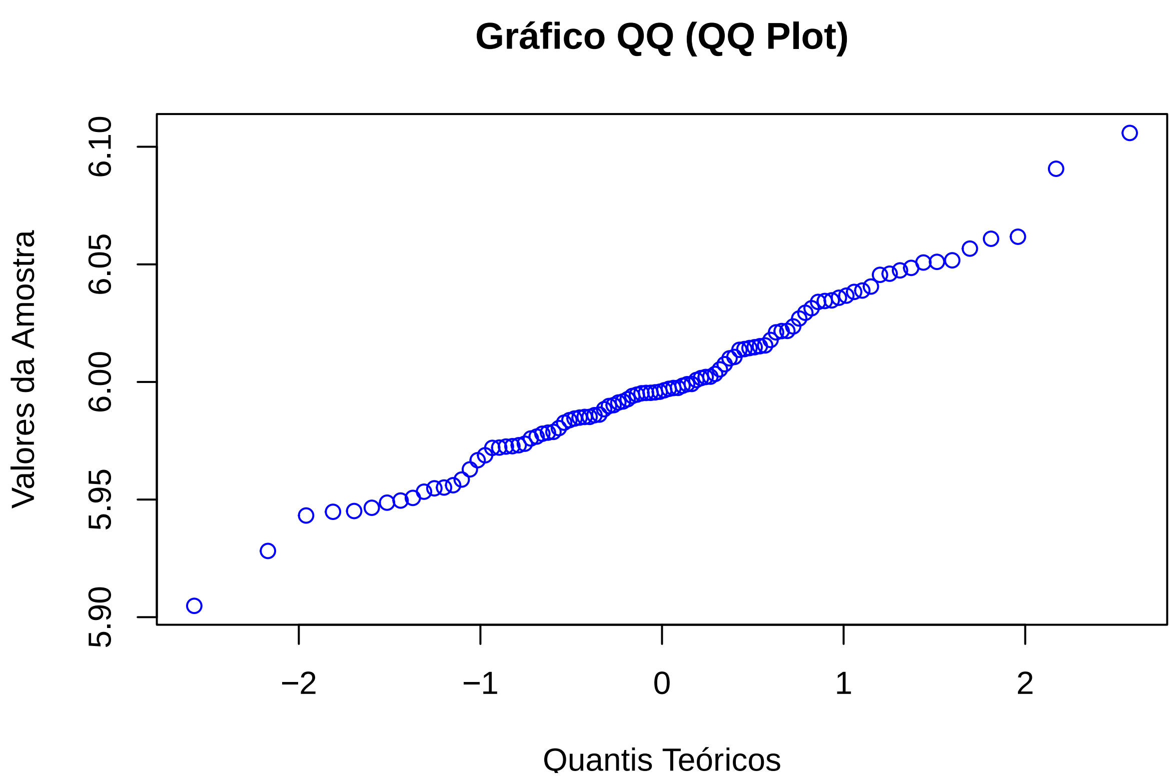
Os seis histogramas nos mostram que, à medida que o tamanho n da amostra cresce, a forma do histograma se aproxima cada vez mais de uma curva normal, comprovando o *Teorema Central do Limite.*

7.2 Teste da Normalidade

Utilizaremos o teste de Shapiro-Wilk para verificarmos a normalidade dos dados.

```
##
##  Shapiro-Wilk normality test
##
## data:  medias.qui
## W = 0.98894, p-value = 0.5796
```

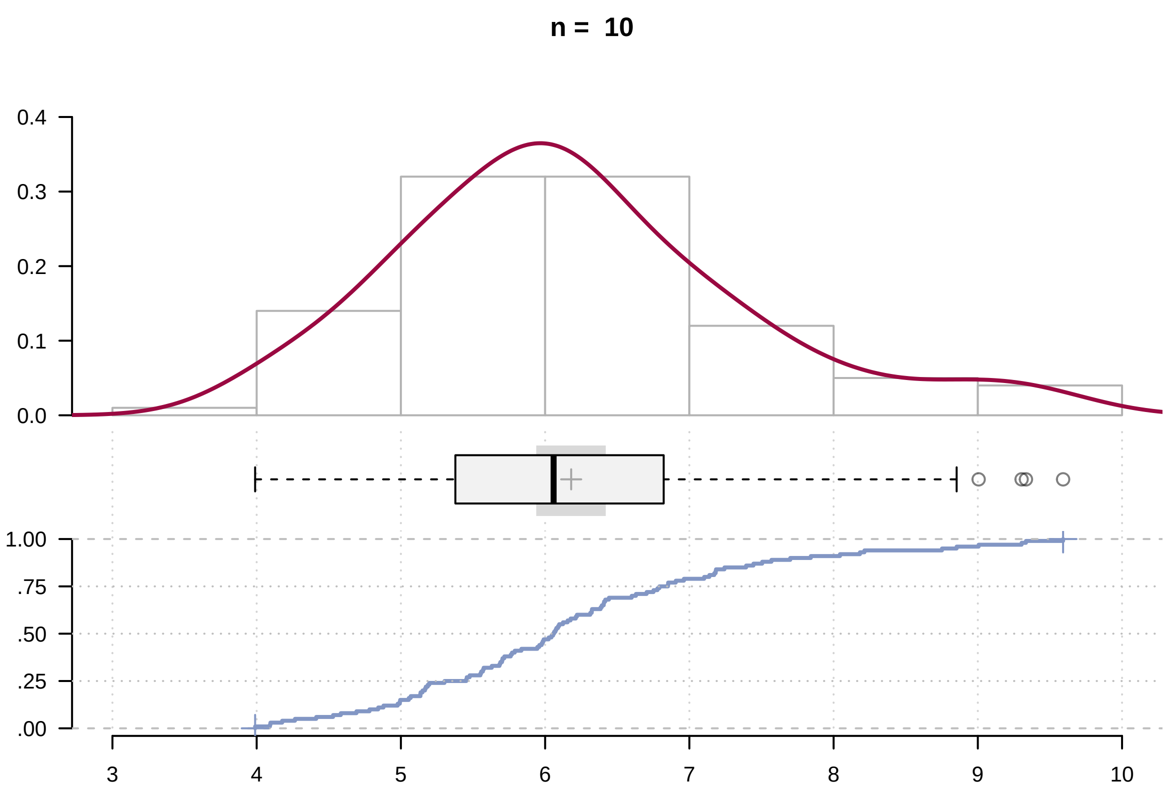
7.3 Plotando o Gráfico



Análise: p-value = 0,9194, indica uma probabilidade de que essa variável veio de uma distribuição normal o que é comprovada pela análise gráfica.

7.4 Estatística descritiva com o pacote *DescTools*

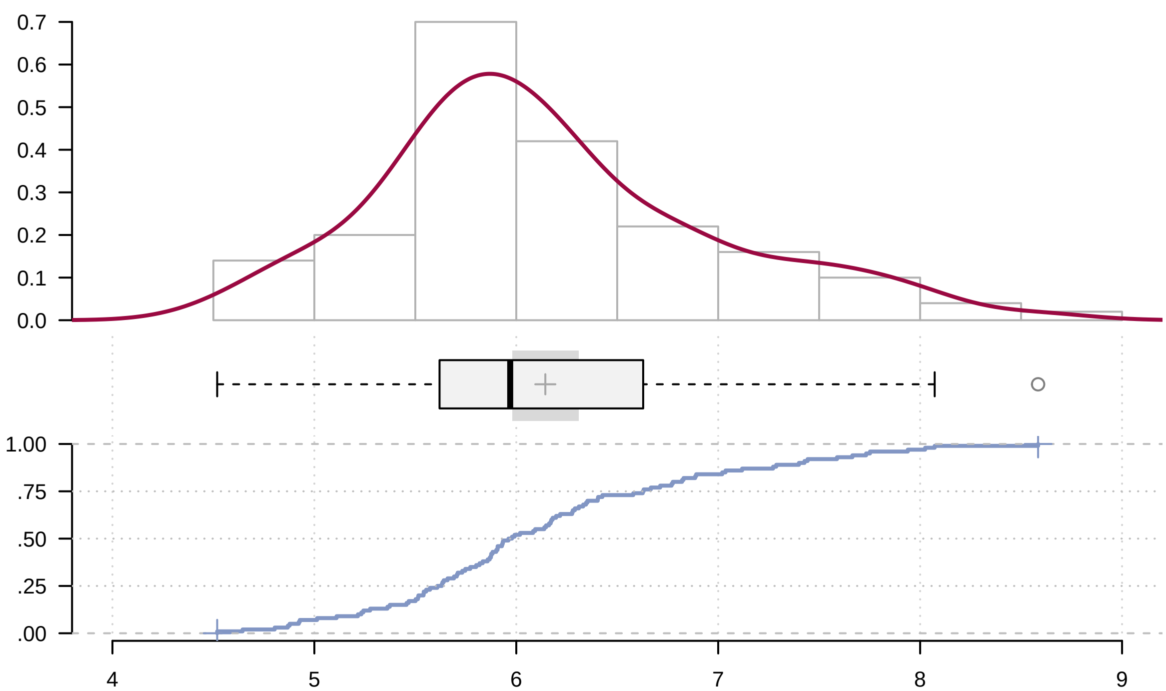
```
## ------------------------------------------------------------------------------
## n =  10
##
##     length         n       NAs    unique        0s       mean    meanCI'
##        100       100         0       = n         0   6.180589   5.941256
##               100.0%      0.0%                 0.0%              6.419923
##
##        .05       .10       .25    median       .75        .90        .95
##   4.405695  4.836662  5.415575  6.058717  6.806827   7.713779   8.756472
##
##      range        sd     vcoef       mad       IQR       skew       kurt
##   5.601868  1.206185  0.195157  1.104287  1.391251   0.687871   0.368256
##
## lowest : 3.989335, 4.092172, 4.095352, 4.17679, 4.265684
## highest: 8.853064, 9.005574, 9.303441, 9.333708, 9.591203
##
## ' 95%-CI (classic)
```



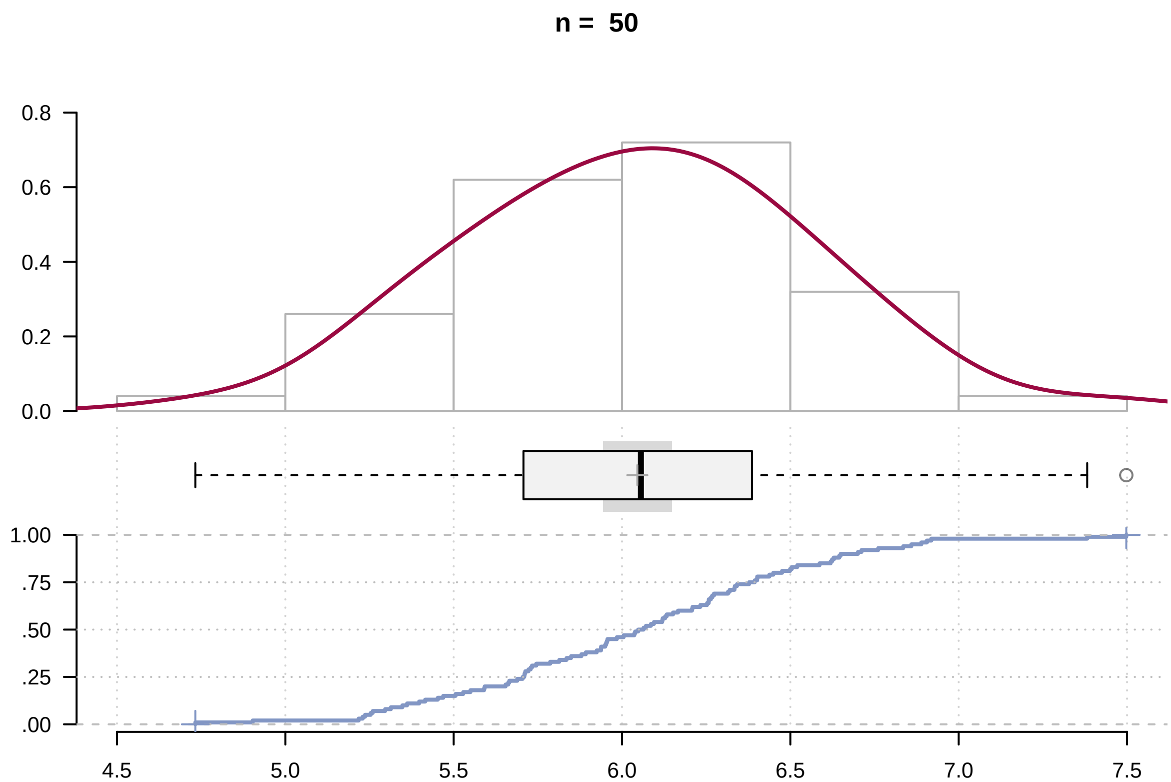

```
## ------------------------------------------------------------------------------
## n =  20
##
##     length         n       NAs    unique        0s       mean    meanCI'
##        100       100         0       = n         0   6.143925   5.980531
##               100.0%      0.0%                 0.0%              6.307318
##
##        .05       .10       .25    median       .75        .90        .95
```

```
##   4.920595  5.231114  5.625614  5.969968  6.627485  7.402505  7.733459
##
##      range        sd     vcoef       mad       IQR      skew      kurt
##   4.065305  0.823468  0.134030  0.639406  1.001871  0.630294  0.113770
##
## lowest : 4.518855, 4.6448, 4.803615, 4.868474, 4.877593
## highest: 7.751769, 7.938972, 8.024685, 8.072048, 8.584159
##
## ' 95%-CI (classic)
```

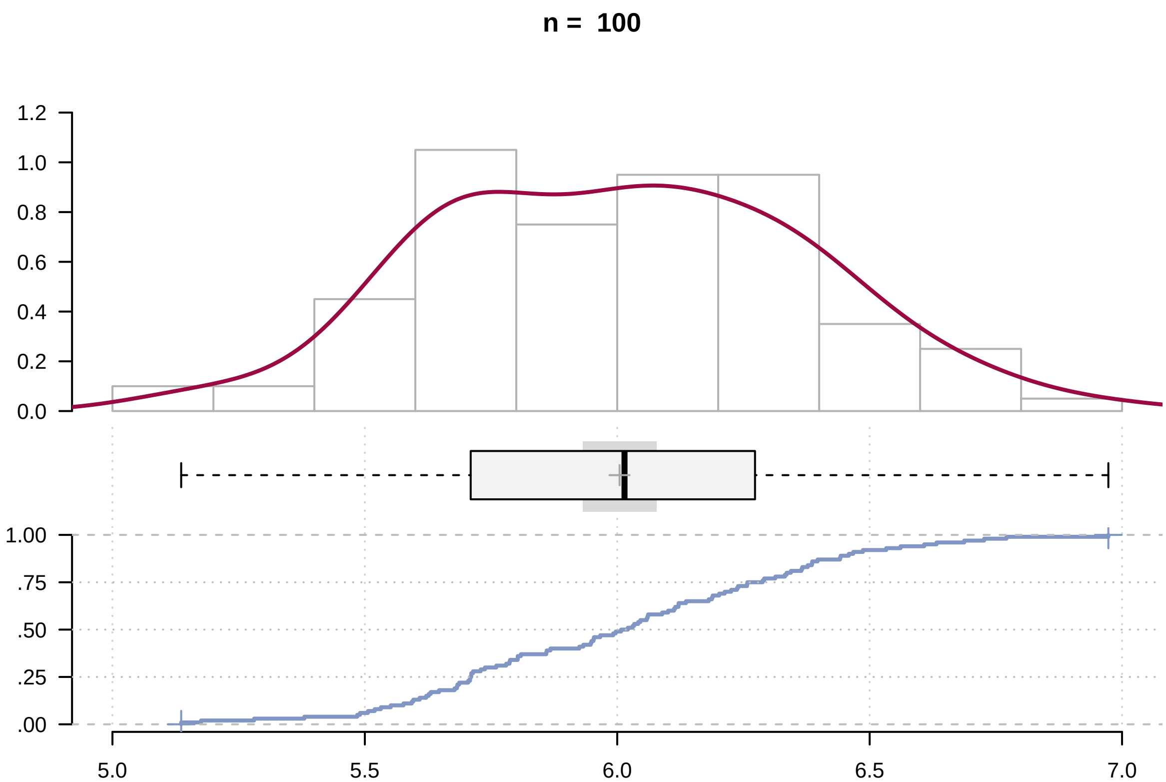
n = 20




```
## ------------------------------------------------------------------------------
## n =  50
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        100       100         0       = n         0  6.045567   5.943780
##               100.0%      0.0%                 0.0%             6.147354
##
##        .05       .10       .25    median       .75       .90        .95
##   5.252855  5.360295  5.708535  6.056047  6.381739  6.654573   6.860707
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   2.765122  0.512984  0.084853  0.511818  0.673203  0.106479  -0.017459
##
## lowest : 4.732683, 4.903193, 5.217731, 5.229889, 5.236883
## highest: 6.888885, 6.905395, 6.918689, 7.381681, 7.497805
##
## ' 95%-CI (classic)
```

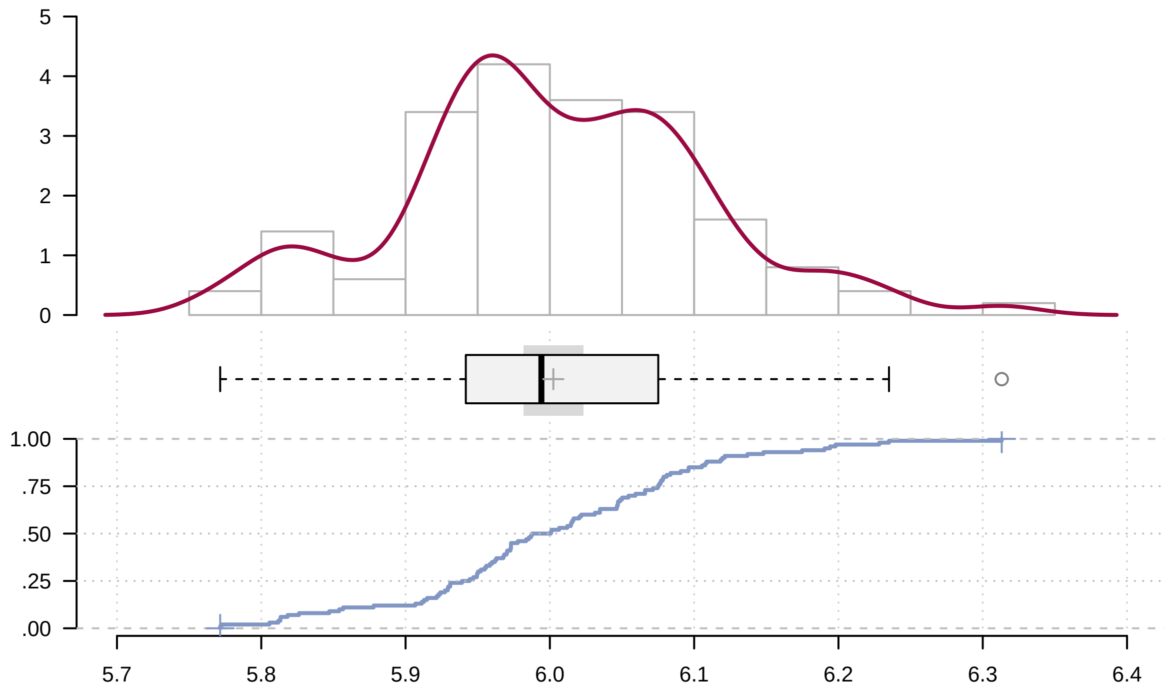



```
## ------------------------------------------------------------------------------
## n =  100
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        100       100         0       = n         0  6.004604   5.931538
##               100.0%      0.0%                0.0%             6.077670
##
##        .05       .10       .25    median       .75       .90        .95
##   5.490196  5.574052  5.710250  6.014391  6.265314  6.459544   6.609217
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   1.836701  0.368236  0.061326  0.443848  0.555063  0.094753  -0.431943
##
## lowest : 5.136169, 5.175572, 5.280582, 5.38014, 5.484931
## highest: 6.632422, 6.687336, 6.726806, 6.771126, 6.97287
##
## ' 95%-CI (classic)
```

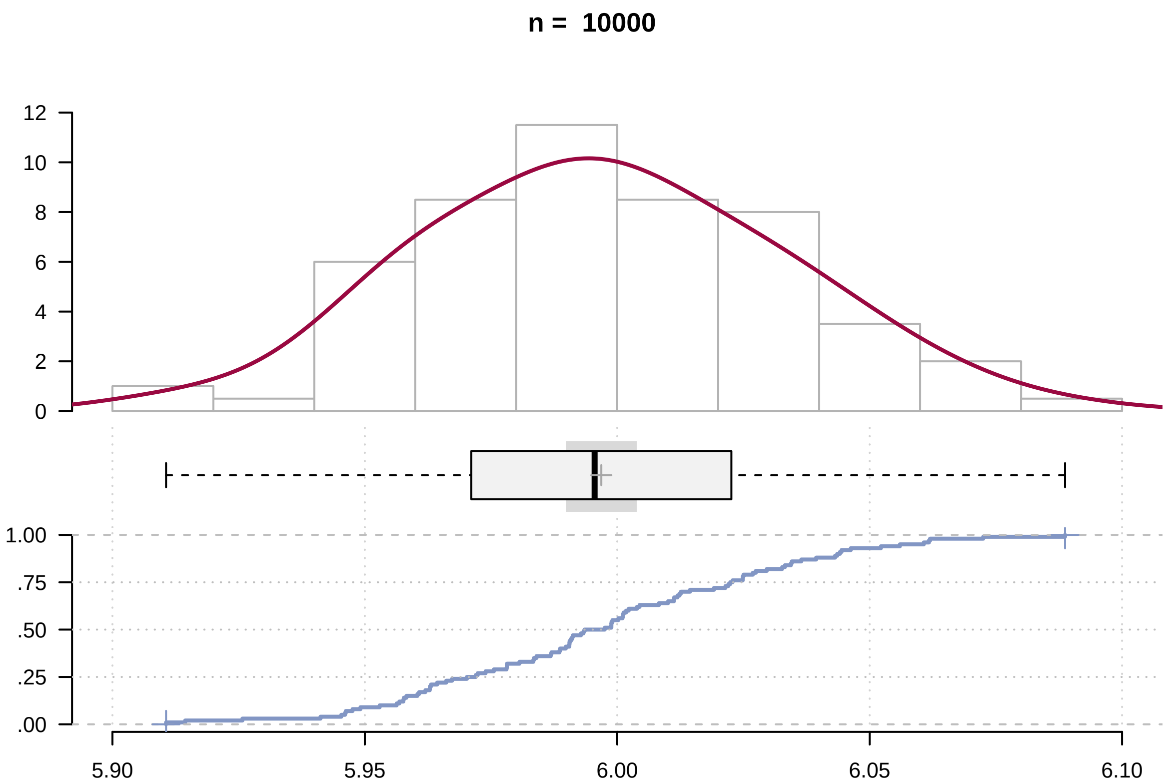



```
## ------------------------------------------------------------------------------
## n =  1000
##
##     length        n      NAs   unique       0s      mean    meanCI'
##        100      100        0      = n        0  6.002410  5.981651
##              100.0%     0.0%             0.0%            6.023170
##
##        .05      .10      .25   median      .75       .90       .95
##   5.813397 5.856454 5.943095 5.994127 6.074956  6.119676  6.190534
##
##      range       sd    vcoef      mad      IQR      skew      kurt
##   0.541629 0.104622 0.017430 0.096395 0.131861  0.149622  0.139202
##
## lowest : 5.77153, 5.772426, 5.805622, 5.811812, 5.813222
## highest: 6.194179, 6.197946, 6.228253, 6.235013, 6.313158
##
## ' 95%-CI (classic)
```

n = 1000




```
## ------------------------------------------------------------------------------
## n =  10000
##
##      length          n        NAs     unique         0s       mean     meanCI'
##         100        100          0        = n          0   5.996847    5.989829
##                 100.0%       0.0%                  0.0%               6.003865
##
##         .05        .10        .25     median        .75        .90         .95
##    5.946040   5.955968   5.971531   5.995514   6.022491   6.043640    6.056232
##
##       range         sd      vcoef        mad        IQR       skew        kurt
##    0.178080   0.035369   0.005898   0.039553   0.050960   0.074521   -0.307291
##
## lowest : 5.910626, 5.91441, 5.925746, 5.94122, 5.945336
## highest: 6.060722, 6.061737, 6.061959, 6.072504, 6.088706
##
## ' 95%-CI (classic)
```



A tabela a seguir expõem as principais métricas estatísticas da distribuição qui-quadrado geradas:

	Médias	sd	Mediana	Curtose
10	6.180589	1.206185	6.058714	0.368256
20	6.143925	0.823468	5.969968	0.113770
50	6.045567	0.512984	6.056047	-0.017459
100	6.004604	0.368236	6.014391	-0.431943
1000	6.002410	0.104622	5.994127	0.139202
10000	5.996847	0.035369	5.995514	-0.307291

8. TCL:Exemplo utilizando a distribuição Gama.

8.1 Definição: A variável aleatória X tem distribuição Gama, se sua densidade for dada por:

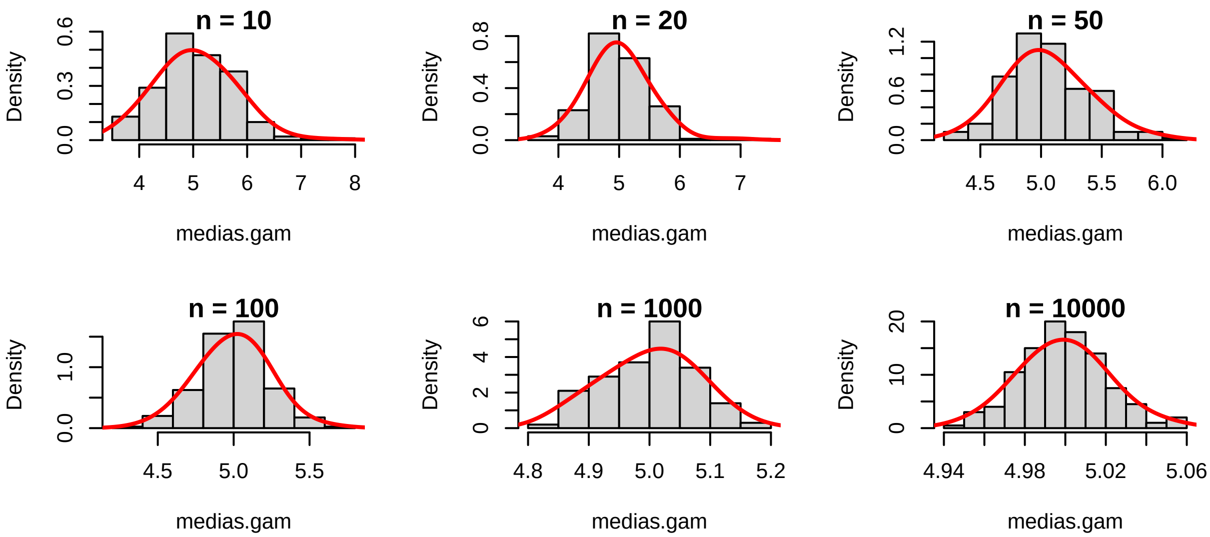
$f(x) = \frac{\beta}{\Gamma_(\alpha)}\ x^(\alpha - 1)\ e^{-\beta x}$, x $\geq$ 0,

sendo α e β dois parâmetros positivos e $\Gamma_(\alpha)$ sendo a função matemática Gama, definida por

$\Gamma_(\alpha) = \int_0^\infty x^{\alpha-1}\ e^{-x}\, dx$, $\alpha > 0$.

Notação: X $\sim$ Gama(α,β).

8.2 TCL: Códigos no R para a elaboração dos gráficos com as simulações - Gama



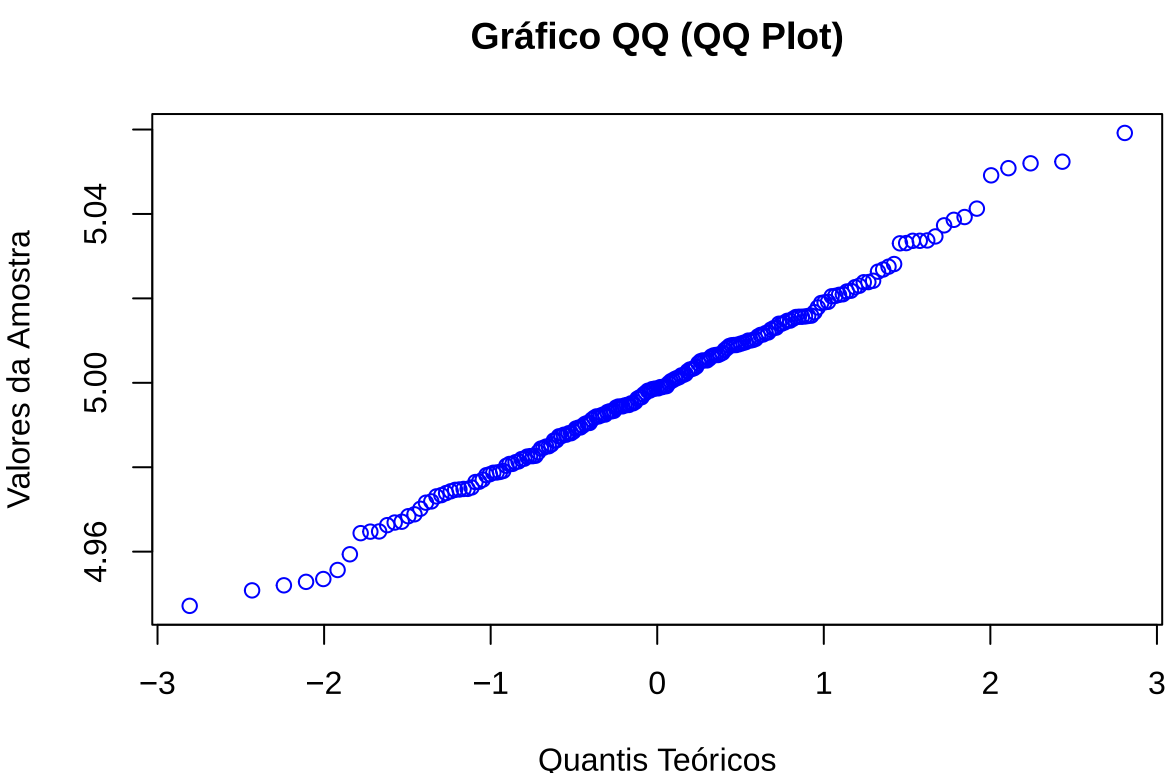
Os seis histogramas nos mostram que, à medida que o tamanho n da amostra cresce, a forma do histograma se aproxima cada vez mais de uma curva normal, comprovando o *Teorema Central do Limite.*

8.3 Teste da Normalidade

Utilizaremos o teste de Shapiro-Wilk para verificarmos a normalidade dos dados.

```
##
##  Shapiro-Wilk normality test
##
## data:  medias.gam
## W = 0.99459, p-value = 0.6884
```

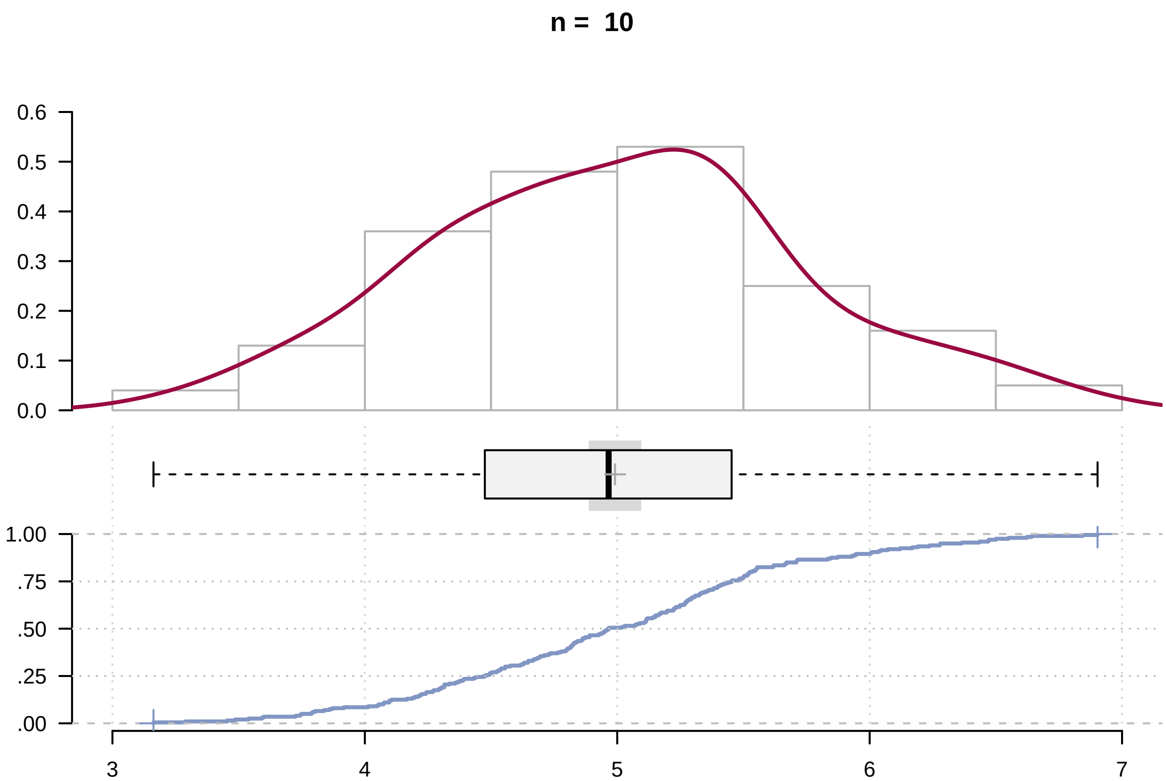
8.4 Plotando o Gráfico



Análise: p-value = 0,7975, indica uma probabilidade de que essa variável veio de uma distribuição normal o que é comprovada pela análise gráfica.

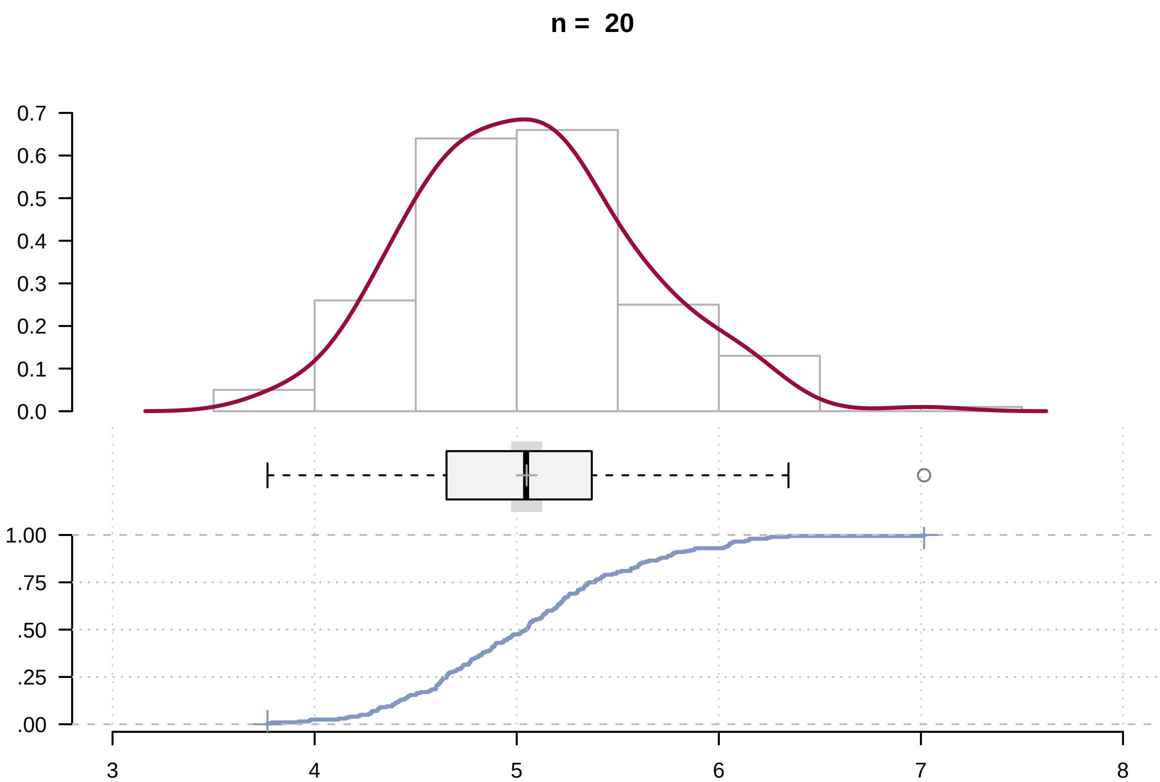
8.5 Estatística descritiva com o pacote *DescTools*

```
## ------------------------------------------------------------------------------
## n =  10
##
##     length         n       NAs    unique        0s      mean     meanCI'
##        200       200         0       = n         0  4.991224   4.887343
##              100.0%      0.0%                 0.0%             5.095104
##
##        .05       .10       .25    median       .75       .90        .95
##   3.786569  4.072046  4.477347  4.965807  5.452388  6.004163   6.283585
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   3.740320  0.744993  0.149260  0.722950  0.975041  0.110659  -0.244905
##
## lowest : 3.162623, 3.289009, 3.45462, 3.485854, 3.539154
## highest: 6.550239, 6.623321, 6.642224, 6.844316, 6.902943
##
## ' 95%-CI (classic)
```

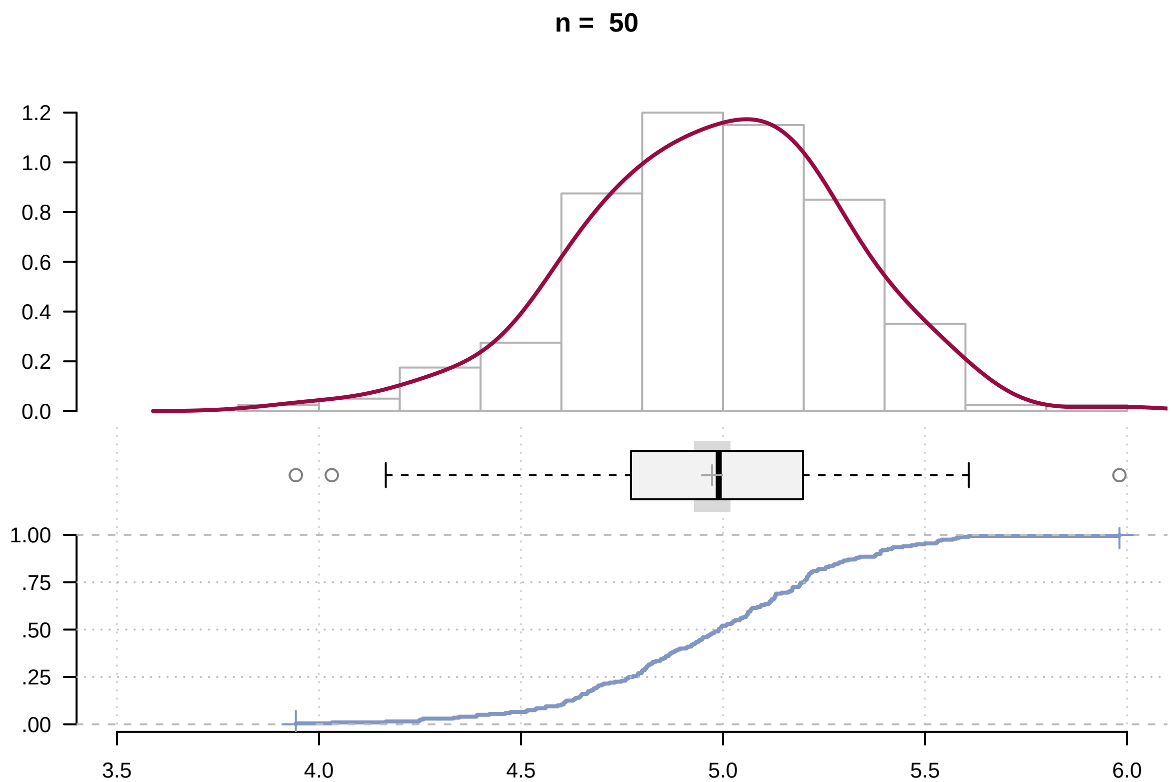



```
## ------------------------------------------------------------------------------
## n =  20
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        200       200         0       = n         0  5.048975  4.972482
##              100.0%      0.0%                 0.0%            5.125468
##
##        .05       .10       .25    median       .75       .90       .95
```

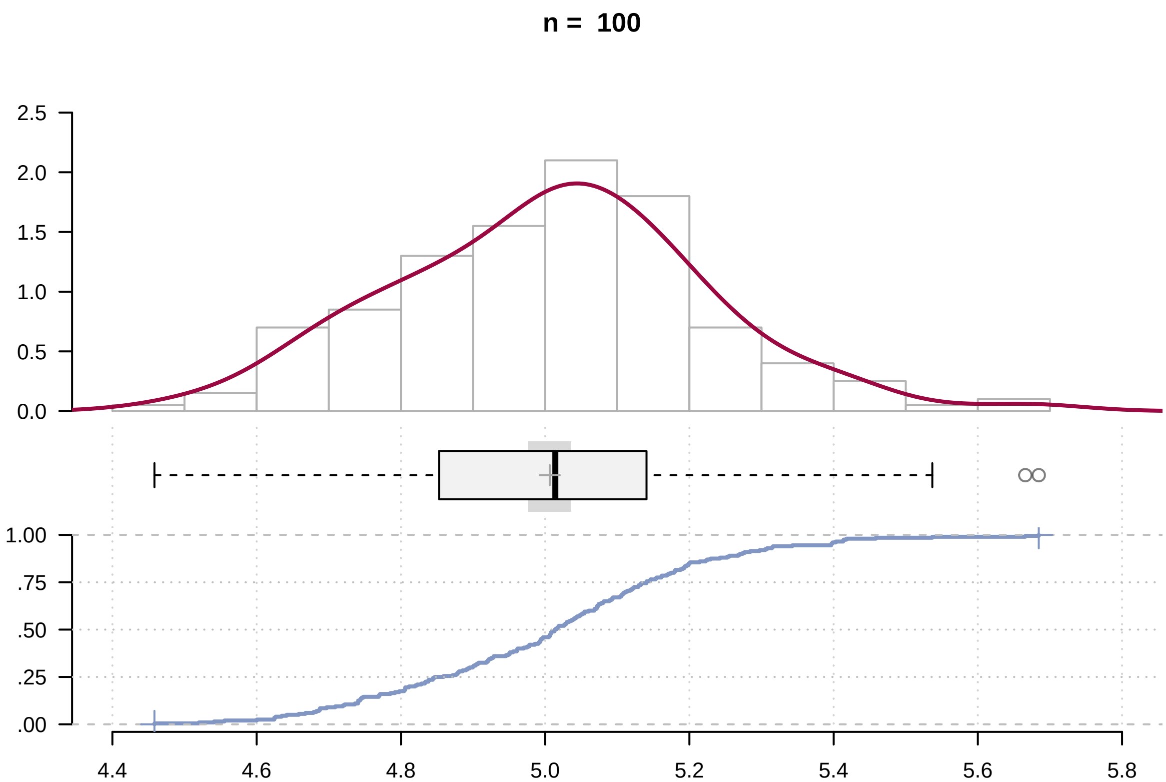
```
##   4.265893  4.385271  4.652695  5.046407  5.363438  5.771639  6.051975
##
##      range        sd     vcoef       mad       IQR      skew      kurt
##   3.248991  0.548580  0.108652  0.552681  0.710743  0.351826  0.132389
##
## lowest : 3.766708, 3.788259, 3.919714, 3.973337, 3.980441
## highest: 6.150516, 6.23911, 6.2544, 6.344614, 7.015699
##
## ' 95%-CI (classic)
```



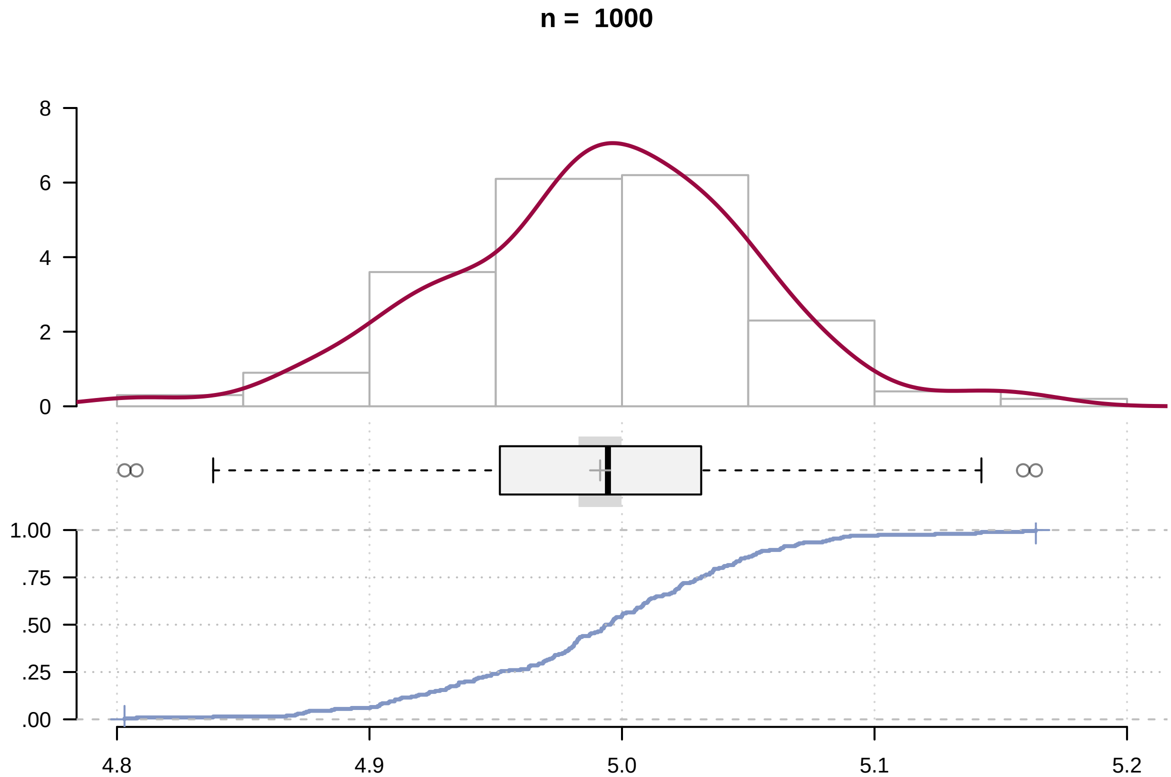

```
## ------------------------------------------------------------------------------
## n =  50
##
##     length         n       NAs    unique        0s       mean    meanCI'
##        200       200         0       = n         0   4.972802   4.927799
##                100.0%      0.0%                0.0%              5.017805
##
##        .05       .10       .25    median       .75        .90        .95
##   4.420279  4.598884  4.774845  4.989514  5.196754   5.381517   5.479187
##
##      range        sd     vcoef       mad       IQR       skew       kurt
##   2.038455  0.322743  0.064902  0.313634  0.421909  -0.255996   0.330556
##
## lowest : 3.942738, 4.031695, 4.165346, 4.247356, 4.250417
## highest: 5.569642, 5.578265, 5.585192, 5.608454, 5.981193
##
## ' 95%-CI (classic)
```



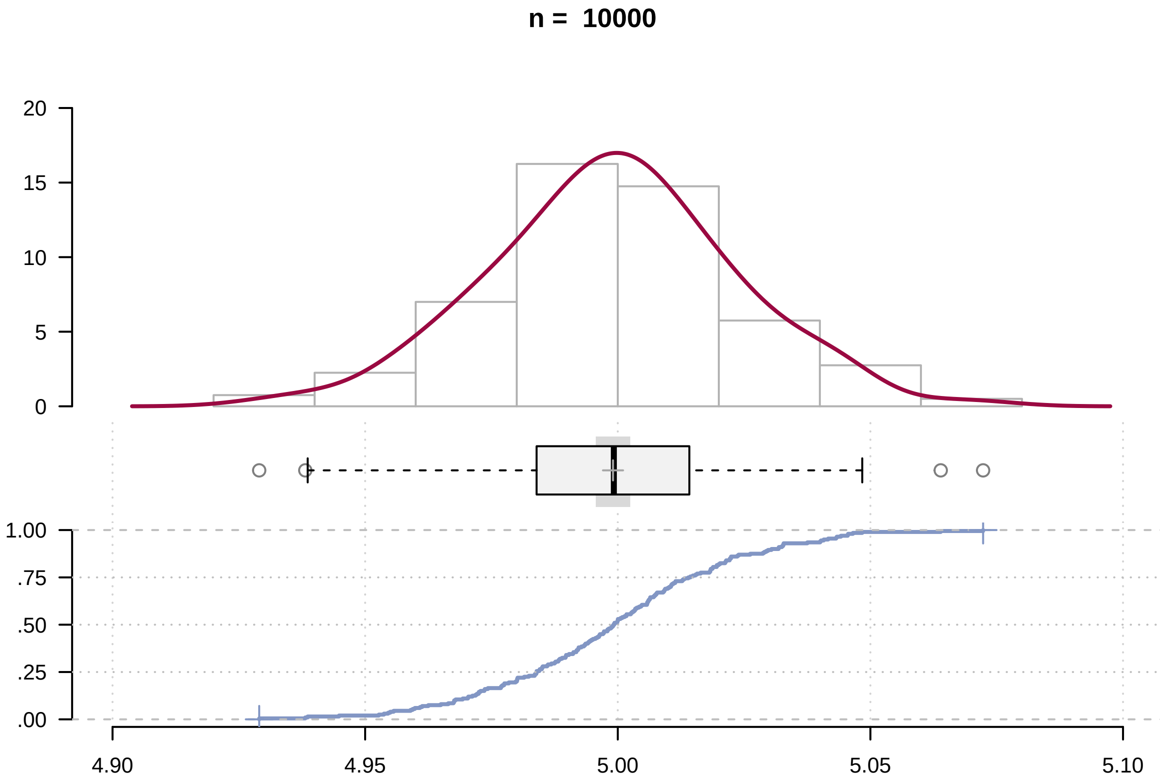

```
## ------------------------------------------------------------------------------
## n =  100
##
##     length         n       NAs    unique        0s      mean    meanCI'
##        200       200         0       = n         0  5.006404   4.976301
##               100.0%      0.0%                 0.0%            5.036507
##
##        .05       .10       .25    median       .75       .90        .95
##   4.657535  4.721756  4.856038  5.014161  5.140509  5.270519   5.396649
##
##      range        sd     vcoef       mad       IQR      skew       kurt
##   1.226159  0.215888  0.043122  0.202676  0.284471  0.141008   0.153516
##
## lowest : 4.458321, 4.52, 4.540865, 4.555334, 4.600204
## highest: 5.417875, 5.458676, 5.536922, 5.665804, 5.68448
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  1000
##
##     length         n       NAs    unique        0s       mean    meanCI'
##        200       200         0       = n         0   4.991335   4.982839
##              100.0%      0.0%                  0.0%              4.999830
##
##        .05       .10       .25    median       .75        .90        .95
##   4.886094  4.910092  4.951842  4.994427  5.031317   5.062590   5.082352
##
##      range        sd     vcoef       mad       IQR       skew       kurt
##   0.360905  0.060927  0.012206  0.057133  0.079475  -0.169762   0.485402
##
## lowest : 4.802993, 4.807723, 4.838098, 4.867087, 4.870634
## highest: 5.123937, 5.140097, 5.142333, 5.158843, 5.163898
##
## ' 95%-CI (classic)
```




```
## ------------------------------------------------------------------------------
## n =  10000
##
##     length         n       NAs    unique        0s       mean    meanCI'
##        200       200         0       = n         0   4.999060   4.995676
##               100.0%      0.0%               0.0%              5.002443
##
##        .05       .10       .25    median       .75        .90        .95
##   4.959401  4.967871  4.983932  4.999221  5.014063   5.030561   5.040807
##
##      range        sd     vcoef       mad       IQR       skew       kurt
##   0.143278  0.024263  0.004854  0.022571  0.030131  -0.000684   0.160981
##
## lowest : 4.92903, 4.938155, 4.938626, 4.944822, 4.9527
## highest: 5.045579, 5.046544, 5.048387, 5.063922, 5.072308
##
## ' 95%-CI (classic)
```



A tabela a seguir expõem as principais métricas estatísticas da distribuição Gama geradas:

	Médias	sd	Mediana	Curtose
10	4.991224	0.784993	4.965807	-0.2449050
20	5.048975	0.548580	5.046407	0.1323890
50	4.972802	0.322743	4.989514	0.3305560
100	5.006404	0.215888	5.014161	0.1153516
1000	4.991335	0.060927	4.994427	0.4854020
10000	4.995676	0.024263	4.999221	0.1609810

Conclusão

Ao utilizarmos amostras, nosso objetivo final é fazer **inferência a respeito dos reais parâmetros da população**. Para isso, usamos ferramentas como intervalos de confiança e teste de hipóteses que partem da suposição da normalidade dos dados. Desse modo, mesmo que os nossos dados sejam de uma população com distribuição desconhecida ou que não seja normalmente distribuída, ainda assim poderemos fazer tais análises uma vez que o Teorema Central do Limite é verdadeiro, (MACIEL, 2022). De forma geral, **quanto maior for a amostra, mais próxima da normal a distribuição amostral das médias será.**

Referências

ALCOFORADO, Luciane Ferreira. Utilizando a Linguagem R: conceitos, manipulação, visualização, modelagem e elaboração de relatórios. Rio de Janeiro: Alta Books, 2021. 377 p.

MACIEL, Prof. Fernanda. Teorema Central do Limite. 2022. Disponível em: https://blog.proffernandamaciel.com.br/teorema_central_limite/. Acesso em: 30 out. 2022.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

VASCONCELLOS, Klaus Leite Pinto. Fundamentos para a Estatística de Convergência de Variáveis Aleatórias. Rio de Janeiro: Sbm, 2021. 380 p. (Coleção Matemática Aplicada; 05).