Inferência Estatística I

Amostra aleatória

Prof. Paulo Cerqueira Jr
Faculdade de Estatística - FAEST
Instituto de Ciências Exatas e Naturais - ICEN

https://github.com/paulocerqueirajr

Introdução

Introdução

- Os avanços científicos são, na maioria das vezes, atribuídos aos experimentos realizados.
- Um pesquisador realiza o experimento e obtém dados.
- Baseado nos dados, algumas conclusões podem ser retiradas.
- Estas conclusões vão, geralmente, além dos que foi observado nos dados.
- Dessa forma, o pesquisador generaliza, partindo de um experimentos para os demais que são similares.
- Esta generalização é denominada de `Inferência`.

> **Uma função da Estatística:**
>
> Fornecer um conjunto de metodologias para realizar a inferência e medir o grau de incerteza dessa inferência, através da `teoria das probabilidades`.

Introdução

Exemplo 1

- Suponha um recepiente com 10 milhões de sementes de flores.
- Cada semente pode produzir flores brancas ou vermelhas.
- Pergunta-se: Qual a porcentagem de flores brancas que serão geradas?
- Para saber o resultado real, teríamos que plantar todas as sementes.
- Seria uma tarefa muito trabalhosa!
- Solução: Plantar algumas sementes, e baseando-se nos resultados podemos obter alguma informação para a porcentagem de flores brancas.

População e amostra

Definição 1 (População) É um conjunto que contém todos os elementos do problema a ser discutido, com pelo menos uma característica em comum. Desejamos obter informação sobre esta população.

Exemplo 2

- Preços da carne em um mês na região metropolitana de Belém.
- Preços do pão em certo dia em Belém.
- Produção de leite por animal em uma fazenda.
- Queremos estudar a proporção de votos para um determinado candidato ao governo do Estado do Pará.
- Queremos estudar o grau de satisfação dos usuários de uma determinada operadora de telefonia celular.

População e amostra

População e amostra

Definição 2 (Amostra aleatória - a.a) Sejam $X_1, X_2, \ldots, X_n$ uma sequência de variáveis aleatórias com distribuição conjunta $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ que fatora como na seguinte igualdade:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \cdots \times f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i),$$

em que $f(\cdot)$ é a função de probabilidade (f.p) ou função de densidade de probabilidade (f.d) para cada X_i . Então, $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de tamanho n retirada de uma população com p.d/f.d.

População e amostra

Exemplo 3 Imagine 10 milhões de sementes em um recepiente e a produção de flores brancas e vermelhas.

- `População:` Sementes dentro do recipiente.
- `Unidade experimental:` Uma semente.
- `Característica:` Flor branca ou vermelha.

- Não temos um valor numérico associado a cada elemento, mas podemos definir o seguinte tipo de resposta:

$$\text{Flor branca} = 1 \text{ e } \text{Flor vermelha} = 0.$$

`Variável aleatória:` $X_i = 1$ ou $X_i = 0$, para $i = 1, 2, \ldots, n$.

Poupulação e amostra

- A variável aleatória X_i é uma representação do valor numérico que a $i-$ésima unidade amostral irá assumir.
- Depois que a amostra $X_1, X_2, \ldots, X_n$ observada os valores serão conhecidos e denotados por $x_1, x_2, \ldots, x_n$.
- Logo:

Suponha que X pode assumir apenas os valores 0 ou 1 com probabilidades $1-\theta$ e θ, respectivamente. Então, sejam $X_1, X_2, \ldots, X_n$ uma a.a de $X \sim Ber(\theta)$, sua distribuição conjunta $P(X_1 = x_1; X_2 = x_2; \ldots; X_n = x_n)$ é dada por

$$= \theta^{x_1}(1-\theta)^{1-x_1} \times \theta^{x_2}(1-\theta)^{1-x_2} \times \cdots \times \theta^{x_n}(1-\theta)^{1-x_n}$$

$$= \theta^{x_1+x_2+\cdots+x_n}(1-\theta)^{(1-x_1)+(1-x_2)+\cdots+(1-x_n)}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{(n-\sum_{i=1}^{n} x_i)}$$

Estatísticas e Parâmetros

Estatísticas e Parâmetros - Introdução

Um dos problemas principais da Estatística envolve o seguinte:

- Estudar uma população com f.p/f.d $f(\cdot \mid \theta)$, onde a forma da f.p/f.d é conhecida com parêmetro desconhecido θ.
- Se θ fosse conhecido f.p/f.d estaria completamente especificada.

Procedimento de inferência envolverá:

- A obtenção de uma amostra aleatória $X_1, X_2, \ldots, X_n$ desta f.p/f.d.
- O uso de uma função $T(x_1, x_2, \ldots, x_n)$ como estimativa para o parâmetro θ (desconhecido).

Estatísticas

- O problema aqui consiste em determinar qual será a melhor função para estimar θ.
- Iremos avaliar certas funções (funções amostrais) de uma amostra aleatória.

Definição 3 (Estatísticas) É uma função da amostra, $T(x) = f(X_1, X_2, \ldots, X_n)$, representando uma característica da amostra.

Importante:

A formulação de uma estatística `não pode envolver quantidades desconhecidas`.

Estatísticas

> Os exemplos mais comuns:

- Média amostral: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.
- Variância amostral: $\hat{\sigma^2} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
- Mediana amostral: $\tilde{X} = \mathrm{med}(X_1, X_2, \ldots, X_n)$.
- Mínimo amostral: $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$.
- Máximo amostral: $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$.
- Ponto médio amostral: $\dfrac{1}{2}(X_{(1)} + X_{(n)})$.

Parâmetros

Definição 4 (Parâmetro) Uma parâmetro é uma medida (desconhecida) usada para descrever uma característica da população.

- As relação das estatísticas com seus respectivos parâmetros:

Medida	**Estatística**	**Parâmetro**
Média	$\overline{X}$	μ
Variância	$\hat{\sigma^2}$	σ^2
N de elementos	n	N
Proporção	$\hat{\theta}$	θ

Estatísticas e Parâmetros

Exemplo 4 Considere uma variável aleatória observável com f.d:

- $f(x) = N\left[x \mid \mu, \sigma^2\right]$, com μ e σ desconhecidos.
- Logo,

$$X - \mu \text{ e } X/\sigma \text{ são Estatísticas??}$$

- `Não são`, pois contém elementos desconhecidos.

- $X, X + 3$ e $X^2 + \log X^2$ `são estatísticas`.

Estatísticas e Parâmetros

Exemplo 5 Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória com f.p/f.d $f(\cdot;\theta)$ então:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad \frac{1}{2}\{\min(X_1,\ldots,X_n) + \max(X_1,\ldots,X_n)\}$$

`são exemplos` de estatísticas.

Exemplo 6 Se $f(x;\theta) = N[x \mid \theta, 1]$, com θ desconhecido.

$$\overline{X}_n - \theta \quad \text{é uma Estatística?}$$

- `Não` é uma estatística, pois depende de θ.

Momentos amostrais

Momentos amostrais

Seja $X_1, \ldots, X_n$ uma a.a com f.p/f.d $f(\cdot)$. O r-ésimo momento amostral em relação à 0 é definido por

$$M_r^{'} = \frac{1}{n} \sum_{i=1}^{n} (X_i - 0)^r = \frac{1}{n} \sum_{i=1}^{n} (X_i)^r.$$

- Em particular, quando $r = 1$, temos a média amostral $\overline{X}$ ou $\overline{X}_n$, em que

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i).$$

O r-ésimo momento em relação à $\overline{X}_n$ é dado por

$$M_r = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^r$$

Momentos amostrais

> Momentos amostrais são exemplos de estatísticas!

Teorema 1 (Momentos Amostrais) Seja $X_1, \ldots, X_n$ uma a.a com f.p/f.d $f(\cdot)$. O valor esperado do r-ésimo momento amostral (em relação à 0) é igual ao r-ésimo momento populacional, isto é,

$$E(M_r^{'}) = \mu_r^{'}, \text{ se } \mu_r^{'} \text{ existir.}$$

- Temos que $\mu_r^{'} = E(X^r)$ é o r-ésimo momento populacional de uma população com f.p/f.d $f(x) = f_X(x)$.
- Além disso:

$$\begin{aligned} Var(M_r^{'}) &= \frac{1}{n}\left[E(X^{2r}) - E^2(X^r)\right] \\ &= \frac{1}{n}\left[\mu_{2r}^{'} - (\mu_r^{'})^2\right]. \end{aligned}$$

Momentos amostrais

Demonstração:(cont.)

> A média:

$$E(M_r^{'}) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i)^r\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i)^r\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}E\left[(X_i)^r\right] = \frac{1}{n}\sum_{i=1}^{n}\mu_r^{'} = \mu_r^{'}.$$

> A variância:

$$Var(M_r^{'}) = Var\left[\frac{1}{n}\sum_{i=1}^{n}(X_i)^r\right] = \frac{1}{n^2}Var\left[\sum_{i=1}^{n}(X_i)^r\right]$$

Momentos amostrais

`Demonstração:(cont.)`

Supondo independência, temos

$$Var(M_r^{'}) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left[(X_i)^r\right] = \frac{1}{n^2}\sum_{i=1}^{n}\left[E(X_i)^{2r} - E^2\left(X_i^r\right)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[\mu_{2r}^{'} - (\mu_r^{'})^2\right] = \frac{1}{n}\left[\mu_{2r}^{'} - (\mu_r^{'})^2\right].$$

Momentos amostrais

Quando $r = 1$, temos o seguinte corolário.

Corolário 1 Seja $X_1, \ldots, X_n$ uma a.a com f.p/f.d $f(\cdot)$ e se $\overline{X}_n = \dfrac{1}{n} \sum_{i=1}^{n} (X_i)$ é a média amostral, então,

$$E(\overline{X}_n) = \mu, \text{ e } Var(\overline{X}_n) = \frac{\sigma^2}{n}.$$

em que μ e σ^2 são a média e a variância de $f(\cdot)$.

Momentos amostrais

- O Teorema 1 fornece a média e a variância, em termos de momentos populacionais, do r-ésimo momento amostral em relação a 0.
- Um resultado similar, porém mais complicado, pode ser derivado para a média e variância do r-ésimo momento amostral em relação a média amostral.
- Considere $r = 2$, tal que $M_2 = \frac{1}{n} \sum_i (X_i - \overline{X})^2$.
- M_2 as vezes é chamado de `variância amostral`.
- Entretanto, definiremos a variância amostral de outra forma.

Momentos amostrais

Definição 5 Seja $X_1, \dots, X_n$ uma a.a com f.p/f.d $f(\cdot)$,

$$S_n^2 = S^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2, \text{ para } n > 1,$$

é definida como **variância amostral**.

> A razão para considerarmos S^2 ao invés de M_2 como variância é devido
>
> $$E(S^2) = \sigma^2 \text{ e } E(M_2) \neq \sigma^2$$

- Revisando: $\mu_r^{'} = E(X^r)$ é o r-ésimo momento de X (em relação a 0).

Momentos amostrais

Definição 6 (Momento central) O r-ésimo momento central de uma variável aleatória X com relação ao ponto $\mathbf{a}$ é definido por

$$\mu_r = E\left[(X - \mathbf{a})^r\right]$$

- Se $\mathbf{a} = E(X) = \mu$, temos $\mu_r = E\left[(X - \mathbf{a})^r\right]$, então

$$\mu_1 = E\left[(X - \mu)^1\right] = 0 \quad \text{e} \quad \mu_2 = E\left[(X - \mu)^2\right] = Var(X) = \sigma^2.$$

Teorema 2 Seja $X_1, \ldots, X_n$ uma a.a com f.p/f.d $f(\cdot)$ e seja

$$S^2 = \frac{1}{n-1} \sum_i (X_i - \overline{X})^2,$$

Então,

$$E(S^2) = \sigma^2 \quad \text{e} \quad Var(S^2) = \frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\sigma^2\right].$$

Momentos amostrais

Prova: Para $E(S^2) = \sigma^2$, temos que $\sigma^2 = E\left[(X-\mu)^2\right]$ e $\mu_r = E\left[(X-\mu)^r\right]$. Note que,

$$
\begin{aligned}
\sum_i (X_i - \mu)^2 &= \sum_i \left(X_i + \overline{X} - \overline{X} - \mu\right)^2 \\
&= \sum_i \left[\left(X_i - \overline{X}\right)^2 - 2\left(X_i - \overline{X}\right)\left(\overline{X} - \mu\right) + \left(\overline{X} - \mu\right)^2\right] \\
&= \sum_i \left(X_i - \overline{X}\right)^2 - 2\left(\overline{X} - \mu\right)\sum_i \left(X_i - \overline{X}\right) + n\left(\overline{X} - \mu\right)^2 \\
&= \sum_i \left(X_i - \overline{X}\right)^2 - 2\left(\overline{X} - \mu\right)\underbrace{\left(n\overline{X} - n\overline{X}\right)}_{0} + n\left(\overline{X} - \mu\right)^2 \\
&= \sum_i \left(X_i - \overline{X}\right)^2 + n\left(\overline{X} - \mu\right)^2
\end{aligned}
$$

Momentos amostrais

Prova (cont.):

Assim,

$$
\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\sum_i (X_i - \overline{X})^2\right] \\
&= E\left[\frac{1}{n-1}\sum_i (X_i - \mu)^2 - n\left(\overline{X} - \mu\right)^2\right] \\
&= \frac{1}{n-1} E\left[\sum_i (X_i - \mu)^2 - n\left(\overline{X} - \mu\right)^2\right] \\
&= \frac{1}{n-1} E\left[\sum_i (X_i - \mu)^2\right] - \frac{n}{n-1} E\left[\left(\overline{X} - \mu\right)^2\right] \\
&= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1} Var(\overline{X}) \\
&= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n} \\
&= \sigma^2
\end{aligned}
$$

Momentos amostrais

- De forma similar,

$$
\begin{aligned}
E(M_2) &= E\left[\frac{1}{n}\sum_i (X_i - \overline{X})^2\right] \\
&= \frac{1}{n}\sigma^2 - \frac{n}{n} Var(\overline{X}) \\
&= \sigma^2 \left(\frac{n-1}{n}\right)
\end{aligned}
$$

- Momentos amostrais são exemplos de **exemplos de estatísticas** que podem ser usados para estimar quantidades populacionais.
- Por exemplo:
  - $M_r^{'}$ para estimar $\mu_r^{'}$;
  - $\overline{X}$ para estimar μ;
  - S^2 para estimar σ^2.

Função de verossimilhança

Função de verossimilhança

Definição 7 (Função de verossimilhança) A f.p/p.d.f conjunta é denominada função de verossimilhança de θ, correspondente a amostra observada $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ e será denotada por

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \theta) = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta).$$

Dada a amostra $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, podemos encontrar o `ponto mais verossímil` para θ.

Função de verossimilhança

Exemplo 7 (Caso discreto) Sejam $X_1, X_2, \ldots, X_n$ uma a.a de $X \sim Pois(\theta)$. Temos que a função de verossimilhança é dada por

$$
\begin{aligned}
L(\theta \mid \mathbf{x}) &= \prod_{i=1}^{n} \frac{\exp\{-\lambda\}\lambda^{x_i}}{x_i!} \\
&= \frac{\exp\{-\lambda\}\lambda^{x_1}}{x_1!} \times \frac{\exp\{-\lambda\}\lambda^{x_2}}{x_2!} \times \cdots \times \frac{\exp\{-\lambda\}\lambda^{x_n}}{x_n!} \\
&= \frac{\exp\{-n\lambda\}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Função de verossimilhança

Exemplo 8 (Caso contínuo) Sejam $X_1, X_2, \ldots, X_n$ uma a.a de $X \sim N(\mu, \sigma^2)$. Temos que a função de verossimilhança é dada por

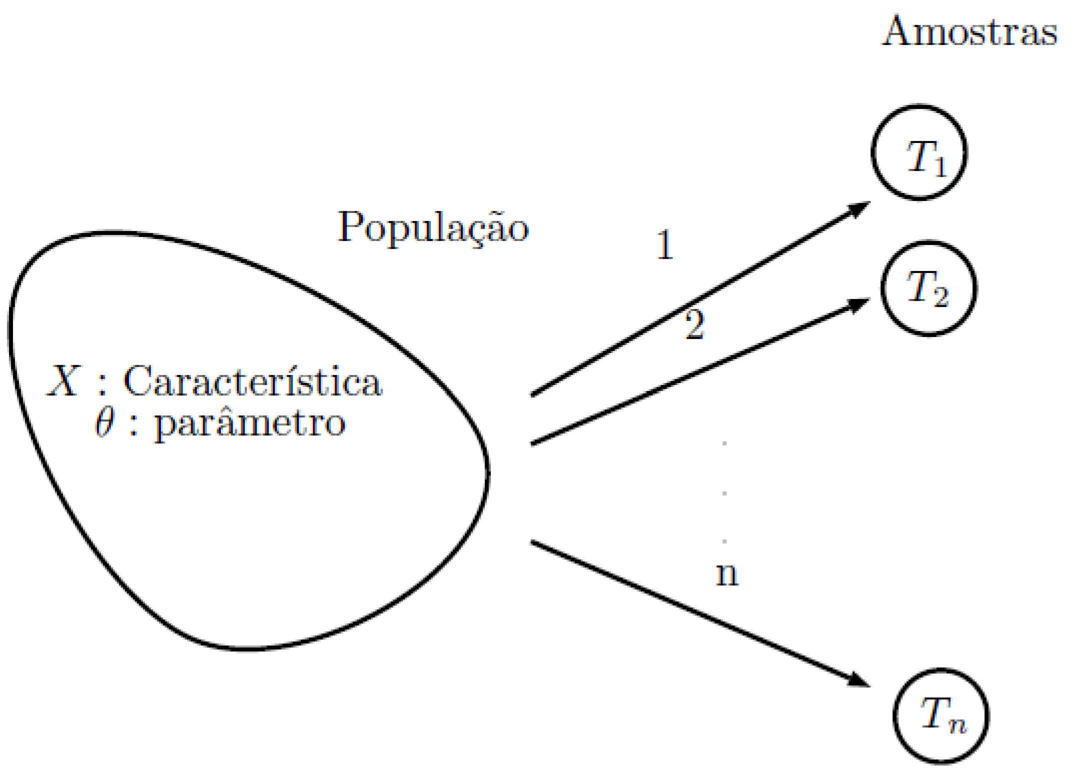
$$
\begin{aligned}
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{1}{2\sigma^2} (x_i - \mu)^2 \right\} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{1}{2\sigma^2} (x_1 - \mu)^2 \right\} \times \cdots \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{1}{2\sigma^2} (x_n - \mu)^2 \right\} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ \frac{1}{2\sigma^2} \textstyle\sum_{i=1}^{n} (x_i - \mu)^2 \right\} \\
&= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{ \frac{1}{2\sigma^2} \textstyle\sum_{i=1}^{n} (x_i - \mu)^2 \right\} \\
&= \left(2\pi\sigma^2 \right)^{-n/2} \exp\left\{ \frac{1}{2\sigma^2} \textstyle\sum_{i=1}^{n} (x_i - \mu)^2 \right\}
\end{aligned}
$$

Distribuição amostral

Introdução

Para este caso temos:

- X : Variável de interesse;
- θ : parâmetro de interesse;
- $T = f(X_1, X_2, \ldots, X_n)$ Função da amostra que vai fornecer informação sobre θ.



T é uma variável aleatória.

Pergunta: Qual a distribuição de T quando $X_1, X_2, \ldots, X_n$ assume valores observados?

Distribuição amostral da média

Distribuição amostral da média

Exemplo 9 Suponha que selecionamos todas as amostras de tamanho 2, com reposição, da população $\{1, 3, 5, 5, 7\}$

X	**1**	**3**	**5**	**7**
$P(X = x)$	1/5	1/5	2/5	1/5

- Encontrar a distribuição conjunta da v.a. $(X_1, X2)$, sendo X_1 sendo o número selecionado na primeira extração e X_2 o número da segunda.
- Encontre a distribuição de $\overline{X} = \dfrac{X_1 + X_2}{2}$.

Distribuição amostral da média

Combinação	Prob.	X_1	X_2	$\overline{X}$
$(1,1)$	1/25	1	1	1
$(1,3)$	1/25	1	3	2
$(1,5)$	2/25	1	5	3
$(1,7)$	1/25	1	7	4
$(3,1)$	1/25	3	1	2
$(3,3)$	1/25	3	3	3
$(3,5)$	2/25	3	5	4
$(3,7)$	1/25	3	7	5
$(5,1)$	2/25	5	1	3
$(5,3)$	4/25	5	3	4
$(5,5)$	2/25	5	5	5
$(5,7)$	1/25	5	7	6
$(7,1)$	1/25	5	7	4
$(7,3)$	1/25	5	7	5
$(7,5)$	1/25	5	7	6
$(7,7)$	1/25	5	7	7

`Distribuição conjunta:`

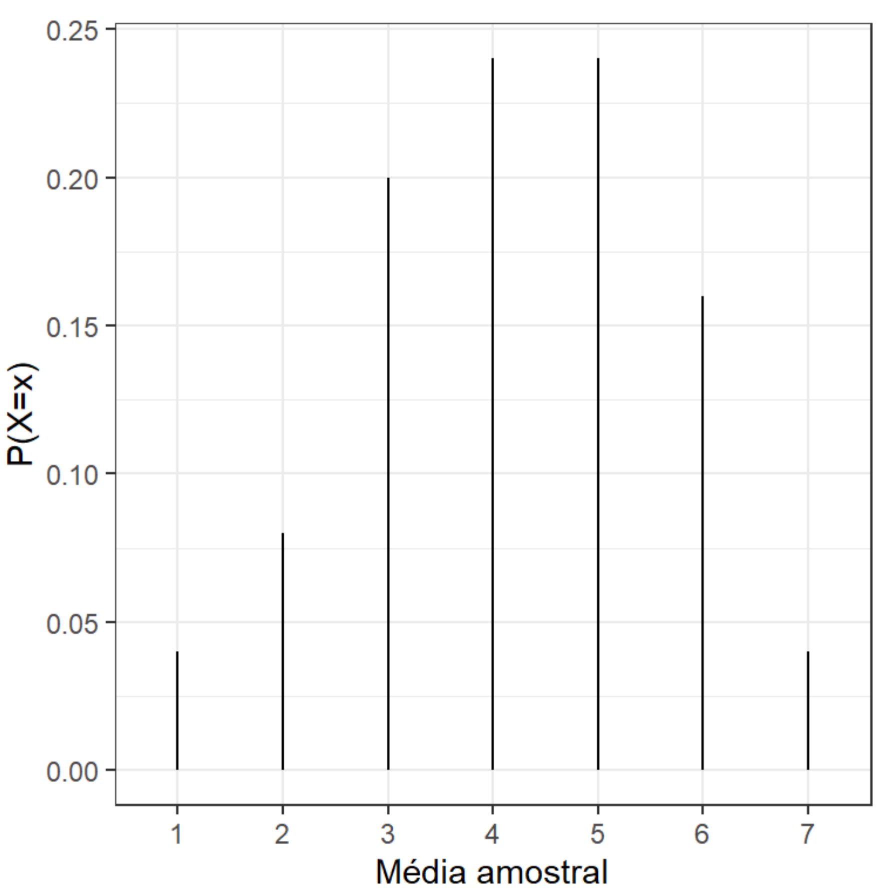
X_1/X_2	**1**	**3**	**5**	**7**	**Total**
1	1/25	1/25	2/25	1/25	1/5
3	1/25	1/25	2/25	1/25	1/5
5	2/25	2/25	4/25	2/25	2/5
7	1/25	1/25	2/25	1/25	1/5
Total	1/5	1/5	2/5	1/5	Total

Distribuição amostral da média

Distribuição amostral da média $\overline{X}$:

$\overline{X}$	1	2	3	4	5	6	7
$P(\overline{X} = \overline{x})$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{5}{25}$	$\frac{6}{25}$	$\frac{6}{25}$	$\frac{4}{25}$	$\frac{1}{25}$

```
require(ggplot2)
mx     <- c(1, 2, 3, 4, 5, 6, 7)
pmx    <-c(1/25, 2/25, 5/25, 6/25, 6/25, 4/25, 1/25)
dados <- data.frame(mx, pmx)
ggplot(data=dados, aes(x = factor(mx), ymin=0, ymax=pmx))+geom_linerange()+
  scale_x_discrete(breaks=1:7)+ylab("P(X=x)")+
  xlab("Média amostral") + theme_bw()
```



Distribuição amostral da média

- O primeiro momento amostral é a média definida como:

$$\overline{X} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i).$$

onde $X_1, X_2, \ldots, X_n$ é uma amostra aleatória com f.p/f.d $f(\cdot)$.

- $\overline{X}$ é função das v.a $X_1, X_2, \ldots, X_n$ e, portanto a distribuição pode ser encontrada teoricamente.
- Pode ser útil pensar na média amostral $\overline{X}$ como uma estimativa da média μ da f.p/f.d $f(\cdot)$ a partir de qual amostra foi selecionada.

> Um dos objetivos da amostragem é estimar μ a partir de $\overline{X}$.

Distribuição amostral da média

Teorema 3 Seja $X_1, X_2, \ldots, X_n$ uma a.a com f.p/f.d $f(\cdot)$, média μ e variância σ^2. Considere:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i) \, .$$

Então, $E(\overline{X}) = \mu_{\overline{X}} = \mu$ e $Var(\overline{X}) = \sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$.

$E(\overline{X}) = \mu$: diz que em média $\overline{X}$ é igual ao parâmetro μ sendo estimado ou que a distribuição de $\overline{X}$ está centrada em μ.

$Var(\overline{X}) = \dfrac{\sigma^2}{n}$: diz que a dispersão dos valoers de $\overline{X}$ em torno de μ é pequena para amostras grandes em comparação com tamanhos menores.

Teoremas de convergência

Teoremas de convergência

- Para amostras grandes, os valores de $\overline{X}$ (que são usados para estimar μ) tendem a estar mais concentrados de μ do que em amostras pequenas.
- Esta noção será definida pela **Lei dos Grandes Números**
- Seja $E(X) = \mu$ para a f.p/f.d $f(\cdot)$. Desejamos estimar μ.
- De maneira não rigorosa, $E(X)$ é a média de um número infinito de valores da variável aleatória X.
- Em qualquer problema real podemos observar apenas um número finito de valores da variável aleatória X.
- Questão: Usando apenas um número finito de valores de X (uma amostra aleatória de tamanho n) pode ser feita qualquer inferência confiável sobre $E(X)$? **A resposta é sim**.
- Usaremos isso através da Lei Fraca dos Grandes Números.

Teoremas de convergência

Teorema 4 (Lei fraca dos Grandes Números) Seja $X_1, X_2, \ldots, X_n$ uma a.a de tamanho n de uma população com variável X, com média $E(X) = \mu$ e $Var(X) = \sigma^2 < \infty$. Sejam, $\epsilon > 0$ e $0 < \delta < 1$. Se, $n > \dfrac{\sigma^2 \epsilon^2}{\delta}$, então,

$$P(\mid \overline{X}_n - \mu \mid < \epsilon) \geq 1 - \delta,$$

ou seja, $\overline{X}_n$ converge em probabilidade para μ.

Teorema de convergência

Seja $X_1, X_2, \ldots, X_n$ uma a.a. de $X \sim Ber(0.5)$. Observe que,

$$X_i = \begin{cases} 0, & \text{fracasso} \\ 1, & \text{sucesso} \end{cases}, i = 1, 2, \ldots, n.$$

A proporção amostral é determinada por

$$\hat{p}_n = \overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

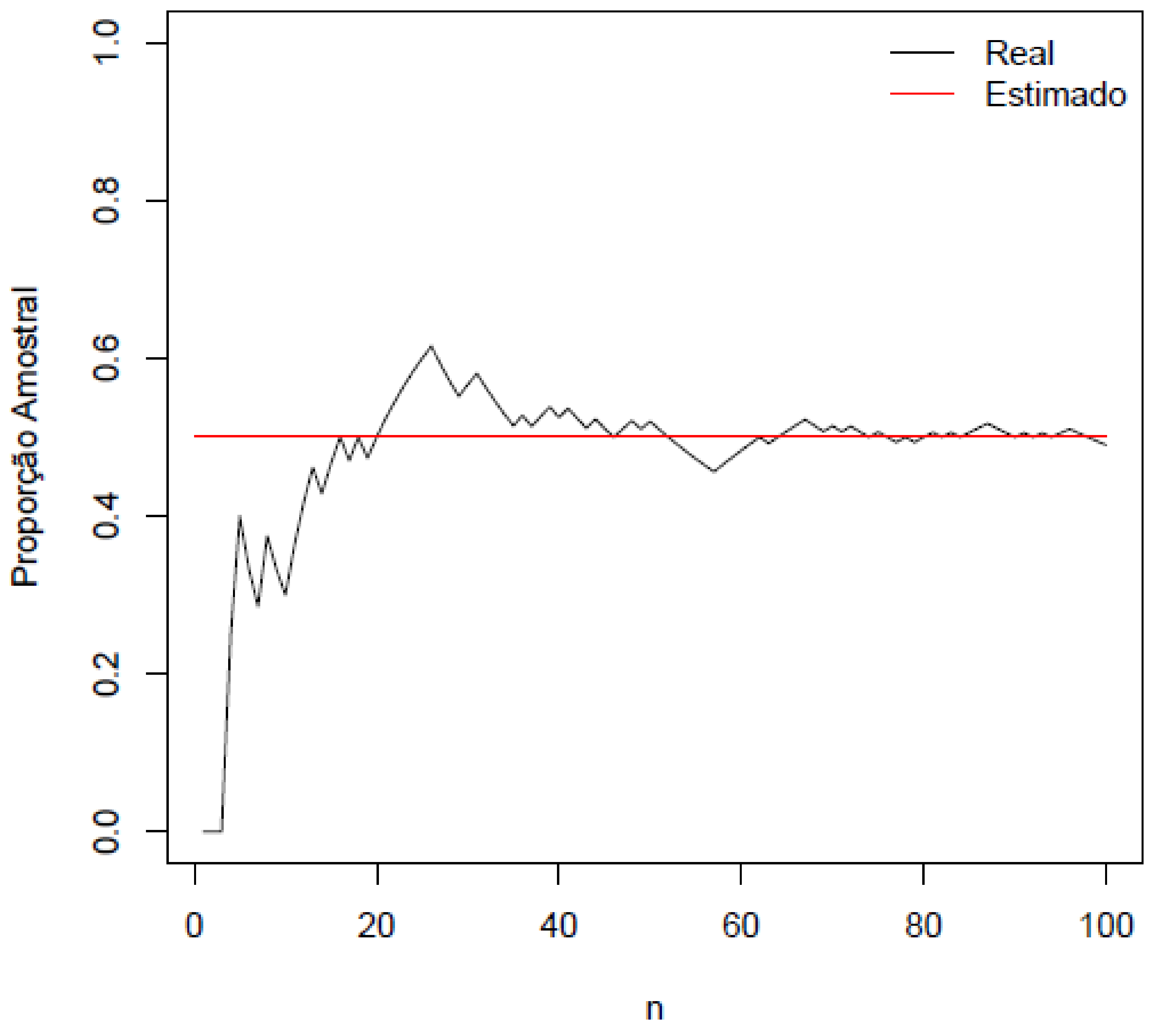
Para $n = 1 \quad \Rightarrow \quad \hat{p}_1 = \dfrac{X_1}{1}.$

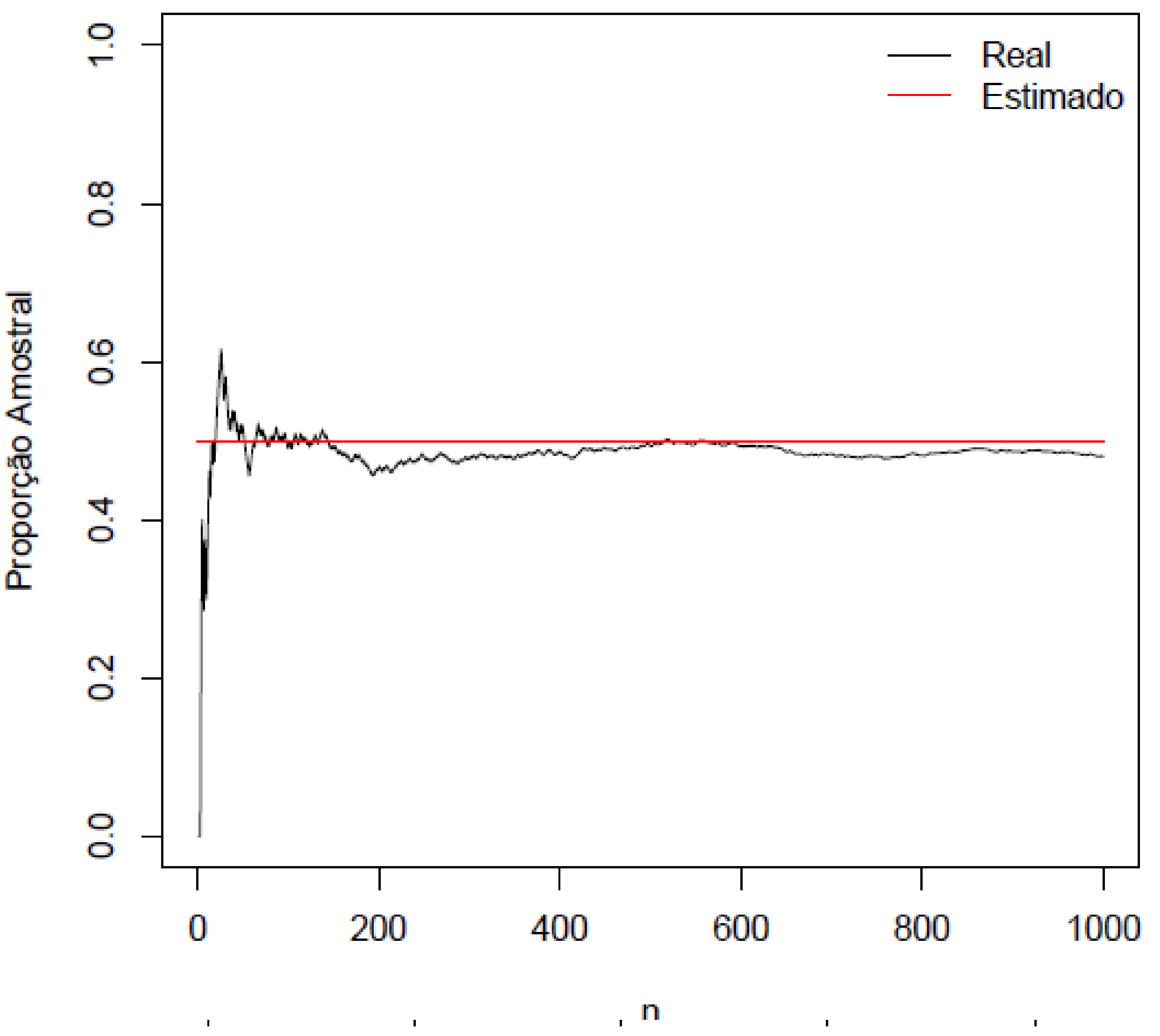
Para $n = 2 \quad \Rightarrow \quad \hat{p}_2 = \dfrac{X_1 + X_2}{2}.$

$\vdots$

Para $n = n \quad \Rightarrow \quad \hat{p}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}.$

Teorema de convergência





Teorema central do limite

> O Teorema Central do Limite é um dos mais importantes resultados em toda área de Probabilidade e Estatística. Ele nos diz aproximadamente como a **média amostral** é distribuída.

Teorema 5 (Teorema Central do Limite - TCL) Seja $X_1, X_2, \ldots, X_n$ uma sequência de v.a.'s independentes com $E(X_i) = \mu_i$ e $Var(X_i) = \sigma^2) < \infty, i = 1, 2, \ldots, n$. Tome $S_n = X_1 + X_2 + \cdots + X_n$, então, sob determinadas condições gerais,

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \to N(0, 1).$$

A distribuição de Z_n se aproxima da $N(0, 1)$ quando $n \to \infty$.

O Teorema 5 nos diz que a ditribuição limite de Z_n (S_n padronizado) será a distribuição $N(0, 1)$.

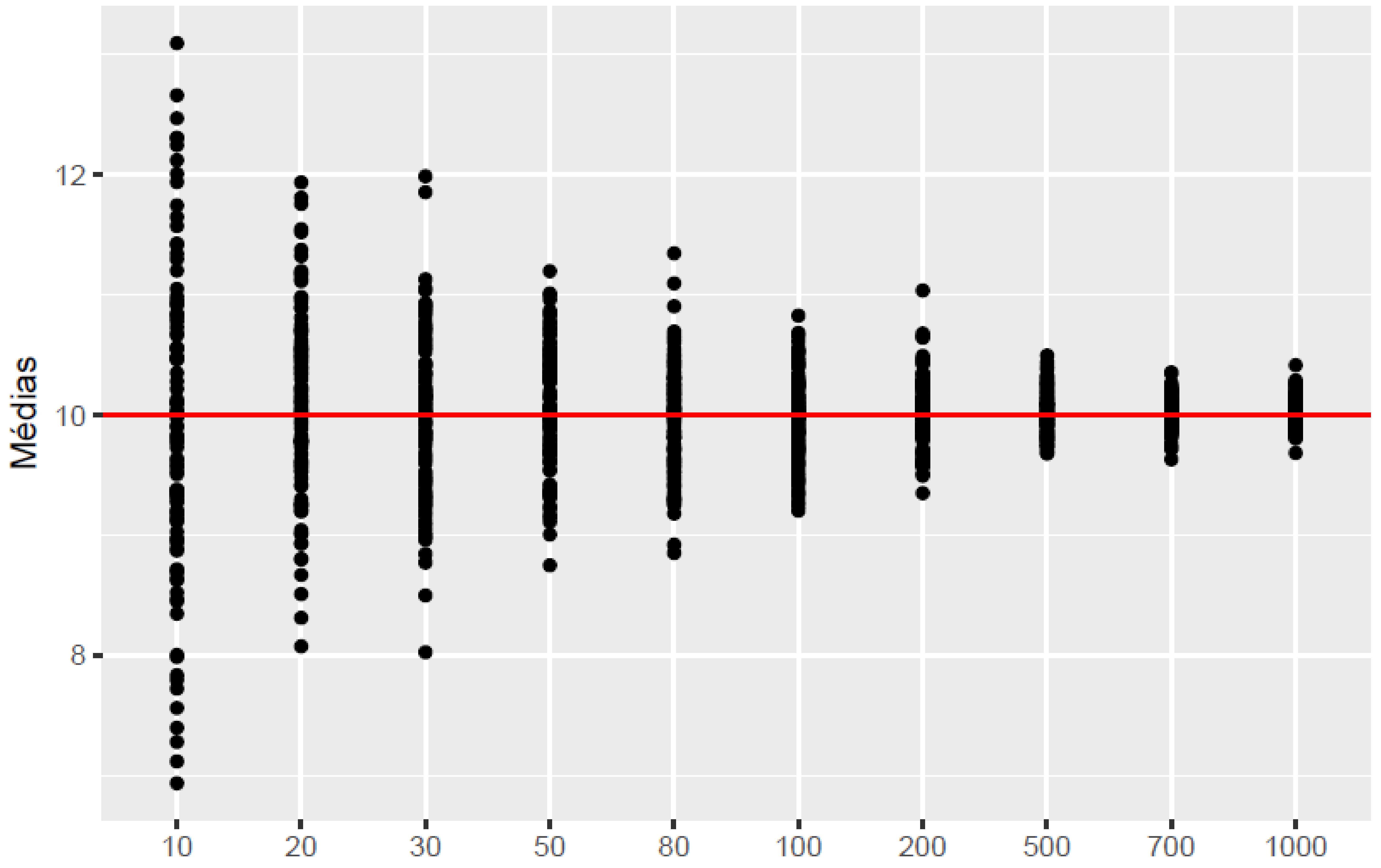
Distribuição amostral da média

Corolário 2 Seja $(X_1, X_2, \dots, X_n)$ uma a.a. de X com $E(X) = \mu$ e $Var(X) = \sigma^2 < \infty$. Então, para $n \to \infty$,

$$Z_n = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \to N(0,1).$$

- Em outras palavras $\overline{X}_n$ é assintoticamente distribuído como uma Normal com média μ e variância σ^2/n.
- Um aspecto importante sobre o Teorema 5 é o fato de que nada é dito sobre a forma da **f.p ou f.d original**. Qualquer que seja a distribuição, dado que possui variância **finita**, a média amostral terá **aproximadamente** distribuição Normal para amostras grandes.

Representação do TCL graficamente



Algumas distribuições exatas

- Se $X \sim Ber(\theta)$, então:

$$P(\bar{X} = \bar{x}_n) = \binom{n}{n\bar{x}_n}\theta^{n\bar{x}_n}(1-\theta)^{n-n\bar{x}_n}, \bar{x}_n = 0, 1/n, 2/n, \ldots, 1.$$

- Se $X \sim Pois(\theta)$, então:

$$P(\bar{X} = \bar{x}_n) = \frac{e^{-n\theta}\theta^{n\bar{x}_n}}{n\bar{x}_n!}, \bar{x}_n = 0, 1/n, 2/n, \ldots$$

- Se $X \sim Normal(\mu, \sigma^2)$, então:

$$\bar{X}_n \sim N(\mu, \sigma^2).$$

- Se $X \sim Exp(\theta)$, então:

$$\bar{X}_n \sim Gama(n, n\theta).$$

Distribuição amostral da proporção

- Seja $X_1, X_2, \ldots, X_n$ uma a.a. de $X \sim Ber(\theta)$.
- Em que θ representa a proporção de elementos com uma determinada característica na população.
- Temos que

$$E(X_i) = \theta \quad \text{e} \quad Var(X_i) = \theta(1-\theta).$$

- Seja $S_n = X_1 + X_2 + \cdots + X_n$, então a proporção amostral é definida por

$$\hat{p} = \frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \bar{X}.$$

Distribuição amostral da proporção

Distribuição exata:

- Temos que $S_n = X_1 + X_2 + \cdots + X_n \sim Bin(n, \theta)$, então

$$P\left(\hat{p} = \frac{k}{n}\right) = \binom{n}{k}\theta^k(1-\theta)^{n-k}, k = 0, 1, 2, \ldots, n.$$

Distribuição aproximada pelo TCL:

- Temos que $\hat{p} = \dfrac{S_n}{n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n} = \bar{X}_n.$, então

$$\hat{p} \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right).$$