ESCOLA
SUPERIOR
DE TECNOLOGIA
E GESTÃO

P.PORTO

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Oficina de Estatística

Curso Técnico Superior Profissional
DWDM/CRSI

Docente: Rosa Silveira

Def. Clássica de Probabilidade

- $P(A) = \frac{n_A}{N}$;
  - Definição à priori por suposição da **equiprobabilidade** dos acontecimentos
  - Lei de Laplace

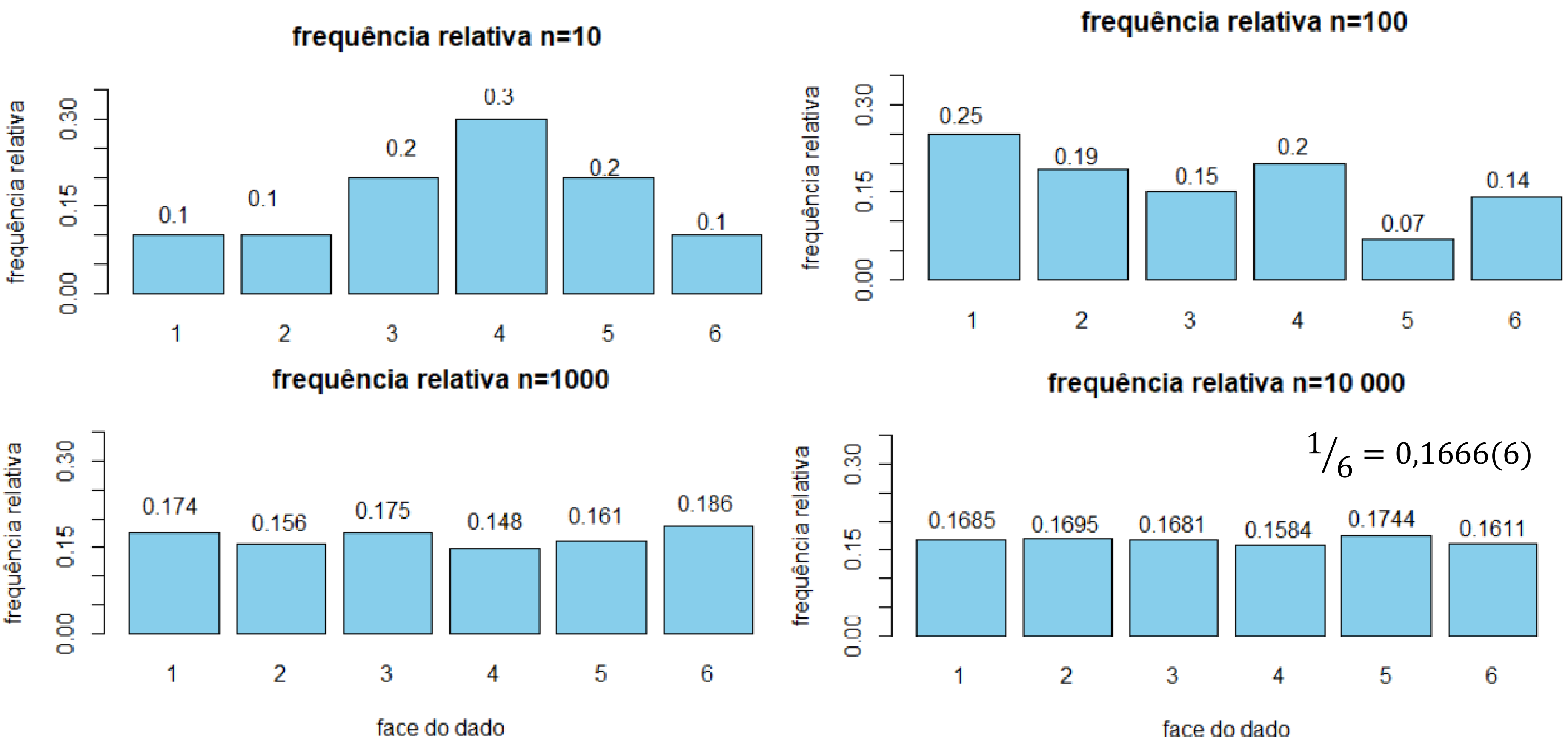
Definição frequencista de Probabilidade

- $P(A) = \lim_{n \to \infty} f_n(A)$
  - Definição à posteriori com base nos resultados observados na experiência
  - **Lei dos grandes números**

Como definir uma probabili-dade?

Exemplo: O dado está viciado?

Considere a experiência aleatória: "lançamento de um dado n vezes"



Dado equilibrado

No caso do dado ser equilibrado a distribuição de probabilidade para a Variável aleatória X, pode ser representada pela tabela:

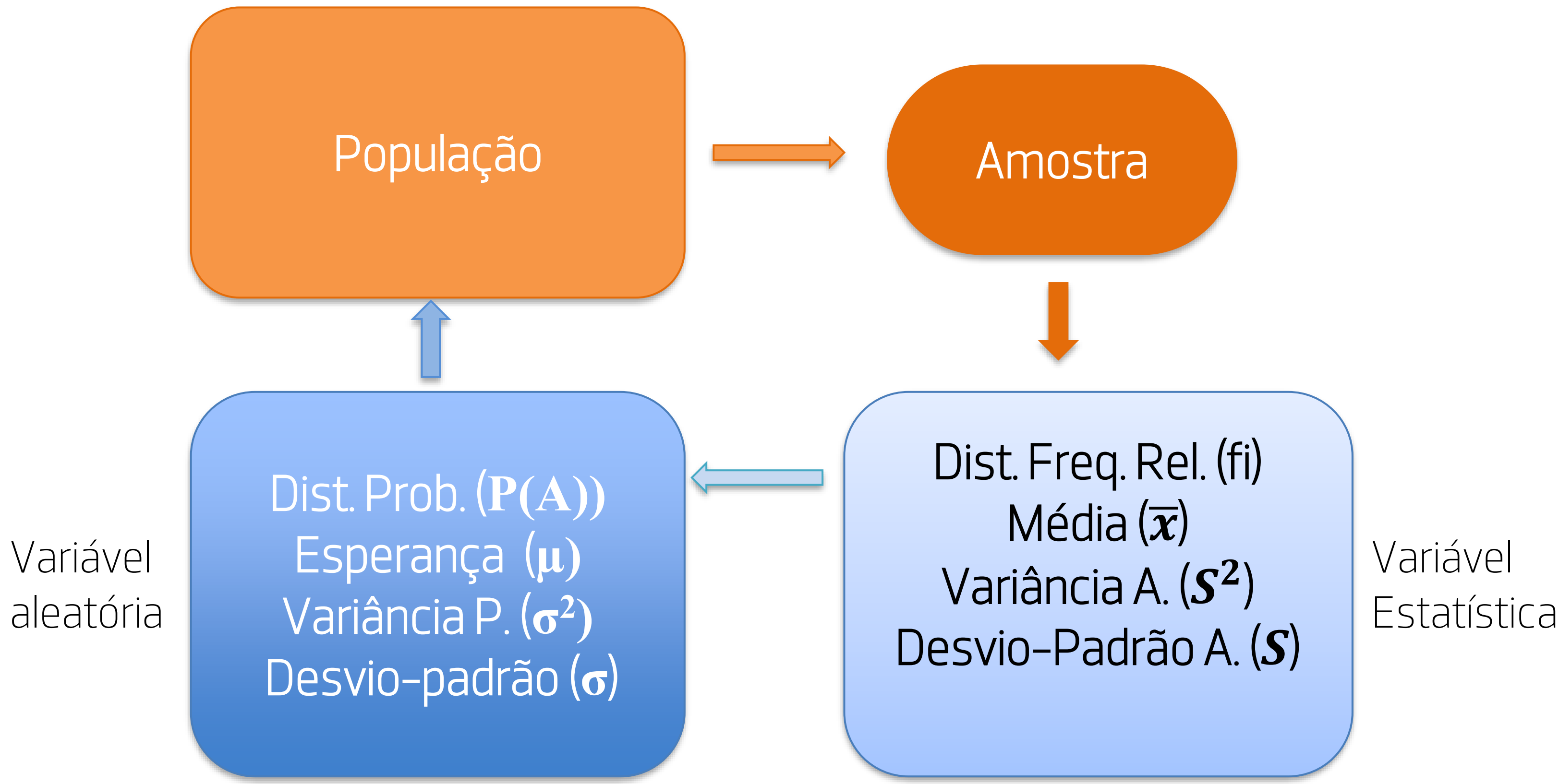
x_i	1	2	3	4	5	6
$P(X = x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

$$^1/_6 = 0{,}1666(6)$$

Neste processo **inferimos** para a população (n ~N) distribuição de probabilidade sugerida pela distribuição de frequências relativas.

Nota: Se ao fim de um elevado número de experiências o valor de alguma das faces prevalecesse, poderíamos considerar o **dado viciado**!

Do mesmo modo se estabelecem outros paralelismos:



Experiência aleatória

é um procedimento que nos leva à obtenção de resultados sujeitos ao **acaso**, dentro de um espaço de resultados possíveis – Ω

- Algumas vezes os elementos de Ω são números reais: contagens, medidas, etc,
- No entanto, existem muitos casos em que Ω não é um conjunto numérico, por exemplo, o resultado do lançamento de uma moeda ao ar.
- Quando Ω não é um conjunto numérico, atribui-se frequentemente, a cada elemento ω do espaço de resultados Ω, um número real, atribuição essa que pode ser meramente convencional.

As variáveis podem ainda classificar-se em discretas, contínuas, mistas ou outras.

Definição

*Se os valores que a variável aleatória pode tomar são em número finito ou infinito numerável a variável diz-se **discreta**.*

Exemplo

Variáveis aleatórias discretas:

- O número de pintas que sai no lançamento de um dado;
- A observação, a intervalos regulares, do número de pessoas em fila de espera na caixa de um supermercado;
- O número de riscos numa superfície;
- A proporção de peças defeituosas em cada 1 000 peças testadas;
- O número de bits recebidos com erro no envio de mensagens por canais digitais;
- A quantidade de artigos encomendados pelos clientes de uma empresa;

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Definição

*Se os valores que a variável aleatória pode tomar são um número real qualquer dentro de um determinado intervalo, ou conjunto infinito não numerável a variável diz-se **contínua**.*

Exemplo

Variáveis aleatórias contínuas:

1. O peso de um indivíduo;
2. O comprimento de uma mesa;
3. O tempo de vida de uma lâmpada;
4. O valor da corrente elétrica gerado por um painel solar;
5. A pressão sanguínea num conjunto de pacientes;
6. A temperatura nas capitais de distrito portuguesas.

Seja X uma v.a. **discreta** que assume valores distintos $x_1, x_2, \ldots, x_n$, para $n \in \mathbb{N}$.

Definição

*Designa-se por **função de probabilidade da variável aleatória** X (probability mass function) a função f que associa a cada valor particular) que a v.a. possa assumir (ou seja, a cada concretização de X), a probabilidade de X ser igual a x, ou seja,*

$$f(x) = \begin{cases} P(X = x) & \text{se} \quad x = x_i \\ 0 & \text{se} \quad x \neq x_i \end{cases}, \quad i = 1, 2, \ldots, n$$

Propriedades

Nas condições apresentadas na definição anterior tem-se:

- $0 \leq f(x_i) \leq 1$;
- $\sum_{i=1}^{n} f(x_i) = 1$.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Definição

Designa-se por ***função de distribuição*** *ou* ***função de probabilidade acumulada da v.a.*** X *(cumulative distribution probability) a função* F *que associa a cada valor* $x \in \mathbb{R}$ *a probabilidade de* X *ser menor ou igual a* x*, isto é:*

$$F(x) = P(X \leq x) = \sum_{i=1}^{k} f(x_i),$$

com x_k *o maior valor da variável aleatória* X *menor do que* x.

Propriedades

- $0 \leq F(x) \leq 1$;
- $F(x_2) \leq F(x_1), \forall x_1, x_2$ com $x_1 > x_2$ (F é uma função monótona não decrescente)
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$
- $P(x_1 < x \leq x_2) = F(x_2) - F(x_1), \forall x_1, x_2$ com $x_2 > x_1$

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Exercício

Considere X a Variável aleatória que representa o número de golos que um famoso jogador marca num jogo e a seguinte distribuição de probabilidade:

x_i	0	1	2	3
$P(X = x_i)$	0,1	0,5	?	0,1

- Indique a probabilidade de o jogador marcar 2 golos
- Determine o valor médio e interprete-o;
- Determine a probabilidade do jogador marcar mais do que um golo por jogo;
- Indique a probabilidade do jogador marcar um golo e meio;
- Defina a função de probabilidade e a função distribuição;
- Esboce, graficamente, ambas.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Variável Aleatória Contínua

- Sendo X uma v.a, **contínua**, toma valores num conjunto contínuo
- A aplicação do conceito de função de probabilidade a uma infinidade não numerável de valores leva a que

$$P(X = x) = 0, \forall x$$

  i.e., **a probabilidade pontual é sempre nula.**
- O que não implica que o acontecimento seja impossível, no entanto, sendo as possibilidades de valores num intervalo infinitas cada uma delas é nula, ou seja, é nula a probabilidade de acertar exatamente no valor $X = x$.
- Não é nula, no entanto, a função de probabilidade em todos os intervalos reais.
- Assim, para v.a. contínuas a **função densidade de probabilidade** $f = F'$, com F a função de distribuição de probabilidade da variável, representa a taxa a que a probabilidade está a aumentar num dado ponto.

Definição

Se X é uma v.a. contínua, então existe uma função $f = F' = \frac{dF}{dx}$ designada por ***função densidade de probabilidade de*** X*, tal que:*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

Teorema

Qualquer função real de variável real que verifique:

$$f(x) \geq 0 \quad \text{e} \quad \int_{-\infty}^{+\infty} f(x)\, dx = 1$$

é função densidade de probabilidade de uma dada v.a.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Distribuições importantes

Binomial

- Variável nominal ou discreta
- Número de sucessos obtidos

Poisson

- Variável nominal ou discreta
- Nº de ocorrências por unidade de tempo ou espaço

Normal

- Variável contínua

Condições para estarmos perante uma distribuição binomial:

- repetição de uma série de n provas de Bernoulli
- a variável deve ter apenas dois resultados possíveis e disjuntos
- um dos resultados é considerado como "sucesso" – A, o outro, "insucesso" – $\bar{A}$
- a probabilidade do sucesso é constante em cada tentativa $(P(A) = p)$
- as tentativas são independentes
- A v.a. representa o número de sucessos obtidos

Distribuição Binomial

A v.a. X: número de sucessos num conjunto de n **provas de Bernoulli**, tem **distribuição de Binomial**, com parâmetro n e p, onde $p = P(Sucesso)$ e $q = 1 - p = P(Insucesso))$.
Escreve-se $X \sim B(n, p)$ e a sua função de probabilidade (f.p.) é dada por:

$$f(x) = P(X = x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x} = C_x^n p^x q^{n-x}$$

- Considere que a v.a. X – número de sucessos num conjunto de n **provas de Bernoulli**, tem distribuição de Binomial, com parâmetro n e p, i.e. $X \sim B(n, p)$
- Média ou Valor Esperado: $\mu_X = E(X) = np$
- Variância: $\sigma_X^2 = V(X) = p(1-p) = npq$
- Desvio padrão: $\sigma_X = \sqrt{V(X)} = \sqrt{npq}$
- A soma de todas as probabilidades é um:
$$\sum_{x=1}^{n} P(X = x) = \sum_{x=1}^{n} C_x^n p^x q^{n-x} = (p + (1-p))^n = 1$$
- Se as v.a. X_i, $i = 1, 2, \ldots, m$, forem independentes e $X_i \sim B(n_i, p)$, então

$$\sum_{i=1}^{m} X_i \sim B\left(\sum n_i, p\right)$$

ou seja, a soma de v.a. com distribuições binomial de probabilidade de sucesso p é uma v.a. com distribuição binomial.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Exemplo

Suponhamos que as peças que saem de uma linha de produção são classificadas em defeituosas (D) e não defeituosas (N).
Admitamos que foram escolhidas ao acaso e classificadas 3 dessas peças.
Suponhamos ainda que:

- a probabilidade de uma peça ser defeituosa é 0,2 e de ser não defeituosa é 0,8;
- as probabilidades são as mesmas para cada peça;
- a classificação de cada peça em particular é independente da classificação de qualquer outra peça.

a) Indique o espaço amostral Ω desta experiência.

b) Indique as probabilidades associadas aos vários resultados de Ω.

c) Calcule a probabilidade de nenhuma peça sair defeituosa.

d) Calcule a probabilidade de apenas uma peça sair defeituosa.

e) Calcule a probabilidade de saírem defeituosas menos de duas peças.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resolução:

Consideramos cada peça defeituosa como um sucesso.

Assim, $n = 3$, $p = 0,2$ e $q = 0,8$ e a v.a. X – número de peças defeituosas segue uma distribuição Binomial $X \sim B(n, p)$, ou seja, $X \sim B(3; 0,2)$.

a) O espaço amostral desta experiência é:

$$\Omega = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}$$

b) As probabilidades associadas aos vários resultados de Ω são:

$P(DDD) = 0,2^3$;
$P(DDN) = 0,2^2 \times 0,8$;
$P(DND) = 0,2 \times 0,8 \times 0,2 = 0,2^2 \times 0,8$;
$P(NDD) = 0,8 \times 0,2 \times 0,2 = 0,2^2 \times 0,8$;
$P(NND) = 0,8 \times 0,8 \times 0,2 = 0,8^2 \times 0,2$;
$P(NDN) = 0,8 \times 0,2 \times 0,8 = 0,8^2 \times 0,2$;
$P(DNN) = 0,2 \times 0,8 \times 0,8 = 0,8^2 \times 0,2$;
$P(NNN) = 0,8^3$

Nota: O Diagrama em árvore facilita a compreensão mas nem sempre é possível enumerar todos os casos!

c) A probabilidade de nenhuma peça sair defeituosa é a probabilidade do resultado NNN, ou seja, $0,8^3 = 0,512$, que corresponde a $P(X=0)$.
Utilizando função de probabilidade da distribuição binomial:

$$f(0) = P(X=0) = \frac{3!}{0!\,(3-0)!} \times 0,2^0 \times 0,8^{3-0} = \frac{3!}{3!} \times 1 \times 0,8^3 = 0,512$$

d) A probabilidade de apenas uma peça sair defeituosa é:

$$f(1) = P(X=1) = \frac{3!}{1!\,(3-1)!} \times 0,2^1 \times 0,8^{3-1} = \frac{3!}{2!} \times 0,2 \times 0,8^2 = 0,384$$

e) A probabilidade de saírem defeituosas menos de duas peças é

$$P(X<2) = P(X=0) + P(X=1) = 0,512 + 0,384 = 0,896$$

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Para determinar as probabilidades de v.a. com distribuição binomial podemos o R:

Função (massa) de probabilidade: $P(X = x)$ `dbinom(x, n , prob sucesso)`

Função de distribuição cumulativa: $P(X \leq x)$ `pbinom(x, n, prob sucesso)`

c) A probabilidade de nenhuma peça sair defeituosa é $P(X = 0) = 0.512$:

```
> dbinom(0, 3, 0.2)
[1] 0.512
```

d) A probabilidade de apenas uma peça sair defeituosa é $P(X = 1) = 0,384$:

```
> dbinom(1, 3, 0.2)
[1] 0.384
```

e) A probabilidade de saírem defeituosas menos de duas peças é $P(X < 2) = 0,896$:

```
> dbinom(0, 3, 0.2)+dbinom(1, 3, 0.2)
[1] 0.896
```

ou

```
> pbinom(1, 3, 0.2)
[1] 0.896
```


NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Exercício

Exemplo

Uma senhora comprou 4 bolbos de narcisos, tendo-lhe o florista garantido que havia uma probabilidade de 75% de cada um florescer para a primavera seguinte.
Estude a variável aleatória X que representa o número de narcisos que a senhora irá obter.

Cofinanciado pela União Europeia

Distribuição de Poisson

A v.a. X número de sucessos num intervalo de amplitude fixa, tem **distribuição de Poisson**, com parâmetro λ e escreve-se $X \sim Poisson(\lambda)$, quando a sua função de probabilidade é dada por:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

(probabilidade de se registarem x ocorrências numa unidade de tempo)

- λ=taxa média de ocorrências (número de ocorrências por unidade de tempo)
- e = número de Neper (aprox. 2,7183)
- Média ou Valor Esperado: $\mu_X = E(X) = \lambda$
- Variância: $\sigma_X^2 = V(X) = \lambda$
- Desvio padrão: $\sigma_X = \sqrt{V(X)} = \sqrt{\lambda}$

Exemplo

Uma linha telefónica de atendimento a clientes recebe em média 4 chamadas por minuto.

a) Qual é a probabilidade de num certo minuto não surgir nenhuma chamada?

b) E qual a probabilidade de se receberem 2 chamadas num dado minuto?

c) E qual a probabilidade de se receberem 3 ou menos chamadas num dado minuto?

Resolução:

Taxa média de chegada: 4 chamadas por minuto $\rightarrow \lambda = 4$

X – número de de chamadas num minuto.

a) $P(X=0) = \frac{e^{-4}4^0}{0!} = e^{-4} \approx 0{,}0183.$

b) $P(X=2) \approx 0{,}1465.$

c) $P(X \leq 3) \approx 0{,}4335$

Aproximação da Binomial à Poisson

Como regra prática, se $n > 20$, $p < 0,1$ e $\lambda = np \leq 20$, fazemos a aproximação pela distribuição de Poisson com parâmetro np.

Exemplo

Um fabricante produz peças das quais cerca de 1 em 1000 são defeituosas, i.e. $p = 0,001$. Efetuando uma amostra de 500 peças, a probabilidade de que nenhuma peça seja defeituosa é $P(X = 0) = 0,999^{500} = 0,6064$.
Se aplicarmos a aproximação de Poisson, vem $P(X = 0) \simeq e^{-0,5} = 0,6065$.

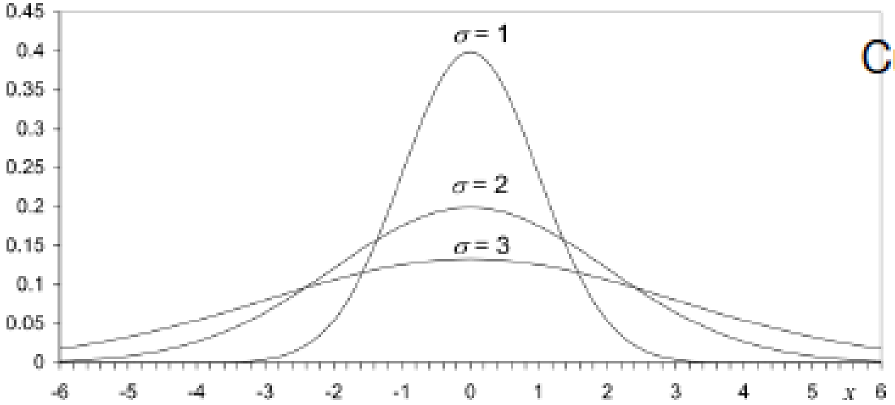
NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Distribuição Normal – *Normal Distribution*

A v.a. X que toma valores reais tem **distribuição normal** ou **gaussiana**, se a função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

onde $\mu \in \mathbb{R}$ é o valor esperado e $\sigma > 0$ é o desvio padrão.
Escrevemos $X \sim N(\mu, \sigma)$.



Curva de Gauss ou curva normal:

- se σ for grande, o gráfico tende a ser "achatado"
- se σ for pequeno, o gráfico a ser "pontiagudo"

Características da Distribuição Normal

- **Curva de Gauss**

  Os valores muito altos e muito baixos são pouco frequentes e a maior parte está distribuída em redor da média

- **Curva simétrica** em relação a $x = \mu$

  50% dos valores da variável em cada lado

  Média, moda e mediana coincidem

  atinge um máximo absoluto no ponto $x = \mu$

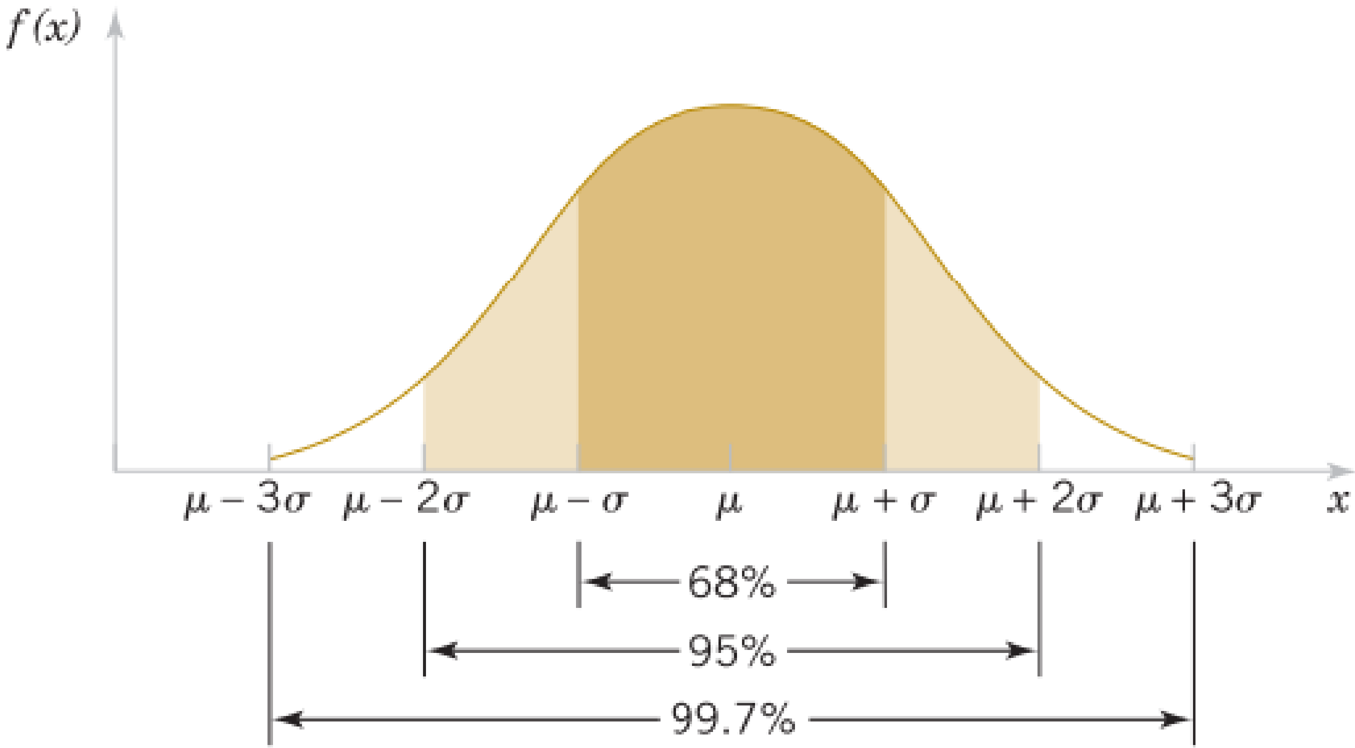
  tem dois pontos de inflexão em $x = \mu \pm \sigma$

- o eixo OX é uma assintota horizontal.

- A curva é completamente caraterizada pela **média**, que indica o centro da distribuição, e pelo **desvio padrão** (ou pela variância), que determina a dispersão da distribuição.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Áreas abaixo da curva normal



NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Exemplo

Uma biblioteca deseja estudar a circulação de documentos por utilizador e ano. Para isso escolheu uma amostra de empréstimos e registou os dados.
Representando os dados graficamente, observaram que a sua distribuição é aproximadamente Normal, com média 5 e desvio padrão 1,2.

a) Indique um intervalo para o número de documentos que são consultados por ano por 95% dos utilizadores.
b) Calcule a probabilidade de um utilizador consultar exatamente 3 documentos num ano.
c) Calcule a probabilidade de um utilizador consultar até 8 documentos num ano.
d) Calcule a probabilidade de um utilizador consultar entre 3 e 6 documentos por ano.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

a) Indique um intervalo para o número de documentos que são consultados por ano por 95% dos utilizadores.

Resolução:

$$[\mu - 2\sigma, \mu + 2\sigma] = [5 - 2 \times 1,2; 5 + 2 \times 1,2] = [2.6, 7.4]$$

Logo, 95% dos utilizadores consultam entre 2.6 e 7.4 documentos.

b) Calcule a probabilidade de um utilizador consultar exatamente 3 documentos num ano.

Resolução:

Seja X v.a.: número de documentos que um utilizador consultar num ano.

Temos que $X \sim N(\mu = 5; \sigma = 1,2)$.

Pretendemos determinar $P(X = 3)$. Como a v.a é contínua $P(X = 3) = 0$.

c) Calcule a probabilidade de um utilizador consultar até 8 documentos num ano.

Resolução:

Pretendemos determinar $P(X \leq 8)$.

0.993790335

Logo, a probabilidade de um utilizador consultar até 8 documentos num ano é de, aproximadamente, 99.38%.

d) Calcule a probabilidade de um utilizador consultar entre 3 e 6 documentos por ano.

Resolução:

Pretendemos determinar $P(3 \leq X \leq 6)$.

Temos que

$$P(3 \leq X \leq 6) = P(X \leq 6) - P(X < 3) \simeq 0.79767 - 0.04779 = 0.74988$$

Logo, a probabilidade de um utilizador consultar entre 3 e 6 documentos por ano é de, aproximadamente, 74.99%.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

- Uma v.a. que segue uma distribuição normal com $\mu = 0$ e $\sigma^2 = 1$ é uma v.a. que segue uma **distribuição normal padrão** (***standard normal distribution***) e denota-se por $Z \sim$N(0, 1).
- A função de distribuição é $\Phi(z) = P(Z \leq z)$ e temos que $\Phi(-z) = 1 - \Phi(z)$
- A função densidade de probabilidade é:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < +\infty$$

- Se $X \sim N(\mu, \sigma)$, então a v.a.

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Z é uma v.a. que segue a **distribuição normal padrão.**

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z)$$

$P(Z \leq 0)$ e $P(Z \leq 1{,}15)$

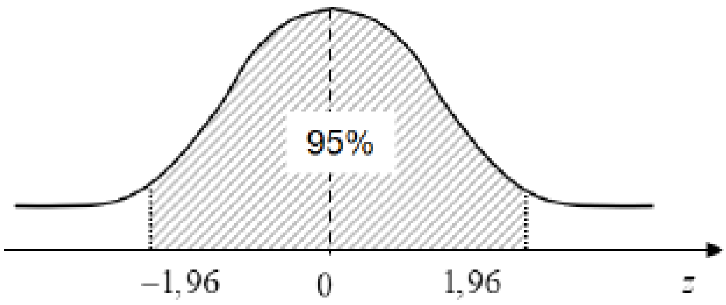
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Distribuição Normal Padrão: $Z \sim N(0, 1)$

Intervalo	Probabilidade
$\mu \pm \sigma$	0,6826
$\mu \pm 2\sigma$	0,9544
$\mu \pm 3\sigma$	0,9973
$\mu \pm 0,6745\sigma$	0,5000
$\mu \pm 1,6450\sigma$	0,9000
$\mu \pm 1,9600\sigma$	0,9500
$\mu \pm 2,5758\sigma$	0,9900

valor de *z*



- Um valor de z refere-se sempre à posição na curva de um determinado valor da v.a. Z relativamente à média
- A média, moda e mediana correspondem a $z = 0$
- O desvio padrão corresponde a $z = 1$
- Os valores de z e respetivas áreas abaixo da curva encontram-se tabelados
- Valores da tabela são cumulativos, i.e., $F(z) = P(Z <= z)$

Exemplo

A variação relativa diária da cotação de fecho de um determinado fundo pode ser razoavelmente aproximada por uma distribuição normal com valor esperado 0,2% e desvio padrão 1,6%. Calcule a probabilidade:

a) da próxima variação do preço de fecho ultrapassar 1%;

b) da próxima variação do preço de fecho se situar entre 1% e 1,16%.

Fazer no quadro.

- Função densidade de probabilidade: $P(X = x)$ — `dnorm(x,mean=0,sd=1)`
- Função de distribuição: $P(X \leq x)$ — `pnorm(x,mean=0,sd=1)`
- Quantis: $P(X \leq x) = p$ — `qnorm(p,mean=0,sd=1)`

Exemplo

Sabe-se que a carga de rutura de um tecido de algodão, X, é normalmente distribuída com $E(X) = 165$ e $V(X) = 9$.
Considere que uma amostra desse tecido é defeituosa se $P(X < 162)$.
Qual é a probabilidade de que um tecido escolhido ao acaso seja defeituoso?

Resolução:

$$P(X < 162) = P\left(\frac{X-\mu}{\sigma} < \frac{162-165}{\sqrt{9}}\right)$$
$$= \Phi(-1) = 1 - \Phi(1) \approx 0,159$$

A probabilidade de um tecido escolhido ao acaso ser defeituoso é de, aproximadamente, 15,9%.

Teorema da aditividade da distribuição normal

Se $(X_1, X_2, \ldots, X_n)$ são n v.a. independentes e se $X_i \sim N(\mu_i\ \sigma_i)$, então

$$Y = \sum_{i=1}^{n} X_i \sim N\left(\mu_Y = \sum_{i=1}^{n} \mu_i , \sigma_Y = \sqrt{\sum_{i=1}^{n} \sigma_i^2} \right).$$

Corolários

Se X_i , $i=1,\ldots,n$ são v.a. i.i.d. (independentes e identicamente distribuídas) a $X \sim N(\mu_X, \sigma_X)$, então

$$Y = \sum_{i=1}^{n} X_i \sim N\left(\mu_Y = n\mu_X , \sigma_Y = \sqrt{n\sigma_X^2} = \sigma_X \sqrt{n} \right)$$

e

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu_{\overline{X}} = \mu_X , \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \right)$$

Introdução à estimação

Definições

- **Estimador** é qualquer estatística usada para estimar o valor de um parâmetro.
- **Estimativa** de um parâmetro é qualquer valor específico de uma estatística desse parâmetro.
- **Estimação** é todo o processo que se baseia em utilizar um estimador para produzir uma estimativa do parâmetro.

Exemplos de problemas de estimação:

- Média μ de uma população
- Variância σ^2 (ou desvio padrão σ) de uma população
- Proporção p de items de uma população que pertencem a uma certa classe de interesse

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Da estimação à inferência

População

Amostra

Estatística Indutiva

Estatística Descritiva

Dist. Prob. ($\mathbf{P(A)}$)
Esperança ($\boldsymbol{\mu}$)
Variância P. ($\boldsymbol{\sigma^2}$)
Desvio-padrão ($\boldsymbol{\sigma}$)

Dist. Freq. Rel. (fi)
Média ($\bar{x}$)
Variância A. (S^2)
Desvio-Padrão A. (S)

Tipo de Estimação

- **Estimação pontual**
  fornece-nos um valor simples altamente sujeito ao erro e que não permite uma avaliação da precisão do estimador, i.e.,
  não permite o cálculo da diferença provável entre a estatística e o parâmetro.
- **Estimação intervalar**
  a qualidade de uma estimativa é definida associando-lhe um intervalo (de confiança) tendo uma probabilidade conhecida de conter o verdadeiro valor de θ.

Um intervalo de confiança pode não conter o verdadeiro valor de θ
porém, em contraste com a estimação pontual,
a probabilidade de erro para o intervalo de confiança pode ser objetivamente determinada.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Estimação Pontual

Estimadores pontuais destes parâmetros são:

- Para μ, o estimador é a **média amostral** $\hat{\mu} = \bar{x}$.
- Para σ^2, o estimador é a **variância amostral** $\hat{\sigma}^2 = s^2$.
- Para p, o estimador é a **proporção amostral** $\hat{p} = x/n$, onde x é o número de items numa amostra de dimensão n que pertencem a uma classe de interesse.

Exemplo

Suponha que uma v.a. X é **normalmente distribuída** com **média desconhecida** μ.

A média amostral é um **estimador pontual** da média populacional desconhecida, i.e., $\hat{\mu} = \bar{X}$.

Depois da amostra ter sido selecionada, o valor numérico $\bar{x}$ é um estimador pontual de μ.

Se $x_1 = 25$, $x_2 = 30$, $x_3 = 29$ and $x_4 = 31$, o **estimador pontual** de μ é

$$\bar{x} = \frac{25+30+29+31}{4} = 28.75$$

i.e. $\hat{\mu} = \bar{x} = 28.75$.

Média amostral: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Exemplo

Suponha que uma v.a. X é **normalmente distribuída** com **média desconhecida** μ.
Seja $x_1 = 25$, $x_2 = 30$, $x_3 = 29$ and $x_4 = 31$

Se a **variância populacional** σ^2 é também **desconhecida**,
um **estimador pontual** para σ^2 é a variância amostral S^2,
cujo valor numérico $s^2 = 6.9$ é calculado usando os dados da amostra.

Variância amostral corrigida:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Estimação Intervalar

- Em geral uma estimativa pontual é **insuficiente**.
- Um **intervalo de confiança** especifica sempre um **grau de confiança** (usualmente 90%, 95%, ou 99%) que corresponde a uma **medida de confiabilidade do procedimento**.

Definição

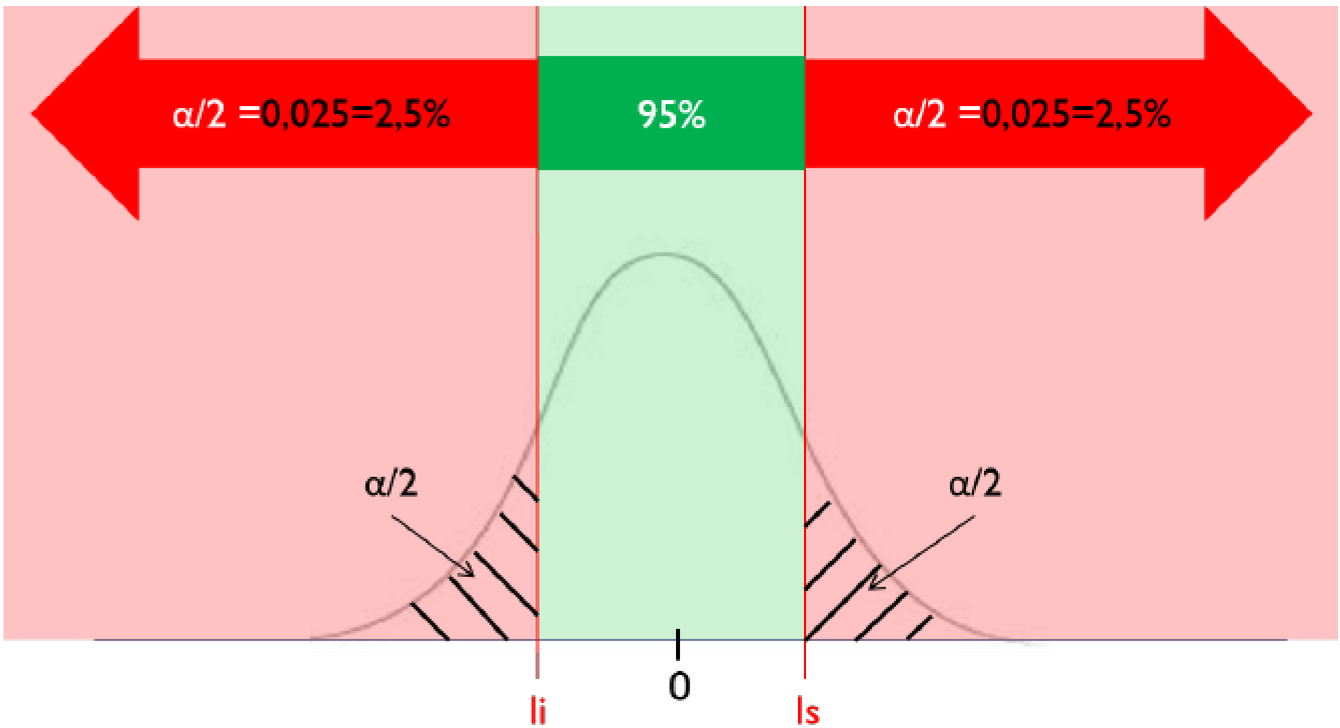
Dada uma amostra $x = (x_1, x_2, \ldots, x_n)$ *de uma v.a.* X *com função de probabilidade* $f(x|\theta)$ *e duas estatísticas* $t_1 = T_1(x_1, x_2, \ldots, x_n)$ *e* $t_2 = T_2(x_1, x_2, \ldots, x_n)$, *com* $t_1 < t_2$, *o intervalo* $]t_1, t_2[$ *tal que*

$$P(t_1 < \theta < t_2) = 1 - \alpha, \quad 0 < \alpha < 1$$

é um ***intervalo de confiança*** *contendo o verdadeiro valor do parâmetro* θ.
α *é o nível de significância e* $1 - \alpha$ *é o grau de confiança.*

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

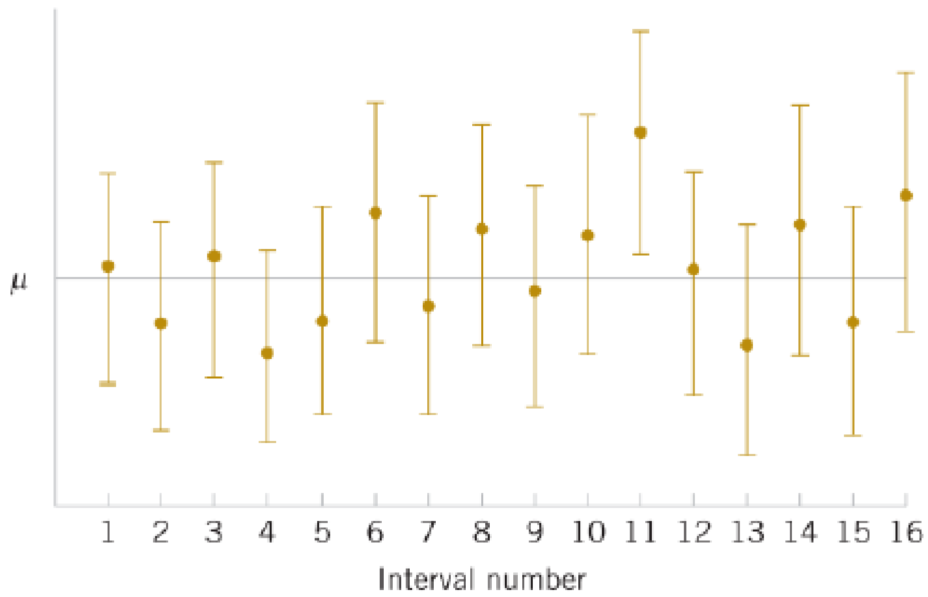
Representação gráfica



O Intervalo de confiança é uma medida de precisão da estimação;

Interpretação

Se $]T_1, T_2[$ é um IC a 95% do parâmetro θ, em 100 intervalos calculado com as mesmas estatísticas e usando 100 amostras aleatórias, 95 incluirão o valor verdadeiro do parâmetro θ.



Teorema do Limite Central

Se estamos a retirar uma amostra de uma população com **distribuição de probabilidade desconhecida** a distribuição amostral da média amostral, $\bar{X}$, terá aproximação aproximadamente normal com média μ e variância σ^2/n **se a amostra tem uma dimensão n grande**:

Teorema do Limite Central (TLC)

Se $X_1, \ldots, X_n$ é uma amostra aleatória de dimensão n extraída de uma população (finita ou infinita) com média μ e variância finita σ^2 e se $\bar{X}$ é a média amostral,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

quando $n \to \infty$, tem distribuição **normal standard (padrão)**.

Seja $x_1, x_2, \ldots, x_n$ uma amostra de uma v.a. X com média μ e desvio padrão σ.

A média aritmética é: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

A distribuição amostral de $\bar{X}$ (cujos valores são $\bar{x}$) tem a mesma média[1] que X e o desvio padrão é dado por:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \equiv \frac{\sigma}{\sqrt{n}}$$

Intervalo de Confiança para a média, com variância conhecida

Se $\bar{x}$ é a média amostral de uma amostra de dimensão n de uma **população normal** com variância conhecida σ^2, um **Intervalo de Confiança** da μ, a $100(1-\alpha)\%$, é dada por:

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

onde $z_{1-\alpha/2}$ é tal que $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$.

[1] Soma de v.a. i.i.d.

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resumo

Variáveis Aleatórias Teóricas

Discretas

- $P(X = x)$
- $P(X < x) \neq P(X \leq x)$

Contínuas

- $P(X = x) \approx 0$
- $P(X < x) = P(X \leq x)$

Resumo

Variáveis Aleatórias em R

Tipo	Variável	Densidade	Acumulada	Quantis
Discreta	$Bi(n,p)$	dbinom(x,n, p)	pbinom(x,n, p)	qbinom(x,n, p)
Discreta	$Po(\lambda)$	dpois(x, λ)	ppois(x, λ)	qpois(x, λ)
Contínua	$N(\mu,\sigma)$	dnorm(x,μ, σ)	pnorm(x,μ, σ)	qnorm(x,μ, σ)

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resumo

Estatísticas das principais distribuições

Variável	E(X) - média	V(X) - Variância	S(X) – Desvio-Padrão
$Bi(n,p)$	np	$np(1-p)$	$\sqrt{npq}$
$Po(\lambda)$	λ	λ	$\sqrt{\lambda}$
$N(\mu,\sigma)$	μ	σ^2	σ

Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resumo

Estatísticas das principais distribuições

Teorema	Variável	E(X) – média	V(X) – Variância	S(X) – Desvio-Padrão
Aditividade	$\sum N(\mu_i, \sigma_i)$	$\sum \mu_i$	$\sum {\sigma_i}^2$	$\sqrt{\sum {\sigma_i}^2}$
Corolário Soma	$\sum N(\mu, \sigma)$	$n\mu$	$n\sigma^2$	$\sqrt{n}\,\sigma$
Corolário Média	$\bar{X} = \frac{\sum N(\mu, \sigma)}{n}$	μ	$\frac{\sigma^2}{\sqrt{n}}$	$\frac{\sigma}{\sqrt{n}}$
Limite Central	$\bar{X} = \frac{\sum X_i}{n}, n > 30$	μ	$\frac{\sigma^2}{\sqrt{n}}$	$\frac{\sigma}{\sqrt{n}}$

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resumo

Objetivos da Estatística

Descritiva

- Descrever a amostra;
- Comparar grupos.

Indutiva

- Testar;
- Estimar parâmetros com base em Intervalos de Confiança;
- Prever valores com base em Modelos uni ou multivariados

NORTE 2030 Programa Regional do Norte
PORTUGAL 2030
Cofinanciado pela União Europeia

Resumo

Principais estimações para o valor esperado de uma população:

Intervalo de Confiança	Nível de Confiança	Nível de significância
$]\mu - 1{,}645\sigma, \mu + 1{,}645\sigma[$	90%	10%
$]\mu - 1{,}96\sigma, \mu + 1{,}96\sigma[$	95%	5%
$]\mu - 2{,}5758\sigma, \mu + 2{,}5758\sigma[$	99%	1%