AULA 3:
DISTRIBUIÇÕES DE PROBABILIDADES

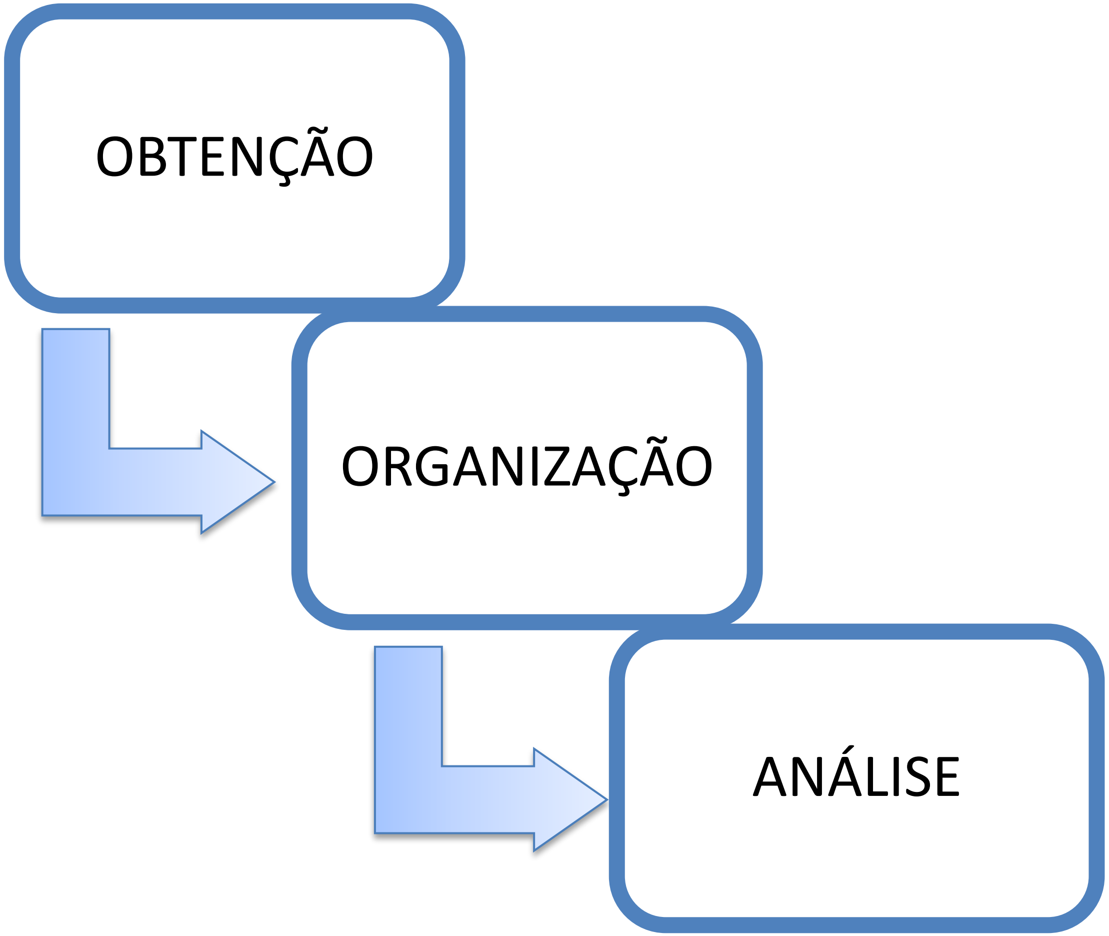
Gleici Castro Perdoná

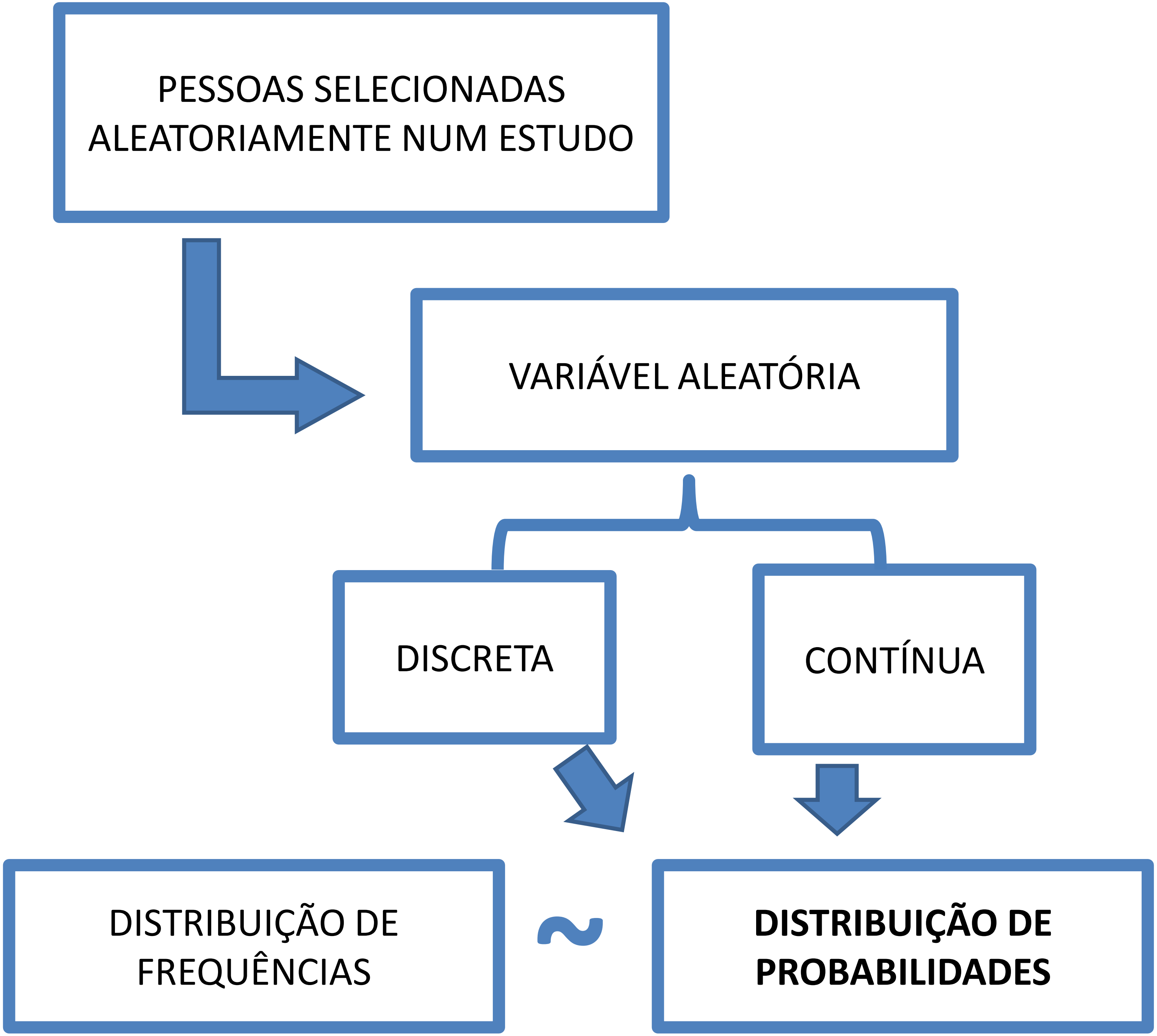
pgleici@fmrp.usp.br

AVALIAÇÃO 1 - (20 min)

- ENTRAR NO STOA E RESPONDER 5 QUESTÕES

Qual objetivo das medidas descritivas?




PESSOAS SELECIONADAS ALEATORIAMENTE NUM ESTUDO
VARIÁVEL ALEATÓRIA
DISCRETA
CONTÍNUA
DISTRIBUIÇÃO DE FREQUÊNCIAS
~
DISTRIBUIÇÃO DE PROBABILIDADES

DISTRIBUIÇÃO DISCRETA DE PROB

- Uma variável aleatória discreta é DO TIPO QUE pode assumir apenas um número contável de valores distintos, como 0,1,2,3,4, ...
- As variáveis aleatórias discretas são geralmente (mas não necessariamente) contagens.
- Se uma variável aleatória pode levar apenas um número finito de valores distintos, então deve ser discreto.
- Exemplos : número de crianças em uma família, a freqüência de sexta-feira à noite em um cinema, o número de pacientes em cirurgia de um médico, o número de lâmpadas defeituosas em uma caixa de dez ,ETC.

EXEMPLOS

- Suponha que X uma variavel pode assumiros valores 1,2,3 OU 4. e cada desfecho é descrito pela probabilidade na tabela.

Desfecho	1	2	3	4
Probabilidade	0.1	0.3	0.4	0.2

- A probabilidade de X ser 2 ou 3 é a soma das probabilidade 0.3 +0.4 =0.7.
- A probabilidade de X ser maior que 1 é 1 P(X=1)=0.9

Tipo de Distribuições

- ✓ Bernoulli
- ✓ Binomial
- ✓ Poisson
- Geométrica
- Hipergeométrica
- Multinomial
- Etc

Variável aleatória de Bernoulli (ou com distribuição de Bernouli) de parâmetro p – representa-se *X ~ Ber(p)*

X = nº de sucessos numa prova de Bernoulli, em que P(sucesso) = p (0 < p <1)

$$P(X=0)=1-p$$
$$P(X=1)=p$$

$$P(X=x)=p^{x}(1-p)^{1-x},\quad x=0,1$$

Esperança e Variância

$$E(X)=0\times(1-p)+1\times p=p$$

$$E(X^{2})=0^{2}\times(1-p)+1^{2}\times p=p$$

$$V(X)=E(X^{2})-E^{2}(X)=p-p^{2}=p(1-p)$$

Variável aleatória binomial (ou com distribuição binomial) de parâmetros n e p – representa-se *X* ~ *Bin(n,p)*

X = nº de sucessos em n provas de Bernoulli independentes e com P(sucesso) = p em cada uma (X = 0, 1, 2, ..., n)

$$P(X=x)=f_X(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x},$$

$$E(X)=\sum_{x=0}^{n} xP(X=x)=np$$

$$V(X)=\sum_{x=0}^{n} x^2P(X=x)-E^2(X)=np(1-p)$$

DISTRIBUIÇÃO DE POISSON

Limite especial da distribuição binomial: *n* aumenta e *p* diminui de modo que se torna uma constante, λ.

Condição: haver um intervalo real com ocorrências aleatórias, divisível em subintervalos, de forma que:

1. P(mais de uma ocorrência num subintervalo) = 0
2. P(ocorrência num subintervalo) = constante e ∝ comprimento do subintervalo.
3. Ocorrências em diversos subintervalos são independentes.

Processo de Poisson

Se o número médio de ocorrências no intervalo total for λ > 0 e se X – número de ocorrências no intervalo total, então X tem distribuição de Poisson com parâmetro λ.

$$P(X=x)=f_X(x;\lambda)=\frac{e^{-\lambda}\lambda^x}{x!},\quad x=0,1,2,\ldots$$

$E(X)=\lambda$ $\quad$ $Var(X)=\lambda$

Ex: nº de acidentes por semana num cruzamento ou numa secção de estrada (excetuando-se acidentes em cadeia).

Ex: nº de clientes que chegam a uma loja ou a um serviço em determinado intervalo de tempo (excetuando-se chegadas em grupo).

RESUMO

- Uma V.A discreta é DO TIPO QUE pode assumir apenas um número contável de valores distintos, como 0,1,2,3,4,....
- BERNOULLI $P(X=x)=p^{x}(1-p)^{1-x}, \quad x=0,1$
- BINOMIAL $P(X=x)=f_X(x;n,p)=\binom{n}{x}p^{x}(1-p)^{n-x},$
- POISSON $P(X=x)=f_X(x;\lambda)=\dfrac{e^{-\lambda}\lambda^{x}}{x!}, \quad x=0,1,2,\ldots$

EXEMPLO

NUMERO DE HIPERTENSOS	DIST PROB P(X=X)	DISTRIBUIÇAO FREQUENCIA (AMOSTRA)
0	0.008	0/100 = 0.000
1	0.076	9/100 = 0.090
2	0.265	24/100= 0.240
3	0.411	48/100= 0.480
4	0.24	19/100 = 0.190

UMA EMPRESA FARMACEUTICA PODE FORNECER UMA DROGA PARA 100 MÉDICOS E PEDIR A CADA UM DELES PARA TRATRAR OS PRIMEIROS 4 PACIENTES QUE NUNCA UTILIZARAM A DROGA. CADA MÉDICO, INFORMOU OS RESULTADOS A EMPRESA E OS RESULTADOS FORAM COMPARADOS COM A DISTRIBUIÇÃO TEORICA DE PROBABILIDADE

ROSNER, B.Fundamentos de bioestatistica. 8ª Edição Norte Americana, Cengage Learning, 2016.

VALOR ESPERADO E VARIANCIA

- O Valor esperado representa o valor “médio” da vam e é obtido multiplicando-se cada valor possivel por sua respectiva probabilidade e somando-se esses produtos com todos os valores que tem prob positiva.
- E(X) = 0(0.008)+1(0.076)+2(0.265)+3(0.411)+4(0.24)=μ
- Assim, seria esperado que cerca de 2,8 hipertensos fossem controlados para 4 que são tratados.

Variância

- A variância (dispersão) é obtida multiplicando-se o quadrado da distância de cada valor possível ao valor esperado por sua respectiva probabilidade e somando-se com todos os valores que possuam prob positiva.
- $VAR(X) = (0-2.8)^2\,(0.008)+(1-2.8)^2\,(0.076)+(2-2.8)^2\,(0.265)+(3-2.8)^2\,(0.411)+(4-2.8)^2\,(0.24)=0.8406$
- Também pode ser soma disso: $X^2P(X=x)-[E(X)]^2$

Variância

- A variância significa dispersão em relação a média. Mas não conseguimos interpretar o valor 0.8406 porque esta em pacientes ao quadrado.
- Já o desvio padrão (raiz (0.8406) = 0.92), podemos 0.92 pacientes em relação a média , e pela lei de Chebyshev 2.8+-2 (0.92) =(0.97 ;4,63) . 95% dos pacientes terão entre 0 e 4 episódios.

Exercício (10min)

- A probabilidade de uma mulher desenvolver cancer de mama ao longoda vida é cerca de 1/9.
- a. Qual a prob de que exatemente duas mulheres ,dentre dez desenvolvam câncer de de mama ao longo da vida?
- b. Qual a prob de que pelo menos duas mulheres, dentre dez, desenvolvam ca mama ao longo da vida?

Intervalo (10 minutos)

Exercício (10min)

- Suponha que o numero de mortes por febre tifoíde ao longo de um periodo de 1 ano tenha distribuição de poisson com parametro μ =4.6, λ =4.6 mortes por ano. Qual é a distribuição de probabiidade do numero de mortes ao longo de um periodo de 6 meses? E de 3 meses? $\mu = \lambda t$. E o valor médio ? E variancia?

Usando o R

- Distribuições de probabilidade
- dbinom(x,size,prob)
- pbinom(q, size, prob)
- qbinom(p,size,prob)
- rbinom(n, size, prob)

- dpois(x, lambda)
- ppois(q, lambda)
- qpois(p, lambda)
- rpois(n, lambda)

- **Informações sobre a distribuição binomial**

dbinom(x, size, prob)	✓ **densidade**
pbinom(q, size, prob)	✓ **probabilidade**
qbinom(p, size, prob)	✓ **quantil**
rbinom(n, size, prob)	✓ **gera valores**

b) Lançamento de um dado perfeito 10 vezes consecutivas

X - número de vezes que sai a face 6

Sucesso=sair a face 6, $p = \frac{1}{6}$

Admitindo que os lançamentos são independentes

$$X \sim Bin\left(10, \frac{1}{6}\right)$$

$$P(X=0) = \left(\frac{5}{6}\right)^{10} \qquad P(X=1) = 10 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^{9}$$

$$P(X=2) = \binom{10}{2} \times \left(\frac{1}{6}\right)^{2} \times \left(\frac{5}{6}\right)^{8} \qquad \dots\dots\dots$$

$$P(X=x) = \binom{10}{x} \times \left(\frac{1}{6}\right)^{x} \times \left(\frac{5}{6}\right)^{10-x}, \quad x = 0,1,\dots,10$$

RGui - [R Console]

File Edit Misc Packages Windows Help

```
[Previously saved workspace restored]

> dbinom(0,10,1/6)
[1] 0.1615056
>
>
>
> dbinom(1,10,1/6)
[1] 0.3230112
>
>
>
> dbinom(2,10,1/6)
[1] 0.2907100
>
>
>
> dbinom(4,10,1/6)
[1] 0.05426588
> 
```

R 1.9.0 - A Language and Environment

Iniciar | Explorando - 2005 | Microsoft PowerPo... | RGui - [R Cons... | 08:27

RGui - [R Console]

File Edit Misc Packages Windows Help

```
>
>
>
>
> pbinom(0,10,1/6)
[1] 0.1615056
>
>
>
> pbinom(1,10,1/6)
[1] 0.4845167
>
>
>
> pbinom(2,10,1/6)
[1] 0.7752268
>
> pbinom(10,10,1/6)
[1] 1
>
>
```

R 1.9.0 - A Language and Environment

RGui - [R Console]

File Edit Misc Packages Windows Help

```
>
> qbinom(0.05,10,1/6)
[1] 0
>
>
>
> qbinom(0.10,10,1/6)
[1] 0
>
>
> qbinom(0.17,10,1/6)
[1] 1
>
>
> qbinom(0.48,10,1/6)
[1] 1
> qbinom(0.49,10,1/6)
[1] 2
>
>
>
```

R 1.9.0 - A Language and Environment

Iniciar | Explorando - 2005 | Microsoft PowerPo... | RGui - [R Cons... | 08:30

Faça então ! (7 min)

- Suponha que a probabilidade de se transmitir um gene responsável por certa doença seja de **¼.** Qual a probabilidade de que, em uma família de 4 crianças, haja um filho são ou nenhum filho são? E qual a probabilidade de que haja exatamente 2 filhos doentes?

DISTRIBUIÇÃO CONTÍNUA DE PROB

- Uma variável aleatória contínua é DO TIPO QUE pode assumir valores na reta.

- Exemplos : medições de pressão, volumes, mediçoes bioquimicas séricas, escores, nível de poluição, peso, etc

- Também chamadas de pdf ou em portugues fdp (função densidade de probabilidade)

Tipo de Distribuições

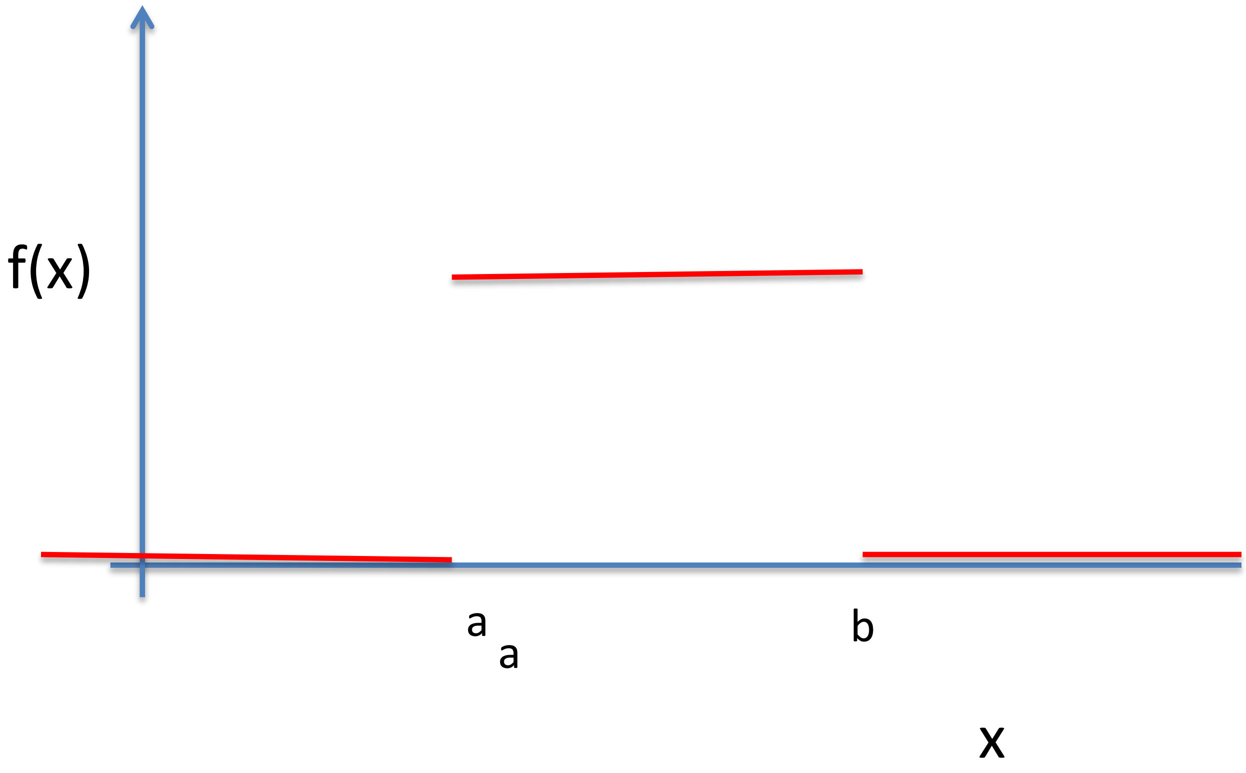
- Uniforme
- ✓ Normal
- ✓ T-Student
- ✓ Qui-quadrado
- Weibull
- Gama
- Beta
- Etc

Uniforme

- Uma variável aleatória tem distribuição Uniforme no intervalo [a,b] se sua função densidade de probabilidade for dada por:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b; \\ 0, & \text{caso contrário} \end{cases}$$

Gráfico



Calculando a probabilidade

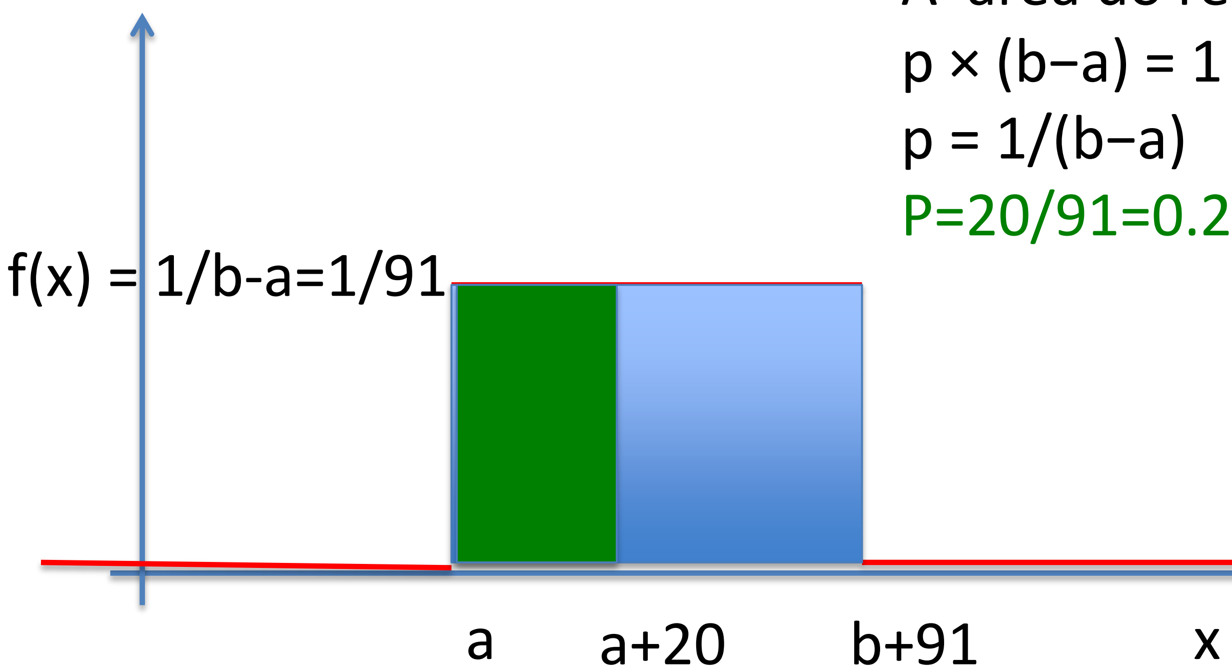
Um Gêiser entra em erupção a cada 91 minutos. Você chega lá aleatoriamente e espera por 20 minutos ... qual é a probabilidade de vê-lo entrar em erupção?

A area do retângulo = 1

$p \times (b-a) = 1$

$p = 1/(b-a)$

$P=20/91=0.22$



Mas lembre-se que isso é uma coisa aleatória! Pode entrar em erupção no momento em que você chegar, ou a qualquer momento nos 91 minutos.

DISTRIBUIÇÃO NORMAL

Mais importante distribuição de probabilidade, aplicada a inúmeros fenômenos e utilizada para o desenvolvimento teórico da Estatística.

Seja X uma variável aleatória. X terá distribuição normal se:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\sigma^2}}$$

Sendo: μ = média da distribuição

σ = desvio-padrão da distribuição

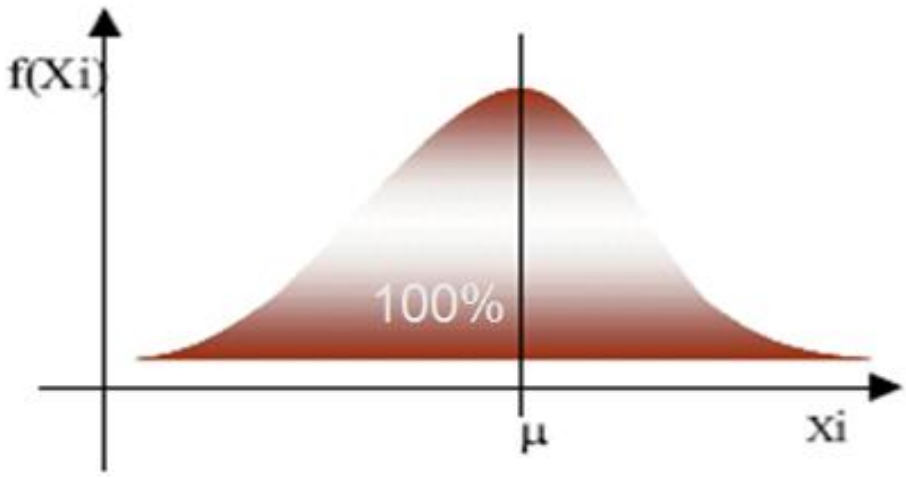
π = 3,1416

e = 2,7

e $-\infty < X < +\infty$

X tem distribuição Normal com média μ e variância σ^2 : $X \approx N(\mu, \sigma^2)$.

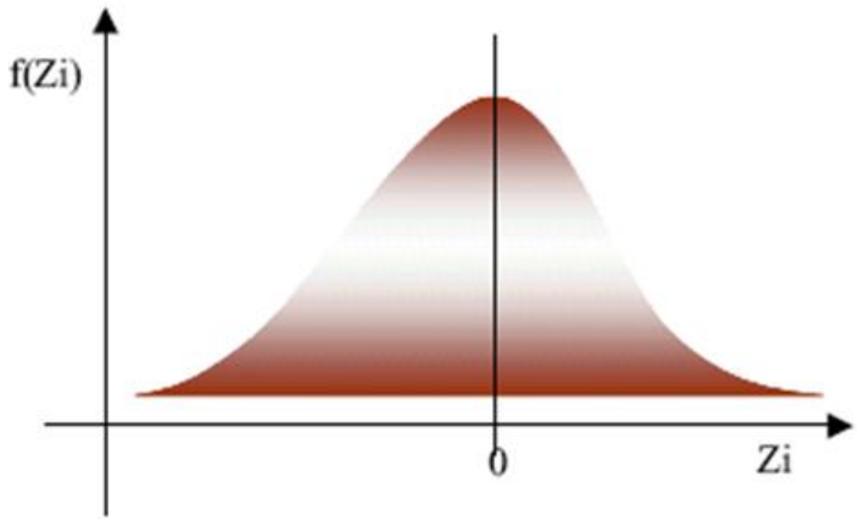
PROPRIEDADES DA DISTRIBUIÇÃO NORMAL



- É simétrica em relação à origem X = μ.
- Possui o máximo para X = μ.
- Tende a 0 quando X tende a + ∞ ou – ∞ (X é assíntota de f(X)).
- Tem dois pontos de inflexão em X = μ – σ e X = μ + σ

  - $\mu \pm 1\ \sigma$ contém 68,27% da área sob a curva;
  - $\mu\ \pm 2\sigma$ contém 95,45% da área soba a curva;
  - $\mu\ \pm 3\sigma$ contém 99,73% da área sob a curva.

PROPRIEDADES DA DISTRIBUIÇÃO NORMAL PADRÃO



- É simétrica em relação à origem Z = 0.
- Possui o máximo para X = 0, onde a ordenada vale 0,39.
- Tende a zero quando Z tende para –∞ ou + ∞ (Z é assíntota de f(Z)).
- Tem dois pontos de inflexão em Z = – 1 e 1.

Existem tabelas que retornam as probabilidades ou funções em softwares.

A DISTRIBUIÇÃO NORMAL PADRÃO

Para calcular probabilidades $P(X \leq x)$, é necessário integrar f(X) para diferentes valores de μ e σ. A solução é transformar a variável X na variável Z, dada por:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

A distribuição de probabilidades associada à variável Z denomina-se distribuição normal padrão:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

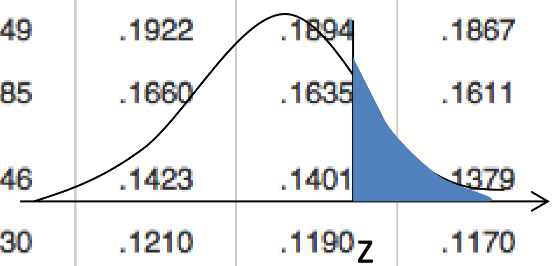
Sendo: média da distribuição = 0

variância = 1

e $-\infty < Z < +\infty$

Z tem distribuição Normal com média 0 e variância 1: $Z \approx N(0,1)$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294



Treinando (7 min)

Dê as probabilidades abaixo.

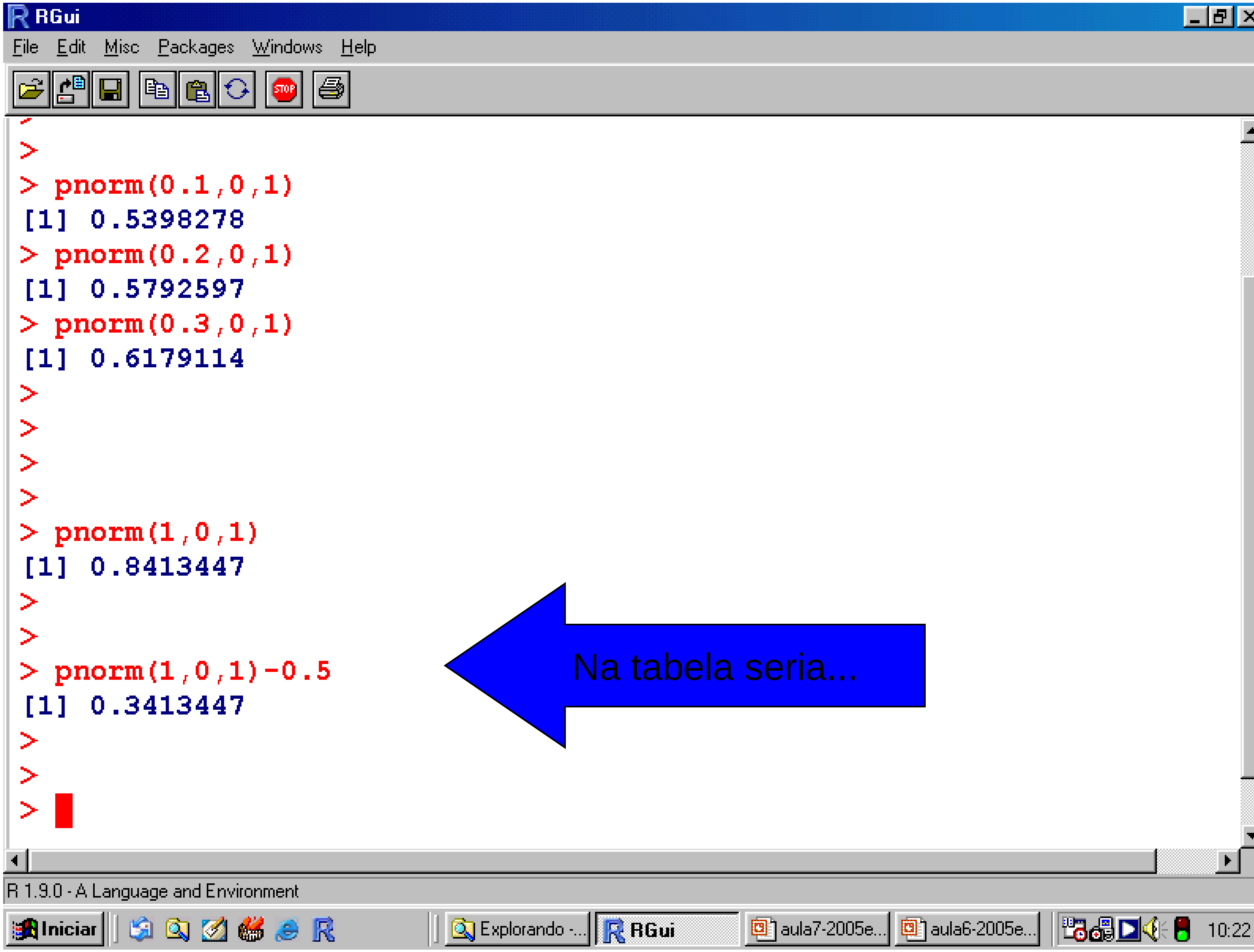
- $P(Z<1,8)$

Obs: probabilidades da cauda superior da curva.

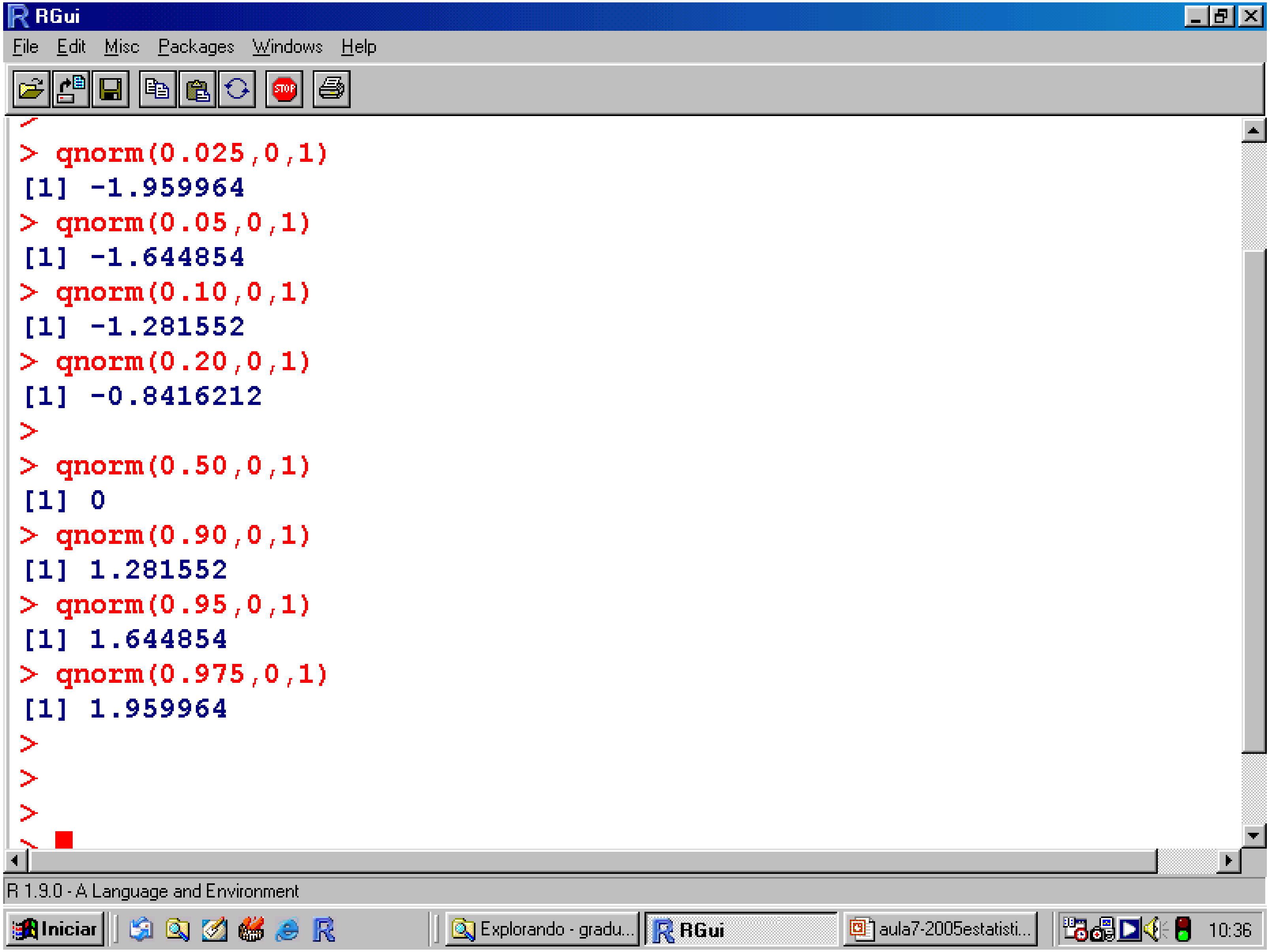
Disponível em: http://www.normaltable.com/ztable-righttailed.html

Intervalo (10 minutos)

➢ O R informa a probabilidade acumulada P(Z<z)



➤ O R o quantil de interesse



T -Student

df	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	1.376	3.078	6.314	12.71	31.82	63.66
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977

Treinando (3 min)

Para probabilidade 5% na cauda direita, encontre o valor de t

- Para n= 12
- Para n=5
- Para n =30

Para probabilidade 2.5% na cauda direita, encontre o valor de t

- Para n= 12
- Para n=5
- Para n =30

Na prática como funciona

1. As alturas dos alunos de determinada escola são normalmente distribuídas com média 1,60 m e desvio-padrão 0,30 m. Encontre a probabilidade de um aluno medir:

a) entre 1,50 m e 1,80 m – ;
b) mais de 1,75 m
c) menos de 1,48 m –

Qual intervalo compreende 90% da população? –

2. Sabe-se que o tempo gasto no exame de um paciente tem distribuição aproximadamente Normal, com média 30 min e desvio padrão de 5 min.
a. Sorteando-se um médico residente ao acaso, qual é a probabilidade dele terminar o exame antes de 24 minutos?

b. Qual deve ser o tempo de exame, de modo a permitir que 95% dos residentes terminem no prazo estipulado?

c. Qual é o intervalo de tempo, simétrico em torno da média tal que 80% dos residentes gastam para completar o exame?

BIBLIOGRAFIA RECOMENDADA

BUSSAB, W.O.; MORETTIN, P**. Estatística básica.** 4 ed. São Paulo, Atual, 1987.

PAGANO, M. e GAUVREAU, K. Princípios de Bioestatística - Tradução da 2ª Edição Norte Americana, Pioneira Thonpson Learning, São Paulo, SP,2004.

MEDRONHO R; CARVALHO DM; BlOCH KV; LUIZ RR; WERNECK GL. Epidemiologia.
Atheneu, 2 ed. São Paulo, 2008

ROSNER, B.Fundamentos de bioestatistica**.** 8ª Edição Norte Americana, Cengage Learning, 2016.

OBRIGADA

ATÉ SEXTA, DIA 25 DE AGOSTO