

BIOESTATÍSTICA

AULA 3

Anderson Castro Soares de Oliveira
Jose Nilton da Cruz

Departamento de Estatística/ICET/UFMT

Probabilidade

PROBABILIDADE

- Probabilidade é o ramo da matemática que estuda fenômenos aleatórios.
- Probabilidade é uma medida que quantifica a sua incerteza frente a um possível acontecimento futuro.
- Probabilidade está associada a estatística, porque sua teoria constitui a base da inferência estatística.
- Conceito de Probabilidade esbarra no conceito da palavra aleatório

Determinístico vs Aleatório

- Experimento aleatório é aquele que, se repetido sobre as mesmas condições, não produz necessariamente o mesmo resultado.
 - Podem ser repetidos indefinidamente sob as mesmas condições.
 - Não se pode adiantar um resultado particular, mas pode-se descrever todos os resultados possíveis
 - Se repetidos muitas vezes apresentarão uma regularidade em termos de frequência de resultados.

Espaço Amostral

- Espaço amostral (Ω) - é o conjunto de todos os possíveis resultados de um experimento.
- Exemplo:
 - Lançamento de um dado não viciado. Neste caso o espaço amostral é

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Lançar uma moeda duas vezes e observar as faces obtidas

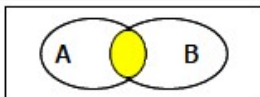
$$\Omega = \{(Ca, Co), (Ca, Ca), (Co, Ca), (Co, Co)\}$$

Eventos

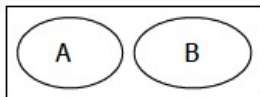
- Evento - é um subconjunto do espaço amostral que representa um resultado definido.
- No lançamento de um dado pode-se interessar, por exemplo, somente na ocorrência de número ímpares.
- O subconjunto $A = \{1, 3, 5\}$ do espaço amostral Ω representa o evento A definido pela ocorrência de números ímpares.

Operação com eventos

- Sejam A e B dois eventos de um mesmo espaço amostral.
- O evento intersecção de A e B, denotado $A \cap B$, e o evento em que A e B ocorrem simultaneamente.

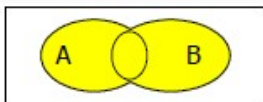


- Dois eventos A e B são mutuamente exclusivos ou disjuntos se eles não podem ocorrer simultaneamente $A \cap B = \emptyset$.

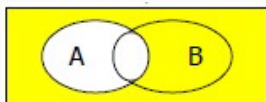


Operação com eventos

- O evento União de A e B, denotado $A \cup B$, é o evento em que A ocorre ou B ocorre (ou ambos).



- O evento complementar de A, denotado A^c , é o evento em que A não ocorre.



PROBABILIDADE

Operação com eventos

- Exemplo: Seja o espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$ e considere os eventos:

$$A = \{1, 3, 5\} \quad B = \{2, 4, 6\} \quad C = \{3, 4, 5, 6\}$$

- Fazendo as operações abaixo temos:

$$A \cap B = \emptyset \quad \text{Conjuntos mutuamente exclusivos ou disjunto}$$

$$A \cap C = \{3, 5\}$$

$$A \cup B = \{1, 2, 3, 4, 5, 6\} = \Omega$$

$$A \cap B^c = \{1, 3, 5\} = A \quad \text{os elementos de } \Omega \text{ que não estão no conj}$$

PROBABILIDADE

- Probabilidade - frequência relativa associada a um variável descritora de uma população.
- Num espaço amostral Ω , a probabilidade de ocorrer um evento A , representado por $P(A)$, é dado pela medida de A em Ω :

$$P(A) = \frac{n(A)}{n(\Omega)}$$

- Exemplo: A probabilidade de ocorrer face ímpar no lançamento de um dado não viciado é

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{3}{6} = \frac{1}{2} = 0,5 = 50\%$$

PROBABILIDADE

- A probabilidade de ocorrência de Ω vale 1, ou seja, $P(\Omega) = 1$
- Probabilidade de em evento certo e de um evento impossível

$$P(\Omega) = 1; \quad P(\emptyset) = 0$$

- A probabilidade de ocorrência do evento A é não negativa, ou seja, $P(A) \geq 0$
- Domínio da Probabilidade

$$0 \leq P(A) \leq 1$$

Propriedades de Probabilidade

PROPRIEDADES DE PROBABILIDADE

- Regra da Adição de probabilidades de dois eventos A e B :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- No exemplo do lançamento de um dado seja os eventos $A = \{2, 4, 6\}$ e $B = \{3, 4, 5, 6\}$.
- A união entre os dois conjuntos daria $\{2, 3, 4, 5, 6\}$. Assim:

$$P(A \cup B) = \frac{5}{6} = 0,83 = 83\%$$

- Utilizando a regra da adição teríamos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6} = 0,83 = 83\%$$

em que $A \cap B = \{4, 6\}$

PROPRIEDADES DE PROBABILIDADE

- Dados dois eventos A e B , a probabilidade condicional de A dado que ocorreu B é representada por $P(A|B)$ e definida por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0.$$

- Sempre que calculamos $P(A|B)$, estamos essencialmente calculando $P(A)$ em relação ao espaço amostral reduzido devido a B ter ocorrido, em lugar de fazê-lo em relação ao espaço amostral original Ω .

PROPRIEDADES DE PROBABILIDADE

- Dados dois eventos A e B , a probabilidade condicional de A dado que ocorreu B é representada por $P(A|B)$ e definida por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0.$$

- Sempre que calculamos $P(A|B)$, estamos essencialmente calculando $P(A)$ em relação ao espaço amostral reduzido devido a B ter ocorrido, em lugar de fazê-lo em relação ao espaço amostral original Ω .

PROPRIEDADES DE PROBABILIDADE

- Das expressões da probabilidade condicional resulta a **regra do produto**, que se refere ao cálculo da probabilidade do evento interseção,

$$P(A \cap B) = P(A|B).P(B)$$

- Dois eventos A e B são **independentes** se a ocorrência de um não altera a probabilidade de ocorrência do outro, isto é, $P(A|B) = P(A)$ ou $P(B|A) = P(B)$, ou ainda, a seguinte forma equivalente:

$$P(A \cap B) = P(A).P(B)$$

PROPRIEDADES DE PROBABILIDADE

Exemplo

- A tabela a seguir mostra a relação entre a aplicação de HCG e o diagnóstico de gestação em éguas.

HCG	Diagnóstico de gestação.		Total
	Sim	Não	
Sim	22	12	34
Não	43	17	60
Total	65	29	94

- Nesse caso temos 4 eventos possíveis:
 1. HCG^+ a égua recebeu recebido HCG
 2. HCG^- a égua não recebeu recebido HCG
 3. G^+ a égua teve diagnostico positivo para gestação
 4. G^- a égua teve diagnostico negativo para gestação

PROPRIEDADES DE PROBABILIDADE

Exemplo

- Qual a probabilidade de uma égua ter recebido HCG?

$$P(HCG^+) = \frac{34}{94} = 0,3617$$

- Qual a probabilidade de uma égua ter diagnóstico positivo e ter recebido HCG?

$$P(G^+ \cap HCG^+) = \frac{22}{94} = 0,2340$$

- Qual a probabilidade de uma égua ter diagnóstico positivo dado que recebeu HCG?

$$P(G^+ | HCG^+) = \frac{P(G^+ \cap HCG^+)}{P(HCG^+)} = \frac{\frac{22}{94}}{\frac{34}{94}} = \frac{22}{34} = 0,6470$$

Testes Diagnósticos

TESTES DIAGNÓSTICOS

- A Probabilidade condicional é muito utilizada em testes diagnósticos
- Diagnóstico é parte essencial na prática clínica, e muitas pesquisas médicas têm por objetivo melhorar os métodos de diagnóstico.
- A questão de interesse é quão bom um particular teste de diagnóstico pode ser.
- A qualidade de um teste diagnóstico pode ser expressa em termos de probabilidade, estudando os conceitos de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo de um teste.

TESTES DIAGNÓSTICOS

- Considerando os seguintes eventos
 - O paciente está doente
 - O paciente não está doente (saudável)
 - O paciente tem o paciente tem resultado positivo no teste de diagnóstico
 - O paciente tem o paciente tem resultado negativo no teste de diagnóstico

RESULTADO	DOENTE	SAUDAVEL	TOTAL
POSITIVO	VERDADEIRO	FALSO	a+b
	POSITIVO (a)	POSITIVO (b)	
NEGATIVO	FALSO	VERDADEIRO	c+d
	NEGATIVO (c)	NEGATIVO (d)	
TOTAL	a+c	b+d	n=a+b+c+d

TESTES DIAGNÓSTICOS

- A sensibilidade de um teste é a proporção de resultados positivos identificados, entre todos os doentes.

$$\text{Sensibilidade} = P(\text{Positivo}|\text{Doente}) = \frac{a}{a + c}$$

- A especificidade de um teste é a proporção de resultados negativos, entre os não doentes.

$$\text{Especificidade} = P(\text{Negativo}|\text{Saudavel}) = \frac{d}{b + d}$$

TESTES DIAGNÓSTICOS

- O valor preditivo positivo (VPP) de um teste é a proporção dos pacientes que têm a doença entre os que apresentam resultado positivo no teste

$$VPP = P(Doente|Positivo) = \frac{a}{a + b}$$

- O valor preditivo negativo (VPN) de um teste é a proporção dos pacientes que não têm a doença entre os que apresentam resultado negativo no teste

$$VPN = P(Saudavel|Negativo) = \frac{d}{c + d}$$

TESTES DIAGNÓSTICOS

- Os valores preditivos (VP) dependem da prevalência da doença na população de estudo
- O VPP aumenta com a prevalência enquanto os VPN diminuem.
- Quando a doença é rara (baixas prevalências) o VPP é baixo e o VPN é alto, assim

$$VPP = \frac{sen \times p}{sen \times p + (1 - esp)(1 - p)}$$

$$VPN = \frac{sen(1 - p)}{esp(1 - p) + (1 - sen)p}$$

TESTES DIAGNÓSTICOS

- Acurácia é a proporção de acertos, ou seja, o total de verdadeiramente positivos e verdadeiramente negativos, em relação a amostra estudada.

$$\begin{aligned}\text{Acurácia} &= P(\text{Verdadeiro Positivo} \cup \text{Verdadeiro Negativo}) \\ &= \frac{a + d}{n}\end{aligned}$$

TESTES DIAGNÓSTICOS

- Razão de Verossimilhança de um resultado de Teste Positivo (RVP):

$$RVP = \frac{\text{sensibilidade}}{1 - \text{especificidade}} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

- Razão de Verossimilhança de um resultado de Teste Negativo (RVN):

$$RVP = \frac{1 - \text{sensibilidade}}{\text{especificidade}} = \frac{\frac{c}{a+c}}{\frac{d}{b+d}}$$

TESTES DIAGNÓSTICOS

- As razões de verossimilhança são uma forma de descrever o desempenho de um teste diagnóstico.
- As razões de verossimilhança expressam quantas vezes é mais provável (ou menos) encontrar um resultado de um teste em pessoas doentes comparadas com as não doentes
- Quanto maior a RVP, melhor o teste. Para ser um bom teste a RVP deve ser maior que 1.
- Quanto menor a RVN, melhor o teste

TESTES DIAGNÓSTICOS

Exemplo

- Um novo teste clínico é usado para diagnosticar uma doença. Os resultados do estudo de 344 indivíduos estão resumidos na tabela

Resultado do teste	Estado do indivíduo		Total
	Doente	Saudavel	
Positivo	231	32	263
Negativo	27	54	81
Total	258	86	344

TESTES DIAGNÓSTICOS

Exemplo

- Assim:

$$\text{Sensibilidade} = P(\text{Positivo}|\text{Doente}) = \frac{231}{258} = 0,8953$$

$$\text{Especificidade} = P(\text{Negativo}|\text{Saudavel}) = \frac{54}{86} = 0,6279$$

$$VPP = P(\text{Doente}|\text{Positivo}) = \frac{231}{263} = 0,8783$$

$$VPN = P(\text{Saudavel}|\text{Negativo}) = \frac{54}{81} = 0,6667$$

TESTES DIAGNÓSTICOS

Exemplo

- Assim:

$$\begin{aligned}\text{Acurácia} &= P(\text{Verdadeiro Positivo} \cup \text{Verdadeiro Negativo}) \\ &= \frac{231 + 54}{344} = 0,8285\end{aligned}$$

$$RVP = \frac{\text{sensibilidade}}{1 - \text{especificidade}} = \frac{0,8953}{1 - 0,6279} = 2,4061$$

$$RVN = \frac{1 - \text{sensibilidade}}{\text{especificidade}} = \frac{1 - 0,8953}{0,6279} = 0,1667$$

Exemplo

- Com o objetivos de diagnosticar a *colecistitis*, use-se o ultra-som.
- Tal técnica tem uma sensibilidade de 91% e uma especificidade de 98%.
- Se a probabilidade de um indivíduo ter *colecistitis* é de 0,20, qual o VPP e VPN dessa técnica.

$$VPP = \frac{sen \times p}{sen \times p + (1 - esp)(1 - p)} = \frac{0,91 \times 0,20}{0,91 \times 0,20 + (1 - 0,98)(1 - 0,20)} = 0,9191$$

$$VPN = \frac{sen(1 - p)}{esp(1 - p) + (1 - sen)p} = \frac{0,98(1 - 0,20)}{0,98(1 - 0,20) + (1 - 0,91)0,20} = 0,9775$$

Usando o R

- O `RcmdrPlugin.EZR` acrescenta uma série de funções relacionadas a estatística médica
- Ao carregar o `RcmdrPlugin.EZR`, a interface gráfica do `R Commander` é alterada.
- Para cálculo da sensibilidade, especificidade, VPP, VPN, Acurácia, RVP e RVN utiliza-se:
 - `Statistical Analysis` ⇒
`Accuracy of diagnostic test` ⇒
`Accuracy of qualitative test`
 - Em seguida deve-se digitar os valores do resultado do teste e do estado indivíduo

Usando o R

- Para o VPP e VPN considerando a prevalência, sensibilidade e especificidade utiliza-se:
 - `Statistical Analysis`⇒
`Accuracy of diagnostic test` ⇒
`Compute positive and negative predictive values`
 - Em seguida deve-se digitar os valores de prevalência, sensibilidade e especificidade

Distribuição de Probabilidade

DISTRIBUIÇÃO DE PROBABILIDADE

- Variável Aleatória - variável descritora de populações, cujos valores são associados a probabilidades de ocorrência.
- As variáveis aleatórias são classificadas em dois tipos:
 - Variável aleatória discreta
 - os possíveis resultados estão contidos em um conjunto finito ou enumerável
 - Variável aleatória contínua
 - os possíveis resultados abrangem todo um intervalo de números reais

DISTRIBUIÇÃO DE PROBABILIDADE

- Cada variável aleatória tem uma distribuição de probabilidade correspondente.
- A distribuição de probabilidade descreve o comportamento da variável.
 - Por meio de uma representação gráfica é possível ter uma noção da distribuição de probabilidade de uma variável

DISTRIBUIÇÃO DE PROBABILIDADE

- As distribuições de frequências dos dados podem ter uma variedade de formas, incluindo formas simétricas e não simétricas.
- Existem vários modelos matemáticos mais comumente usados para representar os dados:
 - Distribuição binomial
 - Distribuição Poisson
 - Distribuição Normal
 - Distribuição t-student
 - Distribuição Qui-quadrado
 - Distribuição F-snedecor

Distribuição Binomial

DISTRIBUIÇÃO BINOMIAL

- A distribuição Binomial é uma distribuição de probabilidade para variável discreta
- Uma variável aleatória discreta é caracterizada por sua função de probabilidade $P(X = x)$ com as propriedades:

1. $P(X = x) \geq 0$;

2. $\sum_{i=1}^n P(X = x_i) = 1$

DISTRIBUIÇÃO BINOMIAL

- Muitas variáveis aleatórias têm apenas dois possíveis resultados ou eventos elementares, por exemplo:
 - sexo (macho, fêmea),
 - desfecho de um tratamento (cura, não cura)
- Situações com alternativas dicotômicas podem ser representadas genericamente por respostas tipo 0 e 1.
- Em geral ao observar um evento dicotômico representa-se 0 como fracasso e 1 como sucesso de ocorrência

DISTRIBUIÇÃO BINOMIAL

- Experimentos com resultados dicotômicos recebem o nome de ensaio de Bernoulli e neste caso, consideramos uma experiência com dois possíveis resultados
 - Sucesso $\rightarrow P(\text{sucesso}) = p$;
 - Fracasso $\rightarrow P(\text{fracasso}) = q$.
- Temos que:

$$\Omega = \{\text{Sucesso}, \text{Fracasso}\} \therefore P(\Omega) = 1$$
$$p + q = 1 \quad q = 1 - p$$

DISTRIBUIÇÃO BINOMIAL

- Na maior parte das vezes, são realizados n ensaios de Bernoulli. O interesse está no número X de ocorrências de sucessos.
- Um experimento binomial é dado da seguinte forma:
 - consiste em n ensaios de Bernoulli;
 - cujos ensaios são independente;
 - a probabilidade de sucesso em cada ensaio é sempre igual a p , $0 < p < 1$

DISTRIBUIÇÃO BINOMIAL

- Uma variável aleatória X tem distribuição binomial com parâmetros n e p se sua função de probabilidade é

$$P(X = x) = C_x^n p^x q^{n-x}$$

- p é a probabilidade de sucesso em cada ensaio;
- $q = 1 - p$ é a probabilidade de fracasso em cada ensaio;
- $C_x^n = \binom{n}{x} = \frac{n!}{x!(n-x)!}$ é a combinação de n valores x a x
- Notação: $X \sim \text{Bin}(n, p)$

Exemplo

- Uma suspensão contendo organismos de Leishmania é preparada e quando uma determinada quantidade é inoculada em ratos, 30% deles se tornam infectados. Se 10 ratos forem inoculados independentemente, qual a probabilidade de um rato ficar infectado?
 - A variável aleatória X representa o número de ratos infectados, ou seja, $X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 - As possibilidades é cada rato seja infectado (sucesso) ou não (fracasso).
 - Temos $n = 10$, $p = 0,30$ $q = 1 - 0,30 = 0,70$
 - Queremos saber a $P(X = 1)$

DISTRIBUIÇÃO BINOMIAL

Exemplo

- Assim,

$$\begin{aligned}P[X = 0] &= C_1^{10} p^1 q^{10-1} \\&= \frac{10!}{1!(10-1)!} (0,3)^1 (0,7)^9 \\&= 10 \times 0,3 \times 0,0404 = 0,1212 \text{ ou } 12,12\%\end{aligned}$$

DISTRIBUIÇÃO BINOMIAL

Exemplo

- Qual a probabilidade de mais de 1 rato ser infectado?

$$P[X > 1] = P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5] + P[X = 6] + P[X = 7] + P[X = 8] + P[X = 9] + P[X = 10]$$

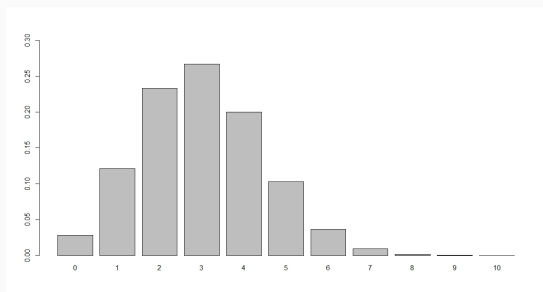
$$P[X > 1] = 1 - P[X \leq 1] = 1 - (P[X = 0] + P[X = 1])$$

$$\begin{aligned} P[X = 0] &= C_0^{10} p^0 q^{10-0} \\ &= \frac{10!}{0!(10-0)!} (0,3)^0 (0,7)^{10} \\ &= 1 \times 1 \times 0,0282 = 0,0282 \text{ ou } 2,82\% \end{aligned}$$

$$\begin{aligned} P[X > 1] &= 1 - P[X \leq 1] = 1 - (0,0282 + 0,1212) \\ &= 1 - 0,1494 = 0,8506 \text{ ou } 85,06\% \end{aligned}$$

DISTRIBUIÇÃO BINOMIAL

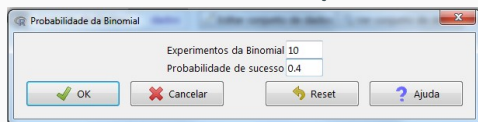
Exemplo



DISTRIBUIÇÃO BINOMIAL

Usando o R

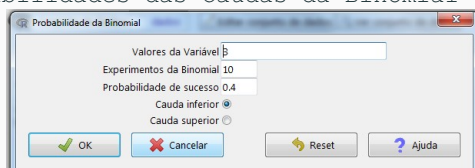
- Para obter probabilidades da distribuições binomial utiliza-se:
 - Distribuições $\Rightarrow$ Distribuições Discretas $\Rightarrow$ Distribuição Binomial $\Rightarrow$ Probabilidades da Distribuição Binomial



DISTRIBUIÇÃO BINOMIAL

Usando o R

- Para obter probabilidades acumulada da distribuições binomial utiliza-se:
 - Distribuições $\Rightarrow$ Distribuições Discretas $\Rightarrow$ Distribuição Binomial $\Rightarrow$ Probabilidades das Caudas da Binomial



- Marque a opção cauda inferior para $P(X \leq x)$ e a opção cauda superior para $P(X > x)$

Distribuição Normal

DISTRIBUIÇÃO NORMAL

- A distribuição Normal corresponde a mais importante distribuição de variáveis aleatórias contínuas, em razão da sua enorme aplicação nos mais variados campos do conhecimento.
- A distribuição Normal é uma distribuição de probabilidade para variável contínua
- Uma variável aleatória contínua é caracterizada por sua função densidade de probabilidade $f(x)$ com as propriedades:
 1. $f(x) \geq 0$ para todo valor de x
 2. A área sob a curva de densidade é 1;
 3. $P(X = x_0) = 0$ sendo x_0 um valor fixo de X

DISTRIBUIÇÃO NORMAL

- Uma variável aleatória X tem distribuição Normal se sua função de densidade de probabilidade é dada por:

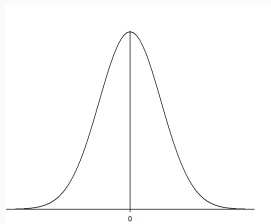
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty$$

em que os parâmetros μ e σ^2 são respectivamente a média e a variância da distribuição.

- Notação: $X \sim N(\mu, \sigma^2)$

DISTRIBUIÇÃO NORMAL

- A distribuição normal apresenta a seguinte propriedades:
 - É simétrica em relação a μ ;



- O ponto máximo de $f(x)$ ocorre em $x = \mu$. Neste ponto as três medidas de posição (média, moda e mediana) se confundem;

DISTRIBUIÇÃO NORMAL

- A distribuição Normal com média $\mu = 0$ e variância $\sigma^2 = 1$ é conhecida como distribuição Normal reduzida ou padronizada.
- Uma variável aleatória com essa distribuição geralmente é simbolizada pela letra Z .
- Notação: $Z \sim N(0, 1)$

DISTRIBUIÇÃO NORMAL

- Se X é uma variável aleatória segue qualquer distribuição Normal então ela pode ser transformada em uma variável normal padrão Z , por meio da expressão

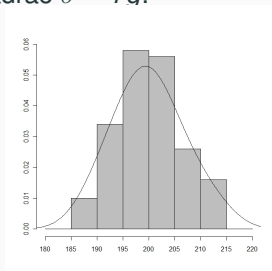
$$Z = \frac{X - \mu}{\sigma}$$

- Assim, para qualquer variável aleatória X com distribuição Normal é possível utilizar a tabela da distribuição normal padrão.

DISTRIBUIÇÃO NORMAL

Exemplo

- Suponha que o peso de ratos machos adultos seja uma variável aleatória com distribuição normal com média $\mu = 200g$ e desvio padrão $\sigma = 7g$.



DISTRIBUIÇÃO NORMAL

Exemplo

- Qual a probabilidade de encontrar um rato com peso superior a 210,5g
- Para o calculo de probabilidade da normal é necessário utilizar tabelas ou softwares
- Variável Padronizada

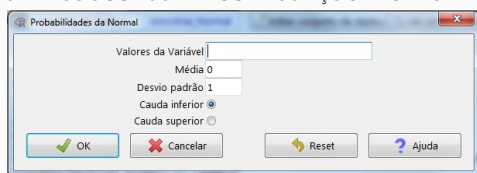
$$z = \frac{x - \mu}{\sigma} = \frac{210,5 - 200}{7} = 1,5$$

DISTRIBUIÇÃO NORMAL

Usando o R

- Para obter probabilidades da distribuições normal utiliza-se:

- Distribuições $\Rightarrow$ Distribuições Contínuas $\Rightarrow$
Distribuição Normal $\Rightarrow$
Probabilidades da Distribuição Normal



- Deve-se informar o valor da variável, a média e o desvio padrão
- Marque a opção `cauda inferior` para $P(X \leq x)$ e a opção `cauda superior` para $P(X \geq x)$