

Universidade Federal do Ceará  
Faculdade de Medicina  
Disciplina de Epidemiologia e Bioestatística

# Conceitos Básicos em Epidemiologia e Bioestatística

---

Hermano Alexandre Lima Rocha  
Eduardo Rebouças Carvalho

Orientação:  
Dr. Luciano Lima Correia

---



A natureza é probabilística  
A informação, incompleta  
Os resultados, essenciais  
Os recursos, limitados  
As decisões, inevitáveis

H.S. Frazier  
Universidade de Havard

Pag 5	_____	Por que estudar Bioestatística?
Pag 7	_____	Conceitos básicos
Pag 9	_____	Medidas de Tendência Central
Pag 12	_____	Conceito - Medidas de Variabilidade
Pag 14	_____	Distribuição normal
Pag 17	_____	Conceito - Medidas de Associação
Pag 19	_____	Classificação e características dos estudos epidemiológicos
Pag 20	_____	Estudo de coorte
Pag 22	_____	Estudos transversais
Pag 23	_____	Estudos de caso-controle
Pag 24	_____	Estudos Ecológicos
Pag 25	_____	Testes de hipótese
Pag 28	_____	Intervalo de confiança
Pag 29	_____	Testes de instrumentos diagnósticos
Pag 30	_____	Apresentando dados em Tabelas e gráficos

## Por que estudar Bioestatística?

---



### Para avaliar a Literatura

---

A leitura da literatura médica começa cedo no treinamento de profissionais da saúde e continua ao longo de suas carreiras. Eles têm que entender bioestatística para decidir se eles podem acreditar ou não os resultados apresentados na literatura. Os editores tentam excluir artigos que são projetados ou analisados imprópriamente, mas poucos têm treinamento estatístico formal e eles focalizam naturalmente mais no conteúdo da pesquisa do que no método. Investigadores de grandes estudos quase sempre consultam estatísticos para ajudar no desenvolvimento do projeto e na análise de dados, especialmente em pesquisas custeadas pelos Institutos Nacionais de Saúde e outras agências nacionais e fundações. Mesmo assim é importante estar atento a possíveis falhas no modo com um estudo é projetado e realizado. Em projetos menores, os investigadores consultam menos freqüentemente os estatísticos, por estarem desavisados da necessidade da estatística ou porque os recursos financeiros não são suficientes. A disponibilidade de programas de computação fácil-de-usares para executar análise estatística foi importante promovendo o uso de métodos mais complexos. Porém, esta mesma acessibilidade permite as pessoas sem o treinamento ou perícia em metodologia estatística a informar análises complicadas quando eles sempre não forem apropriados.



### Para Aplicar os Resultados dos Estudos aos Cuidados de Pacientes

---

Aplicar os resultados de pesquisas nos pacientes é a principal razão pela qual os clínicos leem a literatura médica. Eles querem saber quais os melhores procedimentos diagnósticos, quais métodos de tratamento estão sendo utilizados com maior sucesso, e como o tratamento deve ser projetado e implementado. Eles também leem os artigos para se atualizarem.



### Para Interpretar estatísticas Vitais

---

Médicos devem conseguir interpretar estatísticas vitais para diagnosticar e tratar os pacientes efetivamente. Estatísticas vitais são baseadas em dados colecionados do registro contínuo de eventos vitais, como nascimentos e mortes. Uma compreensão básica de como estatísticas vitais são determinadas, o que eles querem dizer, e como eles são usados facilita muito o seu uso.



### Para entender problemas epidemiológicos

---

Os médicos têm que entender os problemas de epidemiologia porque esta informação lhes ajuda a fazer diagnósticos e desenvolver planos de tratamento para seus pacientes. Dados epidemiológicos revelam a prevalência de uma doença, sua variação por período do ano e local, e sua relação com certos fatores de risco. Além disso, a epidemiologia nos ajuda a entender como vírus e outros agentes infecciosos se disseminam. Esta informações também ajudam a sociedade a tomar decisões conscientes sobre o desenvolvimento de programas de saúde, por exemplo, se uma comunidade deveria começar um programa de vigilância, se um programa de screening deve ser executado e como pode ser projetado para ser eficiente, e se deveriam ser usados recursos da comunidade para problemas de saúde específicos.

### Para interpretar informações sobre drogas e equipamentos

Médicos avaliam continuamente informações sobre drogas e instrumentos médicos. Este material pode ser provido por representantes de companhias, enviados pelo correio, ou podem ser publicado em revistas. Por causa do custo alto de drogas e instrumentos médicos em desenvolvimento, companhias fazem tudo que podem para recuperar os investimentos. Para vender os produtos, tem que convencer os médicos que seus produtos são melhores que os dos concorrentes. Para fazer isso, usa gráficos, quadros, e os resultados de estudos que comparam seus produtos com outros no mercado.

### Para usar técnicas diagnósticas

Identificar o procedimento diagnóstico correto é uma necessidade para tomar decisões sobre os cuidados do paciente. Além de saber a prevalência de uma determinada doença, médicos devem estar atentos a sensibilidade de um teste diagnóstico em descobrir a doença quando estiver presente e a frequência com que o teste não indica nenhuma doença em uma pessoa que não tem nenhuma doença. Estas características são chamadas a sensibilidade e especificidade de um teste diagnóstico.

### Para ficar informado

Se manter atualizado e ser crítico sobre dados são habilidades gerais e difíceis de medir. Estas habilidades também não são fáceis de adquirir porque muitas responsabilidades tomam o tempo de um profissional.

### Para avaliar as Diretrizes(Guidelines)

O número de diretrizes para diagnóstico e tratamento aumentou muito ultimamente. Os médicos avisam que não devem ser aceitas diretrizes sem avaliação crítica do profissional; embora algumas estejam baseadas em evidências médicas, muitas representam a opinião coletiva de peritos.

### Para participar ou dirigir projetos de investigação

Clínicos que participam de pesquisas sabem que conhecimento sobre bioestatística e métodos de investigação é indispensável. Residentes em todas as especialidades devem mostrar evidência de atividade docente, e isto leva frequentemente a forma de um projeto de investigação. Além de ser muito exigido produção científica no currículo de graduados ; )





## População

conjunto de indivíduos ou objetos que apresentam em comum determinadas características definidas para o estudo. Ex.: População de pacientes internados em um dado hospital.

Uma população pode ser finita e pequena, sendo fácil de conhecer todos os seus elementos. Porém, na maioria das vezes, é finita mais incontável ou mesmo infinita. Nestes dois últimos casos, para conhecer uma população, a estatística lança mão de um recurso que é coletar uma amostra desta população e caracterizar alguma variável da população a partir dos resultados obtidos a partir da amostra, ou seja, tirar conclusões sobre a população a partir de resultados obtidos em amostras (inferência estatística).



## Amostra

Conjunto de elementos de uma população, selecionado segundo algum critério. É um subconjunto da população. Tipos de amostras casuais (tipo de amostra na qual os elementos são escolhidos ao acaso, eliminando a tendenciosidade do pesquisador):



### Tipos de amostras casuais

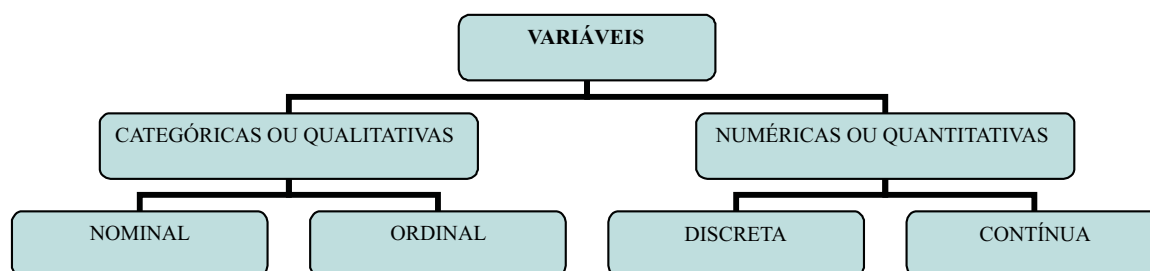
Tipo	Característica
Casual simples ou equiprobabilística	Cada elemento da população tem a mesma chance de entrar na amostra. Por exemplo, se desejamos tomar uma amostra casual de 5 cartas de um baralho de 52 cartas, uma maneira seria embaralhar cuidadosamente e então tirar 5 cartas sem olhar a face das mesmas. Aqui o baralho é a população, e as 5 cartas as amostras.
Sistemática	Na amostragem sistemática, seleciona-se qualquer unidade amostral e, a partir desta, escolhem-se as seguintes segundo o intervalo de seleção. Por exemplo: Imagine uma cidade em que cada quarteirão tem 10 casas. Um pesquisador quer analisar 1 casa de cada quarteirão. Ele pode sortear um número de 1 a 10. Se o número sorteado for 7, ele vai pesquisar somente a sétima casa de cada quarteirão.
Estratificada	Uma amostra pode ser estratificada de acordo com algum fator como sexo, idade ou nível econômico. Para uma amostra estratificada de acordo com o sexo, por exemplo, divide-se a população em dois estratos: homens e mulheres, e em seguida escolhe-se uma amostra casual dentro de cada estrato.

## Variável

É a característica que se deseja estudar de uma dada população. Ex.: Cor dos olhos dos moradores de Fortaleza, altura dos alunos da UFC. As variáveis são classificadas segundo suas características particulares em quatro categorias. Tais classificações não são simplesmente didáticas, mas assumem papel importante na estatística, pois terão tratamentos diferentes como será visto adiante.

## Tipos de Variáveis

Tipo	Característica
Contínuas	São aquelas que podem assumir qualquer valor dentro de um intervalo de interesse. Os dados advindos deste tipo de variável são ditos contínuos. Ex.: peso, estatura, distância percorrida em um teste de esforço etc. Em geral estão associadas a medidas que tenham unidade (m, kg, l, m/s etc.)
Discretas	São aquelas que só podem assumir valores inteiros dentro de um intervalo de interesse. Os dados discretos são resultados da contagem do número de itens referente à variável. Ex.: número de repetições executadas em um exercício, número de gols marcados em uma partida de futebol, quantidade de saltos dados por um jogador em uma partida de voleibol etc.
Nominais	São aquelas que só podem assumir alguns estados ou categorias e geralmente não são numéricas: Os dados nominais surgem quando se definem categorias e se conta suas observações. Ex.: Sexo de uma população (masculino e feminino), queixas de dor lombar (sim e não), cor dos olhos de uma população (azuis, castanhos, pretos, verdes) etc.
Ordinais	São aquelas que se relacionam a avaliações subjetivas segundo preferência ou desempenho. Os dados ordinais constituem valores relativos, atribuídos para denotar ordem. Ex.: Avaliação do estado de saúde de um paciente excelente, bom ou reservado.





 **Conceito**

O conceito de medida de tendência diz respeito a medida de um valor que possa melhor representar a tendência de um conjunto de números, ou uma variável. As três medidas mais utilizadas são a média, a mediana e a moda.



 **Média Aritmética**

Medida de tendência central mais utilizada;  
 É definida como soma dos valores teóricos de todas as observações (observação é um elemento de uma amostra) dividida pelo número de observações;  
 O símbolo  $\mu$  (mu) será usado para denotar média de uma população;  
 O símbolo  $\bar{x}$  será usado para denotar a média de uma amostra;

$$\bar{x} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

Exemplo: Uma amostra constituída de 3 pesos de recém-nascidos: 2,75kg, 3,25kg e 3,80kg. Aqui, n, o tamanho da amostra, é igual a 3.  $x_1$ , primeira observação, é 2,75kg;  $x_2$ , segunda observação é 3,25kg;  $x_3$  é 3,80kg.

$$\bar{x} = \frac{(2,75 + 3,25 + 3,80)}{3} = 9,80/3 = 3,27, \text{ ou seja, peso médio é } 3,27 \text{ kg.}$$

 **Mediana**

Outra medida usada para indicar o centro de uma distribuição;  
 A mediana de uma série de observações é o número que fica exatamente no meio da série quando os dados estão ordenados e o número de observações é ímpar; ou a média aritmética de dois números do meio, quando o número de observações é par;  
 Isto significa que para um conjunto de dados, se os mesmos forem ordenados, a mediana ocupará o centro deste conjunto.  
 Ex.: Dada a variável  $x = \{1, 3, 0, 2, 4\}$ , a mediana é 2.  
 Para se calcular a mediana a mediana de um conjunto de dados deve-se:  
 1) ordenar o conjunto; no exemplo acima:  $x = \{0, 1, 2, 3, 4\}$ ;  
 2) verificar se há um número par ou ímpar de valores no conjunto; no exemplo acima: 5 observações - ímpar;  
 3) se for ímpar a mediana será o valor que ocupa a posição central e se for par será a média entre as duas posições centrais.

 **Exercício**

Dê as medianas:

Conjunto	Mediana	Conjunto	Mediana
2;3;3;4		40;90;81;80;100	
1;18;19;22		5,1;6,5;8,1;9,1;10,6;15,5	
9,2;3,7;11,8;10,1;12,8		1;2;3;3;3;4;7	



## Moda

A moda é o valor que aparece com maior frequência em uma distribuição.

Exemplos:

Seja  $x = \{0\ 1\ 0\ 2\ 3\ 4\ 4\ 0\ 3\ 2\ 5\ 6\}$ , a moda é 0.

Seja  $x = \{3\ 1\ 2\ 3\ 3\ 4\ 5\ 1,5\ 2\ 1,5\ 0\ 4\ 1,5\ 1,5\ 6\}$  a moda é 1,5.



## Exercício

Qual a moda da frequência?

Frequência Semanal	Número de Indivíduos
0	12
1	22
2	46
3	20



## Comparação entre Média, Mediana e Moda

Muitas vezes, precisamos decidir qual a medida de tendência central que mais se adequa aos nossos objetivos. A seguir, segue uma tabela que apresenta vantagens e limitações de cada uma delas.

	Vantagens	Limitações	Tipo de variável aplicável
Média	Reflete todos os valores da amostra. Possui propriedades matemáticas definidas	É influenciada por valores extremos.	Contínua Discreta
Mediana	Menos sensível a valores extremos que a média	Mais difícil de ser determinada para grande quantidade de dados	Contínua Discreta
Moda	Representa um valor típico	Não tem função em termos de cálculo Não tem função em certos conjuntos de dados	Contínua Discreta Categórica Ordinal



## Conclusão

---

Portanto, é importante saber que a média é a melhor e mais importante medida de tendência central, devido a sua maior estabilidade amostral, utilidade e facilidade de cálculos. Porém, em distribuições assimétricas, a média não descreve adequadamente o centro. Nesses casos, prefere-se a mediana. Por exemplo, numa distribuição: ( 2, 4, 6, 9, 10, 11, 2000) na qual um dos valores(2000) destoa bastante dos outros. Nesse caso, temos:

\* média:  $2+4+6+9+10+11+2000 / 7 = 291,7$

\* mediana: temos sete elementos nesse conjunto. Portanto, a mediana é o elemento central desse conjunto ordenado – 4º elemento ( $n+1/2$ , ou seja,  $7 + 1 / 2 = 4$ ). O quarto elemento desse conjunto corresponde à observação 9.

Mediana = 9.

Nesse caso, verifica-se facilmente que a mediana representa melhor o centro da distribuição.





### Conceito - Medidas de Variabilidade

---

As medidas de tendência central nos dão uma idéia da concentração dos dados em torno de um valor. Entretanto, é preciso também conhecer suas características de espalhamento ou dispersão – medidas de variabilidade (ou dispersão).



### Exercício

---

As séries são iguais? Qual a média e a mediana de cada uma das séries?

Série A	Série B	Série C
1	1	1
1	44	8
2	45	11
3	46	14
5	48	28
6	48	30
6	49	37
7	50	48
93	50	52
94	51	62
94	52	70
95	52	72
97	54	84
98	55	91
98	55	92
100	100	100

---



### Amplitude

---

Amplitude é definida como a diferença do menor ao maior valor de um conjunto de dados. Na série 7-3-4-6-1-6-7-6-5, a amplitude é  $6=(7-1)$ ; É considerada uma medida de dispersão muito satisfatória para grupos pequenos de dados.



### Desvio padrão e variância

---

Nem me perguntem o que significa desvio padrão. Porém, o importante nesse momento é saber para que serve essa medida de dispersão.

Considere essa série: 7-3-4-6-1-6-7-6-5. A média  $\bar{x}$  é igual a 5. Para o cálculo do desvio padrão e da variância é necessário diminuir cada observação pela média  $(x-\bar{x})$ . Essa diferença é chamada de desvio médio  $(x-\bar{x})$ . Depois eleva-se o valor do desvio médio ao quadrado  $(x-\bar{x})^2$ . Feito isso, soma-se todos os desvios médios ao quadrado. Divide-se esse valor pelo número de observações - 1 ( $n-1$ ), obtendo-se, portanto a variância de uma amostra. Veja exemplo.



Cálculo da variância e do desvio padrão

x	(x- $\bar{x}$ )	(x- $\bar{x}$ ) <sup>2</sup>
7	+2	4
3	-2	4
4	-1	1
6	+1	1
1	-4	16
6	+1	1
7	+2	4
6	+1	1
5	0	0
45	0	32

$$\text{variância} = s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = 32/8 = 4$$

$$\text{desvio padrão} = s = \sqrt{4} = 2$$

A fórmula da variância em uma amostra (n= número de elementos de uma amostra) é diferente da fórmula de variância em uma população (N=números de elementos de uma população). A variância de uma amostra é usualmente denotada por  $s^2$  e é definida pela fórmula:  $\text{variância} = s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ ,

A variância de uma população é representada por  $\sigma^2$  e é dada pela fórmula:  $\text{variância} = \sigma^2 = \frac{\sum(x-\bar{\mu})^2}{N}$



Conclusão

O desvio padrão nada mais é do que a raiz quadrada(positiva) da variância. O desvio padrão de uma amostra é denotado por s e o da população por  $\sigma$  (sigma).

Portanto, desvio padrão mede dispersão. E o faz mais ou menos do mesmo modo que o desvio médio, ou seja, medindo o afastamento médio dos dados em relação à média do conjunto. A diferença é que, tomando o quadrado dos desvios, o desvio padrão obtém uma espécie de média ponderada desses desvios, onde desvios maiores entram com peso maior que os desvios menores.

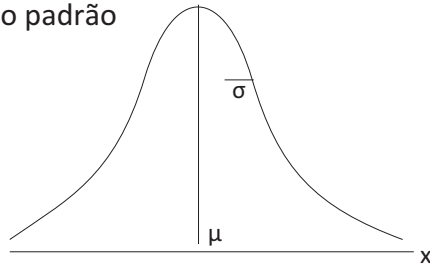


## Distribuição normal

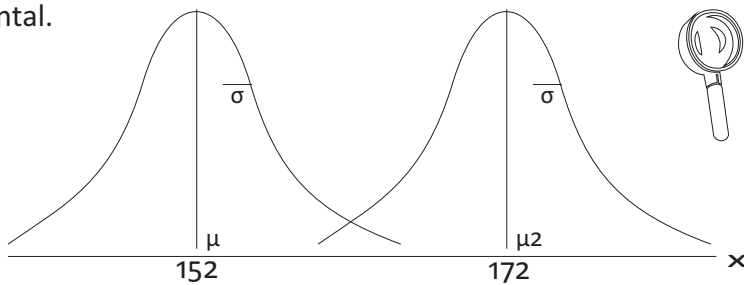
Uma curva de distribuição pode descrever a forma da distribuição de uma população. Conhecendo-se a forma, a média e o desvio padrão, pode-se a caracterizar uma população.

Um tipo de curva de distribuição bastante comum para representar a distribuição de populações de dados biológicos( peso, altura, pressão arterial, taxa de glicose no sangue, etc.....) é a curva normal ou curva de Gauss.

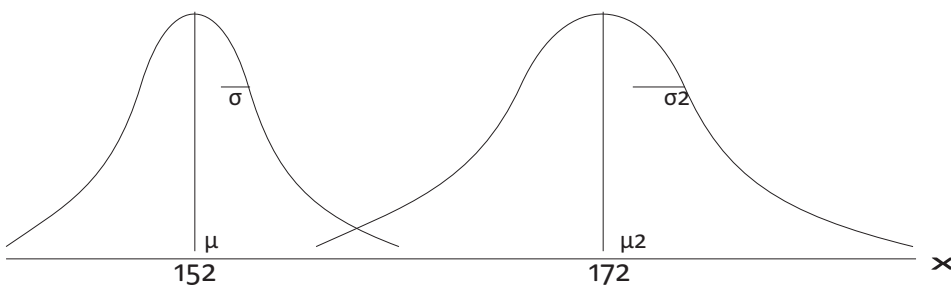
$\mu$  : média populacional     $\sigma$ : desvio padrão



É importante destacar que não há apenas uma curva normal. Para cada valor diferente da média e do desvio padrão, há uma curva normal diferente. Estaturas de adultos do sexo masculino, por exemplo, teriam distribuição normal com uma altura média de 172cm, enquanto as estaturas de meninos de 10 anos de idade seriam ainda normalmente distribuídas mas em torno de uma média de 152cm. Se a dispersão( desvio padrão:  $\sigma$ ) das duas populações de estaturas é igual, as duas curvas terão forma idêntica, ocupando posições diferentes no eixo horizontal.



Ou, se o desvio padrão for diferente, assumindo formas diferentes:



## Propriedades da curva normal

1. A área entre a curva e o eixo horizontal (eixo  $x$ ) é igual a 100%
2. A curva é simétrica em torno do ponto  $\mu$ =média e tem a forma aproximada de um sino.
3. A área entre a média e um ponto qualquer poderá ser medida em termos números de desvio padrão.  
 $z = \frac{x - \mu}{\sigma}$ ,  $z$  é chamado de desvio relativo, pois mede o afastamento médio dos valores de  $x$  em relação a média, em unidades de desvio padrão.





## Áreas sobre a curva normal

Aqui é realmente importante atenção!!

Exemplifiquemos: Suponha uma população de estaturas de homens adultos, com distribuição normal, média 172cm e desvio padrão 5cm.

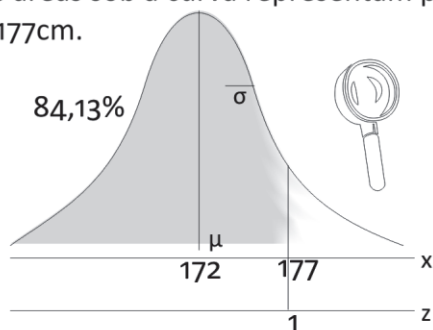
1) Que proporção dessas estaturas são menores de 177cm?

1º passo: Aplicando a fórmula do desvio relativo ( $z = \frac{x - \mu}{\sigma}$ ), teremos:

$$z = \frac{177 - 172}{5} = \frac{5}{5} = 1$$

2º passo: Olhar a tabela com valores de z (lembrar de levar essa tabela pra prova – disponível na última página dessa apostila).

Olhando a tabela, verificamos que para  $z=1,00$  temos 0,3413. Esta é a área sob a curva normal entre a média e a estatura 177cm; somada a 0,500 que é a área à esquerda da média, totaliza 0,8413, que é a área abaixo de  $z=1,00$ . Lembrando que as áreas sob a curva representam proporções, temos que 84,13% das estaturas de homens adultos são menores que 177cm.



2) Que proporção dessas estaturas são maiores que 177cm?

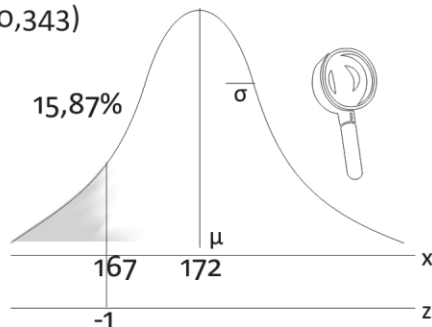
Essa proporção pode ser encontrada subtraindo 0,8413(resposta da questão 1) de 1,000(área total sob a curva). Assim 15,87% dos homens adultos são mais altos que 177cm.

3) Que proporção das estaturas são menores que 167cm?

1º passo: Aplicando a fórmula do desvio relativo (z), teremos:

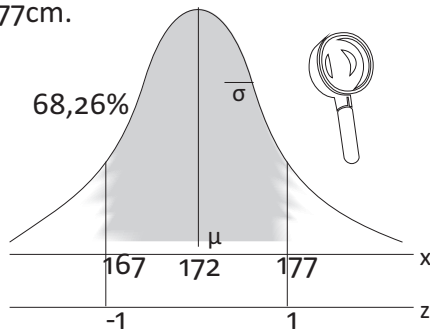
$$z = \frac{167 - 172}{5} = \frac{-5}{5} = -1$$

Obs: A tabela não dá valores negativos de z, mas como a curva é simétrica, a área entre a média e  $z = -1$  é igual a área da media a  $z = 1$ , ou seja, 34,13%. Mas no problema, pede-se a proporção dos indivíduos menores de 167cm  $\rightarrow 15,87\%(0,5000 - 0,343)$



4) Que proporção de estaturas estão no intervalo 167 a 177 cm?

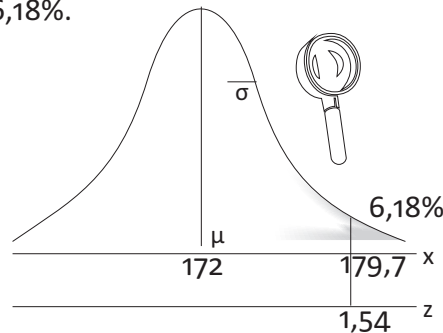
Do resultado anterior temos que, para  $x=167$ ,  $z=-1$  e para  $x=177$ ,  $z=+1$ . A área entre  $z=0$  e  $z=1$  é igual a 34,13%. Deste modo, a área entre  $z=-1$  e  $z=1$  será de 68,26%, ou seja, 68,26% das estaturas de homens adultos estão entre 167 a 177cm.



5) Que proporção de estaturas são maiores que 179,7 cm?

$$Z = \frac{179,1 - 172}{5} = 1,54$$

Na tabela z, a área entre  $z=0$  e  $z=1,54$  é 0,4382. Subtraindo de 0,5000, a área à direita de  $z=1,54$ , ou seja, a proporção de estaturas acima de 179,7 cm será de 6,18%.



### Interpretando áreas como probabilidade

Nos exemplos anteriores, áreas sob partes da curva normal foram interpretadas como porcentagens de homens adultos, cujas estaturas estavam dentro de certos intervalos. Tais áreas podem ser interpretadas como probabilidade. No exemplo 1, onde 84,13% das estaturas dos homens adultos estava abaixo de 177cm, afirmou-se que 84,13% das estaturas de homens adultos estão abaixo de 177cm. Pode-se dizer também que, se um homem adulto é tomado ao acaso, a probabilidade de que sua estatura seja menor que 177cm é 0,8413 ou 84,13%.



### Valores normais

Uma aplicação bastante útil da curva normal é a determinação de valores normais. Como a prática da medicina está se tornando cada vez mais quantitativa, grande importância está sendo atribuída aos exames laboratoriais. Torna-se essencial, portanto, definir os limites de normalidade de modo a diferenciar os possíveis valores anormais. Para isso fazemos medidas para uma série de indivíduos supostamente normais e para essas medidas calculam-se a média e o desvio padrão. A seguir determinam-se intervalos (simétricos em torno da média) que incluam qualquer proporção desejada de indivíduos. Usa-se normalmente um intervalo de 95% que, numa curva normal, inclui valores entre média menos 1,96 desvios padrão até média mais 1,96 desvios padrão (ou  $\sim \mu \pm 2\sigma$ ). Por exemplo, se a temperatura média de um grupo de adultos normais (sadios), do sexo masculino, for calculado em 36,8°C, com um desvio padrão de 0,27°C, poderemos dizer que 95% dos indivíduos normais do sexo masculino têm uma temperatura entre  $36,8 - 2(0,27)$  e  $36,8 + 2(0,27)$ , ou seja, entre 36,3°C e 37,3°C. Esse intervalo é chamado de intervalo de normalidade para temperatura.





## Conceito - Medidas de Associação

---

Buscam responder se existe uma associação entre uma exposição e um desfecho. São medidas do tipo razão que comparam duas medidas de frequência e medem a força da relação estatística entre uma variável e a frequência da doença.

Os estudos longitudinais prospectivos e retrospectivos proporcionam ao investigador a oportunidade de identificar, dentro do sistema composto das variáveis que são selecionadas e estudadas, os fatores de risco e de proteção, considerando as variáveis independentes e dependentes, contidas no modelo de estudo. Um dos primeiros procedimentos no estudo é identificar a variável dependente e as variáveis independentes. Tanto seja, variável discreta ou contínua, podemos realizar a análise dos dados, com o intuito de verificar o risco de ocorrência de um fato, pela presença de um ou mais fatores, considerando inclusive a sua intensidade. Exemplificando, poderíamos verificar a hipótese de que o consumo de cigarros tenha associação com câncer de pulmão, determinando-se o risco para tabagistas, baseando-se em indivíduos não fumantes, tendo-se em conta que os não tabagistas também podem contrair a doença. Conceituar o que seja risco é fundamental. Portanto, considera-se como risco a probabilidade de que pessoas que estão sem a doença, mas expostas a certos fatores, possam adquiri-la. Na verificação da ocorrência de doentes na população, variável chave da investigação epidemiológica, as medidas de incidência, prevalência e outros coeficientes servem como indicadores de risco.

## Prevalência

---

Prevalência é a fração de um grupo de pessoas que possuem uma determinada condição clínica em um dado período de tempo. A prevalência é obtida através de uma população predefinida contendo pessoas saudáveis e com a condição em questão, em um único momento.

## Incidência

---

Incidência é a fração de um grupo que inicialmente não possuía a doença e que a desenvolve em um determinado período de tempo. A incidência então se refere aos novos casos da doença que aparecem em uma população inicialmente saudável. O grupo de pessoas utilizado na medida de incidência é um grupo suscetível à doença em questão.

## Coeficiente de incidência

---

número de casos novos em relação ao total de indivíduos expostos na unidade de tempo.

## Coeficiente de prevalência

---

número de casos novos + antigos, em relação ao total de indivíduos expostos na unidade de tempo.

## Risco relativo (RR)

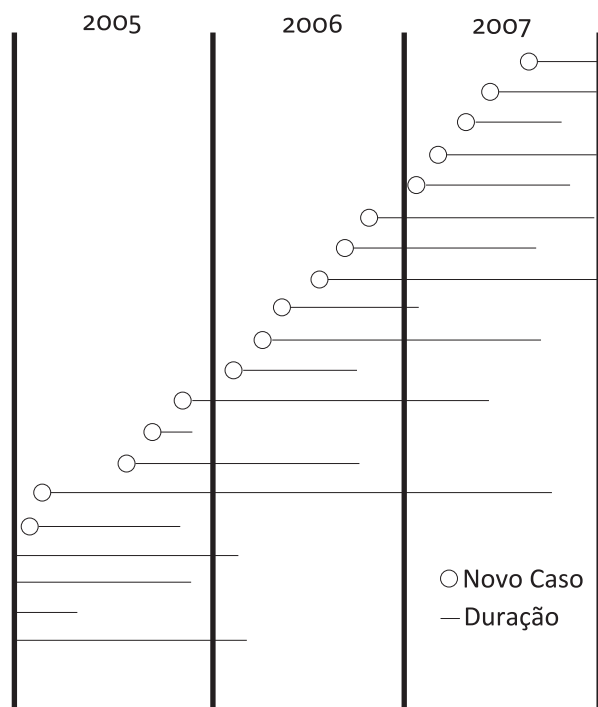
---

relação existente entre o coeficiente de incidência de casos expostos e o coeficiente de incidência de casos não expostos.  $RR = \frac{I_E}{I_{NE}}$



## Exercício

A figura abaixo mostra a ocorrência de doença em um grupo de 100 pessoas durante o decorrer de três anos. Nos três anos, 16 desenvolveram a doença e quatro já a possuíam.



### Perguntas:

01 - Qual a prevalência: no começo de 2005, no começo de 2006, no começo de 2007 e no começo de 2008?

02 - Qual a incidência: No período anual de 2005, no período anual de 2006, no período anual de 2007, no período de 2005 até 2007?

A prevalência considera a situação do paciente em um determinado ponto do tempo, enquanto a incidência considera um intervalo de tempo em que os susceptíveis foram acompanhados.



## Estudos para determinar Prevalência e Incidência

Os estudos para determinar a prevalência de uma doença são os estudos transversais (cortam a população em um só ponto); os estudos para determinar incidência são chamados de estudos de coorte (acompanham um mesmo grupo de pessoas "coorte" durante um determinado período).



## Comparação Incidência x Prevalência



Característica	Incidência	Prevalência
Numerador	Novos casos em um grupo previamente sadio	Todos os casos do grupo
Denominador	Todas as pessoas susceptíveis no começo do período	Todos os indivíduos do grupo, doentes ou não
Tempo	Duração do período	Um único ponto
Medida	Estudo de Coorte	Estudo transversal



## Epidemia

Considera-se que uma doença apresenta um surto epidêmico quando sua incidência em um determinado período é maior que a Incidência Máxima Esperada (IME). A IME é calculada como sendo:  $IME = (\text{Média da incidência}) + 2 \times (\text{Desvio Padrão da Incidência})$ , ou seja, uma epidemia é uma variação de aproximadamente 5% nas pontas da curva normal.



## Classificação e características dos estudos epidemiológicos



### Estudos observacionais

Descritivos

Caso-controle

- Causas e incidência de doenças
- Identificação de fatores de risco

Estudos transversais

- Descrição de doenças
- Diagnóstico e estadiamento
- Mecanismos de doenças

Estudos de coorte

- Causas e incidência de doenças
- Historia natural, prognostico
- Identificação de fatores de risco

Estudos de coorte históricos

### Estudos experimentais

Experimentos controlados

- Controles paralelos
  - Randomizados
  - Não randomizados

Estudos sequenciais

- Auto controlados
- Cruzados

Controles externos

Experimentos sem controle

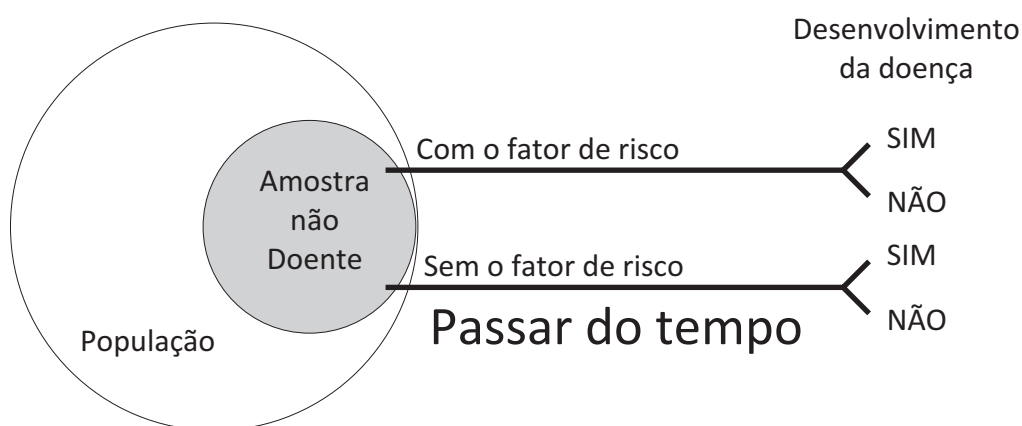
### Análise de dados

## Definição de Estudos de Coorte

Em um estudo de coorte, um grupo de pessoas (coorte) é montado, em que nenhum elemento tenha a doença em estudo, mas lhe sejam susceptíveis. No início do estudo, todas as pessoas são classificadas para determinação de suas características que possam atuar como fatores de risco para a doença. Essas pessoas são então observadas durante um período definido de tempo para se observar qual delas desenvolverá a doença. Servem principalmente para estudar fatores de risco. Estudos de coorte são também chamados de estudos longitudinais, prospectivos e de incidência.



## Diagrama de Estudos de Coorte



## Tipos de estudo de coorte

Os estudos de coorte podem ser feitos de duas maneiras: A coorte pode ser montada no presente e estudada durante o futuro (Estudo de coorte Concorrente) ou pode ser montada no passado e estudada no presente (Estudo de Coorte Histórico). Os estudos de coorte históricos sofrem vários vieses.

## Vantagens e desvantagens do estudo de coorte

Os estudos de coorte são os melhores estudos para descrever fatores de risco. A sua principal desvantagem é que se a doença não é freqüente, um grande número de pessoas deve ser estudado. Outro problema é que as pessoas estudadas podem se mudar durante o estudo, o que torna o estudo trabalhoso e muito caro. Um dos vieses mais freqüentes é a influencia de muitos outros fatores nas observações, que podem também influenciar o aparecimento da doença. Assim, os fatores externos que atuam nas observações devem ser controlados.

## Comparando os fatores de risco

A incidência é a expressão básica do risco de se desenvolver uma determinada doença. Para se comparar a incidência de uma doença em dois grupos de uma coorte que diferem em possuir ou não um fator de risco, usamos as medidas de efeito.



## Medidas de efeito para estudos de coorte

---

Diferença de risco:

Representa qual o valor da incidência atribuível ao fator de risco.

Incidencia nos Expostos ao Fator de Risco menos Incidencia nos não Expostos ao Fator de Risco

$$I_E - I_{NE}$$

Risco relativo:

Quantas vezes mais tendência uma pessoa exposta ao fator de risco tem de desenvolver a doença.

Incidencia nos Expostos ao Fator de Risco dividida pela Incidencia nos não Expostos ao Fator de Risco

$$I_E / I_{NE}$$

Risco na população atribuível ao fator de risco:

Representa a incidência na população da doença com aparecimento associado ao fator de risco.

Diferença de risco vezes a Prevalência da doença

$$(I_E / I_{NE}) \times P$$

Fração do Risco na população atribuível ao fator de risco:

Qual a porcentagem de casos de doenças na população que podem ser atribuídos ao fator de risco.

Risco na população atribuível ao fator de risco sobre incidência total da doença.

$$[(I_E / I_{NE}) \times P] / I_T$$



## Estudos transversais

Nos Estudos Transversais, cada indivíduo é avaliado para o fator de exposição e a doença em determinado momento. Muitas vezes o estudo transversal é realizado apenas com objetivo descritivo sem nenhuma hipótese para ser avaliada. Alguns têm usado o termo levantamento para denominar estudos transversais realizados com essa finalidade. O estudo transversal pode ser usado como um estudo analítico, ou seja para avaliar hipóteses de associações entre exposição ou características e evento. No entanto limitações existem quando se tenta concluir qual a natureza da relação entre exposição e evento nestas situações. Essa limitação relaciona-se, principalmente, com o fato de que a exposição e a doença são avaliados ao mesmo tempo (transversalmente). Uma questão importante que pode ficar sem resposta é sobre o que apareceu primeiro, o fator de exposição ou a doença. Um outro ponto que deve ser observado nestes estudos transversais é a dificuldade em separar os casos novos da doença dos casos já presentes por algum tempo. Desta forma os estudos transversais irão refletir não apenas determinantes de doença mas, também, determinantes de sobrevida. Apesar das limitações, os estudos transversais, pelo fato de incluir indivíduos com e sem o evento e de poder avaliar associações entre o evento e exposições ou características, podem ser considerados um passo adiante na identificação dos determinantes de doenças, quando comparados com relatos ou séries de casos. Em verdade, existem situações em que os estudos transversais podem ser considerados estudos verdadeiramente analíticos. Isto ocorre quando o fator de exposição não sofre influência do tempo. Diversos exemplos serão encontrados entre aqueles fatores presentes desde o nascimento, como tipo sanguíneo, sexo (masculino, feminino) e sistema HLA. Como os estudos transversais descrevem o que ocorre com um determinado grupo e em um determinado momento, eles são importantes guias para tomadas de decisões no setor de planejamento de saúde. Para o profissional que lida diretamente com pacientes, os estudos transversais oferecem informações da maior utilidade ao chamar atenção para características ligadas com a freqüência de uma determinada doença na comunidade ou em determinado serviço assistencial. Estes estudos, portanto, podem influenciar o raciocínio clínico e a tomada de decisões na prática médica. Quando da escolha do desenho de pesquisa deve ser considerado que os estudos transversais, habitualmente, exigem menos recursos financeiros e podem ser realizados mais rapidamente do que os estudos de caso-controle ou coorte.

Os estudos transversais também podem ser:

- Comparados - 4835 indivíduos de 20 a 74 anos foram selecionados aleatoriamente da população adulta do Rio Grande do Sul para, em sua própria residência, responderem a um questionário sobre hábitos alimentares e terem sua pressão arterial medida. Os 4565 indivíduos efetivamente estudados foram então classificados como consumidores excessivos de sal ou não, e em hipertensos e não hipertensos.

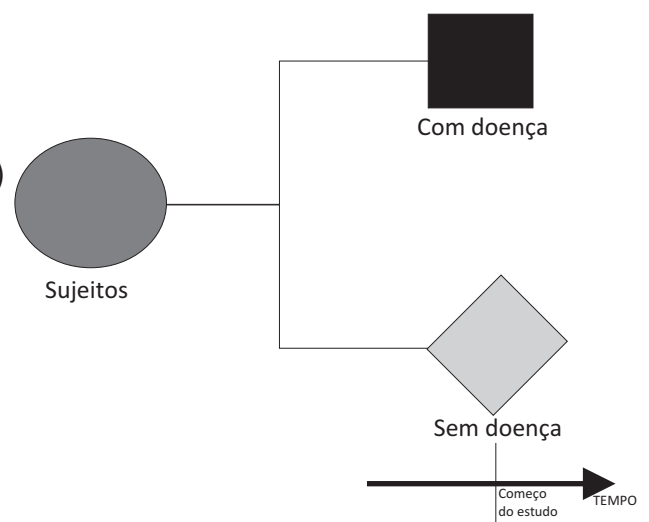
- Não comparados - (estudo de prevalência) 3101 moradores da fronteira sudeste do Rio Grande do Sul foram submetidos a investigação para determinar a prevalência de soro positividade para hidatidose.

Vantagens:

- 1 - Fáceis, rápidos e baratos.
- 2 - Boa fonte de hipóteses.

Inconvenientes:

- 1 - Impossível determinar o que ocorre primeiro (causa - efeito)
- 2 - Desconhecimento da ação dos fatores no passado.
- 3 - Impossibilidade de estabelecer uma prova causal.

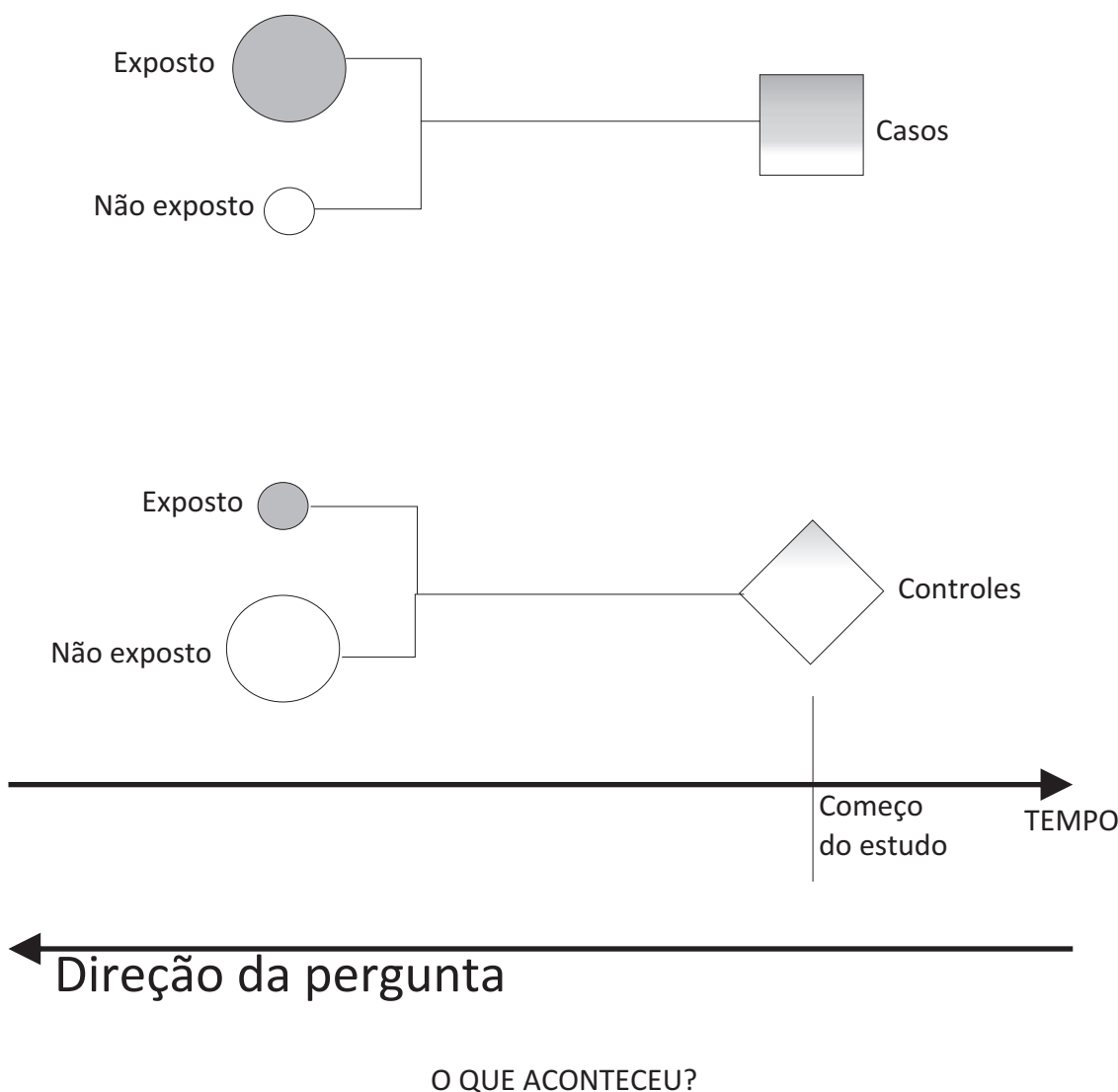


O QUE ESTÁ ACONTECENDO?

 Estudos de caso-controle

O estudo de caso controle começa com a ausência ou a presença de um evento e então procura no passado tentando detectar possíveis causas ou fatores de risco que possam ser encontrados. Os "casos" são indivíduos selecionados de um grupo q possui um determinado evento, os "controles" são indivíduos que não apresentam o evento. São analisados a história de casos e controles em uma tentativa de identificar uma característica ou fator de risco presente nas histórias dos casos mas não nas histórias dos controles.

A figura ilustra que os sujeitos do estudo são escolhidos no começo do estudo depois que eles sejam reconhecidos como portadores de uma doença ou resultado (quadrados) ou controles sem a doença ou resultado (diamantes). São examinadas as histórias de casos e controles em um período prévio para descobrir a presença (áreas sombreadas) ou ausência (áreas não sombreadas) de características predisponentes ou de fatores de risco, ou, se a doença é infecciosa, se o caso foi exposto ao agente infeccioso presumido. Nos estudos de caso-controle, a pergunta está no passado, como mostra a seta apontando para trás na figura, por isso sendo chamado de retrospectivo. Podemos dizer que são os estudos que perguntam: O QUE ACONTECEU?. Os estudos de caso-controle são também longitudinais, pois cobrem um período de tempo.





## Estudos predominantemente descritivos

---

Descreve a ocorrência de um evento (ex. doença) de acordo com diversas exposições ou características das pessoas (sexo ou gênero, idade, raça, nível sócio econômico), local (hospital, bairro, cidade, país etc.) e tempo (ex. a detecção de aumento importante na frequência de um evento em um determinado momento é um dos critérios para diagnosticar epidemias). Os estudos descritivos são especialmente úteis quando pouco é conhecido sobre frequência, história natural ou determinantes de uma doença. Observacional, Predominantemente Analíticos: Tem o objetivo básico de avaliar (não apenas descrever) se a ocorrência de um determinado evento é diferente entre indivíduos expostos e não expostos a um determinado fator ou de acordo com as características das pessoas. Estes são estudos realizados com o objetivo específico de testar hipóteses.



## Estudos Ecológicos

---

Nos Estudos Ecológicos as medidas usadas representam características de grupos populacionais. Portanto a unidade de análise é a população e não o indivíduo. Um exemplo seria um estudo envolvendo diversas cidades brasileiras em que se procurasse correlacionar dados sobre mortalidade infantil a nível de cada município com a renda per capita e índice de analfabetismo do local no sentido de encontrar evidências de que o nível sócio econômico é um dos determinantes de mortalidade infantil.

A limitação principal do estudo ecológico é que a relação entre o fator de exposição e o evento pode não estar ocorrendo ao nível do indivíduo. Desta forma uma associação entre uma exposição e evento ao nível da população não permite afirmar que a exposição está mais presente naqueles que adquirem a doença - (falácia ecológica). Estes estudos, no entanto, ajudam a identificar fatores que merecem uma investigação mais detalhada através de estudo com maior capacidade analítica. Por exemplo, a demonstração de uma relação entre venda de cigarros per capita e mortalidade de doença cardiovascular (DCV) em estudos ecológicos motivou o planejamento de estudos Caso-Controle e Coorte que vierem a demonstrar de forma bem mais convincente que o hábito de fumar é um dos fatores determinantes de DCV, particularmente doença coronariana e acidente vascular cerebral.

Os estudos de séries temporais, em que uma mesma área ou população é estudada em momentos distintos do tempo, são classificados como um subtipo dos estudos ecológicos. Nesse caso, cada unidade de tempo passaria a ser tratada como uma unidade ecológica completa

### Vantagens

- 1 -Facilidade de execução
- 2 -Baixo custo relativo
- 3 - Simplicidade Analítica
- 4 - Capacidade de gerar hipóteses

### Desvantagens

- 1 - Baixo poder analítico
- 2 - Pouco desenvolvimento das técnicas de análise dos dados
- 3 - Vulnerável à chamada "falácia ecológica"





## Testes de hipótese

Quando fazemos uma pesquisa, normalmente usamos amostras, pelos motivos já estudados. Entretanto, o nosso objetivo final é generalizar os resultados para toda a população. O que nos diz se isso é válido ou não são os testes de hipótese.

Chamamos de  $H_0$  (hipótese nula) a hipótese que diz que os nossos resultados são obra do acaso, ou seja, não podem ser generalizados para a população. Chamamos de  $H_1$  a hipótese que diz que os nossos dados são válidos para toda a população.

Os testes de hipótese dizem a probabilidade da hipótese  $H_0$  ser a verdadeira, o que por fim se traduz no valor de  $p$ .



## Teste t de student

DESCOBRIR SE AS MÉDIAS DIFERENTES JUSTIFICAM EXCLUIR  $H_0$


Condições para o uso do teste t:

As distribuições seguem uma distribuição normal; os desvios padrão nas amostras são iguais; as observações são independentes; e o  $n$  da amostra é maior que 30.

Como usar o teste t:

Definem-se quais médias se quer comparar:  $X_1$ =media de idade das mulheres ao entrar na faculdade;  $X_2$ =media de idade dos homens ao entrar na faculdade. Definem-se os tamanhos das amostras:  $N_1$ =numero de mulheres na amostra;  $N_2$ =numero de homens na amostra. Descobre-se o  $t$  da nossa pesquisa:

$$t_{(n_1+n_2-2)} = \frac{(X_1 - X_2)}{SD \sqrt{[(1/n_1) + (1/n_2)]}}$$


 Graus de Liberdade =  $(n_1 + n_2 - 2)$

Descobre-se o  $t$  crítico na tabela, baseado no grau de liberdade da amostra ( $N_1 + N_2 - 2$ ) e no  $\alpha$  desejado e compara-se com o  $t$  obtido. Se o  $t$  calculado for MENOR que o tabelado, a hipótese válida é a  $H_1$ .



## Exercício

Teste t de student:

Em um determinado estudo, ao se compararem duas médias, encontramos um  $t$  calculado de 4,35. O  $t$  tabelado para  $\alpha$  igual a 0,05 é 4,98. Podemos então afirmar:

- A pesquisa tem somente 1% de chance de não ser significativa.
- O valor de  $p$  é 0,05.
- A comparação é significativa com pelo menos 95% de chance.
- O valor de  $t$  independe do tamanho da amostra.

## Teste F

---

### Teste F

DESCOBRIR SE AS VARIÂNCIAS DIFERENTES JUSTIFICAM EXCLUIR  $H_0$

Condições de uso do teste F:

As distribuições seguem uma distribuição normal; os desvios padrão nas amostras são desiguais; as observações são independentes, qualquer  $n$  é válido.

Cálculo do F:

$$(SD_{\text{maior}})^2 / (SD_{\text{menor}})^2$$

Procura-se na tabela o F tabelado para o  $\alpha$  desejado e o respectivo grau de liberdade. Se o F calculado for MAIOR que o F tabelado, rejeita-se  $H_0$ .



## Exercício

---

Teste F:

Em um determinado estudo, ao se compararem duas proporções, foi encontrado um valor de F igual a 4. O F tabelado para um  $\alpha$  igual a 0,01 e o respectivo grau de liberdade foi igual a 5. Logo:

- a) A pesquisa é estatisticamente significativa para  $\alpha = 0,01$ .
- b) É possível que o desvio padrão maior fosse 16 e o menor 4.
- c) É possível que a pesquisa seja estatisticamente significativa para um  $\alpha$  de 0,05.
- d) O teste F pode ser usado com variáveis nominais.

## Teste de Levene

---

Mesma idéia do teste F, mas não precisa ser distribuição normal e é um teste mais exato.

## Teste do chi-quadrado de independência entre variáveis

---

USADO PARA TESTAR A INDEPENDÊNCIA ENTRE DUAS VARIÁVEIS

O teste compara os resultados esperados em uma tabela com os resultados encontrados na pesquisa.

Condições de uso do chi-quadrado:

Distribuição normal;  $n$  mínimo de 20; variáveis nominais.

Como usar o teste do chi-quadrado:

Define-se o que se quer testar: Eu quero saber se existe dependência entre fumar e ter câncer de pulmão (independência das variáveis);

Define o  $\alpha$  (0,05 por exemplo);

Encontra-se o valor crítico (tabelado) para este  $\alpha$  e o grau de liberdade da amostra ( $GL = DF = (\text{linhas} - 1) \times (\text{colunas} - 1)$ );

Calcula-se o  $\chi^2$  para a amostra:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+c)(b-d)(a+b)(c+d)}$$

Fórmula válida para tabelas 2x2 do seguinte estilo:

a	b
c	d

Se o chi-quadrado calculado for MAIOR que o valor crítico (tabelado), rejeita-se a hipótese nula.



### Exercício

Teste chi-quadrado:

Em um determinado estudo, realizou-se uma análise utilizando a fórmula do chi-quadrado em uma tabela que comparava as idades dos participantes com o valor do colesterol encontrado em uma dosagem. Chegou-se ao valor de 53. O valor crítico tabelado para  $\alpha=0,05$  é 3. Podemos afirmar:

- a) O resultado é altamente significativo.
- b) O resultado não é significativo.
- c) O p com certeza é menor que 0,05.
- d) Não podemos afirmar nada, pois o teste foi mal empregado.



### Teste de fisher

Mesma idéia do teste de chi-quadrado, mas calculado de forma exata:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Fórmula válida para tabelas 2x2 do seguinte estilo:

a	b
c	d



### O valor de p (p value)

O valor de p é o alpha mínimo em que ainda se rejeita a hipótese nula.  $p=0,05=5\%$ , quer dizer que esse é o menor alpha que pode ser utilizado para que seja impossível o resultado ser obra do acaso; ou seja, se eu usasse  $p=0,04$ , a hipótese nula já seria considerada. Digamos que tenhamos achado um  $t = 1$ ; o t tabelado para alpha igual a 0,05 seria 1,1, e o t tabelado para alpha igual a 0,04 seria igual a 0,9.



## Intervalo de confiança de uma diferença de proporções

---

### USADO PARA COMPARAR PROPORÇÕES

Exemplo: (P1=proporção de pessoas com câncer de pulmão que fumavam; P2=proporção de pessoas sem câncer de pulmão que fumavam)

Cálculo do intervalo de confiança para a proporção:

$$IC = P1 \pm 1,96(\text{para IC de 95\% (ESSE É O ALPHA DESSE CÁLCULO)}) \times SD$$

Cálculo do intervalo de confiança para a diferença entre as proporções:

$$IC = (P1 - P2) \pm 1,96(\text{para IC de 95\%}) \times SE$$

Se o intervalo de confiança obtido contiver o, não se rejeita a hipótese nula.

Calculo do Erro Padrão:

$$SE_{(p_1 - p_2)} = \sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Onde p é:

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$



### Exercício

---

Em um determinado estudo, uma determinada diferença de proporções foi igual a 1,96, com erro padrão igual a 1. Para um alpha igual a 0,05 (z=1,96), podemos afirmar:

- A diferença entre essas proporções é estatisticamente significativa.
- A diferença entre estas proporções deve estar entre 1,96 e 3,92.
- Para um alpha=0,01, o inverso da diferença entre proporções estudada torna-se possível.
- O n da amostra não influencia o intervalo de confiança.



## Testes de instrumentos de diagnóstico

Os teste de diagnóstico são utilizados para medir a utilidade de novos métodos diagnósticos. Nós testamos um novo método comparando com um método já existente que se assume que nos daria 100% de certeza sobre seu resultado (teste padrão-ouro).

Antes de realizar as questões referentes a esse assunto, construa a seguinte tabela:

	Padrão-ouro positivo	Padrão-ouro negativo	TOTAL
Exame positivo	a	b	a+b
Exame negativo	c	d	c+d
TOTAL	a+c	b+d	a+b+c+d

Podemos avaliar um método diagnóstico utilizando as seguinte razões:

**Sensibilidade:** é a capacidade de um instrumento de reconhecer os verdadeiros positivos em relação ao total de doentes. Fórmula:  $a/a+c$

**Especificidade:** é o poder de distinguir os verdadeiros negativos em relação ao total de doentes. Fórmula:  $d/b+d$

**Valor preditivo positivo (VPP):** é a probabilidade de um caso identificado com um determinado instrumento ser de fato positivo. Fórmula:  $a/a+b$

**Valor preditivo negativo (VPN):** é a probabilidade de um resultado negativo obtido com um determinado instrumento ser de fato negativo. Fórmula  $d/d+c$



## Exercício

Um estudo foi realizado com o objetivo de avaliar um novo exame de sangue para diagnosticar câncer de mama. 100 mulheres foram avaliadas, e destas 20 tiveram o resultado do exame de sangue positivo para câncer de mama. Ao se realizar a biópsia em todas as mulheres, foi encontrado na lâmina células cancerígenas em 30 mulheres, e em 5 das que se queixavam não foram vistas células cancerígenas. Dê a especificidade, sensibilidade e valores preditivos positivo e negativo.



### Apresentando dados em Tabelas e gráficos

---

É chegada a hora de aplicar tudo o que você aprendeu!  
Vamos praticar e fazer nosso próprio trabalho!



### Tabelas de freqüência

---

Os artigos científicos apresentam freqüentemente informação em distribuições de freqüência ou tabelas de freqüência. A escala das observações deve ser primeiro dividida em classes. O número de observações em cada classe é então contado. Os passos para construir uma tabela de freqüência são os seguintes:

Identifique as maiores e menores observações.

Subtraia a observação menor do maior obter a variação.

Determine o número de classes. O bom senso normalmente é suficiente para tomar esta decisão, mas as diretrizes seguintes podem ser úteis.

Entre 6 e 14 classes geralmente é adequado para prover bastante informação sem ser detalhista demais.

O número de classes deveria ser grande o bastante para demonstrar a forma da distribuição mas não as flutuações menos notáveis.

Uma dica é dividir a variação de observações pelo número de classes obter a comprimento das classes. Para algumas aplicações, decidir primeiro o comprimento da classe pode fazer mais sentido; então use depois o comprimento de classe para determinar o número de classes. AS seguintes são algumas diretrizes para determinar o comprimento de classe.

Os limites de classe (números iniciais e finais) não devem se sobrepor. Por exemplo, eles devem ser declarados como "40-49" ou "de 40 até 50," não como "40-50" ou "50-60." pois assim não temos como dizer a que classe o 50 pertence.

Se possível, os comprimentos de classe devem ser iguais. Comprimentos de classe desiguais podem gerar problemas na criação dos gráficos e só devem ser usados quando aberturas grandes acontecerem nos dados.

Se possível devem ser evitadas classes em no inicio ou no fim da variação porque eles não comunicam a variação das observações com precisão.

Se possível, deveriam ser escolhidos limites de classe de forma que a maioria das observações na classe ficasse mais proximo de ponto central da classe do que do final da classe.

Conte o número de observações em cada classe. Se você estiver construindo uma tabela de freqüência, você precisa somente do número de observações que ocorrem dentro da classe.