

---

# CONCEITOS ELEMENTARES DE ESTATÍSTICA

---

[CEQ e Data Mining](#) - Seminário de 2/07/2003

[Palestra para calouros do CSE - 11/05/2006](#)

---

FONTE: arquivo de ajuda do software STATISTICA (StatSoft Inc.)

Este texto contém os principais conceitos necessários para a utilização de métodos estatísticos na resolução dos mais variados problemas. A leitura de TODOS os itens abaixo é altamente recomendável para que você possa compreender e aplicar corretamente métodos estatísticos.

- [Conceitos básicos.](#)
  - [Relacionamento entre variáveis.](#)
  - [O que é "significância" estatística?](#)
  - ["Formato geral" de muitos testes estatísticos.](#)
  - [Por que a distribuição normal é importante?](#)
- 

## **O que são variáveis?**

Variáveis são características que são medidas, controladas ou manipuladas em uma pesquisa. Diferem em muitos aspectos, principalmente no papel que a elas é dado em uma pesquisa e na forma como podem ser medidas.

## **Pesquisa "Correlacional" X Pesquisa "Experimental"**

A maioria das pesquisas empíricas pertencem claramente a uma dessas duas categorias gerais: em uma pesquisa correlacional (Levantamento) o pesquisador não influencia (ou tenta não influenciar) nenhuma variável, mas apenas as mede e procura por relações (correlações) entre elas, como pressão sangüínea e nível de colesterol. Em uma pesquisa experimental (Experimento) o pesquisador manipula algumas variáveis e então mede os efeitos desta manipulação em outras variáveis; por exemplo, aumentar artificialmente a pressão sangüínea e registrar o nível de colesterol. A análise dos dados em uma pesquisa experimental também calcula "correlações" entre variáveis, especificamente entre aquelas manipuladas e as que foram afetadas pela manipulação. Entretanto, os dados experimentais podem demonstrar conclusivamente relações causais (causa e efeito) entre variáveis. Por exemplo, se o pesquisador descobrir que sempre que muda a variável A então a variável B também muda,

então ele poderá concluir que A "influencia" B. Dados de uma pesquisa correlacional podem ser apenas "interpretados" em termos causais com base em outras teorias (não estatísticas) que o pesquisador conheça, mas não podem ser conclusivamente provar causalidade.

## Variáveis dependentes e variáveis independentes.

Variáveis independentes são aquelas que são manipuladas enquanto que variáveis dependentes são apenas medidas ou registradas. Esta distinção confunde muitas pessoas que dizem que "todas variáveis dependem de alguma coisa". Entretanto, uma vez que se esteja acostumado a esta distinção ela se torna indispensável. Os termos variável dependente e independente aplicam-se principalmente à pesquisa experimental, onde algumas variáveis são manipuladas, e, neste sentido, são "independentes" dos padrões de reação inicial, intenções e características dos sujeitos da pesquisa (unidades experimentais). Espera-se que outras variáveis sejam "dependentes" da manipulação ou das condições experimentais. Ou seja, elas dependem "do que os sujeitos farão" em resposta. Contrariando um pouco a natureza da distinção, esses termos também são usados em estudos em que não se manipulam variáveis independentes, literalmente falando, mas apenas se designam sujeitos a "grupos experimentais" baseados em propriedades pré-existentes dos próprios sujeitos. Por exemplo, se em uma pesquisa compara-se a contagem de células brancas (White Cell Count em inglês, WCC) de homens e mulheres, sexo pode ser chamada de variável independente e WCC de variável dependente.

## Níveis de Mensuração

As variáveis diferem em "quão bem" elas podem ser medidas, isto é, em quanta informação seu nível de mensuração pode prover. Há obviamente algum erro em cada medida, o que determina o "montante de informação" que se pode obter, mas basicamente o fator que determina a quantidade de informação que uma variável pode prover é o seu tipo de nível de mensuração. Sob este prisma as variáveis são classificadas como (a) nominais, (b) ordinais e (c) intervalares.

(a) Variáveis nominais permitem apenas classificação qualitativa. Ou seja, elas podem ser medidas apenas em termos de quais itens pertencem a diferentes categorias, mas não se pode quantificar nem mesmo ordenar tais categorias. Por exemplo, pode-se dizer que 2 indivíduos são diferentes em termos da variável A (sexo, por exemplo), mas não se pode dizer qual deles "tem mais" da qualidade representada pela variável. Exemplos típicos de variáveis nominais são sexo, raça, cidade, etc.

(b) Variáveis ordinais permitem ordenar os itens medidos em termos de qual tem menos e qual tem mais da qualidade representada pela variável, mas ainda não permitem que se diga "o quanto mais". Um exemplo típico de uma variável ordinal é o status sócio-econômico das famílias residentes em uma localidade: sabe-se que média-alta é mais "alta" do que média, mas não se pode dizer, por exemplo, que é 18% mais alta. A própria distinção entre mensuração nominal, ordinal e intervalar representa um bom exemplo de uma variável ordinal: pode-se dizer que uma medida nominal provê menos informação do que uma medida

ordinal, mas não se pode dizer "quanto menos" ou como esta diferença se compara à diferença entre mensuração ordinal e intervalar.

(c) Variáveis intervalares permitem não apenas ordenar em postos os itens que estão sendo medidos, mas também quantificar e comparar o tamanho das diferenças entre eles. Por exemplo, temperatura, medida em graus Celsius constitui uma variável intervalar. Pode-se dizer que a temperatura de 40C é maior do que 30C e que um aumento de 20C para 40C é duas vezes maior do que um aumento de 30C para 40C.

---

## Relações entre variáveis

Duas ou mais variáveis quaisquer estão relacionadas se em uma amostra de observações os valores dessas variáveis são distribuídos de forma consistente. Em outras palavras, as variáveis estão relacionadas se seus valores correspondem sistematicamente uns aos outros para aquela amostra de observações. Por exemplo, sexo e WCC seriam relacionados se a maioria dos homens tivesse alta WCC e a maioria das mulheres baixa WCC, ou vice-versa; altura é relacionada ao peso porque tipicamente indivíduos altos são mais pesados do que indivíduos baixos; Q.I. está relacionado ao número de erros em um teste se pessoas com Q.I.'s mais altos cometem menos erros.

## Por que relações entre variáveis são importantes?

Geralmente o objetivo principal de toda pesquisa ou análise científica é encontrar relações entre variáveis. A filosofia da ciência ensina que não há outro meio de representar "significado" exceto em termos de relações entre quantidades ou qualidades, e ambos os casos envolvem relações entre variáveis. Assim, o avanço da ciência sempre tem que envolver a descoberta de novas relações entre variáveis. Em pesquisas correlacionais a medida destas relações é feita de forma bastante direta, bem como nas pesquisas experimentais. Por exemplo, o experimento já mencionado de comparar WCC em homens e mulheres pode ser descrito como procura de uma correlação entre 2 variáveis: sexo e WCC. A Estatística nada mais faz do que auxiliar na avaliação de relações entre variáveis.

## Aspectos básicos da relação entre variáveis

As duas propriedades formais mais elementares de qualquer relação entre variáveis são a magnitude ("tamanho") e a confiabilidade da relação.

(a) Magnitude é muito mais fácil de entender e medir do que a confiabilidade. Por

exemplo, se cada homem em nossa amostra tem um WCC maior do que o de qualquer mulher da amostra, poderia-se dizer que a magnitude da relação entre as duas variáveis (sexo e WCC) é muito alta em nossa amostra. Em outras palavras, poderia-se prever uma baseada na outra (ao menos na amostra em questão).

(b) Confiabilidade é um conceito muito menos intuitivo, mas extremamente importante. Relaciona-se à "representatividade" do resultado encontrado em uma amostra específica de toda a população. Em outras palavras, diz quão provável será encontrar uma relação similar se o experimento fosse feito com outras amostras retiradas da mesma população, lembrando que o maior interesse está na população. O interesse na amostra reside na informação que ela pode prover sobre a população. Se o estudo atender certos critérios específicos (que serão mencionados posteriormente) então a confiabilidade de uma relação observada entre variáveis na amostra pode ser estimada quantitativamente e representada usando uma medida padrão (chamada tecnicamente de nível-p ou nível de significância estatística).

---

## O que é "significância estatística" (nível-p)

A significância estatística de um resultado é uma medida estimada do grau em que este resultado é "verdadeiro" (no sentido de que seja realmente o que ocorre na população, ou seja no sentido de "representatividade da população"). Mais tecnicamente, o valor do nível-p representa um índice decrescente da confiabilidade de um resultado. Quanto mais alto o nível-p, menos se pode acreditar que a relação observada entre as variáveis na amostra é um indicador confiável da relação entre as respectivas variáveis na população. Especificamente, o nível-p representa a probabilidade de erro envolvida em aceitar o resultado observado como válido, isto é, como "representativo da população". Por exemplo, um nível-p de 0,05 (1/20) indica que há 5% de probabilidade de que a relação entre as variáveis, encontrada na amostra, seja um "acaso feliz". Em outras palavras, assumindo que não haja relação entre aquelas variáveis na população, e o experimento de interesse seja repetido várias vezes, poderia-se esperar que em aproximadamente 20 realizações do experimento haveria apenas uma em que a relação entre as variáveis em questão seria igual ou mais forte do que a que foi observada naquela amostra anterior. Em muitas áreas de pesquisa, o nível-p de 0,05 é costumeiramente tratado como um "limite aceitável" de erro.

## Como determinar que um resultado é "realmente" significativo?

Não há meio de evitar arbitrariedade na decisão final de qual nível de significância será tratado como realmente "significante". Ou seja, a seleção de um nível de significância acima do qual os resultados serão rejeitados como inválidos é arbitrária. Na prática, a decisão final depende usualmente de: se o resultado foi previsto a priori ou apenas a posteriori no curso de muitas análises e comparações efetuadas no conjunto de dados; no total de evidências consistentes do conjunto de dados; e nas "tradições" existentes na área particular de pesquisa. Tipicamente, em muitas ciências resultados que atingem nível-p 0,05 são considerados estatisticamente significantes, mas este nível ainda envolve uma probabilidade de erro razoável (5%). Resultados com um nível-p 0,01 são comumente considerados estatisticamente

significantes, e com nível-p 0,005 ou nível-p 0,001 são frequentemente chamados "altamente" significantes. Estas classificações, porém, são convenções arbitrárias e apenas informalmente baseadas em experiência geral de pesquisa. Uma consequência óbvia é que um resultado considerado significativo a 0,05, por exemplo, pode não sê-lo a 0,01.

## **Significância estatística e o número de análises realizadas**

Desnecessário dizer quanto mais análises sejam realizadas em um conjunto de dados, mais os resultados atingirão "por acaso" o nível de significância convencional. Por exemplo, ao calcular correlações entre dez variáveis (45 diferentes coeficientes de correlação), seria razoável esperar encontrar por acaso que cerca de dois (um em cada 20) coeficientes de correlação são significantes ao nível-p 0,05, mesmo que os valores das variáveis sejam totalmente aleatórios, e aquelas variáveis não se correlacionem na população. Alguns métodos estatísticos que envolvem muitas comparações, e portanto uma boa chance para tais erros, incluem alguma "correção" ou ajuste para o número total de comparações. Entretanto, muitos métodos estatísticos (especialmente análises exploratórias simples de dados) não oferecem nenhum remédio direto para este problema. Cabe então ao pesquisador avaliar cuidadosamente a confiabilidade de descobertas não esperadas.

## **Força X Confiabilidade de uma relação entre variáveis**

Foi dito anteriormente que força (magnitude) e confiabilidade são dois aspectos diferentes dos relacionamentos entre variáveis. Contudo, eles não são totalmente independentes. Em geral, em uma amostra de um certo tamanho quanto maior a magnitude da relação entre variáveis, mais confiável a relação.

Assumindo que não há relação entre as variáveis na população, o resultado mais provável deveria ser também não encontrar relação entre as mesmas variáveis na amostra da pesquisa. Assim, quanto mais forte a relação encontrada na amostra menos provável é a não existência da relação correspondente na população. Então a magnitude e a significância de uma relação aparentam estar fortemente relacionadas, e seria possível calcular a significância a partir da magnitude e vice-versa. Entretanto, isso é válido apenas se o tamanho da amostra é mantido constante, porque uma relação de certa força poderia ser tanto altamente significativa ou não significativa de todo dependendo do tamanho da amostra.

## **Por que a significância de uma relação entre variáveis depende do tamanho da amostra?**

Se há muito poucas observações então há também poucas possibilidades de combinação dos valores das variáveis, e então a probabilidade de obter por acaso uma combinação desses valores que indique uma forte relação é relativamente alta. Considere-se o seguinte exemplo:



Há interesse em duas variáveis (sexo: homem, mulher; WCC: alta, baixa) e há apenas quatro sujeitos na amostra (2 homens e 2 mulheres). A probabilidade de se encontrar, puramente por acaso, uma relação de 100% entre as duas variáveis pode ser tão alta quanto 1/8. Explicando, há uma chance em oito de que os dois homens tenham alta WCC e que as duas mulheres tenham baixa WCC, ou vice-versa, mesmo que tal relação não exista na população. Agora considere-se a probabilidade de obter tal resultado por acaso se a amostra consistisse de 100 sujeitos: a probabilidade de obter aquele resultado por acaso seria praticamente zero.

Observando um exemplo mais geral. Imagine-se uma população teórica em que a média de WCC em homens e mulheres é exatamente a mesma. Supondo um experimento em que se retiram pares de amostras (homens e mulheres) de um certo tamanho da população e calcula-se a diferença entre a média de WCC em cada par de amostras (supor ainda que o experimento será repetido várias vezes). Na maioria dos experimentos os resultados das diferenças serão próximos de zero. Contudo, de vez em quando, um par de amostra apresentará uma diferença entre homens e mulheres consideravelmente diferente de zero. Com que frequência isso acontece? Quanto menor a amostra em cada experimento maior a probabilidade de obter esses resultados errôneos, que, neste caso, indicariam a existência de uma relação entre sexo e WCC obtida de uma população em que tal relação não existe. Observe-se mais um exemplo ("razão meninos para meninas", Nisbett et al., 1987):

Há dois hospitais: no primeiro nascem 120 bebês a cada dia e no outro apenas 12. Em média a razão de meninos para meninas nascidos a cada dia em cada hospital é de 50/50. Contudo, certo dia, em um dos hospitais nasceram duas vezes mais meninas do que meninos. Em que hospital isso provavelmente aconteceu? A resposta é óbvia para um estatístico, mas não tão óbvia para os leigos: é muito mais provável que tal fato tenha ocorrido no hospital menor. A razão para isso é que a probabilidade de um desvio aleatório da média da população aumenta com a diminuição do tamanho da amostra (e diminui com o aumento do tamanho da amostra).

### **Por que pequenas relações podem ser provadas como significantes apenas por grandes amostras?**

Os exemplos dos parágrafos anteriores indicam que se um relacionamento entre as variáveis em questão (na população) é pequeno, então não há meio de identificar tal relação em um estudo a não ser que a amostra seja correspondentemente grande. Mesmo que a amostra seja de fato "perfeitamente representativa" da população o efeito não será estatisticamente significativo se a amostra for pequena. Analogamente, se a relação em questão é muito grande na população então poderá ser constatada como altamente significativa mesmo em um estudo baseado em uma pequena amostra. Mais um exemplo:

Se uma moeda é ligeiramente viciada, de tal forma que quando lançada é ligeiramente mais provável que ocorram caras do que coroas (por exemplo uma proporção 60% para 40%). Então dez lançamentos não seriam suficientes para convencer alguém de que a moeda é viciada, mesmo que o resultado obtido (6 caras e 4 coroas) seja perfeitamente representativo

do viesamento da moeda. Entretanto, dez lançamentos não são suficientes para provar nada? Não, se o efeito em questão for grande o bastante, os dez lançamentos serão suficientes. Por exemplo, imagine-se que a moeda seja tão viciada que não importe como venha a ser lançada o resultado será cara. Se tal moeda fosse lançada dez vezes, e cada lançamento produzisse caras, muitas pessoas considerariam isso prova suficiente de que há "algo errado" com a moeda. Em outras palavras, seria considerada prova convincente de que a população teórica de um número infinito de lançamentos desta moeda teria mais caras do que coroas. Assim, se a relação é grande, então poderá ser considerada significativa mesmo em uma pequena amostra.

### **Pode uma "relação inexistente" ser um resultado significativo?**

Quanto menor a relação entre as variáveis maior o tamanho de amostra necessário para prová-la significativa. Por exemplo, imagine-se quantos lançamentos seriam necessários para provar que uma moeda é viciada se seu viesamento for de apenas 0,000001 %! Então, o tamanho mínimo de amostra necessário cresce na mesma proporção em que a magnitude do efeito a ser demonstrado decresce. Quando a magnitude do efeito aproxima-se de zero, o tamanho de amostra necessário para prová-lo aproxima-se do infinito. Isso quer dizer que, se quase não há relação entre duas variáveis o tamanho da amostra precisa quase ser igual ao tamanho da população, que teoricamente é considerado infinitamente grande. A significância estatística representa a probabilidade de que um resultado similar seja obtido se toda a população fosse testada. Assim, qualquer coisa que fosse encontrada após testar toda a população seria, por definição, significativa ao mais alto nível possível, e isso também inclui todos os resultados de "relação inexistente".

### **Como medir a magnitude (força) das relações entre variáveis?**

Há muitas medidas da magnitude do relacionamento entre variáveis que foram desenvolvidas por estatísticos: a escolha de uma medida específica em dadas circunstâncias depende do número de variáveis envolvidas, níveis de mensuração usados, natureza das relações, etc. Quase todas, porém, seguem um princípio geral: elas procuram avaliar a relação comparando-a de alguma forma com a "máxima relação imaginável" entre aquelas variáveis específicas. Tecnicamente, um modo comum de realizar tais avaliações é observar quão diferenciados são os valores das variáveis, e então calcular qual parte desta "diferença global disponível" seria detectada na ocasião se aquela diferença fosse "comum" (fosse apenas devida à relação entre as variáveis) nas duas (ou mais) variáveis em questão. Falando menos tecnicamente, compare-se "o que é comum naquelas variáveis" com "o que potencialmente poderia haver em comum se as variáveis fossem perfeitamente relacionadas". Outro exemplo:

Em uma amostra o índice médio de WCC é igual a 100 em homens e 102 em mulheres. Assim, poderia-se dizer que, em média, o desvio de cada valor da média de ambos (101) contém uma componente devida ao sexo do sujeito, e o tamanho desta componente é 1. Este valor, em certo sentido, representa uma medida da relação entre sexo e WCC. Contudo, este

valor é uma medida muito pobre, porque não diz quão relativamente grande é aquela componente em relação à "diferença global" dos valores de WCC. Há duas possibilidades extremas:

(a) Se todos os valores de WCC de homens são exatamente iguais a 100 e os das mulheres iguais a 102 então todos os desvios da média conjunta na amostra seriam inteiramente causados pelo sexo. Poderia-se dizer que nesta amostra sexo é perfeitamente correlacionado a WCC, ou seja, 100% das diferenças observadas entre os sujeitos relativas a suas WCC's devem-se a seu sexo.

(b) Se todos os valores de WCC estão em um intervalo de 0 a 1000, a mesma diferença (de 2) entre a WCC média de homens e mulheres encontrada no estudo seria uma parte tão pequena na diferença global dos valores que muito provavelmente seria considerada desprezível. Por exemplo, um sujeito a mais que fosse considerado poderia mudar, ou mesmo reverter, a direção da diferença. Portanto, toda boa medida das relações entre variáveis tem que levar em conta a diferenciação global dos valores individuais na amostra e avaliar a relação em termos (relativos) de quanto desta diferenciação se deve à relação em questão.

---

## "Formato geral" de muitos testes estatísticos

Como o objetivo principal de muitos testes estatísticos é avaliar relações entre variáveis, muitos desses testes seguem o princípio exposto no item anterior. Tecnicamente, eles representam uma razão de alguma medida da diferenciação comum nas variáveis em análise (devido à sua relação) pela diferenciação global daquelas variáveis. Por exemplo, teria-se uma razão da parte da diferenciação global dos valores de WCC que podem se dever ao sexo pela diferenciação global dos valores de WCC. Esta razão é usualmente chamada de razão da variação explicada pela variação total. Em estatística o termo variação explicada não implica necessariamente que tal variação é "compreendida conceitualmente". O termo é usado apenas para denotar a variação comum às variáveis em questão, ou seja, a parte da variação de uma variável que é "explicada" pelos valores específicos da outra variável e vice-versa.

## Como é calculado o nível de significância estatístico

Assuma-se que já tenha sido calculada uma medida da relação entre duas variáveis (como explicado acima). A próxima questão é "quão significativa é esta relação"? Por exemplo, 40% da variação global ser explicada pela relação entre duas variáveis é suficiente para considerar a relação significativa? "Depende". Especificamente, a significância depende principalmente do tamanho da amostra. Como já foi explicado, em amostras muito grandes mesmo relações muito pequenas entre variáveis serão significantes, enquanto que em amostras muito pequenas mesmo relações muito grandes não poderão ser consideradas confiáveis (significantes). Assim, para determinar o nível de significância estatística torna-se necessária uma função que represente o relacionamento entre "magnitude" e "significância" das relações entre duas



variáveis, dependendo do tamanho da amostra. Tal função diria exatamente "quão provável é obter uma relação de dada magnitude (ou maior) de uma amostra de dado tamanho, assumindo que não há tal relação entre aquelas variáveis na população". Em outras palavras, aquela função forneceria o nível de significância (nível-p), e isso permitiria conhecer a probabilidade de erro envolvida em rejeitar a idéia de que a relação em questão não existe na população. Esta hipótese "alternativa" (de que não há relação na população) é usualmente chamada de **hipótese nula**. Seria ideal se a função de probabilidade fosse linear, e por exemplo, apenas tivesse diferentes inclinações para diferentes tamanhos de amostra. Infelizmente, a função é mais complexa, e não é sempre exatamente a mesma. Entretanto, em muitos casos, sua forma é conhecida e isso pode ser usado para determinar os níveis de significância para os resultados obtidos em amostras de certo tamanho. Muitas daquelas funções são relacionadas a um tipo geral de função que é chamada de **normal** (ou gaussiana).

---

## Por que a distribuição normal é importante?

A "distribuição normal" é importante porque em muitos casos ela se aproxima bem da função introduzida no item anterior. A distribuição de muitas estatísticas de teste é normal ou segue alguma forma que pode ser derivada da distribuição normal. Neste sentido, filosoficamente, a distribuição normal representa uma das elementares "verdades acerca da natureza geral da realidade", verificada empiricamente, e seu status pode ser comparado a uma das leis fundamentais das ciências naturais. A forma exata da distribuição normal (a característica "curva do sino") é definida por uma função que tem apenas dois parâmetros: média e desvio padrão.

Uma propriedade característica da distribuição normal é que 68% de todas as suas observações caem dentro de um intervalo de 1 desvio padrão da média, um intervalo de 2 desvios padrões inclui 95% dos valores, e 99% das observações caem dentro de um intervalo de 3 desvios padrões da média. Em outras palavras, em uma distribuição normal as observações que tem um valor padronizado de menos do que -2 ou mais do que +2 tem uma frequência relativa de 5% ou menos (valor padronizado significa que um valor é expresso em termos de sua diferença em relação à média, dividida pelo desvio padrão).

## Ilustração de como a distribuição normal é usada em raciocínio estatístico (indução)

Retomando o exemplo já discutido, onde pares de amostras de homens e mulheres foram retirados de uma população em que o valor médio de WCC em homens e mulheres era exatamente o mesmo. Embora o resultado mais provável para tais experimentos (um par de amostras por experimento) é que a diferença entre a WCC média em homens e mulheres em cada par seja próxima de zero, de vez em quando um par de amostras apresentará uma diferença substancialmente diferente de zero. Quão frequentemente isso ocorre? Se o tamanho da amostra é grande o bastante, os resultados de tais repetições são "normalmente distribuídos", e assim, conhecendo a forma da curva normal pode-se calcular precisamente a

probabilidade de obter "por acaso" resultados representando vários níveis de desvio da hipotética média populacional 0 (zero). Se tal probabilidade calculada é tão pequena que satisfaz ao critério previamente aceito de significância estatística, então pode-se concluir que o resultado obtido produz uma melhor aproximação do que está acontecendo na população do que a "hipótese nula". Lembrando ainda que a hipótese nula foi considerada apenas por "razões técnicas" como uma referência contra a qual o resultado empírico (dos experimentos) foi avaliado.

## **Todos os testes estatísticos são normalmente distribuídos?**

Não todos, mas muitos são ou baseados na distribuição normal diretamente ou em distribuições a ela relacionadas, e que podem ser derivadas da normal, como as distribuições t, F ou Chi-quadrado (Qui-quadrado). Tipicamente, estes testes requerem que as variáveis analisadas sejam normalmente distribuídas na população, ou seja, que elas atendam à "suposição de normalidade". Muitas variáveis observadas realmente são normalmente distribuídas, o que é outra razão por que a distribuição normal representa uma "característica geral" da realidade empírica. O problema pode surgir quando se tenta usar um teste baseado na distribuição normal para analisar dados de variáveis que não são normalmente distribuídas. Em tais casos há duas opções. Primeiramente, pode-se usar algum teste "não paramétrico" alternativo (ou teste "livre de distribuição"); mas isso é freqüentemente inconveniente porque tais testes são tipicamente menos poderosos e menos flexíveis em termos dos tipos de conclusões que eles podem proporcionar. Alternativamente, em muitos casos ainda se pode usar um teste baseado na distribuição normal se apenas houver certeza de que o tamanho das amostras é suficientemente grande. Esta última opção é baseada em um princípio extremamente importante que é largamente responsável pela popularidade dos testes baseados na distribuição normal. Nominalmente, quanto mais o tamanho da amostra aumente, mais a forma da distribuição amostral (a distribuição de uma estatística da amostra) da média aproxima-se da forma da normal, mesmo que a distribuição da variável em questão não seja normal. Este princípio é chamado de **Teorema Central do Limite**.

## **Como se conhece as conseqüências de violar a suposição de normalidade?**

Embora muitas das declarações feitas anteriormente possam ser provadas matematicamente, algumas não têm provas teóricas e podem demonstradas apenas empiricamente via experimentos Monte Carlo (simulações usando geração aleatória de números). Nestes experimentos grandes números de amostras são geradas por um computador seguindo especificações pré-designadas e os resultados de tais amostras são analisados usando uma grande variedade de testes. Este é o modo empírico de avaliar o tipo e magnitude dos erros ou viesamentos a que se expõe o pesquisador quando certas suposições teóricas dos testes usados não são verificadas nos dados sob análise. Especificamente, os estudos de Monte Carlo foram usados extensivamente com testes baseados na distribuição normal para determinar quão

sensíveis eles eram à violações da suposição de que as variáveis analisadas tinham distribuição normal na população. A conclusão geral destes estudos é que as consequências de tais violações são menos severas do que se tinha pensado a princípio. Embora estas conclusões não devam desencorajar ninguém de se preocupar com a suposição de normalidade, elas aumentaram a popularidade geral dos testes estatísticos dependentes da distribuição normal em todas as áreas de pesquisa.

---

**[RETORNAR](#)**