TESTE DE HIPÓTESES 3

Dirce Maria Trevisan Zanetta

3.1 Introdução

Nós já vimos como uma amostra pode representar a população da qual é retirada e como, devido à variabilidade entre os indivíduos que compõem essa amostra, dificilmente por meio de sua estimativa são encontrados valores iguais aos da população. Entretanto, vimos também que existe grande chance de uma amostra estimar um valor próximo do real; essa chance aumenta com o aumento do número de indivíduos na amostra. Estudamos que a distribuição de probabilidades de uma variável permite estimar a probabilidade de obtermos um resultado em uma amostra, bem como o que é um erro do tipo I e um erro do tipo II, que podem ocorrer nas conclusões tiradas com base em resultados amostrais.

Esse conhecimento será útil para entender como são feitos os testes de hipóteses que estudaremos nesta aula.

Uma das formas mais utilizadas para a inferência estatística é a realização de **testes de significância estatística** ou os chamados **testes de hipóteses**. Um teste de hipóteses justapõe duas hipóteses contraditórias e, ao final, auxilia na decisão por uma delas: a **hipótese nula** (H_0 – lê-se "agá zero") e a **hipótese alternativa** (H_A – lê-se "agá A").

3.2 Hipótese nula e alternativa

A primeira etapa de um teste de hipóteses consiste **em estabelecer a hipótese nula (H_0)**. Na maioria das vezes, essa é a hipótese de que um parâmetro é igual a um certo valor já conhecido ou de que a diferença entre dois parâmetros é igual a zero. A hipótese nula, então, indica não haver resposta, mudança ou diferença no que está sendo testado. Como exemplo, para testar se os níveis séricos de colesterol (medida da concentração de colesterol no sangue) de pessoas que se exercitam é diferente do de pessoas sedentárias, a hipótese nula seria a de que o nível sérico de colesterol de pessoas ativas é igual ao de pessoas sedentárias. A hipótese nula é aceita ou rejeitada, com base em um critério estabelecido antes da coleta de dados, que veremos a seguir. Como uma regra, lembrar que o termo igual estará quase sempre na hipótese nula. Em um teste de hipóteses, a princípio, **a hipótese nula é considerada a verdadeira**.

A ideia de se estabelecer uma hipótese nula é comum mesmo em um raciocínio não estatístico. É exatamente o que é feito em processos criminais, em que um acusado (réu) é considerado inocente até que se prove o contrário. A pressuposição de inocência é uma hipótese nula. A "inocência" no teste de hipóteses é a ausência de diferença.

Como o nome indica, **a hipótese alternativa (H_A) será a que vai se contrapor à hipótese nula** e, portanto, será a hipótese de que um parâmetro é diferente de certo valor ou de que a diferença entre dois parâmetros é diferente de zero, isto é, que existe diferença entre os dois parâmetros. No exemplo acima, a hipótese alternativa é a de que o nível sérico de colesterol de pessoas que se exercitam é diferente do de pessoas sedentárias. Quando os dados de uma amostra são incompatíveis com a hipótese nula, e ela é rejeitada, conclui-se que os resultados da amostra são compatíveis com a hipótese alternativa.

Estabelecidas H_0 e H_A, o passo seguinte é considerar os erros possíveis envolvidos em uma decisão por testes de hipóteses. Os erros tipo I, também chamado erro α, e tipo II, ou erro β, que vimos na última aula também se aplicam ao teste de hipóteses, como mostrado na **Tabela 3.1**, relacionando-os com a decisão de um teste de hipóteses.

Tabela 3.1: Tipos de erros em testes de hipóteses.

Decisão estatística	Natureza (estado verdadeiro e desconhecido)	
	Hipótese nula verdadeira	Hipótese nula falsa
Aceitar hipótese nula	Acerto	Erro Tipo II (β)
Rejeitar hipótese nula	Erro Tipo I (α)	Acerto

Se a decisão do teste é de que o nível sérico de colesterol de pessoas ativas e sedentárias são diferentes, mas a diferença observada for devida apenas ao acaso, ocorre o erro do tipo I. Por outro lado, se a conclusão é que são iguais, mas na verdade existe diferença entre os dois grupos, ocorre nesse caso o erro do tipo II.

3.3 Etapas de um teste de hipóteses

Primeiro, o pesquisador deve definir quais os **níveis aceitáveis de erro tipo II (β)** quando está sendo estimado o tamanho adequado das amostras de estudo e isso porque existe uma relação inversa entre o tamanho de uma amostra e a possibilidade de cometer esse tipo de erro. Isso quer dizer que, em amostras pequenas, esse erro tem maior chance de ocorrer e, com o aumento do tamanho da amostra, diminui a possibilidade do erro tipo II ou β, que é concluir por aceitar a hipótese nula de que não existe diferença entre os grupos quando ela existe de fato. Em geral, é aceito um erro β de até 20%. O complemento do erro β, que é 1-β, é o que chamamos de **poder de um teste**. Veja que, quanto maior for uma amostra, o teste estatístico terá mais poder de detectar diferenças, se elas existirem e, portanto, menor será o erro β.

A seguir, deve ser estabelecido, como critério para auxiliar a decisão final do teste, qual a **probabilidade máxima** aceitável de cometer um erro do tipo I ao rejeitar a hipótese nula, e ela for verdadeira, isto é, qual é o **nível do erro tipo I (α) aceitável**. Esse valor, determinado pelo pesquisador antes da análise dos dados, idealmente antes da própria coleta dos dados, é denominado **nível alfa (α) ou nível de significância**.

Existem vários tipos de testes estatísticos que utilizam curvas de distribuição com propriedades bem estudadas e que são utilizadas para avaliar a hipótese nula. Essas curvas de distribuição consistem em todos os valores possíveis de um teste estatístico e suas respectivas probabilidades. Como exemplo, existem testes estatísticos cuja distribuição de valores pode ser uma curva normal, ou uma curva *t*, ou uma binomial, entre outras.

A **escolha do teste a ser utilizado é a próxima etapa** e vai depender do parâmetro que se pretende comparar, como médias, proporções etc. Para a análise estatística, então, deve-se procurar o teste que minimize a possibilidade de ocorrer um erro na tomada de decisão.

O princípio para o teste estatístico é comum a todos eles: é feita a comparação do valor do teste estatístico calculado da amostra com o valor crítico da distribuição que corresponde ao **nível alfa** estabelecido, para decidir se a hipótese nula será aceita ou rejeitada.

O valor do teste avalia a chance de ocorrerem os resultados encontrados em amostras, supondo que a hipótese nula seja verdadeira. Quando essa chance for grande, é fácil aceitar a hipótese nula. Por outro lado, se essa chance for muito pequena, também não temos muito problema em rejeitar a hipótese nula estabelecida. Mas, quão pequena deve ser essa probabilidade para rejeitarmos a hipótese nula?

De modo geral, nas áreas de saúde e biológica, o nível de significância ou nível alfa dos testes é tradicionalmente **fixado em 5%**, isto é, probabilidades maiores que 0,05 (ou, se multiplicarmos por 100, 5%) são consideradas compatíveis com hipótese nula, que é então aceita. Quando é igual ou menor que 5%, considera-se que a probabilidade de ter ocorrido o resultado apenas pelo acaso é pequena e rejeita-se a hipótese nula. Nesse caso, é aceita a hipótese alternativa de que existe uma diferença significante entre os grupos comparados.

Dizer que um resultado é **estatisticamente significante** significa que as diferenças encontradas são suficientemente grandes para não serem atribuídas ao acaso (ou que a chance de diferenças tão grandes ocorrerem ao acaso é muito pequena e, portanto, conclui-se que existe diferença entre os grupos).

Uma forma interessante de compreendermos o que representa o nível alfa de significância de um teste é apresentada por meio de uma analogia feita por Jekel et al. (2005, p. 166):

> "Uma analogia do dia a dia pode ajudar a simplificar a lógica do nível alfa e o processo do teste de significância. Suponha que um jovem casal estava dando instruções para comprar um bracelete de prata para um amigo, durante uma viagem, se fosse possível comprar um por $50 ou menos. Se um bracelete que servisse fosse encontrado, ele seria comprado somente se fosse possível ser obtido por $50 ou menos. Qualquer outro seria muito caro. O alfa é o limite do preço, na analogia. Uma vez estabelecido (digamos 0,05), um investigador iria "comprar" a hipótese alternativa de uma diferença verdadeira, ou associação, se o preço (na probabilidade de estar errado) fosse não mais do que 1 em 20 (0,05). O alfa, portanto, é o preço que um investigador está disposto a pagar na probabilidade de estar errado, se ele rejeitar a hipótese nula".

Esse nível de significância é utilizado para determinar valores do teste estatístico que demarcam, na curva de distribuição de probabilidades, a região de aceitação e de rejeição da H_0. Veja no

Gráfico 3.1 um exemplo com uma curva normal padrão (a curva *z*) de distribuição de como o nível alfa é utilizado para essa delimitação. Lembre-se de que uma curva de distribuição de probabilidades representa todos os resultados possíveis de uma variável com suas respectivas probabilidades. É possível delimitar a região que representa 5% de todos os valores possíveis com seus valores mais extremos. A **região de rejeição da curva** é, então, o conjunto de valores do resultado do teste para os quais a hipótese nula é rejeitada.

Observe que a área nos extremos de uma distribuição do teste estatístico contém valores do teste que são relativamente improváveis, dada uma hipótese nula. Uma vez que valores nessa região se desviam bastante de um valor esperado se a hipótese nula fosse verdadeira, nós decidimos por rejeitá-la. Seu complementar é a região de aceitação da hipótese nula. Valores do teste localizados na região central da curva são compatíveis com a hipótese nula, que então será aceita. Como o teste na maioria das vezes procura decidir se existe ou não diferenças entre parâmetros, a região de rejeição da curva terá $\alpha/2$ em cada um dos extremos, uma vez que a diferença pode ocorrer na direção de um aumento ou diminuição em relação ao parâmetro de comparação.

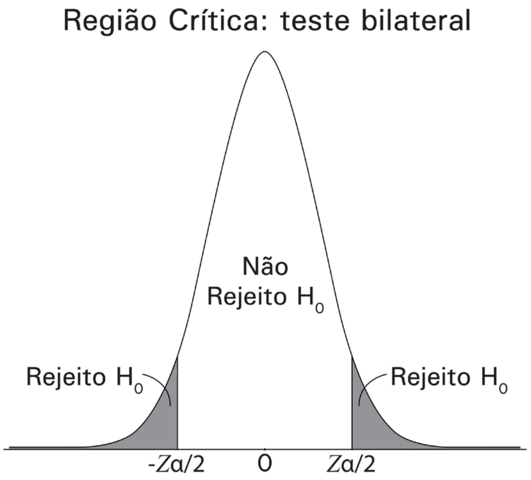

Gráfico 3.1: Região crítica: teste bilateral.

Os valores $-Z_{\alpha/2}$ e $Z_{\alpha/2}$ no **Gráfico 3.1** são os valores da curva que delimitam as áreas à esquerda e à direita, respectivamente, sob a curva normal padrão, para valerem $\alpha/2$, que, como vimos, em geral vai deixar 2,5% da área da curva em cada extremo, para que o total da região de rejeição seja de 5%.

3.4 Testes bicaudais e monocaudais

Esse tipo de teste, denominado teste de hipóteses bicaudal, é feito para detectar se existe alguma diferença (não importa em que direção) entre os parâmetros testados e é o tipo de teste realizado na maioria das vezes.

Em algumas situações deseja-se testar se existe alguma diferença **e em que direção ela está**. Nesse caso, fazemos o chamado **teste de hipóteses monocaudal** para testar se existe diferença **E** se o valor é maior ou se é menor que o parâmetro de referência.

Voltando ao exemplo da comparação dos níveis séricos de colesterol em pessoas ativas e sedentárias, podemos querer testar se pessoas sedentárias tem níveis maiores que pessoas ativas. Nesse caso, as hipóteses a serem testados seriam:

- H_0: o nível sérico de colesterol de pessoas sedentárias é menor ou igual ao de pessoas ativas;
- H_A: o nível sérico de colesterol de pessoas sedentárias é maior que o de pessoas ativas.

> **!** Veja que a igualdade fica na hipótese nula. E lembre-se de que a hipótese nula é a da "inocência", isto é, se o teste é para verificar se um parâmetro é maior, a hipótese nula será a de que ele é igual ou menor ("inocente" de ser maior).

Quando é feito um teste de hipóteses monocaudal, a área de rejeição do teste fica na direção que está sendo testada. Nesse caso, se $\alpha = 0{,}05$, a área de rejeição seria a área correspondente a 5% dos valores localizados no extremo direito da curva, como mostrado no **Gráfico 3.2**. Veja que, neste caso, o nível crítico separa 5% da área da curva apenas em uma rejeição de H_0, pois não há limite para os valores menores pela hipótese nula testada.

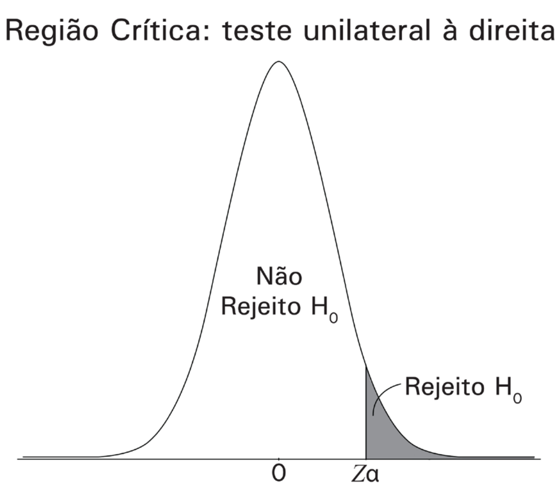

Gráfico 3.2: Região crítica: teste unilateral à direita.

É possível fazer um teste de hipóteses unilateral à esquerda, em que o teste é para verificar se um dos parâmetros é menor que o outro e, neste caso, as hipóteses seriam:

- H_0: o nível sérico de colesterol de pessoas sedentárias é maior ou igual ao de pessoas ativas;
- H_A: o nível sérico de colesterol de pessoas sedentárias é menor que o de pessoas ativas

A região crítica neste caso é representada no **Gráfico 3.3**.

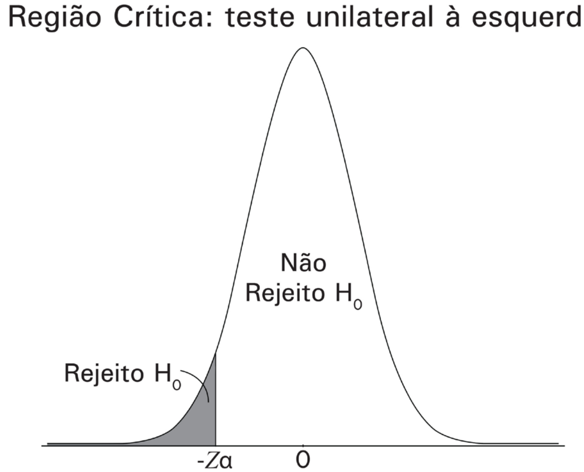

Gráfico 3.3: Região crítica: teste unilateral à esquerda.

Em um teste de hipóteses, a **conclusão é sempre em relação à hipótese nula**, pois a H_A só será aceita se H_0 for rejeitada. Veja que pode haver um erro em qualquer das decisões que se tome, de aceitar ou de rejeitar a hipótese nula. Portanto, a decisão estatística pela hipótese nula ou pela alternativa não prova que não existe ou que existe uma diferença entre grupos, mas apenas diz que a hipótese aceita é confirmada pelos dados e que ela pode ser verdadeira.

As etapas de um teste de hipóteses são resumidas no **Quadro 3.1**.

1. Formular a hipótese nula;
2. Formular a hipótese alternativa;
3. Estabelecer um erro tipo I (nível a) aceitável;
4. Escolher um teste estatístico apropriado para o que se está testando. Depende do tipo de variável e do parâmetro que se pretende comparar;
5. Calcular o valor do teste;
6. Decidir se H_0 é verdadeira ou não.

Quadro 3.1: Etapas de um teste de hipóteses.

A decisão feita na comparação de dois parâmetros é esquematizada na **Figura 3.1**:

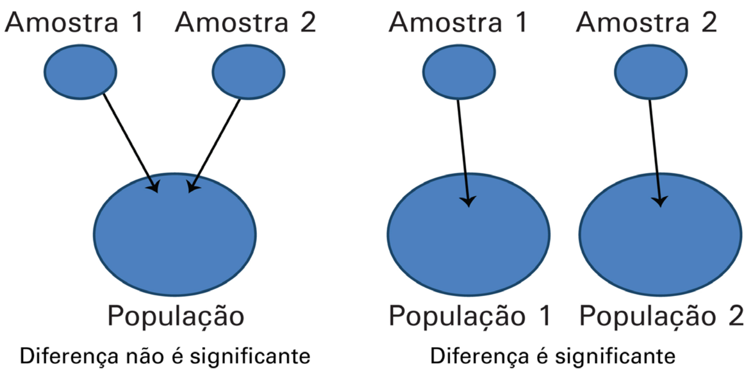

Figura 3.1: Comparação de dois parâmetros.

Quando se concluir que não existe diferença significante entre resultados obtidos em duas amostras, aceita-se que os dados provêm da mesma população e que as diferenças observadas ocorreram pela variabilidade dos indivíduos que compõem a população. Quando a conclusão é a de que existe diferença significante, aceita-se que os dados das duas amostras provêm de duas populações diferentes.

Vamos ver **um exemplo** que pode ajudar na compreensão do processo de um teste de hipóteses. Deseja-se saber se a pressão arterial sistólica (o nível superior da pressão arterial) é diferente em homens e mulheres. Inicialmente, é estabelecida a hipótese nula e, a seguir, a hipótese alternativa:

- H_0: homens têm pressão arterial sistólica média igual à de mulheres;
- H_A: homens têm pressão arterial sistólica média diferente da de mulheres.

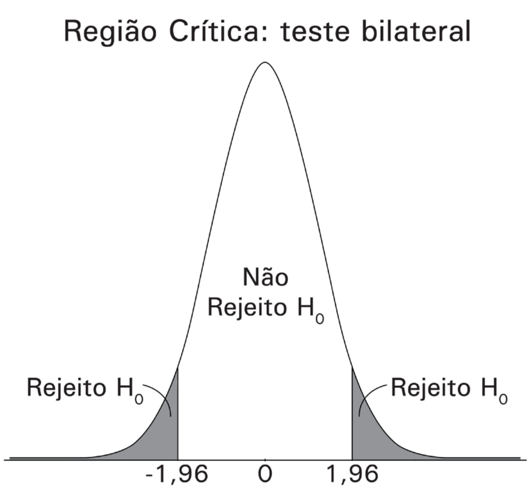

Gráfico 3.4: Região crítica: teste bilateral.

Define-se o nível de significância, geralmente de 5%.

O teste nesse caso é **bilateral**, pois pretende-se avaliar se existe diferença, e essa diferença pode ser por terem os homens a pressão arterial menor ou maior que as mulheres.

Portanto, o nível crítico para definir a região de rejeição e de aceitação da hipótese nula será aquele que deixa 2,5% da curva na região de rejeição de cada lado.

Na curva de distribuição de probabilidades normal padrão, a curva z, 2,5% dos valores no extremo à esquerda são aqueles em que $z \leq -1,96$ e os 2,5%, os que estão no extremo à direita correspondem àqueles em que $z \geq 1,96$.

Para testar a hipótese, mediu-se a pressão arterial sistólica em 100 homens e 100 mulheres. Nos homens, a pressão arterial sistólica média foi de 132 mmHg, com desvio-padrão de 6, e nas mulheres foi de 128 mmHg, com desvio-padrão de 5.

Fazendo o teste para comparação dessas médias, obtém-se como resultado do teste (aceite como correto, não vimos como isso é feito) $z = +5,12$. Como 5,12 é maior que 1,96, esse valor está na região de rejeição da curva (veja no **Gráfico 3.4**: a região de rejeição de H_0 nesse caso compreende todos os valores que forem ≤-1,96 ou ≥1,96). Os dados dessas amostras não apoiam a hipótese de que não existe diferença entre as pressões medidas nos dois sexos e a hipótese nula é rejeitada. Dessa forma, aceita-se a hipótese alternativa e conclui-se que as pressões arteriais médias nos dois sexos são diferentes.

A conclusão de um teste estatístico é feita verificando se o valor calculado do teste está na região de rejeição ou de aceitação da hipótese nula na curva de probabilidades. Essas regiões são delimitadas pelo nível alfa previamente estabelecido. Esse tipo de teste permite concluir se os resultados apoiam ou não a hipótese nula, mas não é possível quantificar quão provável é o resultado de ter ocorrido pelo acaso.

Entretanto, essa quantificação é possível se, com base no resultado do teste estatístico escolhido, for determinado o **nível descritivo do teste ou o valor *p***.

3.5 O valor *p*

O valor *p* é a chance de observar o resultado do teste obtido analisando a amostra, ou qualquer outro resultado mais extremo que ele, supondo que a hipótese nula seja verdadeira. Esta é outra forma de avaliar o resultado de um teste. A etapa final do teste de hipóteses, neste caso, é feita comparando o nível de α estabelecido no início e no valor *p* calculado com o teste estatístico para tomar a decisão de aceitar ou não a hipótese nula.

Se o valor *p* computado do teste for igual ou maior que α, nós aceitamos H_0. Isso significa que a probabilidade de as diferenças observadas serem pelo acaso está entre os níveis que foram estabelecidos no início do teste como aceitáveis para confirmar a hipótese nula. Em outras palavras, a hipótese nula é compatível com os resultados da amostra. Se o valor *p* computado do teste for menor que α, a hipótese nula não é compatível com a amostra e é rejeitada. A hipótese alternativa é então aceita como verdadeira.

No exemplo acima, a região da curva com valores abaixo de −5,12 e acima de +5,12, que são os valores extremos do resultado do teste de hipóteses bicaudal, corresponde a um valor $p < 0{,}001$, isto é, a chance de esse resultado ter ocorrido pelo acaso é menor que 0,01%. Por essa forma de avaliar o teste estatístico, vemos que valor *p* é menor que o nível alfa (estabelecido em 0,05 ou 5%) e, portanto, a decisão é semelhante: rejeita-se H_0 e aceita-se H_A, concluindo que existe diferença entre a pressão arterial sistólica em homens e mulheres.

3.6 Conclusão

Nesta aula, você aprendeu o que é e quais as etapas de um teste de hipóteses, muito utilizado para fazer a inferência estatística. Inicialmente, são estabelecidas as hipóteses nula e alternativa, sendo a nula considerada a verdadeira. Estabelece-se o nível alfa, que em geral na área da saúde é fixado em 5%. Ele permite estabelecer os valores do teste estatístico que demarcam, na curva de distribuição de probabilidades, a região de aceitação e de rejeição da hipótese nula.

Quando se testa se existe diferença, é feito o teste bicaudal, mais frequente. Quando se pretende verificar se existe alguma diferença e em que direção ela está, é feito o teste monocaudal. O resultado do teste estatístico escolhido permite também determinar o valor *p*. A hipótese nula é rejeitada se o valor *p* é menor que o **nível alfa** e não é rejeitada se o valor *p* for igual ou maior que alfa.

Na próxima aula, nós vamos estudar outro tipo de inferência estatística, que é **a estimativa**.

Referências

Bonita, R.; Beaglehole, R.; Kjellström, T. **Epidemiologia Básica**. 2. ed. São Paulo: Santos, 2010.

Dawson-Sanders, B.; Trapp, R. G. **Bioestatística Básica e Clínica**. 3. ed. Rio de Janeiro: Lange-Appleton & Lange/ McGraw-Hill, 2001.

Jekel, J. F.; Katz D. L.; Elmore, J. G. **Epidemiologia, Bioestatística e Medicina Preventiva**. 2. ed. Porto Alegre: Artmed, 2005.

Lopes, A. P. **Probabilidades e Estatística**. Rio de Janeiro: Reichmann & Affonso, 2000.

Magalhães, M. N. **Noções de Probabilidade e Estatística**. 6. ed. São Paulo: Edusp, 2008.

Massad, E. et al. **Métodos Quantitativos em Medicina**. São Paulo: Manole, 2004.

Pagano, M. et al. **Princípios de Bioestatística**. Tradução da 2.ed. norte-americana. São Paulo: Thompson Learning, 2006.

Glossário

Hipótese alternativa (H_A): a que vai se contrapor à hipótese nula, ou a hipótese de que há uma resposta, ou uma mudança ou diferença entre dois parâmetros.

Hipótese nula (H_0): a que indica não haver resposta, mudança ou diferença entre dois parâmetros.

Nível alfa (α) ou nível de significância: a probabilidade máxima que se aceita de cometer um erro do tipo I se rejeitar a hipótese nula, e ela for verdadeira.

Nível descritivo do teste ou valor p: a chance de observar o resultado obtido ou mais extremo que ele com a amostra, supondo que a hipótese nula seja verdadeira.

Região de aceitação: A região de aceitação é o conjunto de valores assumidos pela estatística de teste para os quais a hipótese nula é aceita. Seu complementar é a região de rejeição.

Região de Rejeição: A região de rejeição ou região crítica é o conjunto de valores assumidos pela estatística de teste para os quais a hipótese nula é rejeitada. Seu complementar é a região de aceitação.

Teste de Hipóteses bicaudal: teste feito para detectar se existe alguma diferença (não importa em que direção) entre os parâmetros testados.

Teste de Hipóteses monocaudal: testa se existe alguma diferença entre os parâmetros testados **E** em que direção ela está.

Teste de hipóteses: utilizado para a inferência estatística, justapõe duas hipóteses contraditórias para decidir por uma delas: a **hipótese nula** e a **hipótese alternativa**.