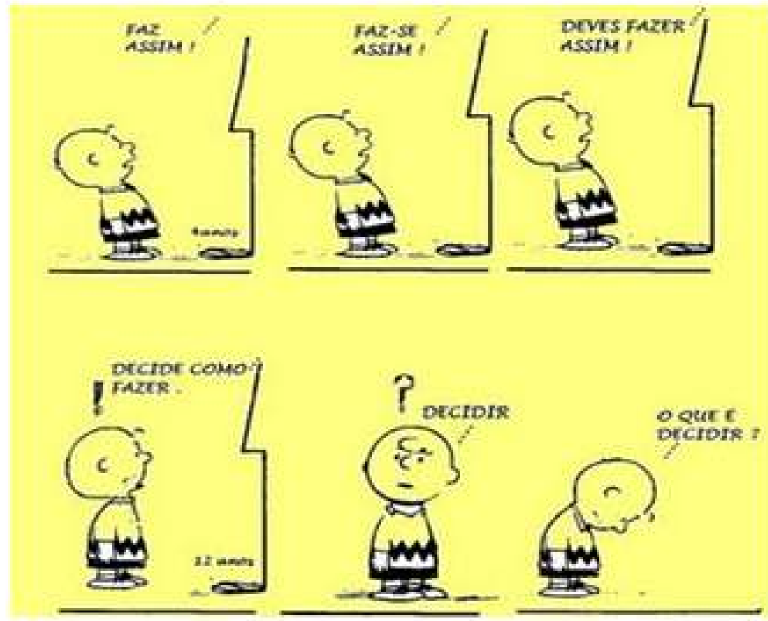
CAPÍTULO 10: INTRODUÇÃO À TEORIA DOS TESTES DE HIPÓTESE

Conceitos e resultados a serem apresentados neste capítulo:

- 10.1 – Conceitos Básicos
  - Hipótese nula e Hipótese alternativa
  - Região de Aceitação e Região de Rejeição ou Região Crítica
  - Ponto de corte ou valor crítico, Nível de Significância, Estatística de teste
  - Erro I e Erro II. Probabilidades de erro: α e β
- 10.2 – Esclarecendo melhor alguns conceitos
  - Como devemos especificar H_0 e H_1
  - O que são Hipótese simples e hipótese composta
  - H_0 vista como uma Hipótese simples
- 10.2 – Rotina para obtenção do Critério de Decisão
- 10.4 -Teste sobre a média populacional
- 10.5 -p-valor ou Nível crítico
- 10.7 -Teste sobre a proporção populacional



"O alvo do cientista não é descobrir uma certeza absoluta, mas descobrir teorias cada vez melhores, capazes de serem submetidas a testes cada vez mais severos, conduzindo a novas experiências, que iluminam a nossa visão."

Karl Popper, filósofo

Nos dois Capítulos anteriores vimos como, por meio do uso de uma amostra aleatória, é possível estimar um ou mais parâmetros populacionais desconhecidos, associados à distribuição de probabilidades de uma variável aleatória X.

No presente Capítulo também lidamos com **amostras aleatórias** e **parâmetros populacionais**. Porém, agora formularemos uma hipótese sobre o valor do parâmetro de interesse e **a informação obtida a partir dos dados amostrais será usada para confirmar ou refutar essa hipótese**.

Neste Capítulo 10, além de apresentarmos conceitos e resultados gerais da teoria dos Testes de Hipótese (TH), discutiremos também dois procedimentos específicos de TH, a saber, **Testes sobre a média populacional** e **Testes sobre a proporção populacional**. Esses dois estão entre os procedimentos mais básicos de TH. Optamos então por apresentá-los ainda no Capítulo 10, até para podermos utilizá-los como exemplos na discussão geral sobre TH.

A titulo de ilustração consideremos o seguinte exemplo.

Exemplo 10.1: Está correta a afirmação do fabricante?

Consideremos novamente a companhia do Exemplo 7.4, que produz cabos náuticos de determinado tipo. O fabricante garante que a carga de ruptura média de seu produto é de pelo menos 90 kg. Um potencial consumidor, interessado na compra de grande quantidade do produto decide fazer ensaios de carga de ruptura com 20 espécimes, obtendo uma média amostral de 88,4 kg.

Com base nesses resultados o consumidor conclui que a carga de ruptura média do produto é inferior à apregoada pelo fabricante, e desiste da compra. Se você fosse o candidato a comprador procederia de modo análogo?

Exemplo 10.1: Está correta a afirmação do fabricante?

Consideremos novamente a companhia do Exemplo 7.4, que produz cabos náuticos de determinado tipo. O fabricante garante que a carga de ruptura média de seu produto é de pelo menos 90 kg. Um potencial consumidor, interessado na compra de grande quantidade do produto decide fazer ensaios de carga de ruptura com 20 espécimes, obtendo uma média amostral de 88,4 kg.

Sabemos que quando usamos amostras para estimar parâmetros populacionais, devemos sempre considerar uma margem de erro.

É possível que, mesmo com uma média populacional da carga de ruptura de 90 kg, a média amostral baseada em uma amostra de tamanho 20 seja de 88,4 kg.

Tanto o fabricante quanto o consumidor podem estar com a razão.

Aqui estão sendo postas a prova duas hipóteses (ou afirmações) opostas.

A hipótese que denotaremos por H_0, e que reflete o ponto de vista do fabricante, pode ser enunciada assim:

H_0: A carga de ruptura média dos cabos produzidos é pelo menos de 90 kg.

Já a hipótese que denotaremos por H_1, e que espelha a posição do consumidor teria o seguinte enunciado:

H_1: A carga de ruptura média dos cabos produzidos é menor do que 90 kg

Se denotarmos por μ a carga de ruptura média populacional dos cabos produzidos pelo fabricante, essas hipóteses podem ser formuladas como $H_0: \mu \geq 90$ e $H_1: \mu < 90$.

A decisão a ser tomada poderá ser:

Aceitar a hipótese H_0, concluindo que o fabricante está certo (rejeitar H_1); ou

Aceitar H_1 (e, portanto, rejeitar H_0), dando a razão ao consumidor.

10.1 Conceitos Básicos de TH

Um procedimento de teste de hipótese permite avaliar a validade (ou não) de uma afirmação sobre uma determinada característica da população, **usando para isso os dados de uma amostra dessa população**. Essa característica pode ser representada por uma variável aleatória contínua X, cujo comportamento probabilístico é expresso por sua função de densidade f. Admita que f depende de um parâmetro θ, cujo valor é desconhecido.

Deseja-se, **com base nas observações** obtidas a partir de uma coleta de dados amostrais, **decidir entre duas afirmações mutuamente excludentes** a respeito do valor correto de θ:

a hipótese nula **H_0** e a hipótese alternativa **H_1**.

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da v.a. X. Como antes, os valores observados das variáveis $X_1, X_2, \ldots, X_n$ correspondem aos dados levantados por amostragem.

Deseja-se estabelecer um critério de decisão que nos leve a:

Aceitar H_0 ou Rejeitar H_0 (em favor de H_1),

com base nos valores observados de $X_1, X_2, \ldots, X_n$.

Usualmente, a escolha desse critério passa por eleger uma função T de X_1, X_2, ..., X_n chamada **estatística de teste** e dividir o conjunto dos valores possíveis de $T(X_1, X_2, ..., X_n)$ em duas partes denominadas

Região de Aceitação **A**

e

Região de Rejeição ou Região Crítica **R.**

Assim, conforme o valor de $T(x_1, x_2, ..., x_n)$, calculado a partir dos dados coletados, pertença a **A** ou pertença a **R**, aceita-se H_0 ou rejeita-se H_0 .

Obs.: A mesma formulação acima se aplica também ao caso discreto. Para isso basta substituirmos a função de densidade f por uma função de probabilidade p.

Exemplo 10.2: No caso da situação acima, sobre a afirmação do fabricante temos:

- População: todos os cabos fabricados pela Companhia
- Propriedade da população a ser analisada: carga de ruptura, em kg, representada pela v.a. X
- Função de densidade: podemos supor distribuição Normal
- Parâmetro: média populacional μ da carga de ruptura dos cabos, em kg.
- Hipótese Nula: A carga de ruptura média dos cabos fabricados pela Companhia é pelo menos de 90 kg. Simbolicamente, H_0: $\mu \geq 90$.
- Hipótese Alternativa: A carga de ruptura média dos cabos fabricados pela Companhia é menor do que 90 kg. Simbolicamente, H_1: $\mu < 90$
- Amostra aleatória X_1, X_2, ..., X_n: carga de ruptura em cada espécime desse tipo de cabo selecionado para análise

Na seção 10.3 veremos como estabelecer o critério para construir a estatística de teste $T(X_1, X_2, ..., X_n)$ e as Regiões de Aceitação (A) e de Rejeição (R) da hipótese nula.

Ao aplicarmos um teste de hipótese, nos cabe tomar uma decisão com base em **informação apenas parcial** (a que está nos dados amostrais) sobre a realidade. Portanto, há dois tipos possíveis de erro de decisão que poderão vir a ser cometidos, e que seria conveniente evitar:

Erro I - Rejeitar H_0, quando H_0 é verdadeira

Erro II - Aceitar H_0, quando H_0 é falsa.

Cada um desses erros tem uma probabilidade de ocorrer:

α = P [Erro I] = **P[Rejeitar H_0] , se H_0 é verdadeira**

β = P [Erro II] = **P[Aceitar H_0] , se H_0 é falsa**

O problema pode ser resumido no seguinte quadro:

Decisão	***Situação Real***	
	H_0 é verdadeira	**H_0 é falsa**
H_0 é aceita	Decisão correta	Erro tipo II
H_0 é rejeitada	Erro tipo I	Decisão correta

10.2 Esclarecendo melhor alguns conceitos

Antes de vermos como se estabelece o critério de decisão que define um procedimento de Teste de Hipótese, vamos tentar esclarecer os conceitos de hipótese nula, hipótese alternativa, hipótese simples, hipótese composta, erro de decisão e probabilidades de erro.

Como devemos especificar o que são H_0 e H_1?

É usual que na formulação de um teste de hipótese, no momento em que se define o que serão H_0 e H_1, seja reservado para **H0 o papel da hipótese mais conservadora** e para **H1 o papel da hipótese mais inovadora**.

No exemplo do cabo, certamente o consumidor tem bons motivos para querer se certificar de que o produto a ser comprado está de acordo com as especificações anunciadas pelo fabricante. Isto porque pode ocorrer que um produto de qualidade inferior represente um perigo quanto a sua utilização.

Então **a hipótese nula H_0**: "A carga de ruptura média dos cabos fabricados pela Companhia é pelo menos de 90 kg" pode ser entendida como: "Os cabos atendem às especificações anunciadas pelo fabricante".

Já **a hipótese H_1**: "A carga de ruptura média dos cabos fabricados pela Companhia é menor do que 90 kg" pode ser entendida como "A carga de ruptura média dos cabos não segue as especificações divulgadas pelo fabricante". Aqui **H_0 é mais conservadora**, no sentido de que, se ela for verdadeira, ninguém perde nada. Enquanto isso, **H_1 é mais inovadora**, no sentido de que, se **H_1** for verdadeira e o consumidor comprar o produto, ele pode estar exposto a prejuízos substanciais quando esse cabo for utilizado.

O que são hipótese simples e hipótese composta?

Dizemos que H_0 (ou H_1) é uma hipótese simples se a ela corresponde um único valor do parâmetro θ. Neste caso fica também definida de forma **única** a distribuição de probabilidade comum às v.a.'s X_1, X_2,..., X_n . Caso contrário, trata-se de uma hipótese composta. Ou seja, **a igualdade $\theta = \theta_0$ define uma hipótese simples**, enquanto **as desigualdades $\theta \leq \theta_0$, $\theta < \theta_0$, $\theta \geq \theta_0$ e $\theta > \theta_0$ definem hipóteses compostas**. Por exemplo, as hipóteses acima enunciadas sobre os cabos são ambas compostas.

A hipótese nula H_0 vista como uma hipótese simples

Admita que as hipóteses H_0 e H_1 fossem ambas simples, ou seja, H_0: $\theta = \theta_0$ e H_1: $\theta = \theta_1$, sendo $\theta_1 > \theta_0$; e que quiséssemos encontrar um procedimento de teste para decidir por H_0 ou por H_1 com base em um conjunto de dados amostrais.

Quanto mais afastados estiverem θ_0 e θ_1, mais fácil seria obter uma solução adequada para esse problema.

Quanto mais próximos entre si estivessem θ_0 e θ_1, mais difícil seria discriminar entre H_0 e H_1.

Suponhamos agora que as hipóteses H_0 e H_1 sejam ambas compostas e do tipo H_0: $\theta \leq \theta_0$ e H_1: $\theta > \theta_0$. Então um bom procedimento de teste de H_0 contra H_1 deveria ser capaz de discriminar qualquer hipótese simples H_0: $\theta = a$, onde $a \leq \theta_0$, de qualquer alternativa simples H_1: $\theta = b$, onde $b > \theta_0$.

Admita agora que consigamos construir um procedimento eficiente para discriminar entre H_0': $\theta = \theta_0$ (o valor mais desfavorável a H_0) e H_1': $\theta = \theta_1$, onde θ_1 é um valor qualquer $> \theta_0$. Esse será um bom procedimento para testar H_0: $\theta \leq \theta_0$ contra H_1: $\theta > \theta_0$.

Sendo assim, **a hipótese H_0, mesmo que ela seja composta, sempre será tratada como se fosse uma hipótese simples**. Já a hipótese alternativa H_1 é, em geral, uma hipótese composta, ou seja, inclui mais de um valor de θ e, conseqüentemente, contempla mais de uma distribuição de probabilidade (de fato, usualmente são infinitas distribuições de probabilidade).

Em resumo, mesmo que as hipóteses sejam do tipo:

H_0: $\theta \leq \theta_0$ vs. H_1: $\theta > \theta_0$, ou H_0: $\theta \geq \theta_0$ vs. H_1: $\theta < \theta_0$,

a hipótese nula será considerada como se fosse H_0: $\theta = \theta_0$, já que em ambas as situações θ_0 é o valor limite entre H_0 e H_1 e portanto é o caso mais desfavorável a H_0, além de especificar uma única distribuição de probabilidade.

O significado dos erros de decisão em um Teste de Hipótese

Quando alguém está começando a travar contato com o que seja um Teste de Hipótese, às vezes surge um questionamento do tipo: "Se H_0 é verdadeira, por que eu iria rejeitá-la? E se H_0 é falsa, por que eu iria aceitá-la?"

A resposta para isso é que tais erros podem ser realmente cometidos, simplesmente pelo fato de que **a decisão é tomada** com base em apenas um nível de **informação parcial** e incompleto sobre a população em estudo: **aquele que está na amostra**. E é justamente por não dispormos de informação total e completa que podemos ser traídos pelos dados e levados a uma decisão incorreta, qual seja, cometer o Erro I ou o Erro II.

O que podemos dizer sobre as probabilidades de erro α e β?

Vimos acima que há dois tipos possíveis de erro de decisão:

- o Erro I, rejeitar H_0 quando ela é verdadeira, tem probabilidade α de ocorrer
- o Erro II, aceitar H_0 quando ela é falsa, tem probabilidade β de ocorrer

O ideal seria que ambas as probabilidades de erro, α e β, fossem tão pequenas quanto possível, mas infelizmente, se o tamanho n da amostra está fixado, uma diminuição de β traz como conseqüência um aumento em α e vice-versa. Por quê? Para diminuirmos a probabilidade do Erro II, teríamos que fazer com que a região de aceitação **A** diminuísse. Conseqüentemente, a região de rejeição **R** aumentaria, o que inevitavelmente provocaria um aumento na probabilidade do Erro I. Assim, **a única forma de fazer com que ambos, α e β, diminuíssem, seria aumentar o tamanho n da amostra**. (Ver Seção 10.6)

O Erro I no centro das atenções

Por definição, α é calculada supondo-se H_0 verdadeira, ou seja, a partir da distribuição de probabilidade única postulada por H_0. Por outro lado, β é calculada supondo H_1 verdadeira. Então, em geral há um número infinito de valores possíveis para β, correspondentes a infinitas distribuições de probabilidade. Assim, o cálculo de β é em geral bem mais complexo que o de α. Levando em conta este fato, os experimentadores tratam o Erro I como o mais importante a ser evitado, isto é, as hipóteses são formuladas de forma a que H_0 seja aquela hipótese cuja rejeição equivocada constitui o erro de maior importância.

O que costuma ser feito usualmente é fixar-se um valor pequeno para α (as escolhas mais comuns são α=0,05 e α=0,01) e, ao mesmo tempo, procurar usar procedimentos de teste para os quais os valores de β sejam os menores possíveis.

A formulação adequada da hipótese H_0 vem a ser, nestas circunstâncias, vital na realização de um experimento. A questão é que o julgamento da importância de um erro de decisão pode ser bastante subjetivo.

Consideremos, por exemplo, o caso da qualidade dos cabos. A hipótese H_0, tal como foi formulada acima, atende aos interesses do fabricante dado que para ele o erro que mais interessa evitar é o de se concluir (erradamente) que a carga de ruptura média de seus cabos é menor do que o afirmado por ele.

Entretanto, do ponto de vista do consumidor a situação pode ser diametralmente oposta. Com efeito, o erro que este quer minimizar é que se conclua que a carga de ruptura média dos cabos é maior do que 90 kg se de fato ela for menor. Assim, para o consumidor, a hipótese H_0 deveria ser: “A carga de ruptura média dos cabos é menor ou igual a 90 kg”.

A nossa atenção estará preferencialmente voltada para o Erro I, e naturalmente admitiremos que a hipótese H_0 foi corretamente formulada.

Exemplo 10.3: Baterias originais ou falsificadas?

A duração da carga das baterias recarregáveis para Notebook de uma certa marca pode ser encarada como uma variável aleatória com **distribuição Normal de média 180 minutos e desvio padrão de 40 minutos**. Existe uma falsificação do mesmo produto quase perfeita, mas nesse caso, a duração da carga da bateria é uma variável aleatória com distribuição Normal de **média 150 minutos e o mesmo desvio padrão anterior**. Um montador de Notebooks recebe 25 baterias dessa marca. Entretanto, ele tem dúvidas quanto à procedência das baterias. **Como ele poderia decidir se as baterias são originais ou falsificadas?**

Uma hipótese será que μ = 180 (originais), enquanto a outra será que μ = 150 (falsificadas), mas qual deve ser a hipótese nula e qual deve ser a hipótese alternativa?

Neste caso está em jogo a reputação do fornecedor das baterias. Se ele de fato estiver agindo corretamente, devemos esperar que o produto entregue ao montador seja original. Por outro lado, se for comprovado que as baterias são falsificadas, estaríamos diante de um delito. Sendo assim, o erro a minimizar deve ser o de decidir que as baterias são falsificadas quando elas são, de fato, originais. Portanto a hipótese nula deverá ser μ = 180 (conservadora) e a hipótese alternativa μ = 150 (inovadora).

Dizer que as baterias são originais equivale a dizer que suas durações, em minutos, provêm de uma Normal $(180; 40^2)$. Já se elas forem falsificadas, suas durações obedecem a uma Normal $(150; 40^2)$.

Como as hipóteses são relativas à média populacional, uma boa medida (**estatística**) para se usar na decisão entre as duas populações é a média da amostra, que será, portanto, a nossa **estatística de teste**.

As 25 baterias recebidas podem ser consideradas como uma amostra aleatória da população de baterias originais ou da população de baterias falsificadas. Suponha que após submeter a prova as 25 baterias foi encontrada, para a duração de suas cargas, uma média amostral $\overline{x}_{obs}$ = 167,4 min.

Qual seria o critério de decisão se quisermos fixar o valor de α em 0,05?

Denotemos por $\overline{x}_c$ o valor que serve como fronteira de decisão entre as regiões de aceitação **A** e de rejeição **R**. Esse ponto é usualmente chamado ponto de corte ou valor crítico. Então, aceitamos H_0, se $\overline{x} \geq \overline{x}_c$; e decidimos pela rejeição de H_0 (ou pela aceitação de H_1), se $\overline{x} < \overline{x}_c$.

Suponhamos que H_0 é verdadeira (ou seja, usemos a Normal com μ = 180 e σ = 40) e calculemos, para essa curva, o valor crítico $\overline{x}_c$ tal que $P(\overline{x} < \overline{x}_c) = 0,05$.

$P\left(Z < \frac{\overline{x}_c - 180}{40/\sqrt{25}} \right) = 0,05$, o que nos leva a $\frac{\overline{x}_c - 180}{40/\sqrt{25}} = -1,64$, ou seja, $\overline{x}_c = 166,88$.

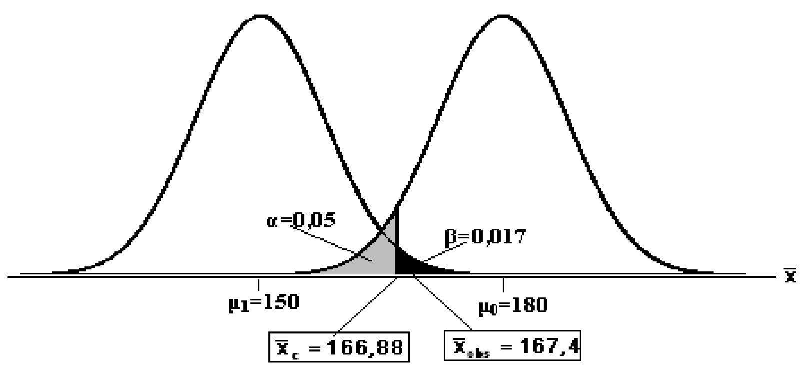
Assim, o critério de decisão seria:

Rejeitar H_0, se $\overline{x} < 166,88$. Caso contrário, aceitar H_0.

$$\beta = P(\overline{X} > \overline{x}_C) = P\left(Z > \frac{166,88 - 150}{40/\sqrt{25}}\right) =$$

$$= P(Z > 2,11) = 0,0174 \text{ ou } 1,74\%.$$

Obs.: A hipótese H_1 do exemplo acima está baseada em uma única distribuição de probabilidade, ou seja, é uma hipótese simples. Esta condição tornou fácil o cálculo do erro tipo II. Entretanto este não é o caso usual nos testes de hipóteses, já que geralmente H_1 é uma hipótese composta.



10.3 Rotina para Obtenção do Critério de Decisão

A forma usual de se propor um critério para testar a hipótese nula H_0 contra a hipótese alternativa H_1 pode ser descrita pela seguinte seqüência de passos :

1. Especificar H_0 e H_1, onde H_0 deverá conter sempre a igualdade (ou seja, deve conter o valor do parâmetro θ que está exatamente na fronteira entre as regiões correspondentes a H_0 e a H_1.).
2. Escolher um valor para α = P[Erro I], também chamado nível de significância do teste (usualmente escolhe-se α igual a 0,01 ou 0,05).
3. Eleger a estatística de teste $T(X_1, X_2, ..., X_n)$, uma variável aleatória que depende de $X_1, X_2, ..., X_n$ e que, supostamente, "resume toda a informação relevante" para que se decida por H_0 ou por H_1.
4. Determinar a distribuição de probabilidade de T correspondente a H_0, lembrando que H_0 será sempre considerada como uma hipótese simples.
5. Especificar a região de rejeição R, ou seja, o conjunto de valores de T que nos levarão a rejeitar H_0, de tal forma a que o nível de significância do teste seja igual ao α escolhido no 2º passo e trabalhando com a distribuição de T determinada no 4º passo. Automaticamente estará também especificada a Região de Aceitação A, complementar de R dentro do conjunto dos valores possíveis de T.
6. Coletar os dados $x_1, x_2, ..., x_n$, calcular valor de $T(x_1, x_2, ..., x_n)$ e decidir pela rejeição ou pela aceitação de H_0, conforme o critério especificado no 5º passo.

Obs.: Recomenda-se que a coleta da amostra só seja feita no último passo, para que o pesquisador não sofra uma possível influência proveniente do conhecimento prematuro dos dados, comprometendo assim a postura de neutralidade que dele se espera.

10.4 Teste para a Média Populacional

Quando o parâmetro envolvido no teste é a média populacional μ da variável de interesse X, então na construção do critério de decisão será usada a média amostral $\overline{X}$, como no exemplo anterior. Suponha que X tenha distribuição Normal, ou que n seja suficientemente grande de forma a que a distribuição de $\overline{X}$ se aproxime da Normal (TCL).

Seja μ_0 uma constante conhecida. O roteiro para se construir um teste de hipótese para μ, quando σ é conhecido, é o seguinte:

1. Especificar H_0 e H_1, usando uma dentre as 3 possibilidades abaixo:

   $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ (teste bilateral) ou

   $H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$ (teste unilateral) ou

   $H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$ (teste unilateral).

   Nos 2 testes unilaterais acima, trabalharemos como se a hipótese nula fosse $H_0 : \mu = \mu_0$, que é o valor extremo das respectivas desigualdades.

2. Escolher o nível de significância α.

3. Usar como estatística de teste a expressão $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.

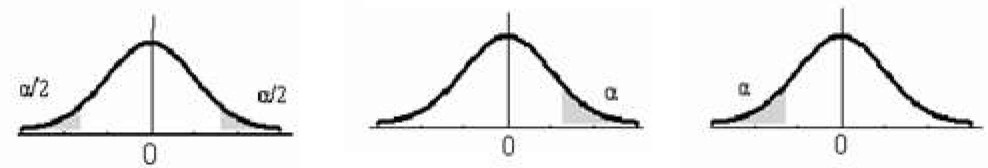
4. A distribuição de probabilidade de Z correspondente a H_0 é N(0;1).

5. Obter a região de rejeição **R** (ou seja, o conjunto de valores de Z que nos levarão a rejeitar H_0), levando em conta os passos anteriores.

6. Coletar os dados $x_1, x_2, \ldots, x_n$, calcular valor de $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ e decidir pela rejeição ou pela aceitação de H_0, conforme o critério especificado no 5º passo.

Abaixo estão os gráficos e as respectivas regiões de rejeição para cada uma das 3 possibilidades mencionadas no 1º passo.

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ $\qquad$ $H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$ $\qquad$ $H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$



Região de Rejeição: $Z < -z_{1-\alpha/2}$ ou $Z > z_{1-\alpha/2}$

Região de Rejeição: $Z > z_{1-\alpha}$

Região de Rejeição : $Z < z_{\alpha} = -z_{1-\alpha}$

Observe que, no caso de teste bilateral, a região de rejeição corresponde às duas caudas simétricas da Normal padrão, ambas de área α/2. Já no caso de teste unilateral, a região de rejeição corresponde a uma única cauda de área α da Normal padrão, ora à direita, ora à esquerda.

É claro que a regra de decisão, ao invés de ser formulada em termos da variável Z, também pode ser expressa em termos da própria média amostral $\overline{X}$. Por exemplo, se as hipóteses a serem testadas forem $H_0: \mu \geq \mu_0$ vs. $H_1: \mu < \mu_0$, então H_0 deve ser rejeitada se $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha$, condição essa que é equivalente a $\overline{X} < \mu_0 - z_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}$.

(Observe que $z_\alpha = -z_{1-\alpha}$.)

Exemplo 10.4: Está correta a afirmação do fabricante? (cont.)

Construir o teste de hipóteses para o exemplo da carga de ruptura média dos cabos, supondo que o desvio padrão populacional da carga de ruptura seja conhecido e igual a 10 kg.

1. As hipóteses H_0 e H_1 já foram especificadas no exemplo 10.1:

   $H_0 : \mu \geq 90$ e $H_1 : \mu < 90$.

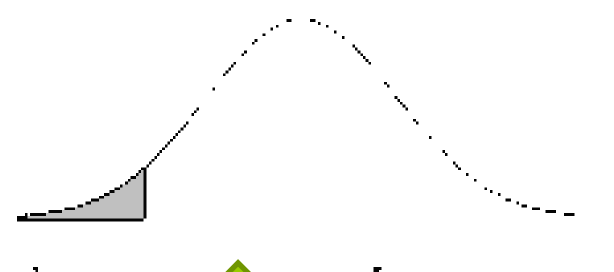
   Como já foi explicado anteriormente, a hipótese H_0 será descrita simplesmente pelo seu valor limite, ou seja, $H_0 : \mu = 90$
2. Fixaremos o nível de significância em $\alpha = 0,05$.
3. Como $\mu_0 = 90$, $\sigma = 10$ e $n = 20$, a estatística de teste apropriada é $Z = \dfrac{\overline{X} - 90}{10/\sqrt{20}}$
4. No exemplo 10.1 foi feita a suposição de que a carga de ruptura dos cabos tem distribuição Normal. Portanto, sob H_0, Z tem distribuição N(0; 1).
5. Para especificar a Região de Rejeição **R** do teste, inicialmente observemos que, quanto menor for a média amostral $\overline{x}$ (e, consequentemente, quanto menor for o valor de Z), maiores serão as razões para se rejeitar H_0. Seguindo o passo 4 do Roteiro acima, rejeitaremos H_0 se $Z < -z_{1-\alpha}$, onde $\alpha = 0,05$. Consultando a tabela da Normal ou um software adequado, vemos que $z_{0,95} = 1,64$. Concluindo, o critério de decisão adotado é definido por:

- Região de rejeição **R**: $Z < -1,64$, ou equivalentemente,

  $\overline{X} < = 90 - 1,64 \times \dfrac{10}{\sqrt{20}} = 86,33\text{kg}$
- Região de aceitação **A**: $Z \geq -1,64$, ou equivalentemente,

  $\overline{X} \geq 86,33\text{kg}$.

$\overline{X}_{obs} = 88,4$ e $Z_{obs} = \dfrac{88,4 - 90}{10/\sqrt{20}} = -0,72$



E se σ for desconhecido?

Se o desvio padrão populacional não fosse conhecido, uma alternativa possível seria trabalharmos com **o desvio padrão amostral** (estimado a partir dos dados) em seu lugar.

A estatística de teste sob H_0 seria uma t de *Student* com (n – 1) graus de liberdade.

Exemplo 10.5: É correta a afirmação do fabricante (**supondo σ desconhecido**)?

Conforme o comentário acima, vamos substituir o desvio padrão populacional σ pelo desvio padrão amostral S e **alterar os passos de 3 a 6** da seguinte forma.

Passo 3:.A estatística de teste será **T =** $\frac{\overline{X} - 90}{S/\sqrt{20}}$.

Passo 4: A distribuição de T sob H_0 é uma t de *Student* com (n–1) = 19 graus de liberdade.

Passo 5 : A região de Rejeição será $T < -t_{1-\alpha}$. Para (n–1) = 19 g.l. e α=0,05, consultando a tabela da distribuição t ou usando um software apropriado, temos que **$t_{0,95}$ = 1,729**. Portanto:

Região de rejeição R: T < -1,729, ou equivalentemente, $\overline{x} < 90 - 1{,}729 \times \frac{S}{\sqrt{20}}$;

Região de aceitação A: T ≥ -1,729, ou equivalentemente, $\overline{x} \geq 90 - 1{,}729 \times \frac{S}{\sqrt{20}}$.

Passo 6 : Temos $\overline{X}_{obs}$ = 88,4 e S_{obs}= 11,0. Logo, $t_{obs} = \frac{88{,}4 - 90}{11{,}0/\sqrt{20}}$ = - 0,650.

Portanto, estamos na região de Aceitação **A.** A decisão a ser tomada continua sendo: Aceitar a hipótese nula de que a carga de ruptura média dos cabos é pelo menos igual a 90 kg, ao nível de significância de 5%.

10.5 O conceito de p-valor

Vamos agora definir o conceito de nível crítico ou p-valor associado a uma certa metodologia de Teste de Hipótese e a uma determinada amostra $x_1, x_2, \ldots, x_n$

> O nível crítico ou p-valor é o menor valor de α (nível de significância do teste) para o qual, ao usar essa metodologia (ou seja, esse procedimento de teste) e trabalhar com os valores observados $x_1, x_2, \ldots, x_n$, ainda rejeitaríamos H_0. O p-valor costuma ser representado pelo símbolo $\tilde{\alpha}$.

Comentários sobre o nível crítico:

Embora a definição acima seja simples, o conceito de p-valor é, ao mesmo tempo:

- muito usado em pesquisas e apresentado nas publicações científicas, quando se trata de reportar os resultados de um teste de hipótese;
- a informação usualmente fornecida pelos softwares estatísticos nas situações do uso de testes de hipótese;

Observações:

1. Usar o **p-valor** para tomar a decisão relativa ao teste de hipóteses **é mais informativo** do que usar o nível de significância, já que ao se rejeitar (ou aceitar) a hipótese nula considerando o valor do nível crítico, **este valor (por si só) nos dá uma idéia do quão longe estamos da veracidade ou da falsidade da hipótese nula**.
2. Para mais detalhes sobre o conceito de p-valor, ver o Capítulo 12.

Exemplo 10.6: Aumentando a pureza de um produto químico

Um produto químico tem seu teor de pureza distribuído conforme uma Normal com média 0,72 e desvio padrão 0,02. A fim de aumentar a pureza, o produto é submetido a um tratamento. Dezesseis unidades amostrais do produto são selecionadas de forma aleatória e submetidas a esse tratamento. Em seguida, a pureza de cada unidade é determinada obtendo-se, para elas, uma média aritmética de 0,73. Podemos dizer que o tratamento contribuiu para o aumento da pureza?

(a) Qual seria a sua conclusão, ao nível de significância de 5%?

(b) Qual o p-valor?

Obs.: Está implícito aqui que o efeito do tratamento poderia ser apenas uma alteração da média, mas não do desvio padrão da pureza.

Solução

Se o tratamento dado ao produto não contribui para o aumento do grau de pureza do mesmo continuaremos a falar numa distribuição Normal com média 0,72 e desvio padrão 0,02. A questão é então: Será razoavelmente possível que uma amostra de 16 unidades amostrais de produto extraída dessa população apresente uma média amostral de 0,73?

A hipótese nula aqui pode ser formulada como **H_0: $\mu = 0,72$**. Como desejamos saber se o tratamento de fato aumenta a pureza do produto, a hipótese alternativa deve ser **$H_1 : \mu > 0.72$**

Temos então uma amostra aleatória X_1, X_2, ..., X_n, de tamanho n=16, onde X_i é Normal(μ, σ^2), para todo i e sabemos que $\sigma = 0,02$. Como o desvio padrão é conhecido, a distribuição de probabilidade da estatística de teste sob H_0 será uma Normal.

Vamos então seguir os passos da Rotina acima recomendada:

1. As hipóteses nula e alternativa podem ser expressas como:

$H_0 : \mu = 0{,}72 \qquad H_1 : \mu > 0{,}72$

2. Escolher $\alpha = 0{,}05$.

3. Segundo o roteiro, a estatística de teste aqui é $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - 0{,}72}{0{,}02/\sqrt{16}} = \dfrac{\overline{X} - 0{,}72}{0{,}005}$.

4. A distribuição de Z sob H_0 é a Normal Padrão. No entanto, como Z e $\overline{X}$ estão ligados através desta relação, também podemos usar $\overline{X}$ como estatística de teste. Isto é então o que será feito neste exemplo.

5. Para especificar a Região de Rejeição **R** do teste, inicialmente observemos que quanto maior for a média amostral $\overline{X}$, maiores serão as razões para se rejeitar H_0. Então, um critério de decisão razoável seria:

Rejeitar H_0, se $\overline{X} > \overline{x}_c$, onde $\overline{x}_c$ é uma constante a ser especificada.

Caso contrário, aceitar (não rejeitar) H_0.

Qual deve ser o valor do ponto de corte $\overline{x}_c$?

a escolha de $\alpha = 5\%$, foi feita no 1º passo.

$0{,}05 = \alpha = P(\text{Erro I}) = P(\text{Rejeitar } H_0)$, quando H_0 é verd.

Logo, $0{,}05 = P(\overline{X} > \overline{x}_c)$, sendo $\overline{X} \sim N(0{,}72;\, 0{,}005^2)$.

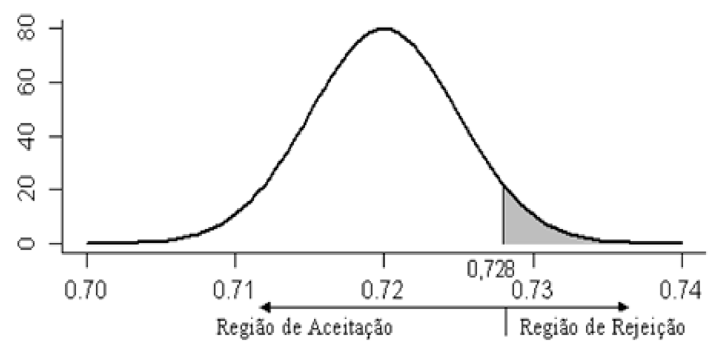
$$0{,}05 = P\left(\frac{X - 0{,}72}{0{,}005} > \frac{\overline{x}_c - 0{,}72}{0{,}005}\right) = P\left(Z > \frac{\overline{x}_c - 0{,}72}{0{,}005}\right),$$

$\dfrac{\overline{x}_c - 0{,}72}{0{,}005} = 1{,}64$, ou seja, $\overline{x}_c = 0{,}72 + 1{,}64 \times 0{,}005 = 0{,}728$.

6. Uma vez coletados os dados, obtivemos $\overline{x}_{obs} = 0{,}73$.

Logo, o "valor observado" de $\overline{x}$ está na Rejeição **R**

A decisão é rejeitar a hipótese H_0, o tratamento não contribui para o aumento da pureza do produto.



(b) Cálculo do p-valor:

Usando α = 0,05, vimos que H_0 foi rejeitada com base na amostra coletada.

E se tivéssemos usado α = 0,01 , qual seria a nossa decisão ?

Reproduzindo o raciocínio acima vemos que o novo ponto de corte seria então,

$\bar{x}_c = 0{,}72 - 2{,}33 \times 0{,}005 = 0{,}732$.

Região de rejeição **R**: $\bar{X} > 0{,}732$; Região de aceitação **A**: $\bar{X} \leq 0{,}732$

Como $\bar{X}_{obs}$ = 0,73 < 0,732, vemos que ao nível de significância α = 0,01, a decisão indicada seria aceitar H_0.

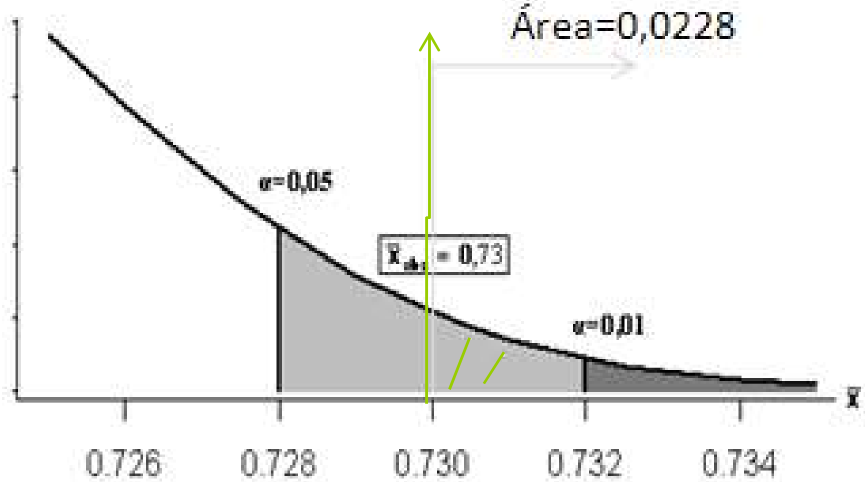
Então, se formos reduzindo progressivamente o valor de α desde 0,05 até 0,01, **vai existir um ponto em que a nossa decisão passa de rejeição para aceitação**. Este é o **p-valor $\tilde{\alpha}$ associado à nossa amostra**. Resolvendo então o problema "de trás para a frente", fixaremos o ponto de corte igual ao valor da média amostral 0,73 e calcularemos $\tilde{\alpha}$. tal que :

$\tilde{\alpha}. = P[\bar{X} > 0{,}73]$, sob a hipótese H_0.

Assim, $\bar{X} \sim N(0{,}72; 0{,}005^2)$, ou seja,

$\tilde{\alpha}. = P\left[Z > \frac{0{,}73 - 0{,}72}{0{,}005}\right]$ = P[Z > 2,0] = 0,0228.

o p-valor do teste é 0,0228 ou 2,28% .



Exemplo 10.7: Novamente a afirmação do fabricante de cabos

Determinar o nível crítico para o exemplo da carga de ruptura média dos cabos. Suporemos distribuição Normal com média 90 e desvio padrão populacional conhecido e igual a 10 kg e uma amostra aleatória de tamanho 20.

As hipóteses H_0 e H_1 já foram especificadas no exemplo 10.1:

$H_0 : \mu \geq 90$ e $H_1 : \mu < 90$

A média amostra obtida para os 20 espécimes examinados foi de 88,4 kg.

Como a região de rejeição encontra-se à esquerda, o nível crítico (ou p-valor) $\tilde{\alpha}$ é igual a P ($\bar{X} < 88{,}4$), quando H_0 é verdadeira, ou seja, quando $\mu = 90$. Portanto,

$$\tilde{\alpha} = P(\bar{X} < 88{,}4) = P\left(Z < \frac{88{,}4-90}{10/\sqrt{20}}\right) = P(Z < -0{,}72) = 0{,}2371 \quad (23{,}71\%)$$

Portanto, o valor mínimo de α que levaria à rejeição da hipótese nula seria de 23,71%. Como esta é uma probabilidade de erro muito grande, optamos por não rejeitar (aceitar) a hipótese de que a carga média de ruptura dos cabos é maior ou igual a 90 kg.

10.7 Teste para proporções

Vamos rever os conceitos apresentados neste capítulo através de outro teste de hipóteses básico, trata-se do teste para uma proporção.

Seja p a proporção populacional (desconhecida) dos elementos de uma população que possuem um determinado atributo. Queremos testar hipóteses sobre os valores possíveis do parâmetro p a partir da realização de um experimento onde será sorteada uma amostra aleatória com n elementos dessa população.

Se X é o número de elementos da amostra que possuem esse atributo, a proporção amostral $\hat{p} = X/n$ será a nossa estatística de teste. A variável aleatória X se comporta segundo uma lei Binomial(n;p) e a distribuição de $\hat{p}$ pode ser aproximada por uma Normal(p; $\frac{p(1-p)}{n}$), se n for suficientemente grande.

Seja p_0 um determinado número entre 0 e 1. Então p_0 e $\sigma_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$ são, respectivamente, a média e o desvio padrão de $\hat{p}$, quando $p = p_0$. (Conforme as considerações feitas na sub-seção 6.7.1 quanto à viabilidade de se aproximar a Binomial por uma Normal, suporemos aqui que $np_0(1-p_0) \geq 3$).

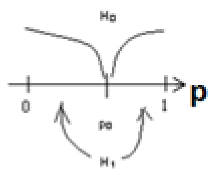
Temos três possibilidades a considerar:

H_0: $p = p_0$ vs H_1: $p \neq p_0$	H_0: $p \leq p_0$ vs H_1: $p > p_0$	H_0: $p \geq p_0$ vs H_1: $p < p_0$
Região de Rejeição: $\hat{p} > p_0 + z_{1-\alpha/2} \cdot \sigma_0$ ou $\hat{p} < p_0 - z_{1-\alpha/2} \cdot \sigma_0$.	Região de Rejeição: $\hat{p} > p_0 + z_{1-\alpha} \cdot \sigma_0$	Região de Rejeição: $\hat{p} < p_0 - z_{1-\alpha} \cdot \sigma_0$.

Em cada uma das figuras abaixo apresentamos dois gráficos: o primeiro mostra o que ocorre no espaço do parâmetro p (não observável) e o segundo mostra o que ocorre com a estatística de teste $\hat{p}$ (observável).

Hipóteses nula e alternativa em termos do parâmetro p

Regiões de aceitação e de rejeição em termos da estatística de teste $\hat{p}$



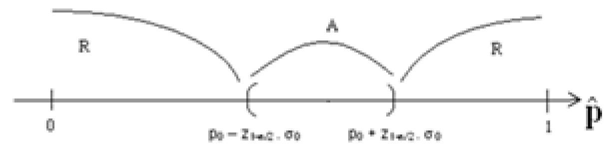
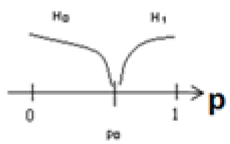


Figura 10.6 – Teste bilateral de proporção - H_0: $p = p_0$ contra H_1: $p \neq p_0$



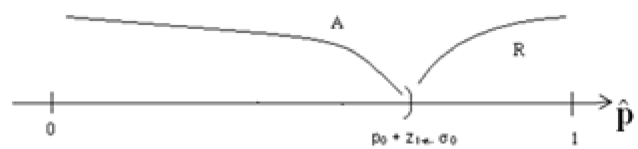
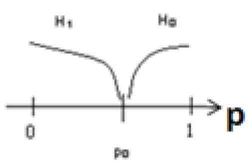


Figura 10.7 – Teste unilateral de proporção - H_0: $p \leq p_0$ contra H_1: $p > p_0$



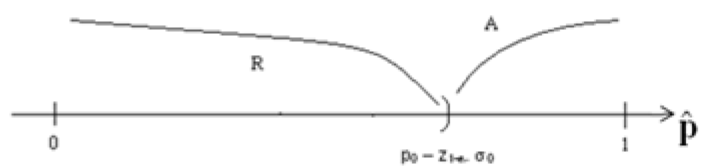


Figura 10.8 – Teste unilateral de proporção - H_0: $p \geq p_0$ contra H_1: $p < p_0$

Exemplo 10.10 Pentes de memória: Estão corretas as especificações?

As especificações dos pentes de memória RAM para computadores fabricados pela Companhia Boa Memória indicam que a porcentagem de pentes defeituosos não excede 5%. Uma amostra de 100 desses pentes apresentou 7 defeituosos.

Com base nesse resultado podemos afirmar que as especificações estão incorretas?

a) Qual a decisão a ser tomada, ao nível de significância de 0,01?
b) Qual o nível crítico do teste?
c) Qual o poder do teste, se p = 0,12?

Solução:

Considerando como hipótese nula o indicado pelas especificações , temos

H_0: $p \leq 0{,}05$ vs. H_1:$p > 0{,}05$, $p_0 = 0{,}05$ e $\sigma_0 = \sqrt{\frac{0{,}05 \times 0{,}95}{100}} = 0{,}022$

a) Como $np_0(1-p_0) = 4{,}75 > 3$, são obedecidas as condições para se usar a aproximação da distribuição Binomial pela Normal.

Vemos que se trata de um teste unilateral. Além disso, $z_{0,99} = 2{,}33$.

Então H_0 deve ser rejeitada se $\hat{p}_{obs} \geq 0{,}05 + 2{,}33 \times 0{,}022 = 0{,}10$.

Como $\hat{p}_{obs} = 7/100 = 0{,}07$ é menor que 0,10 não é possível rejeitar H_0 ao nível de significância $\alpha = 0{,}01$.

Isto significa que, com base nos dados, não há evidências de que a porcentagem de pentes defeituosos nessa linha de produção seja superior a 5%

b) $\tilde{\alpha} = P(\hat{p} > \hat{p}_{obs}) = P\left(Z > \frac{0{,}07-0{,}05}{0{,}022}\right) = P(Z > 0{,}91) = 0{,}1814$ (18,14%)

Isto é, o menor valor de α para se rejeitar a hipótese H_0 é 18,14%. Como é um valor que podemos considerar grande demais a decisão é pela não rejeição da hipótese nula.

c) Se p = 0,12, então $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}12 \times 0{,}88}{100}} = 0{,}0325$ e $\hat{p} \sim N(0{,}12;\ 0{,}0325^2)$. Portanto,

$\Pi = 1 - \beta = P(\hat{p} > 0{,}10) = P\left(\frac{\hat{p}-0{,}12}{0{,}0325} > \frac{0{,}10-0{,}12}{0{,}0325}\right) = P(Z > -0{,}615) = 0{,}73$.

Isso significa que o procedimento de teste utilizado tem 73% de chance de detectar um aumento na proporção de pentes defeituosos de p = 0,05 para p = 0,12.

<u>Condições para aplicabilidade do teste de proporções:</u>

Os procedimentos de teste, nos moldes em que eles foram aqui apresentados, se aplicam somente ao caso em que o tamanho n da amostra é suficientemente grande para que a aproximação da Binomial pela Normal já esteja dentro de um nível de precisão adequado. Se n não é suficientemente grande para que se possa usar a aproximação pela Normal (por exemplo, se $n \leq 20$), devemos trabalhar com a própria Binomial.