ACH4513 - Inferência Estatística
Testes de Hipóteses Clássicos

Marcelo S. Lauretto

Escola de Artes, Ciências e Humanidades,
Universidade de São Paulo

marcelolauretto@usp.br

www.each.usp.br/lauretto

Problema: Decidir se uma moeda é honesta

- Um juiz dará início a uma partida de futebol.
- Pelas regras, o juiz dará a posse inicial de bola através do lançamento de uma moeda honesta:
  - Se der cara, a equipe à sua esquerda (Time A) inicia com a bola;
  - Se der coroa, é a equipe à sua direita (Time B) quem inicia com a bola.
- Todavia, o juiz se dá conta de que esqueceu a moeda.
- O capitão do time B rapidamente retira uma moeda do bolso e a oferece para o sorteio.
- O time A somente concorda se houver evidências de que a moeda seja de fato honesta, ou seja, $Pr(\text{cara}) = 50\%$
- Para isso, deve-se "testar" a moeda de seu lançamento oficial para decidir a posse de bola.

Problema: Decidir se uma moeda é honesta

- O experimento consiste em lançar a moeda 20 vezes sob aproximadamente as mesmas condições e contar a quantidade de caras e coroas.
- Em quais dos resultados abaixo o juiz deveria desconfiar da procedência da moeda?
    1. 10 caras e 10 coroas?
    2. 8 caras e 12 coroas?
    3. 2 caras e 18 coroas?
    4. 0 caras e 20 coroas?
- Uma pergunta mais geral: Para quais dos possíveis resultados o juiz deveria considerar que a moeda não é honesta?
    - Para responder a essa questão: *Procedimento de teste de hipótese.*

Problema: Decidir se uma moeda é honesta

- Sob a abordagem de estatística clássica, um procedimento de teste de hipótese depende da definição dos seguintes elementos:
  1. Condição do experimento e respectiva estatística. Em nosso exemplo:
     - Experimento: n lançamentos independentes da moeda (sob aproximadamente as mesmas condições)
     - X: número de caras nos n lançamentos
  2. Parâmetro sobre o qual se quer fazer inferência e seu respectivo espaço:
     - Parâmetro p: probabilidade da moeda dar cara em um lançamento.
     - *Espaço paramétrico* Ω: $p \in [0, 1]$
  3. Hipótese a ser testada (hipótese *nula*) e hipótese alternativa:
     - $H_0 : p = 0.5$ (moeda honesta)
     - $H_1 : p \neq 0.5$ (moeda tende a dar mais caras ou mais coroas)
       Importante: H_0 e H_1 devem formar uma partição de Ω, ou seja:
       $H_0, H_1 \neq \emptyset; \quad H_0 \cap H_1 = \emptyset; \quad H_0 \cup H_1 = \Omega$

Problema: Decidir se uma moeda é honesta

- (cont.)

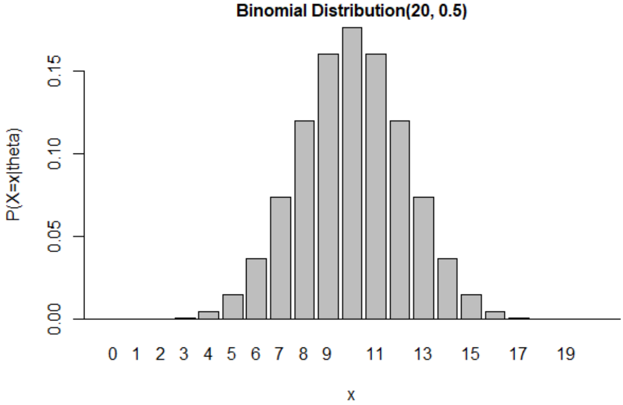
  4. Distribuição de probabilidade dos possíveis resultados do experimento:
     - $P(X = x|p)$: probabilidade de x caras em n lançamentos, dado o parâmetro p:

     $$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

  5. *Região de rejeição* (ou *região crítica*) do teste: Determinado a partir de:
     - $P(X = x|p)$
     - H_0 e H_1
     - *Nível de significância* α

Problema: Decidir se uma moeda é honesta

- Distribuição de probabilidade: $P(X = x|p = 0.5)$
  (X: número de caras em *n* lançamentos)



Como interpretar (e especificar) α?

- A especificação do valor de α leva em conta dois tipos de erro em testes de hipótese:
  - Erro do Tipo I: Probabilidade de *rejeitar* a hipótese quando esta é verdadeira
  - Erro do Tipo II: Probabilidade de *não rejeitar* a hipótese quando esta é falsa

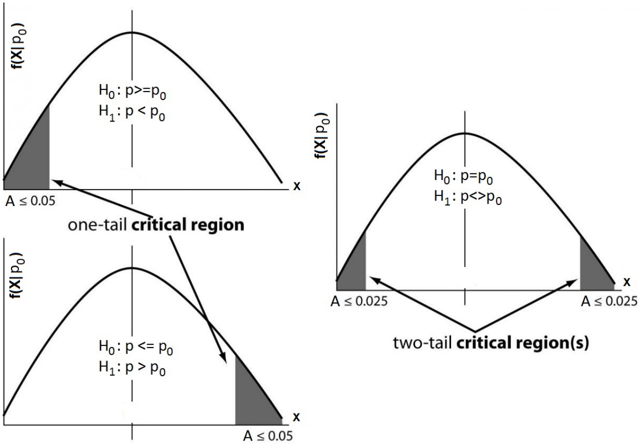
		Population Condition	
		H_0 True	H_a True
Conclusion	Don't reject H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

- Objetivos conflitantes: Baixo Erro do Tipo I implica em alto Erro do Tipo II e vice-versa

Como interpretar (e especificar) α?

- O valor de α, chamado *nível de significância*, corresponde ao Erro do Tipo I tolerado, e deve ser estipulado de acordo com o problema e com as consequências do erro de rejeitar uma hipótese verdadeira
  - Valores usuais: $\alpha = 0.10, 0.05, 0.01, 0.001$
  - Se as consequências de um Erro do Tipo I são moderadas, pode-se usar $\alpha = 0.1$
    (p.ex a moeda da partida de futebol)
  - Se as consequências de um Erro do Tipo I são sérias, deve-se adotar valores mais baixos de α
    P.ex. em um julgamento: um réu só pode ser condenado se houver forte evidência contra a hipótese de sua inocência (baixo valor de α)
- A *Região crítica do teste* corresponde ao conjunto de valores de X para os quais a hipótese H_0 será rejeitada, condicionado a
  Erro do tipo I $\leq \alpha$

Regiões críticas - representação geral



Voltando ao problema da moeda:

- Como definir a hipótese nula H_0 e a alternativa H_1?
  (ou seja, como definir se a região crítica é uni ou bilateral?)
- Relembrando:
  - A posse inicial de bola é decidida através do lançamento de uma moeda:
    – se a moeda der cara, a equipe A inicia com a bola
    – se der coroa, é a equipe B quem inicia com a bola
  - O time B ofereceu a moeda para decidir a posse inicial

Voltando ao problema da moeda:

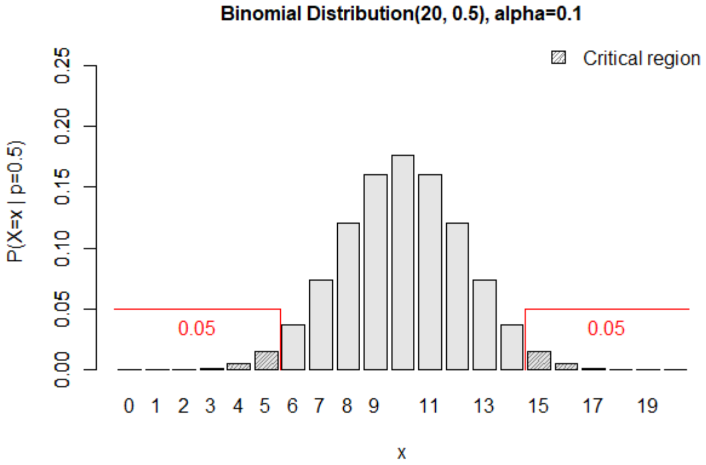
- Logo, juiz deve escolher uma das três hipóteses (e respectivas regiões de rejeição):
  - $H_0 : p = 1/2$, $H_1 : p \neq 1/2$: alta proporção de caras ou de coroas é considerada suspeita
    – Posição mais neutra: moeda é rejeitada se qualquer um dos times puder ser prejudicado por eventual vício na moeda
  - $H_0 : p \geq 1/2$, $H_1 : p < 1/2$: baixa proporção de caras é considerada suspeita
    – Moeda é rejeitada somente se o time *A* puder ser prejudicado por eventual vício na moeda
  - $H_0 : p \leq 1/2$, $H_1 : p > 1/2$: alta proporção de caras é considerada suspeita
    – Moeda é rejeitada somente se o time *B* puder ser prejudicado por eventual vício na moeda

Voltando ao problema da moeda:

- Possibilidade 1: Região crítica bilateral (ou bicaudal):

  1. Hipótese: $H_0 : p = 1/2$ contra $H_1 : p \neq 1/2$
  2. Nivel de significância: $\alpha = 0.1$
     Rejeitamos a moeda se ela fornecer um número de caras muito abaixo ou muito acima do esperado sob a hipótese.

Voltando ao problema da moeda:



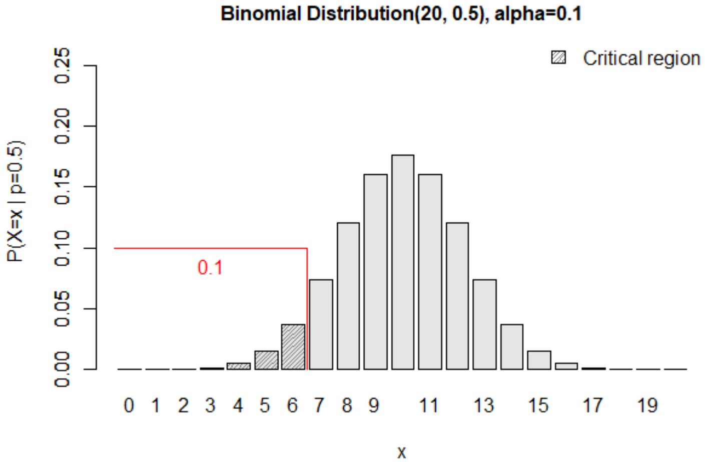
$$
\begin{aligned}
C &= \{x \mid P(X \leq x|p) \leq \alpha/2\} \cup \{x \mid P(X \geq x|p) \leq \alpha/2\} \\
&= \{0, 1, 2, 3, 4, 5, 15, 16, 17, 18, 19, 20\}
\end{aligned}
$$

Voltando ao problema da moeda:

- Região crítica unilateral (ou unicaudal):

  1. Nivel de significância: $\alpha = 0.1$
  2. Hipótese: $H_0 : p \geq 1/2$ contra $H_1 : p < 1/2$
     Rejeitamos a hipótese da moeda ser honesta se esta fornecer um número de caras muito abaixo do esperado.

Voltando ao problema da moeda:



$$C = \{x \mid P(X \leq x|p) \leq \alpha\} = \{0, 1, 2, 3, 4, 5, 6\}$$

Falseabilidade (ou Refutabilidade) de Popper

- Testes de hipóteses frequentistas são baseados no princípio da falseabilidade
- Karl Raimmund Popper (1902–1994): "Racionalismo Crítico"
  - Oposição ao método indutivo (Dados $\rightarrow$ Teoria)
- Postulados:
  - Ciência é uma sequência de conjecturas
  - Teorias científicas não podem ser diretamente provadas
  - Teorias são propostas como hipóteses, substituídas por novas hipóteses quando refutadas experimentalmente ("falseadas")
  - O que diferencia as teorias científicas de outras formas de crença é que as primeiras podem ser falseadas
    $\longrightarrow$ formulação em termos precisos, que definem os resultados esperados.

Falseabilidade (ou Refutabilidade) de Popper

- Tribunais modernos:
  - *In dubio pro reo*: o réu é considerado inocente até que seja provada sua culpa (benefício da dúvida).
  - O benefício da dúvida torna mais difícil condenar um réu.
  - Por outro lado, o veredito de um julgamento nunca pode ser *inocente*, apenas *culpado* ou *não culpado*.
- Na metáfora do tribunal:
  - Uma lei científica é (provisoriamente) aceita pelo tribunal como verdadeira, até que esta seja refutada ou provada errônea por evidência pertinente.
  - Evidência para refutar uma teoria tem a forma de observações empíricas que discordam das conseqüências ou previsões feitas pela teoria em julgamento.

Falseabilidade (ou Refutabilidade) de Popper

- Um julgamento justo no tribunal científico:
  - *pode* assegurar a validade das deduções que levaram a uma prova de falsidade;
  - *não pode* dar uma certificação ou garantia referente à validade da teoria.
- Pelos mesmos princípios acima, um teste de hipótese tem duas conclusões possíveis:
  - *Rejeição* da hipótese;
  - *Não rejeição* da hipótese (mas não *aceitação*).

Procedimento geral de testes de hipóteses

- A construção de um teste de hipóteses, para um parâmetro populacional, pode ser colocada do seguinte modo.
  - Existe uma variável X associada a dada população e tem-se uma hipótese sobre determinado parâmetro θ dessa população
    Por exemplo, afirmamos que o verdadeiro valor de θ é θ_0
  - Colhe-se uma amostra aleatória de elementos dessa população, e com ela deseja-se comprovar ou não tal hipótese.
- Como já vimos anteriormente, iniciamos nossa análise explicitando claramente qual a hipótese que estamos colocando à prova e a chamamos de *hipótese nula*, cuja forma mais geral é:

$$H_0 : \theta = \theta_0$$

Procedimento geral de testes de hipóteses

- Em seguida, convém explicitar também a hipótese que será considerada aceitável, caso H_0 seja rejeitada

  A essa hipótese chamamos *hipótese alternativa* e denotamos H_1 ou H_a

  - Usualmente, H_1 é o complemento de H_0, ou seja, corresponde ao conjunto de todos os valores possíveis de θ, Ω, excluindo-se o conjunto definido por H_0

  A alternativa mais geral seria
  $$H_1 : \theta \neq \theta_0,$$
  embora pudéssemos ter alternativas da forma
  $$H_1 : \theta < \theta_0 \text{ o que implica } H_0 : \theta \geq \theta_0 \text{ ou}$$
  $$H_1 : \theta > \theta_0 \text{ o que implica } H_0 : \theta \leq \theta_0,$$
  dependendo das informações que o problema traz.

Procedimento geral de testes de hipóteses

- Qualquer que seja a decisão tomada, vimos que estamos sujeitos a cometer dois tipos de erros:
  - *Erro do tipo I:* rejeitar a hipótese nula quando essa é verdadeira. Chamamos de α a probabilidade de cometer esse erro, ou seja, $\alpha = P($ erro do tipo I$) = P($rejeitar $H_0)|H_0$ é verdadeira$)$
  - *Erro do tipo II:* não rejeitar H_0 quando H_0 é falsa. A probabilidade de cometer esse erro é denotada por β, logo $\beta = P($ erro do tipo II$) = P($não rejeitar $H_0)|H_0$ é falsa$)$

Procedimento geral de testes de hipóteses

- O objetivo do teste é dizer, usando uma estatística $\hat{\theta}$, se a hipótese H_0 é ou não aceitável
  - Operacionalmente, essa decisão é tomada através da consideração de uma *região crítica* (ou *região de rejeição*) RC
  - Caso o valor observado da estatística pertença a essa região, rejeitamos H_0; caso contrário, não rejeitamos H_0
  - Essa região é construída de modo que $P(\hat{\theta} \in \text{RC}|H_0\text{é verdadeira})$ seja igual (ou menor) a α, fixado *a priori*
  - Note que a região crítica é sempre construída sob a hipótese de H_0 ser verdadeira
  - A determinação do valor de β é mais difícil, pois usualmente não especificam-se valores fixos para o parâmetro sob a hipótese alternativa
    - Futuramente abordaremos essa situação, ao considerarmos o poder de um teste.

Procedimento geral de testes de hipóteses

- Abaixo é sumarizada a sequência geral de passos:
  1. Fixe qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1;
  2. Use a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para testar H_0; obter as propriedades de interesse dessa estatística (distribuição média, desvio padrão);
  3. Fixe a probabilidade α de cometer o erro do tipo I e use esse valor para construir a região crítica (construída sobre os valores do parâmetro hipotetizados por H_0);
  4. Use as observações da amostra para calcular o valor da estatística do teste;
  5. Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite H_0; caso contrário, rejeite H_0.

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Seja $X_1, \ldots, X_n$ uma AAS proveniente de uma população com média μ desconhecida e variância σ^2, e considere a hipótese $H_0 : \mu = \mu_0$.
- Se além, da premissa acima, uma das condições abaixo for satisfeita:
  - A amostra é proveniente de uma distribuição normal; OU
  - O tamanho da amostra é considerado suficientemente grande (usualmente, a partir de $n > 30$) de forma que o Teorema do Limite Central seja válido;

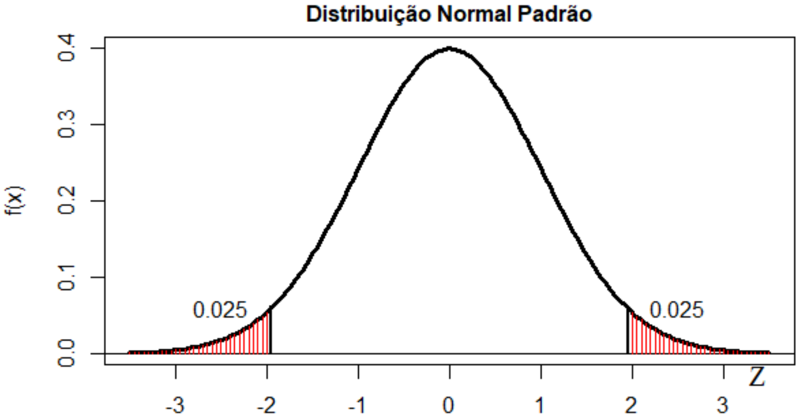
  Então a média amostral $\overline{X}$ segue aproximadamente uma distribuição normal com média μ e variância σ^2/n.
- Se a hipótese for verdadeira $\mu = \mu_0$, então $\overline{X} \sim N(\mu_0, \sigma^2/n)$.
- Logo, a *estatística* $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$!!

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Assim, para testar a hipótese original, basta verificar em qual região da distribuição normal padrão a *estatística* Z se encontra.
  - Note que Z indica quantos erros padrões $\overline{X}$ está distante de μ_0, para mais ou para menos
  - Logo, a região crítica é dada em termos da distância tolerada entre $\overline{X}$ e μ_0, em erros padrões

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

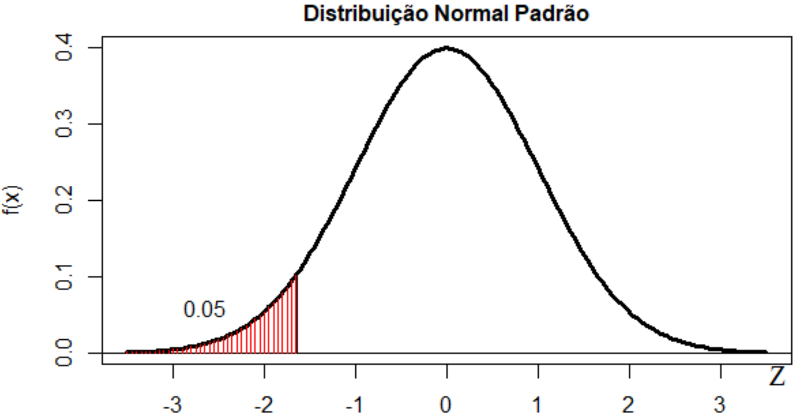
- Ex: $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, $\alpha = 0.05$:



Se $|Z| > 1.96$: rejeite H_0 (se $\overline{X}$ estiver a mais de 1.96 erros padrões acima ou abaixo de μ_0, rejeite H_0)

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

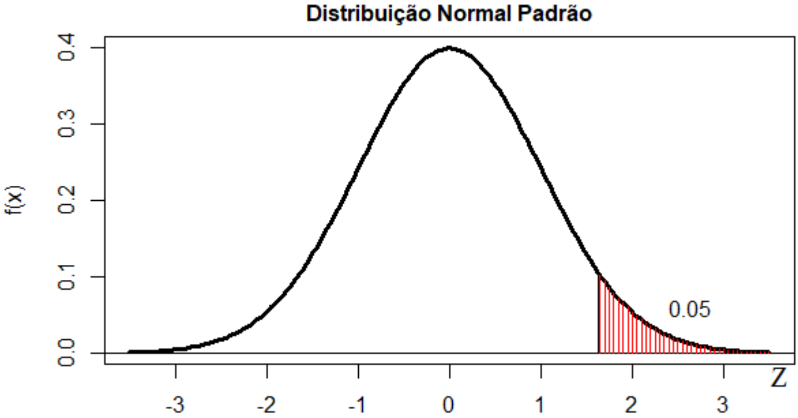
- Ex: $H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$, $\alpha = 0.05$:



Se $Z < -1.64$: rejeite H_0 (se $\overline{X}$ estiver a mais de 1.64 erros padrões abaixo de μ_0, rejeite H_0)

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Ex: $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$, $\alpha = 0.05$:



Se $Z > 1.64$: rejeite H_0 (se $\overline{X}$ estiver a mais de 1.64 erros padrões acima de μ_0, rejeite H_0)

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Exemplo 12.2 (adaptado de Morettin & Bussab):
  - Uma máquina automática para encher pacotes de café enche-os segundo uma distribuição normal, com média μ e desvio padrão assumido constante com valor $\sigma = 24g$.
  - A máquina foi regulada para $\mu = 500g$.
  - Desejamos, periodicamente, colher uma amostra de 36 pacotes e verificar se a produção está sob controle, isto é, se $\mu = 500g$ ou não.
  - Se uma dessas amostras apresentasse uma média $\overline{x} = 492g$, deveríamos ou não parar a produção para regular a máquina?
  - Vejamos como testar essa hipótese.

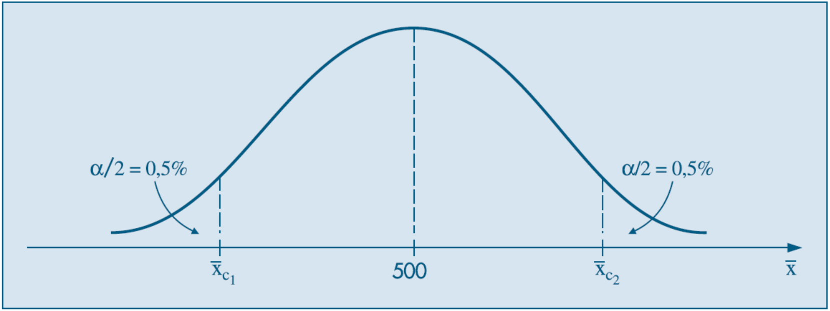
Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Exemplo 12.2 (cont):
  - Passo 1: Indiquemos por X o peso de cada pacote; então, $X \sim N(\mu, 400)$. As hipóteses que nos interessam são:
    $$H_0 : \mu = 500g$$
    $$H_1 : \mu \neq 500g$$
    pois a máquina pode se desregular para mais ou para menos.
  - Passo 2: Pela afirmação do problema, $\sigma = 20g$ será sempre a mesma; logo, para todo μ, a média $\overline{X}$ de 16 pacores terá distribuição $N(\mu, 400/16)$, de modo que o desvio padrão (ou erro padrão) de $\overline{X}$ é $\sigma_{\overline{X}} = 5$. Em particular, se H_0 for verdadeira, $\overline{X} \sim N(500, 25)$.
  - Passo 3: Vamos fixar $\alpha = 1\%$; pela hipótese alternativa, vemos que H_0 deve ser rejeitada quando $\overline{X}$ for muito pequeno ou muito grande (teste bilateral). Portanto, nossa região crítica será definida como a da figura abaixo.

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Exemplo 12.2 (cont):

Figura 12.5: Região crítica para o teste $H_0 : \mu = 500$ vs $H_1 : \mu \neq 500$ do Exemplo 12.2.



Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Exemplo 12.2 (cont):
  - Passo 3 (cont)
    Podemos usar o Excel para calcular os quantis 0.005 e 0.995 da distribuição normal padrão:
    INV.NORMP.N(0.005) e INV.NORMP.N(0.995)
    Assim, obtemos que:
    $$z_1 = -2.58 = (\overline{X}_{c_1} - 500)/4 \Rightarrow \overline{X}_{c_1} = 489.68,$$
    $$z_2 = 2.58 = (\overline{X}_{c_2} - 500)/4 \Rightarrow \overline{X}_{c_2} = 510.32.$$
    Segue-se que a região crítica é
    $$\mathrm{C} = \{\overline{x} \in \mathbb{R} | \overline{x} \leq 489.68 \text{ ou } \overline{x} \geq 510.32\}.$$
  - Passo 4: A informação pertinente à amostra é sua média, que nesse caso particular é $\overline{x}_0 = 492$.
  - Passo 5: Como $\overline{x}_0$ **não pertence** à região crítica, nossa conclusão será **não rejeitar** H_0. Ou seja, o desvio da média da amostra em relação à média proposta por H_0 poderia ser considerado como devido apenas ao sorteio aleatório dos pacotes.

Teste Z para a média de uma população (distribuição normal OU amostras grandes; variância conhecida)

- Exemplo 12.2 (cont):
  - Note que, se ao invés de aplicarmos o nível de significância $\alpha = 0.01$ usarmos $\alpha = 0.05$, deveremos considerar os quantis 0.025 e 0.975 da distribuição normal padrão, que correspondem aos valores -1.96 e 1.96.
  - Nesse caso, a região crítica é dada pelos limites
  $$z_1 = -1.96 = (\overline{X}_{c_1} - 500)/4 \Rightarrow \overline{X}_{c_1} = 492.16,$$
  $$z_2 = 1.96 = (\overline{X}_{c_2} - 500)/4 \Rightarrow \overline{X}_{c_2} = 507.84.$$
  - Ou seja:
  $$\text{C} = \{\overline{x} \in \mathbb{R} | \overline{x} \leq 492.16 \text{ ou } \overline{x} \geq 507.84\}.$$
  - Nesse caso, $\overline{x}_0$ **pertence** à região crítica e portanto **rejeitaremos** H_0 ao nível de significância 0.05.
  - Logo, $\alpha = 0.05$ implica em maior probabilidade de rejeição da hipótese, e portanto fornece um procedimento mais rigoroso para a manutenção da regulagem da máquina.

O nível descritivo ou *p-valor*

- Voltemos ao Exemplo 12.2:
  - Suponha que, no Exemplo 12.2, o especialista que realizou os experimentos e o teste de hipótese tivesse concluído, em seu relatório, que H_0 não foi rejeitado ao nível de significância 0.01, mas não desse nenhuma outra informação.
  - Se o gerente de operações que lesse o relatório fosse mais rigoroso e quisesse a máquina o mais calibrada possível, talvez considerasse o nível de significância $\alpha = 0.05$ ao invés de $\alpha = 0.01$.
  - Sem a informação da estatística z efetivamente obtida e da probabilidade de obter-se valores de Z menos prováveis do que z sob a hipótese H_0, o gerente não teria qualquer base para sua tomada de decisão.

O nível descritivo ou *p-valor*

- Voltemos ao Exemplo 12.2 (cont):
  - Para dirimir esse problema, o correto seria o pesquisador, ao invés de construir a região crítica, reportar essas duas informações:
    - Valor da estatística z obtida:
      Lembre-se que a estatística z é simplesmente a média $\overline{x}$ padronizada:
      $$z = \frac{\overline{x} = \mu_0}{\sigma/\sqrt{n}} = \frac{492 - 500}{4} = -2.0$$
    - Probabilidade de obter valores de Z menos prováveis do que z, assumindo H_0 verdadeira:
      Como estamos considerando o teste bicaudal ($H_1 : \mu \neq \mu_0$), os valores mais extremos do que z correspondem aos intervalos $(\,-\infty, -|z|\,)$ e $(\,|z|, \infty\,)$, e portanto a probabilidade desejada é
      $$\begin{aligned} Pr(|Z| > |z|) &= \Phi(-|z|) + (1 - \Phi(|z|)) = 2\Phi(-|z|) \qquad (1) \\ &= 2(0.023) = 0.046, \end{aligned}$$
      onde Φ denota a f.d.a da distribuição normal padrão.

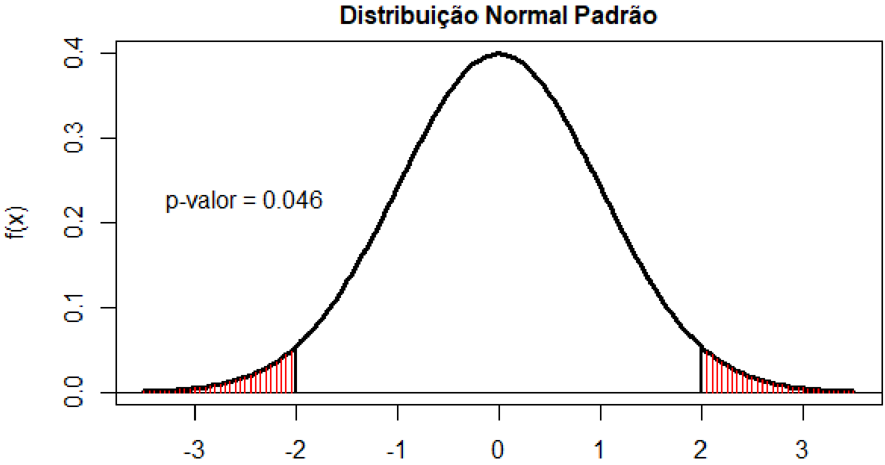
O nível descritivo ou *p-valor*

- A probabilidade apresentada na Eq.(1) é o *p-valor* ou *nível descritivo* do teste conduzido no exemplo 12.2.
- Uma definição mais geral para o *p-valor* é o menor nível α_0 tal que rejeitaríamos a hipótese nula ao nível α com os dados observados.
- O cálculo do p-valor depende da distribuição da estatística, de H_0 e de H_1.
- No Exemplo 12.2, consideramos o teste bicaudal $H_0 : \mu = 500$ contra $H_1 : \mu \neq 500$ (p-valor$= 0.46$). Consideremos as duas versões dos testes monocaudais:
  - $H_0 : \mu \geq 500,\ H_1 : \mu < 500$: $z = -2.0$
    p-valor $= Pr(Z < z) = \Phi(-z) = \Phi(-2) = 0.023$
  - $H_0 : \mu \leq 500,\ H_1 : \mu > 500$: $z = -2.0$
    p-valor $= Pr(Z > z) = 1 - \Phi(-z) = 1 - \Phi(-2) = 0.977$

As representações do p-valor nos três casos são apresentados nos slides a seguir.

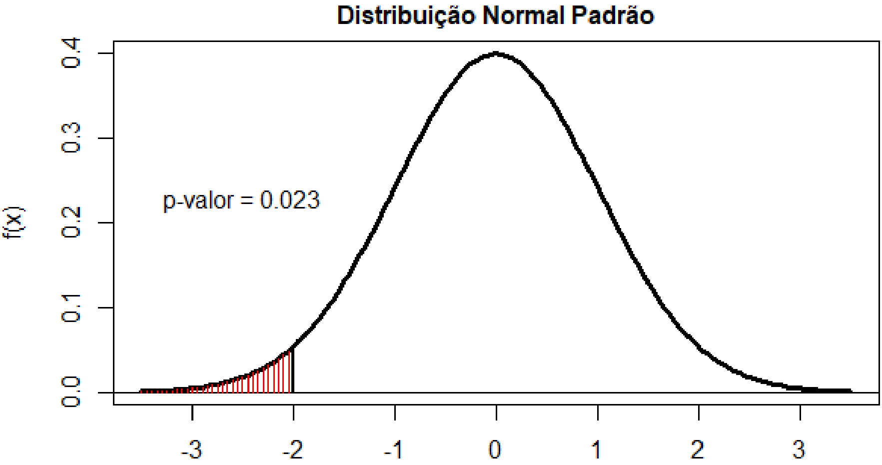
O nível descritivo ou *p-valor*

- Exemplo 12.2: $H_0 : \mu = 500$, $H_1 : \mu \neq 500$, $z = -2$:



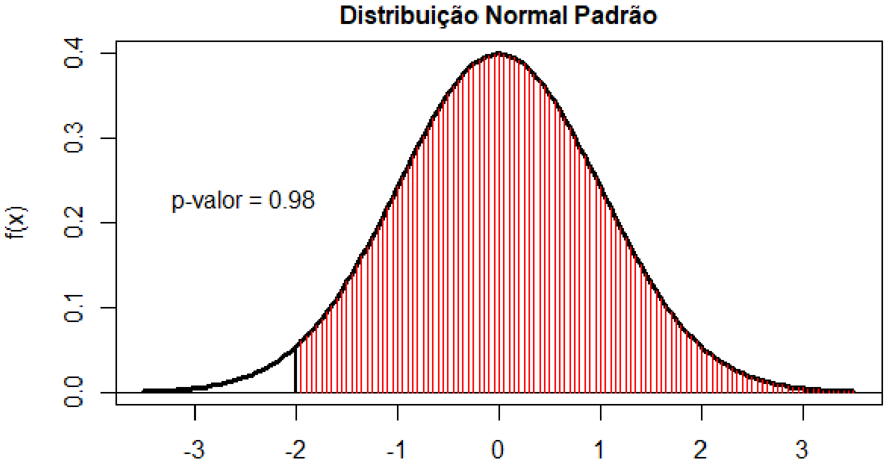
O nível descritivo ou *p-valor*

- Exemplo 12.2: $H_0 : \mu \geq 500$, $H_1 : \mu < 500$, $z = -2$:



O nível descritivo ou *p-valor*

- Exemplo 12.2: $H_0 : \mu \leq 500$, $H_1 : \mu > 500$, $z = -2$:



Teste z sobre a proporção de uma população

- Denotemos por p a proporção (desconhecida) de uma certa característica na população e p_0 um valor particular a ser considerado como hipótese.
- As três formas para um teste de hipótese sobre a proporção em uma população são:

$$\begin{array}{lll} H_0 : p \geq p_0 & H_0 : p \leq p_0 & H_0 : p = p_0 \\ H_1 : p < p_0 & H_1 : p > p_0 & H_1 : p \neq p_0 \end{array}$$

  (Duas primeiras formas são unicaudais; terceira forma é bicaudal.)

- Vimos anteriormente que uma proporção estimada $\overline{P}$ obtida a partir de uma AAS pode ser vista como a média de uma sequência de Bernoulli $X_1, X_2, \ldots, X_n$, onde X_i indica se o i-ésimo indivíduo da amostra possui ($X_i = 1$) ou não ($X_i = 0$) a característica: $\overline{P} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - Logo, o Teorema do Limite Central aplica-se a proporções amostrais, de forma análoga à média amostral de uma variável aleatória qualquer

Teste z sobre a proporção de uma população

- Testes de hipóteses sobre a proporção de uma população são baseados na diferença entre a proporção amostral $\overline{P}$ e o valor sob hipótese p_0.
- Pelo TLC, podemos usar uma variante do teste z para realizar inferências sobre a proporção de uma população.

  Únicas diferenças:

  - Usamos a proporção amostral $\overline{P}$ e seu erro padrão $\sigma_{\overline{p}}$ para calcular a estatística do teste
    - Para calcular $\sigma_{\overline{p}}$, usamos p_0 ao invés de $\overline{P}$, como veremos adiante
  - Ao invés de usar apenas o critério $n \geq 30$ para considerar o tamanho da amostra como adequado, deve-se lembrar que a distribuição amostral de $\overline{p}$ só pode ser aproximada por uma distribuição normal se ambos np_0 e $n(1 - p_0)$ forem ≥ 5
    (Ou seja, as quantidades mínimas esperadas de indivíduos na amostra com a característica e de indivíduos sem a característica devem ser ambas superiores a 5)

Teste z sobre a proporção de uma população

- A estatística do teste é
$$Z = \frac{\overline{P} - p_0}{\sigma_{\overline{p}}}$$
- Como calcular $\sigma_{\overline{p}}$?
  - Para calcular intervalos de confiança para p, vimos que podemos usar $\sqrt{\overline{p}(1-\overline{p})/n}$ como um estimador para $\sigma_{\overline{p}}$.
  - Porém, em testes de hipóteses, o valor p_0 é assumido na hipótese H_0. Logo, quando H_0 é verdadeiro, $\sigma_{\overline{p}}$ deve ser calculada usando p_0 ao invés de $\overline{P}$:
$$\sigma_{\overline{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$
- Observada a amostra e obtido o valor z da estatística, pode-se calcular o p-valor conforme H_0, de maneira análoga ao teste sobre a média:
  - $H_0 : p = p_0 \Rightarrow pv = Pr(|Z| > |z|) = 2\Phi(-|z|)$
  - $H_0 : p \leq p_0 \Rightarrow pv = Pr(Z > z) = 1 - \Phi(z)$
  - $H_0 : p \geq p_0 \Rightarrow pv = Pr(Z < z) = \Phi(z)$
    onde Φ denota a f.d.a da distribuição normal padrão

Teste z sobre a proporção de uma população

- Exemplo: A direção de uma escola de golfe identificou que, nos últimos 12 meses, apenas 20% dos jogadores eram mulheres. Em um esforço para incrementar a proporção de mulheres, a escola realizou uma promoção especial para atrair mais mulheres. Após um mês, uma amostra aleatória de 200 jogadores foi coletada, na qual 150 eram homens e 50 mulheres. A direção deseja saber se os novos dados suportam a conclusão de que a proporção de mulheres aumentou.

Teste z sobre a proporção de uma população

- Exemplo (cont):
  - A hipótese a ser testada e respectiva hipótese alternativa são $H_0 : p \leq 0.20$, $H_1 : p > 0.2$
  - Devemos verificar se np e $n(1-p)$ são ≥ 5; neste exemplo, $np = 200(0.20) = 40$ e $n(1-p) = 200(0.80) = 160$, e portanto a aproximação pela distribuição normal é apropriada.
  - O erro padrão de $\sigma_{\overline{p}}$ é $\sigma_{\overline{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{(0.20)(0.80)}{200}} = 0.0283$
  - A estatística z é dada por $z = \frac{\overline{p}-p_0}{\sigma_{\overline{p}}} = \frac{0.25-0.20}{0.0283} = 1.768$
  - Finalmente, o p-valor é dado por $pv = Pr(Z > z) = 1 - \Phi(z) = 1 - \Phi(1.768) = 0.039$
  - Portanto, pode-se concluir, sob o nível de significância de 5%, que houve um incremento na proporção de mulheres. (O incremento estimado na participação de mulheres foi de $\overline{p} - p_0 = 25\% - 20\% = 5\%$)

Teste z sobre a proporção de uma população

- Voltemos ao exemplo inicial da moeda:
  - Suponha que nosso interesse seja decidir se uma moeda é honesta.
  - Para isso, realizamos 20 lançamentos independentes da moeda, e obtemos 6 caras.
  - Com base nessa informação, considerando um nível de significância de 10%, devemos rejeitar essa moeda?

Teste z sobre a proporção de uma população

- Resposta ao problema da moeda:
  - A hipótese a ser testada e respectiva hipótese alternativa são $H_0 : p = 0.5$, $H_1 : p \neq 0.5$
  - Devemos verificar se np e $n(1-p)$ são ≥ 5; neste exemplo, $np = 20(0.50) = 10$ e $n(1-p) = 20(0.50) = 10$, e portanto a aproximação pela distribuição normal é apropriada.
  - O erro padrão de $\sigma_{\overline{p}}$ é $\sigma_{\overline{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{(0.50)(0.50)}{20}} = 0.1118$
  - A estatística z é dada por $z = \frac{\overline{p}-p_0}{\sigma_{\overline{p}}} = \frac{0.30-0.50}{0.1118} = -1.789$
  - O p-valor é dado por $pv = Pr(|Z| > |z|) = 2\Phi(-|z|) = 2\Phi(-1.789) = 0.0736$
  - Portanto, é mais prudente considerar, sob o nível de significância de 10%, que a moeda não seja honesta.

Distribuição t

- Também conhecida pelo nome *t de Student*, em homenagem a William S. Gosset, que em 1908 publicou seus estudos sobre essa distribuição sob o pseudônimo "Student".

- Definição: Considere duas variáveis aleatórias independentes $Z \sim N(0,1)$ e $Y \sim \chi^2(n)$.

  Seja T a variável aleatória definida pela equação

  $$T = \frac{Z}{\sqrt{Y/n}}\ .$$

  Então a distribuição de T é denominada distribuição t (de Student) com n graus de liberdade.

- Função de densidade de probabilidade:

  $$f(t|n) = \frac{\Gamma(n+1)/2}{\sqrt{n\pi}\ \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \qquad -\infty < x < \infty.$$

- Média e Variância: Se $T \sim t(n)$:

  $\mathrm{E}(T) = 0$ (para $n > 1$), $\mathrm{Var}(T) = n/(n-2)$ (para $n > 2$).

Relação entre a distribuição t e amostras aleatórias de distribuições normais

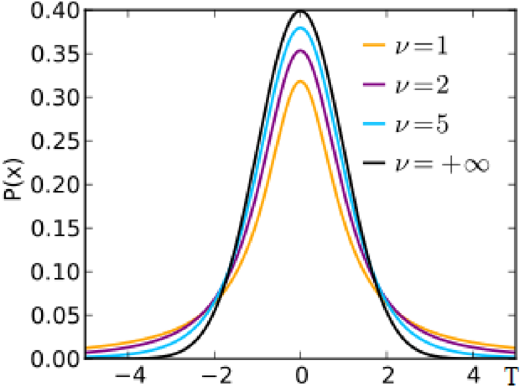
- Suponha que $X_1, \ldots, X_n$ seja uma amostra aleatória de uma distribuição normal com média μ e variância σ^2.
- Sejam $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$ e $Y = SS_X^2/\sigma^2$, onde $SS_X^2 = \sum_{i=1}^{m}(X_i - \overline{X})^2$ (soma dos quadrados dos desvios em relação à média).
- Então:
  - Z e Y são são independentes;
  - $Z \sim N(0, 1)$;
  - $Y \sim \chi^2(n-1)$.
- Logo, da definição da distribuição t segue que a variável

$$T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{\overline{X}-\mu}{\sqrt{s^2/n}}, \quad \text{onde} \quad s^2 = \frac{SS_X^2}{n-1},$$

segue uma distribuição t com $n-1$ graus de liberdade (DeGroot 1986, p.396).

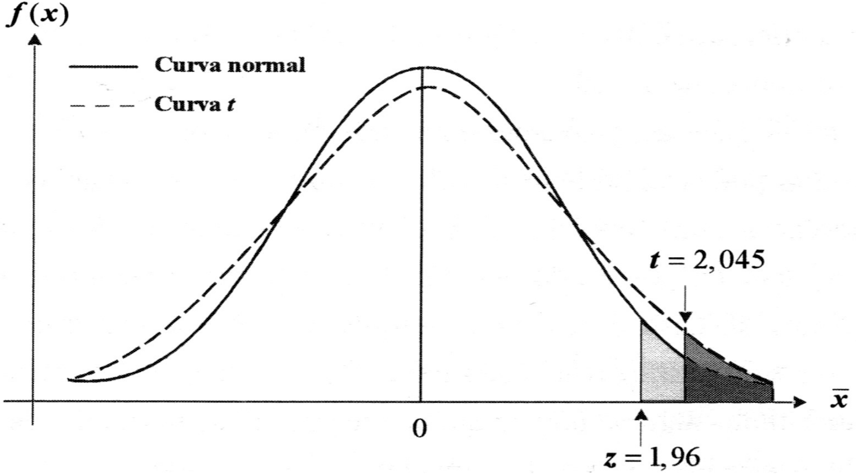
Distribuição t - Exemplos

- $\nu \to +\infty$: a distribuição t converge para a distribuição normal padrão.



Distribuição t - Exemplos

- Comparação entre a distribuição normal padrão e a distribuição *t* de Student para uma amostra com $n = 30$. Note a diferença dos valores críticos que determinam a região de significância de 0.05, bilateral.



Distribuição t - Intervalos de confiança para a média

- Em situações nas quais o desvio padrão σ da população é desconhecido e o tamanho n da amostra é pequeno ($n < 100$), a distribuição t de Student é mais apropriada do que a distribuição normal padrão para obter intervalos de confiança para μ

- Neste caso, usaremos a estatística T,

$$T = \frac{\overline{X} - \mu}{\sqrt{s^2/n}}, \quad \text{onde} \quad s^2 = \frac{SS_X^2}{n-1},$$

  a qual, como vimos, segue uma distribuição t com $n - 1$ graus de liberdade.

- Dado um coeficiente de confiança qualquer γ, devemos encontrar o valor crítico t_γ sob a distribuição t tal que $Pr(-t_\gamma < T < t_\gamma)$.

- O valor de t_γ é calculado por

$$t_\gamma = -F^{-1}_{t(n-1)}\left(\frac{1-\gamma}{2}\right),$$

  onde $F^{-1}_{t(n-1)}$ denota a função quantil (ou inversa da f.d.a.) da distribuição t com $n - 1$ graus de liberdade.

Distribuição t - Intervalos de confiança para a média

- No Excel: $F^{-1}_{t(n-1)}(p) = \text{INV.T}(p, n-1)$
- O intervalo de confiança para μ será:

$$(\overline{X} - t_\gamma s/\sqrt{n},\ \overline{X} + t_\gamma s/\sqrt{n}).$$

- Considere um exemplo no qual temos uma amostra proveniente de uma distribuição normal com média e variância desconhecidas onde $n = 30$, $\overline{X} = 25.9;\quad s = 15;\quad s/\sqrt{n} = 2.74$. Desejamos obter um intervalo de confiança Desejamos obter um intervalo de confiança de 95% para μ:
  - $t_\gamma = -F^{-1}_{t(n-1)}([1-\gamma]/2) = -F^{-1}_{t(29)}(0.025) = 2.045$
  - $t_\gamma s/\sqrt{n} = (2.045)(2.74) = 5.6$
  - Intervalo de confiança: $(\overline{X} - t_\gamma s/\sqrt{n},\ \overline{X} + t_\gamma s/\sqrt{n}) = (25.9 - 5.6,\ 25.9 + 5.6) = (20.3, 31.5)$

Teste t de Student para a média de uma população (distribuição normal, amostras pequenas, variância desconhecida)

- Seja $X_1, \ldots, X_n$ uma amostra de uma população com distribuição normal (possivelmente com $n \leq 30$) com média μ e desvio padrão σ desconhecidos
- Neste caso, a distribuição t pode ser usada para se realizar inferências sobre a média da população
- Estimamos σ segundo o estimador não viesado

$$s^2 = \frac{\sum_{i=1}^{m} (X_i - \overline{X})^2}{n-1}, \quad s = \sqrt{s^2}$$

- Utilizando-se a distribuição t para testes de hipóteses, a estatística de teste é

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

Esta estatística tem uma distribuição t com $n - 1$ graus de liberdade.

Teste t de Student para a média de uma população (distribuição normal, amostras pequenas, variância desconhecida)

- Logo, o teste de hipótese sobre μ é similar ao teste Z, com a exceção de que, nesse caso, usamos a distribuição t de Student para a estatística T, ao invés da distribuição normal padrão.
- Considere um exemplo no qual temos uma amostra proveniente de uma distribuição normal com média e variância desconhecidas onde $n = 25$, $\overline{X} = 25.9$; $s = 15$; $s/\sqrt{n} = 3$. Estamos interessados na hipótese $H_0 : \mu = 20$.

  Neste caso, $T = \frac{25.9-20}{3} = 1.97$

  Consideremos dois testes distintos:

  - $H_0 : \mu = 20$ contra $A : \mu \neq 20$:
    $pv = Pr(|T| > |t|) = 2Pr(T < -|t|) = 2Pr(T < -1.97) = 2F(-1.97; 24)) = 0,06$,
    onde $F(x; \nu)$ denota a f.d.a. t com ν graus de liberdade.
  - $H_0 : \mu \leq 20$ contra $A : \mu > 20$:
    $pv = Pr(T > t) = Pr(T > 1.97) = 1 - F(1.97; 24) = 0,03.$

Teste t de Student para a média de uma população (distribuição normal, amostras pequenas, variância desconhecida)

- Outro exemplo: TCB $\times$ uso de contraceptivo
  - Um pesquisador deseja saber se o uso de contraceptivos orais tem efeito sobre a temperatura corporal basal[1] (TCB) de mulheres na faixa de 18 a 25 anos.
  - Para tal finalidade, ele seleciona uma amostra de 20 mulheres que usam contraceptivos orais, e encontra uma temperatura média $\overline{X} = 36.7^oC$, com desvio $\hat{\sigma} = 0.5^oC$.
  - Ele deseja comparar esses dados com aqueles da população de mulheres na mesma faixa etária que não usam contraceptivos orais. A TCB média dessa população (μ_0) é assumida como 36.3^oC.
  - Considerando que os dados sejam normalmente distribuídos, existe diferença estatisticamente significativa entre a TCB média de mulheres com uso de contraceptivos orais (μ) e a TCB média de mulheres da população, na mesma faixa etária?

Teste t de Student para a média de uma população (distribuição normal, amostras pequenas, variância desconhecida)

- Exemplo: TCB $\times$ uso de contraceptivo (cont)
- $H_0 : \mu = \mu_0 = 36.3$
  $\overline{X} = 36.7$; $s^2 = 0.25$; $\sqrt{s^2/20} = 0.09$; $\nu = n - 1 = 19$
  $T = \frac{36.7 - 36.3}{0.09} = 4.44$
  Teste bicaudal ($A : \mu \neq 36.3$): $P(|T| > |4.44|) \approx 2.8\text{E-}4 (= 0.00028)$
  Teste monocaudal ($A : \mu > 36.3$): $P(T > 4.44) \approx 1.4\text{E-}4$

[1] Temperatura do corpo medida imediatamente após a pessoa acordar, antes de qualquer atividade física

Teste t para duas amostras independentes (distribuições normais, mesma variância)

- Sejam $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ amostras aleatórias independentes *com mesma variância*,

  $X_1, \ldots, X_m \sim N(\mu_1, \sigma^2), \;\; Y_1, \ldots, Y_n \sim N(\mu_2, \sigma^2)$

  (todos os parâmetros desconhecidos).
- Denotemos por $SS_X^2 = \sum_{i=1}^m (X_i - \overline{X})^2$ e $SS_Y^2 = \sum_{j=1}^m (Y_j - \overline{Y})^2$ (somas dos quadrados dos desvios em relação às médias).
- Note que $\overline{X} \sim N(\mu_1, \sigma^2/m)$ e $\overline{Y} \sim N(\mu_1, \sigma^2/n)$.
- Como $\overline{X}$ e $\overline{Y}$ são independentes, segue que a diferença $\overline{X} - \overline{Y}$ segue uma distribuição normal com média $\mu_1 - \mu_2$ e variância $\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2$.
- Logo, sob a hipótese $H_0 : \mu_1 = \mu_2$ e se σ fosse conhecida, a variável

$$Z_1 = \frac{\overline{X} - \overline{Y}}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2} \sigma}$$

  seguiria uma distribuição normal padrão.

Teste t para duas amostras independentes (distribuições normais, mesma variância)

- Adicionalmente, para quaisquer valores de μ_1, μ_2, σ^2, as variáveis aleatórias SS_X^2/σ^2 e SS_Y^2/σ^2 são independentes e possuem distribuições qui-quadrado com $m-1$ e $n-1$ graus de liberdade, respectivamente.

- Logo, a variável aleatória
$$Z_2 = \frac{SS_X^2}{\sigma^2} + \frac{SS_Y^2}{\sigma^2}$$
possui uma distribuição de qui-quadrado com $m+n-2$ graus de liberdade.

- Pelo fato de $\overline{X}, \overline{Y}, SS_X^2, SS_Y^2$ serem independentes (DeGroot, 1986, pg 509), segue que Z_1 e Z_2 são independentes.

- Portanto, quando $\mu_1 = \mu_2$, pela da definição da distribuição t, a estatística
$$T = \frac{Z_1}{[Z_2/(m+n-2)]^{1/2}} = \frac{(m+n-2)^{1/2}\,(\overline{X}-\overline{Y})}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}\left(SS_X^2+SS_Y^2\right)^{1/2}}$$
possui uma distribuição t com $m+n-2$ graus de liberdade.

Teste t para duas amostras independentes (distribuições normais, mesma variância)

- Exemplo: Um pesquisador deseja saber se a concentração de lipídios da espécie de peixe mapará é influenciada por dois diferentes métodos de medição.
- 10 amostras foram medidas pelo método 1, e 12 amostras foram medidas pelo método 2. Assume-se que as amostras são distintas (ou seja, feitas em espécimes diferentes).
- Dados são apresentados na tabela a seguir.
- Para um nível de significância de 0.05, há diferença significativa entre os dois métodos? Em outras palavras, as medidas médias são similares?

Teste t para duas amostras independentes (distribuições normais, mesma variância)

- Valores da concentração de lipídios da espécie de peixe mapará, medidos por dois diferentes métodos.

Amostra	Soxhlet (g/100g)	Blight Dyer (g/100g)
1	14.8	15.8
2	15.2	16.7
3	16.5	15.9
4	15.9	17.2
5	16.8	16.2
6	14.7	15.3
7	14.6	15.1
8	15.4	15.7
9	15.5	16.6
10	16.9	17.1
11		15.5
12		16.7

$H_0 : \mu_1 = \mu_2, \; A = \mu_1 \neq \mu_2$
$m = 10, \; n = 12$
$\overline{X} = 15.6, \; \overline{Y} = 16.2$
$SS_X^2 = 6.7, \; SS_Y^2 = 5.5$
$s_X^2 = 0.74, \; s_Y^2 = 0.50$
$T = -1.56$
$pv = Pr(|T| \leq -|1.56|) = 0.135$
$\Rightarrow$ diferenças não significativas

Teste t para duas amostras independentes (distribuições normais, variâncias distintas)

- Sejam $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ amostras aleatórias independentes *com variâncias distintas*,

  $X_1, \ldots, X_m \sim N(\mu_1, \sigma_1^2), \;\; Y_1, \ldots, Y_n \sim N(\mu_2, \sigma_2^2)$

  (todos os parâmetros desconhecidos).
- Sejam $s_X^2 = \sum_{i=1}^m (x_i - \overline{X})^2/(m-1)$ e $s_Y^2 = \sum_{j=1}^m (y_j - \overline{Y})^2/(n-1)$ (estimadores não viesados para variâncias).
- Note que $\overline{X} \sim N(\mu_1, \sigma^2/m)$ e $\overline{Y} \sim N(\mu_1, \sigma^2/n)$.
- Estatística T é dada por:

$$T = \frac{\overline{X} - \overline{Y}}{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^{1/2}}.$$

- Graus de liberdade estimados:

$$\hat{\nu} = \frac{(g_X + g_Y)^2}{g_X^2/(m-1) + g_Y^2/(n-1)}, \;\; \text{onde } g_X = \frac{s_X^2}{m}, g_Y = \frac{s_Y^2}{n}.$$

Teste t para duas amostras independentes (distribuições normais, variâncias distintas)

- Valores da concentração de lipídios da espécie de peixe mapará, medidos por dois diferentes métodos.

Amostra	Soxhlet (g/100g)	Blight Dyer (g/100g)
1	14.8	15.8
2	15.2	16.7
3	16.5	15.9
4	15.9	17.2
5	16.8	16.2
6	14.7	15.3
7	14.6	15.1
8	15.4	15.7
9	15.5	16.6
10	16.9	17.1
11		15.5
12		16.7

$H_0 : \mu_1 = \mu_2,\ A = \mu_1 \neq \mu_2$
$m = 10,\ n = 12$
$\overline{X} = 15.6,\ \overline{Y} = 16.2$
$s_X^2 = 0.74,\ s_Y^2 = 0.50$
$T = -1.53,\ \hat{\nu} = 17$
$pv = Pr(T \leq -1,53) + Pr(T \geq 1,53) = 0.144$
$\Rightarrow$ diferenças não significativas

Teste t para duas amostras pareadas

- Sejam $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ amostras aleatórias pareadas - medidas observáveis sobre os mesmos indivíduos ou sobre as mesmas condições - onde μ_1 e μ_2 são as médias (desconhecidas) das medidas X e Y na população.
- Considere as variáveis aleatórias $D_1 = X_1 - Y_1, \ldots, D_n = X_n - Y_n$. Denote por $\overline{D}$ e por s_D^2 a média e a variância amostrais de $D_1, \ldots, D_N$, respectivamente.
- Se $D_1, \ldots, D_n \sim N(\mu_D, \sigma_D^2)$, então sob a hipótese $H_0 : \mu_1 = \mu_2 \equiv H_0 : \mu_D = 0$, a estatística

$$T = \frac{\overline{D} - 0}{\sqrt{s_D^2/n}}$$

segue uma distribuição t com $n - 1$ graus de liberdade.

Teste t para duas amostras pareadas

- Valores da concentração de lipídios da espécie de peixe mapará, medidos por dois diferentes métodos sobre os mesmos espécimes.

Amostra	Soxhlet (g/100g)	Blight Dyer (g/100g)	D
1	14.8	15.8	-1.0
2	15.2	16.7	-1.5
3	16.5	15.9	0.6
4	15.9	17.2	-1.3
5	16.8	16.2	0.6
6	14.7	15.3	-0.6
7	14.6	15.1	-0.5
8	15.4	15.7	-0.3
9	15.5	16.6	-1.1
10	16.9	17.1	-0.2

$H_0 : \mu_1 = \mu_2,\ A = \mu_1 \neq \mu_2$
$m = 10,\ n = 12$
$\overline{X} = 15.6,\ \overline{Y} = 16.2,\ \overline{D} = -0.53$
$s_X^2 = 0.74,\ s_Y^2 = 0.52,\ s_D^2 = 0.53$
$T = -2.30$
$pv = Pr(T \leq -2.30) + Pr(T \geq 2.30) = 0.047$
$\Rightarrow$ diferenças significativas para $\alpha = 0.05$.

Distribuição qui-quadrado

- A distribuição qui-quadrado é um caso particular da distribuição gama.
- Função de densidade de probabilidade (fdp) da distribuição gama:

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{\beta x}, \; x > 0$$

  onde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}$ (função gama).
  $\alpha, \beta > 0$: parâmetros de forma e de escala.

- Distribuição qui-quadrado: para qualquer inteiro positivo k, a distribuição gama com $\alpha = k/2$ e $\beta = 1/2$ é denominada a distribuição qui-quadrado (χ^2) com k graus de liberdade:

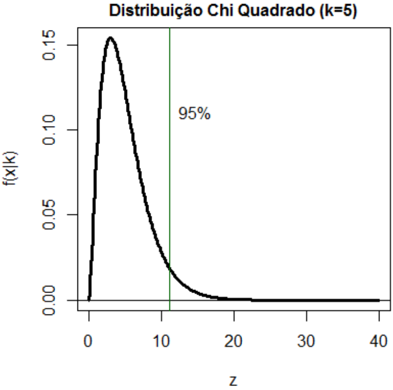
$$f(x|k) = \frac{1}{\Gamma(\alpha)} x^{(k/2)-1 e^{-x/2}}, \; x > 0.$$

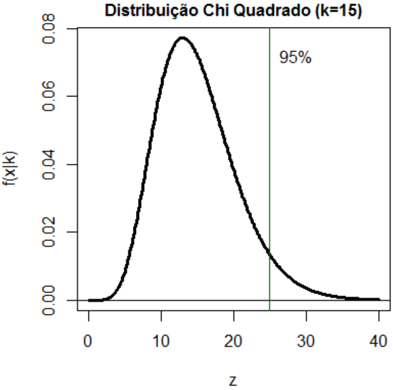
Distribuição qui-quadrado

- Principais propriedades:
  - Se $Y \sim \chi^2(n)$, então $\mathrm{E}(Y) = n$ e $\mathrm{Var}(Y) = 2n$.
  - Se $Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2), \ldots, Y_k \sim \chi^2(n_k)$, então $Y_1 + Y_2 + \ldots + Y_k \sim \chi^2(n_1 + n_2 + \ldots + n_k)$.
  - Se $Y_1, Y_2, \ldots, Y_k \sim N(0, 1)$, então $Y_1^2 + Y_2^2 + \ldots + Y_k^2 \sim \chi^2(k)$.
- Teorema: Suponha que $X_1, \ldots, X_n$ formam uma amostra aleatória de uma distribuição normal com média μ e variância σ^2. Então:
  - A média amostral $\overline{X}$ e a variância amostral SS_X^2/n são independentes[2];
  - $\overline{X} \sim N(\mu, \sigma^2/n)$;
  - $SS_X^2/\sigma^2 \sim \chi^2(n-1)$.

[2] $SS_X^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$

Distribuição qui-quadrado





Teste de qui-quadrado - Ideia Geral

- $X_n = x_1, x_2, \ldots, x_n$: amostra observada

  $E_n = e_1, e_2, \ldots, e_n$: valores esperados para $x_1, x_2, \ldots, x_n$ assumindo que a hipótese H_0 fosse verdadeira.

- Estatística qui-quadrado:

$$
\begin{aligned}
T &= \frac{(x_1 - e_1)^2}{e_1} + \frac{(x_2 - e_2)^2}{e_2} + \ldots + \frac{(x_n - e_n)^2}{e_n} \\
&= \sum_{i=1}^{n} \frac{(x_i - e_i)^2}{e_i}
\end{aligned}
$$

- Sob a hipótese H_0, T segue uma distribuição χ^2 com k graus de liberdade.

  Logo, uma vez calculada T, pode-se verificar se T está ou não na região crítica de rejeição sob χ^2.

- Como obter $e_1, \ldots, e_n$? Como obter k?
  - Depende de cada problema

Testes em tabelas de contingência

- Dados categóricos, categorias excludentes.
- Notação: X: matrix de frequências observadas; p: parâmetros

$$
\begin{array}{|cccc|c}
\hline
x_{11} & x_{12} & \dots & x_{1c} & x_{1\bullet} \\
x_{21} & x_{22} & \dots & x_{2c} & x_{2\bullet} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{r1} & x_{r2} & \dots & x_{rc} & x_{r\bullet} \\
\hline
x_{\bullet 1} & x_{\bullet 2} & \dots & x_{\bullet c} & n
\end{array}
,\qquad
\begin{array}{|cccc|c}
\hline
p_{11} & p_{12} & \dots & p_{1c} & p_{1\bullet} \\
p_{21} & p_{22} & \dots & p_{2c} & p_{2\bullet} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
p_{r1} & p_{r2} & \dots & p_{rc} & p_{r\bullet} \\
\hline
p_{\bullet 1} & p_{\bullet 2} & \dots & p_{\bullet c} & n
\end{array}
$$

$x_{i\bullet} = \sum_{j=1}^{c} x_{ij}$, $x_{\bullet j} = \sum_{i=1}^{r} x_{ij}$;
idem para $p_{i\bullet}, p_{\bullet j}$

Testes de qui-quadrado em tabelas de contingência

$$
\begin{array}{|cccc|c}
\hline
x_{11} & x_{12} & \dots & x_{1c} & x_{1\bullet} \\
x_{21} & x_{22} & \dots & x_{2c} & x_{2\bullet} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{r1} & x_{r2} & \dots & x_{rc} & x_{r\bullet} \\
\hline
x_{\bullet 1} & x_{\bullet 2} & \dots & x_{\bullet c} & n
\end{array}
,
\qquad
\begin{array}{|cccc|c}
\hline
p_{11} & p_{12} & \dots & p_{1c} & p_{1\bullet} \\
p_{21} & p_{22} & \dots & p_{2c} & p_{2\bullet} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
p_{r1} & p_{r2} & \dots & p_{rc} & p_{r\bullet} \\
\hline
p_{\bullet 1} & p_{\bullet 2} & \dots & p_{\bullet c} & n
\end{array}
$$

- Independência:
  - Duas variáveis categóricas são consideradas simultaneamente.
  - p_{ij}: Probabilidade do indivíduo pertencer à $i-$ésima categoria na 1a variável e à j categoria na 2a variável.
  - x_{ij}: Frequência observada de indivíduos pertencentes simultaneamente à categoria i (1a variável) e j (2a variável)
  - Hipótese: independência entre variáveis. $H_0 : p_{ij} = p_{i\bullet} \times p_{\bullet j}$
    $e_{ij} = x_{i\bullet} \times x_{\bullet j}/n$
    $k = (r-1) \times (c-1)$

Referências

DeGroot M.H. (1986). Probability and Statistics, 2nd Ed. Menlo Park, CA: Addison-Wesley

G.B.Drummond and B.D.Tom (2011). How can we tell if frogs jump further? *Br J Pharmacol* **164**(2): 209 –212.

Mitchell, T.M. (1997). Machine Learning. McGraw-Hill.

POPPER, K. (1953). Science: Conjectures and Refutations. `http://poars1982.files.wordpress.com/2008/03/science-conjectures-and-refutations.pdf`

Stern, J.M. (2011). Constructive Verification Empirical Induction, and Falibilist Deduction: A Threefold Contrast. *Information* **2**, 635–650.