Testes de Hipóteses: Abordagem Clássica

Marcelo S. Lauretto

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

marcelolauretto@usp.br

São Paulo - Brasil

Problema: Decidir se uma moeda é honesta I

- Suponha que você seja juiz de uma partida de futebol. Antes do início da partida, chama os capitães das duas equipes, para sorteio de quem iniciará a partida com a bola.
- Pelas regras do torneio, a posse inicial de bola é decidida através do lançamento de uma moeda: se a moeda 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.
- Ao colocar a mão no bolso, você 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 com a condição de que a moeda seja "validada" antes de ser oficialmente lançada para decidir a posse de bola.

Problema: Decidir se uma moeda é honesta II

- O sorteio consiste em lançar a moeda 10 vezes sob aproximadamente as mesmas condições e contar a quantidade de caras e coroas.
- Em quais dos resultados abaixo você desconfiaria da procedência da moeda?
  1. 5 caras e 5 coroas?
  2. 6 caras e 4 coroas?
  3. 4 caras e 6 coroas?
  4. 2 caras e 8 coroas?
  5. 1 cara e 9 coroas?
  6. 0 caras e 10 coroas?
- Uma pergunta mais geral: Para quais dos possíveis resultados você consideraria que a moeda não é honesta?
  - Para responder a essa questão: *Procedimento de teste.*

Problema: Decidir se uma moeda é honesta III

- Sob a abordagem de estatística clássica, o procedimento de teste para decidir se a moeda é honesta depende da definição dos seguintes elementos:

  1. Condição do experimento e respectiva estatística:
     - 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 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:
       $H_0, H_1 \neq \emptyset;\ \ H_0 \cap H_1 = \emptyset;\ \ H_0 \cup H_1 = \Omega$

Problema: Decidir se uma moeda é honesta IV

- (cont.)

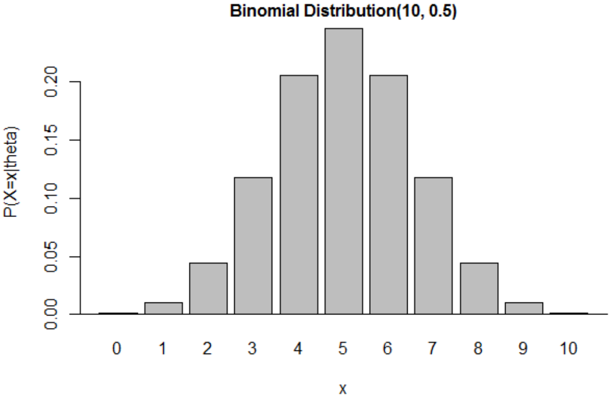
4. Distribuição de probabilidade dos possíveis resultados do experimento:
    - $P(X = x|p)$: probabilidade de x caras em n lançamentos, dado o parâmetro p:

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

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 V

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



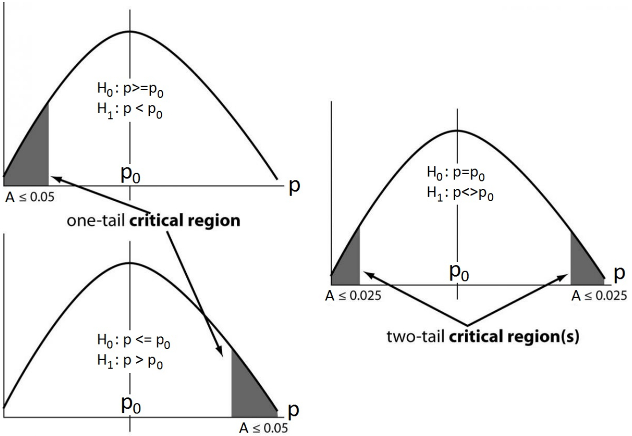
Como interpretar (e definir) α? I

- 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
    Poder do teste: Complemento do erro do tipo II, representa a probabilidade de rejeitar corretamente uma hipótese falsa
- Objetivos conflitantes: Baixo Erro do Tipo I implica em alto Erro do Tipo II e vice-versa

Como interpretar (e definir) α? II

- 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 - Distribuições simétricas



Voltando ao problema da moeda: I

- Como definir as hipóteses nula e alternativa?

  (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: II

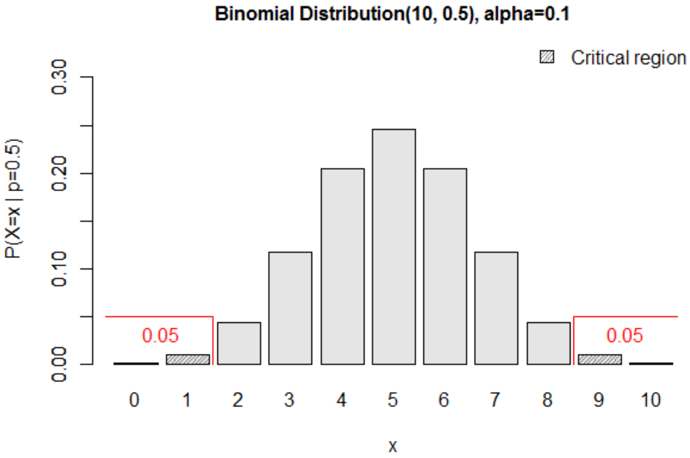
- Logo, juiz deve escolher uma das três hipóteses (e respectivas regiões de rejeição):
  - $H_0 : p = 1/2$, $H_1 : p \neq 1/2$: alta proporção de caras ou de coroas é considerada suspeita
    – Posição mais neutra: moeda é rejeitada se qualquer um dos times puder ser prejudicado por eventual vício na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

- 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: IV



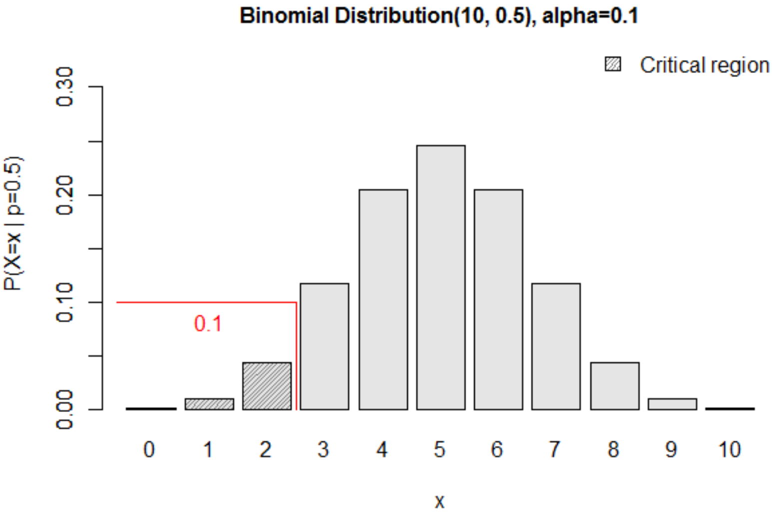
$$C = \{x \mid P(X \leq x|p) \leq \alpha/2 \vee P(X \geq x|p) \leq \alpha/2\} = \{0, 1, 9, 10\}$$

Voltando ao problema da moeda: V

- 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: VI



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

Falseabilidade (ou Refutabilidade) de Popper

- 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.
  - 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 ou boa qualidade da teoria.

- Distribuição Normal - Importância:
  - Capaz de descrever vários fenômenos físicos e biológicos
  - Teorema do Limite Central
- Função de densidade de probabilidade (pdf):

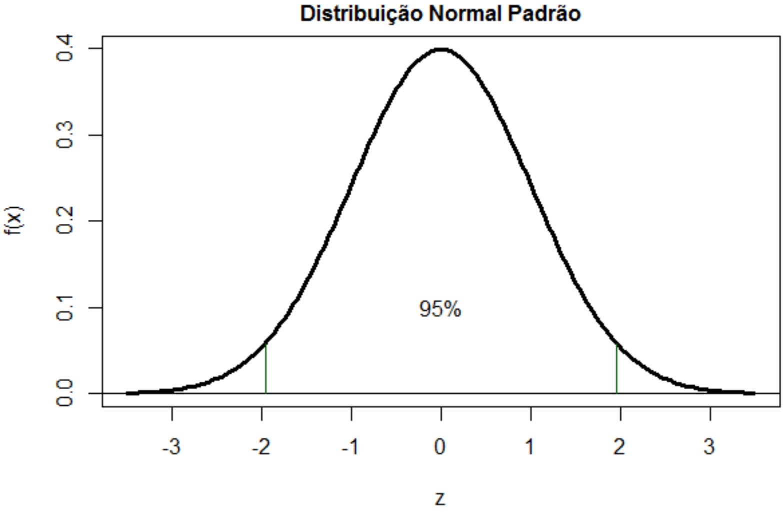
$$f(x|\mu,\sigma) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Parâmetros:
  - μ: média da população;
  - σ^2: variância da população; $\quad \sigma = \sqrt{\sigma^2}$: desvio padrão
- Se as variáveis aleatórias $X_1, X_2, \ldots, X_k$ forem independendes e $X_i \sim N(\mu_i, \sigma_i^2)$, $(i = 1, \ldots, k)$, então

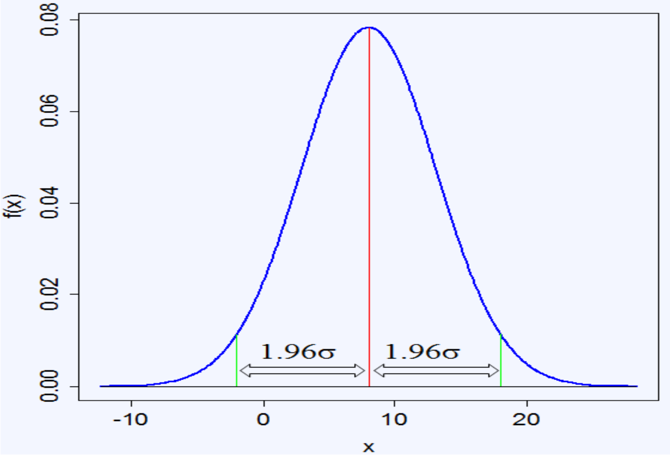
$$X_1 + \ldots + X_k \sim N(\mu_1 + \ldots + \mu_k,\ \sigma_1^2 + \ldots + \sigma_k^2).$$

- Se x segue uma distribuição normal, isto é, se $X \sim N(\mu, \sigma^2)$, então
$$z = \frac{(x - \mu)}{\sigma} \sim N(0, 1).$$
$N(0, 1)$: Distribuição normal padrão.
- É conhecido que
$P(-1.96 < z < +1.96) = 95\%$ e portanto
$P(\mu - 1.96\sigma < x < \mu + 1.96\sigma) = 95\%$
- Logo, o intervalo $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ fornece um *intervalo de previsão para x*.

- Distribuição normal padrão



- Distribuição normal com $\mu = 8, \sigma = 5$



Distribuição Normal - Estimadores

- Valores verdadeiros de μ e σ são quase sempre desconhecidos.
- Suponha que $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$ seja uma amostra aleatória de uma distribuição Normal com média μ e variância σ^2 (desconhecidos).
- Estimadores de máxima verossimilhança para μ e σ^2 são dados por

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i;\ \hat{\sigma}^2 = \frac{S_X^2}{n},\ \text{onde } S_X^2 = \sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2.$$

- Estimador *não viesado*[1] para σ^2:

$$s^2 = \frac{S_X^2}{n-1}. \tag{1}$$

[1] $E(s^2) = \sigma^2$

Intervalo de confiança para a média

- *Teorema do Limite Central*:

  Se $\boldsymbol{X}$ é uma amostra aleatória de tamanho n de uma distribuição qualquer com média μ e variância σ^2, então a distribuição da média amostral seguirá aproximadamente uma distribuição Normal com média μ e variância σ^2/n (DeGroot, 1986, p. 275).

- Logo, para amostras grandes ou oriundas de uma população com distribuição normal, pode-se obter um intervalo de 95% de confiança para μ por:

$$CI_{95\%} = [li(X_n), ls(X_n)]$$

  onde

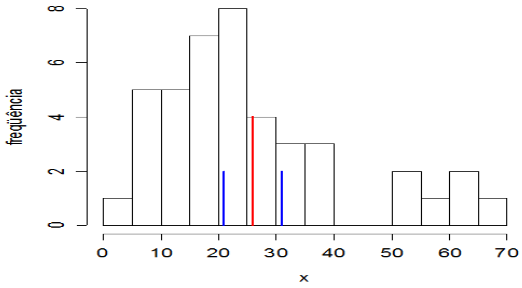
  - se σ^2 é conhecida:
    $li(\boldsymbol{X}) = \overline{X} - 1.96\sqrt{\sigma^2/n}, \quad ls(\boldsymbol{X}) = \overline{X} + 1.96\sqrt{\sigma^2/n}$
  - se σ^2 é desconhecida:
    $li(\boldsymbol{X}) = \overline{X} - 1.96\sqrt{s^2/n}, \quad ls(\boldsymbol{X}) = \overline{X} + 1.96\sqrt{s^2/n}$
    s^2: estimador não viesado (Eq. 1)

Intervalo de confiança para a média

- Exemplo: Considere a amostra representada pelo histograma abaixo. A linha vertical central (em vermelho) representa a média amostral, e as linhas horizontais azuis representam o intervalo de 95% de confiança para μ.

  $n = 42, \quad \overline{X} = 25.9; \quad \hat{\sigma} = 16.49; \quad \hat{\sigma}/\sqrt{42} = 2.54$

  $CI_{95\%} = [20.9, 30.9]$



Intervalo de confiança: definição informal

- Seja $\boldsymbol{X}$ uma amostra aleatória oriunda de uma distribuição de probabilidade com parâmetro p a ser estimado. Um *intervalo de confiança* para o parâmetro p, com *nível de confiança* ou *coeficiente de confiança* γ, é um intervalo $[li(\boldsymbol{X}), ls(\boldsymbol{X})]$ determinado pelo par de variáveis aleatórias $li(X)$ e $ls(X)$, com a propriedade:

$$P\left(li(\boldsymbol{X}) < p < ls(\boldsymbol{X})\right) = \gamma, \quad \text{para todo } p.$$

- Interpretação do IC: se coletássemos indefinidamente amostras aleatórias da mesma população e, para cada amostra coletada $\boldsymbol{X}$, calculássemos $li(\boldsymbol{X})$ e $lu(\boldsymbol{X})$, em $100\gamma\%$ das repetições o valor verdadeiro de p estaria dentro dos intervalos obtidos.

Intervalos de confiança $\times$ testes de hipóteses

- Seja X_n uma amostra de uma população com distribuição normal, e considere a hipótese

  $H_0 : \mu = \mu_0$ contra a alternativa $A : \mu \neq \mu_0$.

- Uma maneira simples para testar H seria construir um intervalo de confiança com coeficiente γ para μ, e verificar se μ_0 pertence a esse intervalo.

  - Se $\mu_0 \in [li(X_n), ls(X_n)]$: não rejeitamos H_0;
  - Se $\mu_0 \notin [li(X_n), ls(X_n)]$: rejeitamos H_0, com nível de significância α.

- O valor de gama é o complementar do nível de significância desejado, ou seja,

  $\gamma = 1 - \alpha$.

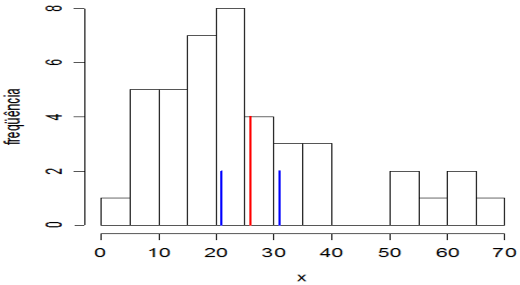
- Atenção: somente vale para testes bicaudais ($A : \mu \neq \mu_0$).

Teste Z para a média de uma população (Distribuição normal, variância conhecida)

- Outra forma (um pouco mais geral) de resolver o problema anterior:
- Seja X_n uma amostra de uma população com distribuição normal com média μ desconhecida e variância σ^2, e considere a hipótese $H_0 : \mu = \mu_0$.
- Se a hipótese for verdadeira $\mu = \mu_0$, então $\overline{X}_n \sim N(\mu_0, \sigma^2/n)$.
- Logo, a *estatística* $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$!!
- Assim, para testar a hipótese original, basta verificar em qual região da distribuição normal padrão a *estatística* Z se encontra.

Teste Z (Distribuição normal, variância conhecida)

- Exemplo anterior (assumindo $\sigma = 16.49$):
  $n = 42, \quad \overline{X} = 25.9; \quad \sigma/\sqrt{42} = 2.54$



- Supondo $H_0 : \mu = 20$.
  $Z = \frac{25.9-20}{2.54} = 2.32$
  Teste bicaudal ($A : \mu \neq 20$): $P(|Z| \geq 2.32) \approx 0.02$
  Teste monocaudal ($A : \mu > 20$): $P(Z \geq 2.32) \approx 0.01$

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 $Y \sim N(0,1)$ e $Z \sim \chi^2(n)$.

  Seja X a variável aleatória definida pela equação

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

  Então a distribuição de X é denominada distribuição t com n graus de liberdade.
- Função de densidade de probabilidade:

  $$f(x|n) = \frac{\Gamma(n+1)/2}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \qquad -\infty < x < \infty.$$
- Média e Variância: Se $X \sim t(n)$:

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

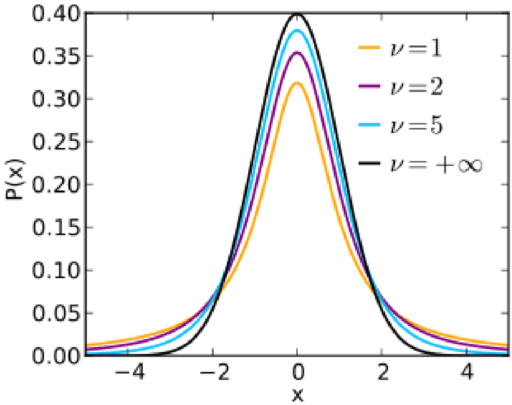
Relação entre a distribuição t e amostras aleatórias de distribuições normais

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

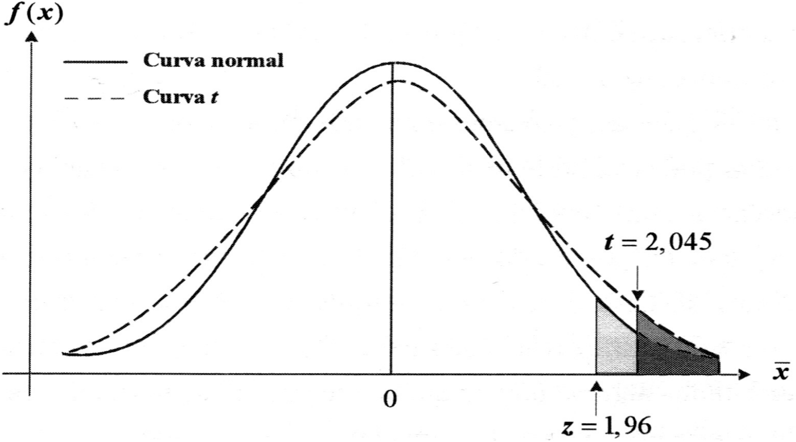
$$T = \frac{Y}{\sqrt{Z/(n-1)}} = \frac{\overline{X} - \mu}{\sqrt{s^2/n}}, \quad \text{onde} \quad s^2 = \frac{S_X^2}{n-1} \ ,$$

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

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



- 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.



Teste t para uma amostra (Distribuição normal, variância desconhecida)

- Seja $X_1, \ldots, X_n$ uma amostra de uma população com distribuição normal com média μ desconhecida e variância σ^2, e considere a hipótese $H_0 : \mu = \mu_0$.
- Se a hipótese for verdadeira $\mu = \mu_0$, então $\overline{X} \sim N(\mu_0, \sigma^2/n)$.
- Logo, pelo resultado anterior, a *estatística* $T = \frac{\overline{X}-\mu_0}{\sqrt{s^2/n}}$ segue uma distribuição t de Student com $n-1$ graus de liberdade.
- Assim, para testar a hipótese original, basta verificar em qual região da distribuição t de student a *estatística* T se encontra.
- No exemplo apresentado anteriormente, supondo $H_0 : \mu = 20$.
  $n = 42, \quad \overline{X} = 25.9; \quad s^2 = 271.92; \quad \sqrt{s^2/n} = 2.54, \quad \nu = 41$
  $T = \frac{25.9-20}{2.54} = 2.32$
  Teste bicaudal ($A : \mu \neq 20$): $P(|T| \geq 2.32) \approx 0.026$
  Teste monocaudal ($A : \mu > 20$): $P(T \geq 2.32) \approx 0.013$

Teste t para uma amostra (Distribuição normal, 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[2] (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 $\overline{X} = 36.7^o C$, com desvio $\hat{\sigma} = 0.5^o C$.
  - 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 $36.3^o C$.
  - 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?

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

Teste t para uma amostra (Distribuição normal, 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 ($A : \mu \neq 36.3$): $P(|T| \geq 4.44) \approx 2.8E - 4$
  Teste monocaudal ($A : \mu > 36.3$): $P(T \geq 4.44) \approx 1.4E - 4$

Teste t para duas amostras (mesma variância)

- Sejam $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ amostras aleatórias independentes, $X_1, \ldots, X_m \sim N(\mu_1, \sigma^2)$, $\;Y_1, \ldots, Y_n \sim N(\mu_2, \sigma^2)$ (todos os parâmetros desconhecidos).
- Denote por $S_X^2 = \sum_{i=1}^m (x_i - \overline{X})$ e $S_Y^2 = \sum_{j=1}^m (y_j - \overline{Y})$.
- 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 $\left(\frac{1}{m} + \frac{1}{m}\right)\sigma^2$.
- Logo, quando $\mu_1 = \mu_2$, a variável

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

segue uma distribuição normal padrão.

Teste t para duas amostras (mesma variância)

- Adicionalmente, para quaisquer valores de μ_1, μ_2, σ^2, as variáveis aleatórias S_X^2/σ^2 e S_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
$$Z_2 = \frac{S_X^2 + S_Y^2}{\sigma^2}$$
possui uma distribuição de qui-quadrado com $m+n-2$ graus de liberdade.
- Pelo fato de $\overline{X}, \overline{Y}, S_X^2, S_Y^2$ 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
$$T = \frac{Z_1}{[Z_2/(m+n-2)]^{1/2}} = \frac{(m+n-2)^{1/2}\left(\overline{X} - \overline{Y}\right)}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}\left(S_X^2 + S_Y^2\right)^{1/2}}$$
possui uma distribuição t com $m+n-2$ graus de liberdade.

Teste t para duas amostras (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 (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

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

Teste t para duas amostras (variâncias distintas)

- Sejam $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ amostras aleatórias independentes, $X_1, \ldots, X_m \sim N(\mu_1, \sigma_1^2)$, $\;Y_1, \ldots, Y_n \sim N(\mu_2, \sigma_2^2)$ (todos os parâmetros desconhecidos).
- Problema conhecido como *problema de Behrens-Fisher*.
- Sejam $s_X^2 = \frac{1}{m-1}\sum_{i=1}^m (x_i - \overline{X})$ e $s_Y^2 = \frac{1}{n-1}\sum_{j=1}^m (y_j - \overline{Y})$ (variâncias amostrais).
- 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 (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

- Distribuição Gama:

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

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

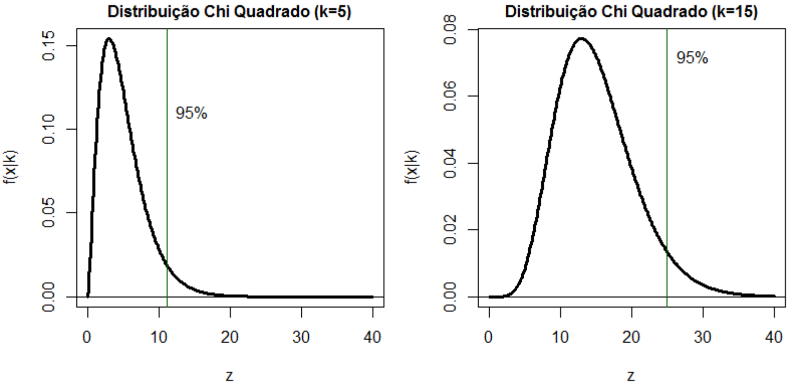
- Distribuição qui-quadrado: para qualquer inteiro positivo k, a distribuição gama com $\alpha = k/2$ e $\beta = 1/2$ é denominada a distribuição qui-quadrado (χ^2) com k graus de liberdade:

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

- 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 S_X^2/n são independentes[3];
  - $\overline{X} \sim N(\mu, \sigma^2/n)$;
  - $S_X^2/\sigma^2 \sim \chi^2(n-1)$.

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

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

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

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