

Aula 0.1: Revisão de Inferência Estatística

Econometria I - IBMEC

Marcus L. Nascimento

29 de outubro de 2025

1. População x Amostra

2. Estimação

3. Testes de Hipóteses

População x Amostra

- **Inferência Estatística:** Conjunto de técnicas que objetiva estudar uma população com base em informações obtidas a partir de uma amostra.
- **População** é um conjunto completo de elementos que compartilham uma ou mais características em comum.
 - Em pesquisas, corresponde ao conjunto total de indivíduos, objetos ou eventos que atendem aos critérios do estudo.
- **Amostra** é um subconjunto da população que contém os elementos observados.
 - A partir da amostra, quantidades de interesse podem ser medidas.

Exemplo

- Como exemplo, podemos considerar a análise da proporção de alunos de economia do IBMEC que desejam atuar no mercado financeiro.
 - População: alunos de economia do IBMEC;
 - Amostra: alunos da unidade Barra da Tijuca, alunos de Econometria I, alunos do 4º período.
- Supondo uma amostra de alunos de Econometria I, que cuidados precisamos ter na escolha e na interpretação dos resultados?

Exemplo

- Como exemplo, podemos considerar a análise da proporção de alunos de economia do IBMEC que desejam atuar no mercado financeiro.
 - População: alunos de economia do IBMEC;
 - Amostra: alunos da unidade Barra da Tijuca, alunos de Econometria I, alunos do 4º período.
- Supondo uma amostra de alunos de Econometria I, que cuidados precisamos ter na escolha e na interpretação dos resultados?
 - A amostra precisa ser “representativa” dos alunos de economia do IBMEC.

- Uma forma simples de coletar uma amostra seria selecionar aleatoriamente n alunos.

- Uma forma simples de coletar uma amostra seria selecionar aleatoriamente n alunos.
 - Se dois analistas seguirem tal procedimento, as amostras sorteadas serão iguais?

- Uma forma simples de coletar uma amostra seria selecionar aleatoriamente n alunos.
 - Se dois analistas seguirem tal procedimento, as amostras sorteadas serão iguais?
 - E se vários analistas realizarem o procedimento?

- Uma forma simples de coletar uma amostra seria selecionar aleatoriamente n alunos.
 - Se dois analistas seguirem tal procedimento, as amostras sorteadas serão iguais?
 - E se vários analistas realizarem o procedimento?
 - Apesar de diferentes, podemos ter respostas próximas ou iguais nas diversas amostras?

- Uma forma simples de coletar uma amostra seria selecionar aleatoriamente n alunos.
 - Se dois analistas seguirem tal procedimento, as amostras sorteadas serão iguais?
 - E se vários analistas realizarem o procedimento?
 - Apesar de diferentes, podemos ter respostas próximas ou iguais nas diversas amostras?
- Devido à natureza aleatória geralmente envolvida no processo amostral, não podemos garantir que repetições de amostras sempre produzam resultados idênticos.
- As quantidades associadas à amostra terão caráter aleatório e, portanto, devem receber tratamento probabilístico.

Estimação

Parâmetro, estimador e estimativa

- Denotaremos por X_1, \dots, X_n uma amostra de tamanho n extraída da população.
- **Parâmetro:** Quantidade da população, em geral desconhecida, sobre a qual temos interesse.
 - Usualmente, representado por letras gregas tais como θ , β , μ e σ ;
 - μ e σ são utilizados como notação para a média e o desvio padrão populacionais.
- **Estimador:** Combinação de elementos da amostra construída para estimar um parâmetro de interesse na população.
 - Geralmente, representado por letras gregas com acento circunflexo: $\hat{\theta}$, $\hat{\beta}$, $\hat{\mu}$ e $\hat{\sigma}$.
- **Estimativa:** Valores numéricos assumidos pelos estimadores.

Exemplo

- Note que um **estimador** $\hat{\theta}$ é uma função das variáveis aleatórias constituintes da amostra, $\hat{\theta} = f(X_1, \dots, X_n)$ e, portanto, também é uma **variável aleatória**.
- Retomando o exemplo no qual desejamos estimar a proporção de alunos de economia do IBMEC (**população**) que desejam atuar no mercado financeiro a partir da turma de Econometria I (**amostra**).
 - $X_i, i = 1, \dots, n$, é uma variável aleatória que assume o valor 1 caso o aluno deseje atuar no mercado financeiro e 0 caso contrário.
 - Qual função dos valores amostrais (**estimador**) podemos utilizar?

Exemplo

- Note que um **estimador** $\hat{\theta}$ é uma função das variáveis aleatórias constituintes da amostra, $\hat{\theta} = f(X_1, \dots, X_n)$ e, portanto, também é uma **variável aleatória**.
- Retomando o exemplo no qual desejamos estimar a proporção de alunos de economia do IBMEC (**população**) que desejam atuar no mercado financeiro a partir da turma de Econometria I (**amostra**).
 - $X_i, i = 1, \dots, n$, é uma variável aleatória que assume o valor 1 caso o aluno deseje atuar no mercado financeiro e 0 caso contrário.
 - Qual função dos valores amostrais (**estimador**) podemos utilizar?
 1. $\hat{\theta}_1 = f_1(X_1, \dots, X_n) = 100 \times \frac{(X_1 + X_n)}{2}$;
 2. $\hat{\theta}_2 = f_2(X_1, \dots, X_n) = 100 \times X_1$;
 3. $\hat{\theta}_3 = f_3(X_1, \dots, X_n) = 100 \times \frac{\sum_{i=1}^n X_i}{n}$.

Propriedades dos estimadores

- Diante de diferentes estimadores para um mesmo parâmetro, como decidir qual deles utilizar?

Propriedades dos estimadores

- Diante de diferentes estimadores para um mesmo parâmetro, como decidir qual deles utilizar?
 - Estudando as propriedades dos estimadores.

Propriedades dos estimadores

- Diante de diferentes estimadores para um mesmo parâmetro, como decidir qual deles utilizar?
 - Estudando as propriedades dos estimadores.
- **Vício:** Um estimador $\hat{\theta}$ é *não viciado* ou *não viesado* para um parâmetro θ se $E(\hat{\theta}) = \theta$.
- **Consistência:** Um estimador $\hat{\theta}$ é consistente se, à medida que o tamanho da amostra aumenta ($n \rightarrow \infty$), seu valor esperado converge para o valor de interesse e sua variância converge para 0.

$$i) \lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta;$$

$$ii) \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0.$$

Propriedades dos estimadores

- Observações:
 - Na definição de consistência, o estimador necessita ser não viciado apenas para valores grandes de n ;
 - Na definição de vício, o resultado deve valer para qualquer n .
- **Eficiência:** Dados dois estimadores não viciados $\hat{\theta}_1$ e $\hat{\theta}_2$ para um parâmetro θ , dizemos que $\hat{\theta}_1$ é mais eficiente do que $\hat{\theta}_2$ se $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$.
- Supondo X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com média μ e variância σ^2 , o que pode ser dito sobre os estimadores $\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ (média) e $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$ (variância)?

Estimação por intervalos

- Por serem variáveis aleatórias, os **estimadores** possuem distribuição de probabilidade.
- Podemos apresentar uma estimativa mais informativa do parâmetro, incluindo uma medida de imprecisão para a **estimativa pontual**.
- **Intervalo de confiança** incorpora, à estimativa pontual do parâmetro, informações a respeito da variabilidade do estimador.
- Intervalos de confiança são obtidos a partir da distribuição amostral de seus estimadores.

Estimação por intervalos (Caso Normal)

Suponha que a variável de interesse seja $X \sim N(\mu, \sigma^2)$ e que X_1, X_2, \dots, X_n seja uma amostra aleatória cujos elementos são i.i.d com densidade Normal de média μ e variância σ^2 .

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, \dots, n;$$

X_i é independente de X_j para todo $i \neq j$.

Qual a distribuição amostral de $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$?

Estimação por intervalos (Caso Normal)

A partir da distribuição amostral de $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, temos:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Fixando um valor γ tal que $0 < \gamma < 1$, é possível encontrar um valor $z_{\gamma/2}$ tal que

$$P(|Z| < z_{\gamma/2}) = P(-z_{\gamma/2} < Z < z_{\gamma/2}) = \gamma.$$

O valor de $z_{\gamma/2}$ pode ser obtido através da tabela da Normal Padrão.

Estimação por intervalos (Caso Normal)

O intervalo de confiança para μ com coeficiente de confiança γ é dado por

$$\text{IC}(\mu, \gamma) = \left[\bar{X} - z_{\gamma/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\gamma/2} \frac{\sigma}{\sqrt{n}} \right].$$

Interpretação:

- γ representa a probabilidade do intervalo conter o verdadeiro valor da média populacional μ ;
- Supondo várias amostras de mesmo tamanho, ao calcularmos os intervalos de confiança, esperamos que a proporção de intervalos que contenham o valor de μ seja igual a γ .

Estimação por intervalos (Teorema Central do Limite)

Suponha uma amostra aleatória simples de tamanho n retirada de uma população com média μ e variância σ^2 . Representando tal amostra por n variáveis aleatórias independentes X_1, X_2, \dots, X_n e denotando sua média por \bar{X} , temos que

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{D} Z.$$

onde $Z \sim N(0, 1)$.

Observações:

- Note que o modelo da variável aleatória não é especificado;
- O Teorema garante que para n grande a distribuição da média amostral, devidamente padronizada, se comporta segundo um modelo Normal com média 0 e variância 1.

Estimação por intervalos (Caso Bernoulli)

Seja Y uma variável aleatória tal que

$$Y = \begin{cases} 1, & \text{se determinada característica é observada;} \\ 0, & \text{caso contrário.} \end{cases}$$

A proporção amostral pode ser escrita da seguinte forma:

$$\hat{p} = \frac{\sum_{i=1}^n Y_i}{n} = \hat{Y}.$$

Supondo que Y_1, Y_2, \dots, Y_n forma uma sequência de variáveis aleatórias independentes com distribuição Bernoulli.

- Para um tamanho de amostra grande, qual a distribuição aproximada de \hat{p} ?
- Construa o intervalo de confiança para p .

Testes de Hipóteses

Conceitos Básicos

- O objetivo de um teste de hipóteses é verificar se há evidência suficiente contra a hipótese nula, ou seja, para rejeitá-la.
 - Hipótese nula (H_0): Representa o *status quo*. Esta hipótese é tida como verdadeira a menos que haja evidência suficiente sugerindo o contrário;
 - Hipótese alternativa (H_1): Representa a hipótese que o analista deseja concluir.
- Supondo que a área de *marketing* de uma companhia deseje avaliar a eficácia de uma nova campanha publicitária. Como as hipóteses podem ser definidas:

Conceitos Básicos

- O objetivo de um teste de hipóteses é verificar se há evidência suficiente contra a hipótese nula, ou seja, para rejeitá-la.
 - Hipótese nula (H_0): Representa o *status quo*. Esta hipótese é tida como verdadeira a menos que haja evidência suficiente sugerindo o contrário;
 - Hipótese alternativa (H_1): Representa a hipótese que o analista deseja concluir.
- Supondo que a área de *marketing* de uma companhia deseje avaliar a eficácia de uma nova campanha publicitária. Como as hipóteses podem ser definidas:

H_0 : A campanha não é eficaz;

H_1 : A campanha é eficaz.

- Em um teste de hipóteses, duas decisões são possíveis:
 1. Rejeitar a hipótese nula;
 2. Não rejeitar a hipótese nula.
- Por que não dizemos “aceitar a hipótese nula”?

- Em um teste de hipóteses, duas decisões são possíveis:
 1. Rejeitar a hipótese nula;
 2. Não rejeitar a hipótese nula.
- Por que não dizemos “aceitar a hipótese nula”?
 - Estamos assumindo que a hipótese nula é verdadeira e desejamos verificar se há evidência contra a mesma. Logo, a conclusão deve referir-se à rejeição da hipótese nula.

Conceitos Básicos

- Dois erros podem ser cometidos ao realizarmos um teste de hipóteses:
 - Erro do Tipo I: Rejeitar a hipótese H_0 quando a mesma é verdadeira;
 - Erro do Tipo II: Não rejeitar a hipótese H_0 quando esta é falsa.

Decisão	Situação Real	
	H_0 Verdadeira	H_0 Falsa
Rejeitar H_0	Erro tipo I	Decisão Correta
Não Rejeitar H_0	Decisão Correta	Erro tipo II

- Como estamos tratando de eventos em que há incerteza associada, desconhecemos a “verdade” acerca da hipótese nula. Em outras palavras, não sabemos se a decisão tomada foi correta ou se cometemos um erro.
- Em situações reais, podemos apenas definir as probabilidades associadas a tais eventos:

$$\alpha = P(\text{erro Tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ verdadeira});$$

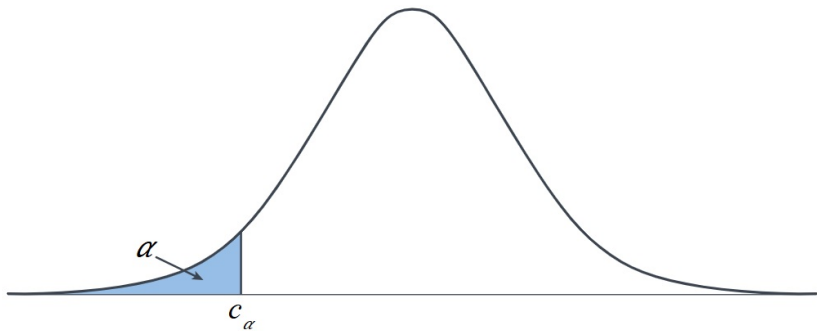
$$\beta = P(\text{erro Tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ falsa}).$$

- **Poder** do teste: Probabilidade de rejeitar H_0 quando esta é falsa ($1 - \beta$).

- O erro mais importante a ser evitado é o erro do Tipo I.
- **Nível de significância** do teste corresponde à probabilidade α de erro do Tipo I.
- Com base no nível de significância α podemos determinar:
 1. **Região de rejeição** ou **Região crítica**: Conjuntos de valores para a estatística de teste que levam à rejeição de H_0 ;
 2. **Valores críticos**: Valores que separam as regiões de rejeição e de não rejeição.
- A região de rejeição baseia-se na hipótese alternativa.

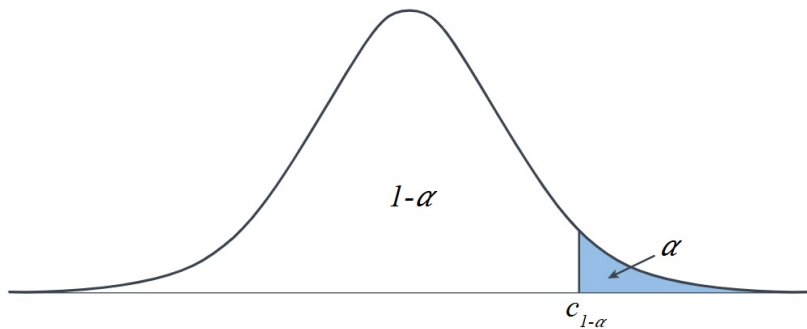
Conceitos Básicos (Teste unilateral à esquerda)

- Hipóteses:
 $H_0: \theta \geq \theta_0$;
 $H_1: \theta < \theta_0$.
- Rejeita H_0 se a estatística de teste for menor que o valor crítico (c_α).



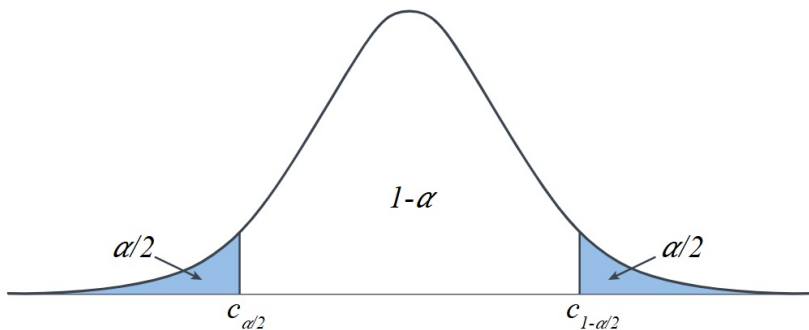
Conceitos Básicos (Teste unilateral à direita)

- Hipóteses:
 $H_0: \theta \leq \theta_0$;
 $H_1: \theta > \theta_0$.
- Rejeita H_0 se a estatística de teste for maior que o valor crítico ($c_{1-\alpha}$).



Conceitos Básicos (Teste bilateral)

- Hipóteses:
 $H_0: \theta = \theta_0;$
 $H_1: \theta \neq \theta_0.$
- Rejeita H_0 se a estatística de teste é maior que o valor absoluto do valor crítico ($c_{\alpha/2}$).



Exemplo (Teste bilateral)

Um pesquisador deseja estudar o efeito da ingestão de bebida alcoólica sobre o tempo de reação a um determinado tipo de estímulo. Um experimento é desenvolvido com voluntários que são inoculados com a substância, submetidos a um estímulo elétrico e seus tempos de reação (em segundos) são registrados. Os seguintes valores são obtidos: 9,1; 9,3; 7,2; 7,5; 13,3; 10,9; 7,2; 9,9; 8,0; 8,6.

Admite-se que o tempo de reação segue, em geral, o modelo Normal com média $\mu = 8$ e desvio padrão $\sigma = 2$ segundos. O pesquisador desconfia, no entanto, que o tempo médio sofre alteração por influência da substância. Neste caso, as hipóteses de interesse são:

H_0 : os voluntários apresentam tempo de reação padrão;

H_1 : os voluntários têm tempo de reação alterado.

a) Descreva as hipóteses em termos estatísticos;

Exemplo (Teste bilateral)

Um pesquisador deseja estudar o efeito da ingestão de bebida alcoólica sobre o tempo de reação a um determinado tipo de estímulo. Um experimento é desenvolvido com voluntários que são inoculados com a substância, submetidos a um estímulo elétrico e seus tempos de reação (em segundos) são registrados. Os seguintes valores são obtidos: 9,1; 9,3; 7,2; 7,5; 13,3; 10,9; 7,2; 9,9; 8,0; 8,6.

Admite-se que o tempo de reação segue, em geral, o modelo Normal com média $\mu = 8$ e desvio padrão $\sigma = 2$ segundos. O pesquisador desconfia, no entanto, que o tempo médio sofre alteração por influência da substância. Neste caso, as hipóteses de interesse são:

H_0 : os voluntários apresentam tempo de reação padrão;

H_1 : os voluntários têm tempo de reação alterado.

- a) Descreva as hipóteses em termos estatísticos;
- b) Encontre a região crítica para $\alpha = 0,6$;

Exemplo (Teste bilateral)

Um pesquisador deseja estudar o efeito da ingestão de bebida alcoólica sobre o tempo de reação a um determinado tipo de estímulo. Um experimento é desenvolvido com voluntários que são inoculados com a substância, submetidos a um estímulo elétrico e seus tempos de reação (em segundos) são registrados. Os seguintes valores são obtidos: 9,1; 9,3; 7,2; 7,5; 13,3; 10,9; 7,2; 9,9; 8,0; 8,6.

Admite-se que o tempo de reação segue, em geral, o modelo Normal com média $\mu = 8$ e desvio padrão $\sigma = 2$ segundos. O pesquisador desconfia, no entanto, que o tempo médio sofre alteração por influência da substância. Neste caso, as hipóteses de interesse são:

H_0 : os voluntários apresentam tempo de reação padrão;

H_1 : os voluntários têm tempo de reação alterado.

- a) Descreva as hipóteses em termos estatísticos;
- b) Encontre a região crítica para $\alpha = 0,6$;
- c) Calcule a probabilidade de erro do Tipo II.