Inferência Estatística

Professor: Pedro M. Almeida-Junior

29 de novembro de 2021

Departamento de Estatística (UEPB)

Ementa

Unidade I:

- Introdução a inferência estatística
- População e amostra.
- Estatísticas e Parâmetros.
- Distribuições amostrais.
- Estimação paramétrica pontual: métodos de estimação, propriedades dos estimadores, suficiência e ancilaridade, família exponencial, estimadores não viesados de variância mínima.

Ementa

Unidade II:

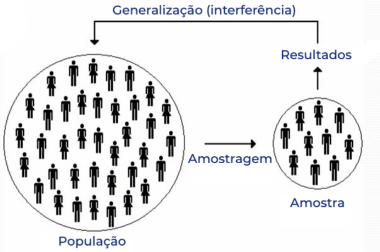
- Estimação paramétrica intervalar: métodos da quantidade pivotal.
- Intervalos de Confiança para uma e duas amostras de populações normais.
- Testes de hipóteses: conceitos iniciais, poder do teste, métodos de construção, testes mais poderosos, testes uniformemente mais poderosos.
- Testes para uma e duas amostras de populações normais.

Revisão Inferência estatística

Inferência estatística

- Um tópico central da estatística é a inferência estatística.

- Na inferência estatística, procuramos tirar conclusões sobre um grande número de eventos com base na observação de apenas uma parte deles;



- A inferência estatística aborda dois tipos de problemas fundamentais
  - Estimação de parâmetros(Pontual e Intervalar);
  - Testes de hipóteses.

Inferência estatística

- Um problema comum em inferência estatística consiste em determinar, em termos de probabilidades, se as diferenças observadas entre duas amostras significam que sejam realmente diferentes entre si.

- Os processos de inferência permite determinar se a diferença pertence em um intervalo de valores que possam ser atribuídos ao acaso ou fatores aleatórios.

- Podemos fazer isso através de testes de hipóteses (ou intervalo de confiança) que podem ser classificados em paramétricos e não paramétricos.

Revisão de alguns conceitos

População

População e Amostra são dois conceitos básicos e necessários para o desenvolvimento da inferência estatística.

Definição: O conjunto de todos os elementos ou resultados sob investigação é denominado **população**.

Parâmetros da população

Em relação ao interesse em observar elementos de uma população, temos

- Características quantitativas (expressas por variáveis numéricas)

- Características qualitativas (expressas por variáveis nominais ou categóricas)

Aos valores numéricos caracterizando globalmente uma população dá-se o nome de parâmetro da população ou parâmetros populacionais.

Amostra

Definição: Uma parte ou sub-conjunto de uma população é denominada de amostra.

Definição: Chamamos de tamanho amostral o número de elementos da amostra.

O objetivo da inferência estatística é produzir afirmações sobre dada característica da população na qual estamos interessados, a partir de informações coletadas de uma parte dessa população (amostra). Essa característica na população pode ser representada por uma variável aleatória.

Amostragem

Tipos de amostragem:

- **Amostragem probabilística:** Existe uma probabilidade conhecida de cada elemento da população vir a participar da amostra.
- **Amostragem não probabilística:** Não existe nenhum mecanismo probabilístico na relação da amostra.

Amostragem probabilísticas

1. Amostragem Aleatória Simples (AAS):
    - População homogênea
    - Rotular os elementos da população e sortear os indivíduos que farão parte da amostra

2. Amostragem Sistemática (AS):
    - População homogênea
    - os elementos da população são ordenados e a retirada dos elementos é feita periodicamente

Amostragem probabilísticas

3. Amostragem Estratificada (AE):
    - População heterogênea. No entanto, a sub-divisão da mesma é feita em grupos homogêneos.
    - A seleção dos elementos de cada estrato é feita de forma aleatória

4. Amostragem por conglomerado (AC):
    - População é dividida em sub-populações heterogêneas.
    - A amostragem é realizada em cima do conglomerado e não mais sobre o indivíduo da população.

Distribuição amostral

Interesse: Uma medida que descreva certa característica da população, parâmetro, geralmente desconhecido.

Solução: A partir da amostra podemos obter uma medida que descreva tal característica, denominada estatística.

Observação: Como os valores da amostra são aleatórios, qualquer quantidade calculada em função dos elementos da amostra também será uma variável aleatória. Assim, as estatísticas sendo v.a. terão alguma distribuição de probabilidade.

Conceito de Estatística

Definição: Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística.

Exemplo: Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X, com f.d.p. ou f.p. $f(x \mid \theta)$. Exemplos de estatísticas são

(i) $X_{(1)} = \min(X_1, \ldots, X_n)$

(ii) $X_{(n)} = \max(X_1, \ldots, X_n)$

(iii) $\tilde{X} = \text{med}(X_1, \ldots, X_n)$

(iv) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(v) $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

Formalização do problema

Seja $X_1, \ldots, X_n$ uma AAS de uma população de tamanho n. Para realizarmos uma afirmação sobre algum parâmetro (θ) da população, tais como, média, variância, etc. Utilizaremos uma estatística (T) que seja uma função da amostra de tamanho n.

Definição: Uma sequência $X_1, \ldots, X_n$ de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função de densidade (f.d.p.) ou, no caso discreto, função de probabilidade (f.p.) $f(x \mid \theta)$ é dita ser uma amostra aleatória de tamanho n da distribuição de X. Nesse caso, temos,

$$f\left(x_1, \ldots, x_n \mid \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right) = f\left(x_1 \mid \theta\right) \ldots f\left(x_n \mid \theta\right) \qquad (1)$$

Função de verossimilhança

Concluímos, a partir da Definição anterior, que usamos a amostra $X_1, \ldots, X_n$ para obter informação sobre o parâmetro θ.

A função de densidade (ou de probabilidade) conjunta dada na Equação (1) é denominada <u>função de verossimilhança</u> de θ, correspondente à amostra observada $\boldsymbol{x} = \left(x_1, \ldots, x_n\right)^\top$ e será denotada por

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right)$$

Definição: O conjunto Θ em que θ toma valores é denominado espaço paramétrico.

Exemplo: Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(\mu, \sigma^2)$.

(i) Se $\sigma^2 = 1$, então $\theta = \mu$ é o parâmetro desconhecido e

$$\Theta = \{\mu, -\infty < \mu < \infty\}$$

(ii) Se $\mu = 0$, então $\theta = \sigma^2$ é o parâmetro desconhecido e

$$\Theta = \left\{\sigma^2, \quad \sigma^2 > 0\right\}$$

(iii) Se μ e σ^2 são desconhecidos então $\theta = \left(\mu, \sigma^2\right)$ e

$$\Theta = \left\{\left(\mu, \sigma^2\right), -\infty < \mu < \infty \quad \text{e} \quad \sigma^2 > 0\right\}$$

Revisão Teorema Central do Limite

Teorema Central do Limite para variáveis aleatórias i.i.d.: Sejam $\{X_n; n \geq 1\}$ variáveis aleatórias independentes, identicamente distribuídas e com esperança μ e variância σ^2, com $0 < \sigma^2 < \infty$. Então, para $S_n = X_1 + \cdots + X_n$, temos,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Teorema Central do Limite de De Moivre-Laplace: Sejam $\{X_n; n \geq 1\}$ variáveis independentes seguindo o modelo Bernoulli com parâmetro p. Assim, $\mu = \mathbb{E}(X_n) = p$ e $Var(X_n) = p(1-p)$. Para $S_n = X_1 + X_2 + \cdots X_n$, temos , pelo teorema anterior,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

OBS: Uma consequência direta do Teorema Central do Limite de De Moivre-Laplace é que podemos calcular, de forma aproximada, probabilidades binomiais com o uso da distribuição Normal.

Teorema Central do Limite de Liapunov: Seja $\{X_n; n \geq 1\}$ uma sequência de variáveis aleatórias independentes, com valor esperado μ_n e variância σ_n^2, com $\sigma_n^2 < \infty$ e pelo menos um dos σ_n^2's maior que zero. Sejam $S_n = X_1 + \cdots + X_n$ e $s_n^2 = \sum_{k=1}^{n} \sigma_k^2$. Se a condição de Liapunov estiver satisfeita, isto é, se existir $\delta > 0$, tal que

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left(\mid X_k - \mu_k \mid^{2+\delta} \right) = 0,$$

então,

$$\frac{S_n - \mathbb{E}(S_n)}{s_n} \xrightarrow{d} N(0,1)$$

Distribuições Amostrais

Distribuição amostral da média

- Consideraremos uma população identificada pela variável aleatória X, cujos parâmetros, média populacional (μ) e variância populacional (σ^2), são supostamente conhecidos.

Proposição: Seja X uma v.a. com média μ e variância σ^2 e seja $X_1, \ldots, X_n$ uma AAS de X. Então,

$$\mathbb{E}(\bar{X}) = \mu \text{ e } V(\bar{X}) = \sigma^2/n$$

Teorema central do Limite

- Sejam $X_1, \ldots, X_n$ uma AAS da v.a. X com distribuição comum com $\mathbb{E}(X_i) = \mu$ e $V(X_i) = \sigma^2$, sendo os X_i's independentes e se n é grande, temos que
$$\bar{X} \overset{aprox.}{\sim} N(\mu, \sigma^2/n)$$
ou seja,
$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}, \quad Z \sim N(0, 1)$$

Obs: No caso em que a distribuição de X seja normal mesmo para valores pequenos de n.

Distribuição amostral da proporção

- Seja X uma v.a. com distribuição Bernoulli com parâmetro p, isto é, $P(X = 1) = p$ e $P(X = 0) = 1 - P$. Temos que, $\mathbb{E}(X) = p$ e $V(X) = p(1 - p)$.

- Considere uma AAS de tamanho n dessa população. Seja $S_n = \sum_{i=1}^{n} X_i$, o número de indivíduos com característica de interesse na amostra, temos que $S_n \sim \text{Binomial}(n, p)$.

- Pelo TCL temos que $\bar{X}$ tem distribuição normal, para n suficientemente grande. Seja $\hat{p} = \bar{X}$ a proporção amostral, temos que $\hat{p} \sim N(p, \frac{p(1-p)}{n})$ ou $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$.

Distribuição amostral da diferença entre médias

- Em vários problemas práticos, deseja-se comparar duas populações de interesse.
- Por exemplo, podemos estar interessados em avaliar a diferença de performance entre duas linhas de produção.
- Suponha que duas populações de interesse, X_1 e X_2, com médias μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2, respectivamente.
- Considere AAS independentes de tamanhos n_1 e n_2, respectivamente, das duas populações.

Distribuição amostral da diferença entre médias

- Pelo TCL, a distribuição amostral da diferença $(\bar{X}_1 - \bar{X}_2)$, para n_1 e n_2 suficientemente grandes, será dada por:

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$$

Distribuição amostral da diferença de proporções

- Suponhamos que duas as duas populações de interesse apresentam distribuição binomial com proporções p_1 e p_2. Considere as AAS independentes de tamanhos n_1 e n_2 são retiradas das populações

- A distribuição amostral da diferença entre proporções $(\hat{p}_1 - \hat{p}_2)$, para n_1 e n_2 suficientemente grande. Pelo TCL é dada por,

$$(\hat{p}_1 - \hat{p}_2) \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

Função Característica

- A função geradora de momentos, tem a desvantagem de que a integral que a define pode, nem sempre ser finita e, portanto, nem sempre existirá.

- Definiremos uma nova transformada que tem a vantagem de sempre existir.

- Entretanto, teremos o inconveniente de trabalhar com uma função de valores complexos

- Diremos que uma variável aleatória X é complexa, se pode ser escrita como $X = X_1 + iX_2$ em que $i = \sqrt{-1}$, e X_1 e X_2 são variáveis aleatórias reais.

- Para verificar que uma função complexa é variável aleatória, precisamos verificar propriedades da imagem inversa nas suas duas partes. Para o valor esperado de X, exige-se que as duas partes sejam finitas. Assim, temos

$$\mathbb{E}(X) = \mathbb{E}(X_1) + i\mathbb{E}(X_2)$$

em que $\mathbb{E}(X_1)$ e $\mathbb{E}(X_2)$ são ambas finitas.

- Para efeitos práticos, quando realizamos integração de funções complexas, podemos operar como se estivéssemos com funções reais (basta tratar o i como sendo uma constante real)

A função característica de uma variável aleatória X é a função,

$$\varphi_X(t) = \mathbb{E}\left(e^{i\,t\,x}\right) = \mathbb{E}\left(\cos(tX)\right) + i\mathbb{E}\left(\sin(tX)\right), \quad t \in \mathbb{R}$$

quando esta esperança existe, em que t é o argumento da função característica e $i = \sqrt{-1}$.

Toda v.a. contínua ou discreta possui função característica que é dada, respectivamente, por

$$\mathbb{E}\left(e^{i\,t\,X}\right) = \int_{-\infty}^{\infty} e^{i\,t\,X} f(x) dx$$

e

$$\mathbb{E}\left(e^{i\,t\,X}\right) = \sum_{x} e^{i\,t\,X} p_X(x) dx$$

Propriedades da função característica

P1. A função característica é limitada por 1:

$$| \varphi_X(t) | < 1, \quad t \in \mathbb{R}$$

P2. A função característica assume o valor 1 no ponto 0:

$$\varphi_X(0) = 1$$

P3. $\overline{\varphi_X(t)} = \varphi_X(-t)$, em que $\bar{c}$ é o complexo conjugado de c. (Se $c = x + i\,y$, o seu complexo conjugado é $\bar{c} = x - i\,y$)

P4. Se X e Y são independentes, então

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

P5. A função característica de uma v.a. X determina a função de distribuição, ou seja,

$$F_X = F_Y \Leftrightarrow \varphi_X(t) = \varphi_Y(t)$$

Distribuições amostrais

Se a população for normal, então $\bar{X}$ terá distribuição exata normal.

Teorema: Sejam $X_1, \ldots, X_n$ uma AAS da v.a. $X \sim N(\mu, \sigma^2)$, então a distribuição amostral de $\bar{X} \sim N(\mu, \sigma^2/n)$.

Revisando algumas distribuições de probabilidade

Distribuição normal

- Dizemos que X é uma variável aleatória normal, com parâmetros μ e σ^2, se a função densidade de X é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \text{ para } -\infty < x < \infty$$

com,

$$\mathbb{E}(X) = \mu \quad e \quad Var(X) = \sigma^2$$

Uma variável aleatória contínua cuja função densidade é dada, para algum $\lambda > 0$, por

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

é chamada de variável aleatória exponencial com parâmetro λ, diga-se $X \sim \exp(\lambda)$.

Com,

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad e \quad Var(X) = \frac{1}{\lambda^2}$$

Distribuição acumulada

A função distribuição acumulada $F(x)$ de uma variável aleatória exponencial é dada por

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= \int_0^x \lambda e^{-\lambda t} dt \\ &= -\left. e^{-\lambda t} \right|_0^x \\ &= 1 - e^{-\lambda x}, \quad x \geq 0 \end{aligned}$$

Distribuição Gama

Uma variável aleatória tem distribuição gama com parâmetros (α, λ), $\lambda > 0, \alpha > 0$, se sua função densidade é dada por

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

em que $\Gamma(\alpha)$, chamada de função gama, é definida como

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$$

Se resolvermos $\Gamma(\alpha)$, vamos obter

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

Para valores inteiros de α, digamos $\alpha = n$, se aplicamos o mesmo procedimento repetidamente na equação anterior, então,

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1)\\ &= (n-1)(n-2)\Gamma(n-2)\\ &= \cdots\\ &= (n-1)(n-2)\cdots\Gamma(1)\end{aligned}$$

Como $\Gamma(1) = \int_0^\infty e^{-x}dx = 1$, tem-se que, para valores inteiros de n

$$\Gamma(n) = (n-1)!$$

Valor esperado e variância

$$\mathbb{E}(X) = \frac{\alpha}{\lambda} \quad e \quad Var(X) = \frac{\alpha}{\lambda^2}$$

Casos Particulares

A distribuição Gama tem como casos particulares outras distribuições, como :

- Distribuição Exponencial
- Distribuição Erlang
- Distribuição Qui-Quadrado ($\chi^2_{(n)}$)

Quando, na distribuição gama, o parâmetro de forma α é igual a $\frac{n}{2}$, com n inteiro positivo, e o parâmetro λ é $\frac{1}{2}$ a distribuição é chamada de distribuição Qui-quadrado com n graus de liberdade e a sua função densidade, para $x > 0$ é definida por:

$$f_X(x) = \frac{1}{\Gamma(n/2)} x^{(n/2)-1} \left(\frac{1}{2}\right)^{n/2} e^{-(1/2)x}$$

$$= \frac{1}{\Gamma(n/2)2^{n/2}} x^{(n/2)-1} e^{-x/2}$$

Distribuição Qui-Quadrado - χ^2

Se X tem distribuição Qui-quadrado com n graus de liberdade (notação: $X \sim \chi^2_{(n)}\Big)$, a sua função densidade é

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(n/2)2^{n/2}} x^{(n/2)-1} e^{-x/2} & \text{, se } x > 0 \\ 0 & \text{, caso contrário.} \end{cases}$$

Além disso,

$$\mathrm{E}(X) = \frac{n/2}{1/2} = n \quad \text{e} \quad \mathrm{Var}(X) = \frac{n/2}{(1/2)^2} = 2n$$

Relação entre $\chi^2_{(1)}$ e $N(0,1)$

Suponha que $Z \sim N(0,1)$, então, $X = Z^2 \sim \chi^2_{(1)}$. Sejam $Z_1, \ldots, Z_n$ v.a. i.i.d., ou seja , $Z_i \sim N(0,1)$, $\forall i$. Então,

$$Y = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$$

Distribuição T-Student

Considere que as v.a. $Z \sim N(0,1)$ e $V \sim \chi^2_{(n)}$ sejam independentes. Assim, a variável aleatória

$$X = \frac{Z}{\sqrt{V/n}} \sim t_{(n)}$$

ou seja, segue uma distribuição t de Student com n graus de liberdade.

Definição: Uma variável aleatória X contínua tem distribuição t de Student com $n \in \mathbb{Z}$, graus de liberdade, $X \sim t_{(n)}$, se sua f.d.p. for

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi\Gamma\left(\frac{n}{2}\right)}}\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}, \quad n \in \mathbb{Z}_+ \quad e \quad -\infty < x < \infty$$

Se $X \sim t_{(n)}$, então, $\mathbb{E}(X) = 0$ e $Var(X) = \frac{n}{n-2}$, se $n > 2$

Exercício: Seja $X \sim t_{(5)}$, calcule o valor de c, dado que $P(X > c) = 0,05$

Distribuição F de Snedecor

Considere que as variáveis aleatórias $U \sim \chi^2_{(m)}$ e $V \sim \chi^2_{(n)}$ sejam independentes. Assim, a v.a.

$$X = \frac{U/m}{V/n} \sim F(m, n)$$

ou seja, segue uma distribuição F de Snedecor com m e n graus de liberdade.

Definição: Uma v.a. contínua tem distribuição F de Snedecor com $m \in \mathbb{Z}_+$ e $n \in \mathbb{Z}_+$ graus de liberdade, $X \sim F(m, n)$ se sua f.d.p. for da forma

$$f(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \frac{x^{\frac{m-2}{2}}}{(1 + \frac{m}{n}x)},$$

Se $X \sim F(m, n)$, então, $\mathbb{E}(X) = \frac{n}{n-2}$, se n$>2$ e $V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

Exercício: Seja $X \sim F(2,4)$. Calcule $P(X \leq 18)$.

Teorema: Seja $X_1, \ldots, X_n$ AAS da v.a. populacional $X \sim N(\mu, \sigma^2)$. Então, a distribuição amostral

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Distribuição amostral das médias quando a variância é desconhecida

Como não se conhece o valor de σ^2, uma possibilidade é substituí-lo pela v.a. S^2 (variância amostral).

Teorema: Seja $X_1, \ldots, X_n$ uma AAS da v.a. populacional $X \sim N(\mu, \sigma^2)$, a distribuição de

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{(n-1)}$$

Distribuição amostral da relação entre duas variâncias

Dadas duas amostras aleatórias independentes $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_m)$, com variâncias amostrais dadas, respectivamente por S_1^2 e S_2^2 e variâncias populacionais dadas respectivamente por σ_1^2 e σ_2^2. Então,

$$\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(n-1, m-1)$$