Inferência I
Inferências com amostras grandes

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Sumário

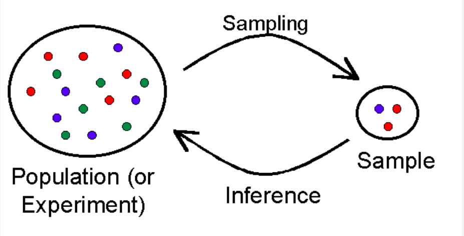
Princípios de Inferência

Definition

Inferência Estatística é o conjunto de técnicas que permite fazer afirmações sobre as características de uma população baseado em em dados obtidos de uma amostra.

Princípios de Inferência

Amostra $\rightarrow$ inferência $\rightarrow$ População



Relembrando

Definition

Um parâmetro é uma variável numérica que representa uma característica da população.

Definition

Uma estatística é uma variável numérica que representa uma característica da amostra.

Relembrando

Inferência I

Felipe Figueiredo

População

$$\mu = \frac{\sum x_i}{N}$$

$$\sigma^2 = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Amostra

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Estimadores

Inferência I

Felipe Figueiredo

- Um estimador pontual é uma estatística que será usada para inferir o valor do parâmetro
- Geralmente usamos um ˆ para designar o estimador. Assim $\hat{\theta}$ é o estimador de θ
- É uma função (qualquer) dos dados: $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$

Estimadores

Inferência I

Felipe Figueiredo

Características de um bom estimador são:

- Não-tendencioso (não-enviesado, não-viciado)
- Consistência
- Eficiência

Estimadores

Definition

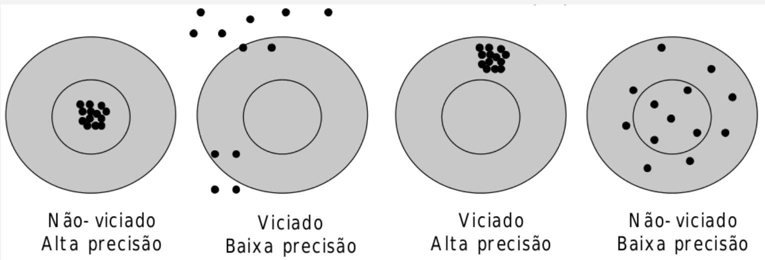
Um estimador é não viesado (não tendencioso, não viciado) quando sua média (ou esperança) é o próprio valor do parâmetro.

Estimadores

Definition

Dados dois estimadores, o mais eficiente é o que tem a menor variância.

Estimadores



Estimadores pontuais para a média

O estimador $\hat{\mu}$ menos tendencioso para a média populacional μ é a média amostral $\bar{x}$.

$$\hat{\mu} = \bar{x}$$

Example

Uma estimativa pontual para a quantidade diária de cigarros por dia em uma população de fumantes pode ser obtida de uma amostra com 30 fumantes.

$$\hat{\mu} = \bar{x} = \frac{\sum x_i}{30} = 12.4$$

Margem de erro para a estimativa da Média

Inferência I

Felipe Figueiredo

- Precisamos considerar o erro do estimador $\epsilon(\mu) = \mu - \hat{\mu}$
- Mas se tivéssemos μ, não precisaríamos de $\hat{\mu}$!
- Assim, precisamos de um outro tipo de estimador, que leve em conta uma margem de erro em torno da estimativa pontual

Estimadores Intervalares para a Média

Inferência I

Felipe Figueiredo

Definition

Chamamos de nível de confiança *c* a probabilidade de que o parâmetro esteja dentro do intervalo

Definition

Um estimador intervalar é um intervalo torno do estimador pontual, considerando uma margem de erro *E* e o nível de confiança *c* da estimativa.

Intervalos de Confiança para a Média

Inferência I

Felipe Figueiredo

- Níveis de confiança usuais: 90%, 95% e 99%.
- Associados a esses níveis de confiança temos os respectivos valores críticos z_c da distribuição normal padrão
- Valores tabelados: $z_c(0.90) = 1.645$, $z_c(0.95) = 1.96$ e $z_c(0.99) = 2.575$.

Intervalos de Confiança para a Média

Se a amostra é grande ($n \geq 30$) temos boas condições analíticas! Pelo Teorema Central do Limite (TCL):

- podemos aproximar uma distribuição normal (contínua) pela binomial (discreta)
- podemos aproximar o desvio-padrão populacional σ por pelo desvio-padrão amostral s
- Calculamos assim a margem de erro E como

$$E = \frac{z_c \cdot s}{\sqrt{n}}$$

- O Intervalo de Confiança fica então

$$\bar{x} \pm E = (\bar{x} - E, \bar{x} + E)$$

Interpretação

Dizemos que o intervalo tem, por exemplo, 95% de chance de conter o verdadeiro valor da média populacional.

- Obs: A média é um valor fixo, está contido ou não.

Exercício

Exercício

Num estudo para descrever o perfil dos pacientes adultos atendidos no ambulatório de um posto de saúde, uma amostra de 70 pacientes adultos foi selecionada ao acaso entre o total de pacientes atendidos no posto durante os últimos três anos, coletando-se dos prontuários desses pacientes dados relativos à idade, à escolaridade e a outros fatores de interesse.

Para a variável idade, observou-se uma média amostral de 36.86 anos com um desvio padrão amostral de 17.79 anos.

Exercício

Exercício

1. Defina a população e a amostra.
2. Forneça uma estimativa pontual, um intervalo de 90% de confiança e um intervalo de 95% de confiança para a idade média dos adultos atendidos neste ambulatório nos últimos três anos. Interprete e compare os intervalos de confiança.

$$E = \frac{z_c s}{\sqrt{n}} \qquad \bar{x} = 36.86$$

$$z_c(95\%) = 1.96 \qquad s = 17.79$$

$$z_c(90\%) = 1.645 \qquad n = 70$$

Exercício

Solução

- IC de 90% (c=0.90)

$$E = \frac{z_c s}{\sqrt{n}} = \frac{1.645 \times 17.79}{\sqrt{70}} \approx 3.50$$

$$IC_{0.90} = \bar{x} \pm E = 36.86 \pm 3.50 = (33.36, 40.36)$$

- IC de 95% (c=0.95)

$$E = \frac{z_c s}{\sqrt{n}} = \frac{1.96 \times 17.79}{\sqrt{70}} \approx 4.17$$

$$IC_{0.95} = \bar{x} \pm E = 36.86 \pm 4.17 = (32.69, 41.03)$$

Exercício

Inferência I

Felipe Figueiredo

Comparando os ICs

$$IC_{0.90} = (33.36, 40.36)$$

$$IC_{0.95} = (32.69, 41.03)$$

Pergunta: Qual estimativa intervalar tem maior precisão?
Ou: Para qual nível de confiança o IC é menor?

Estimadores pontuais para proporções

Inferência I

Felipe Figueiredo

- Para variáveis categóricas, é conveniente considerar a proporção da amostra que satisfaz o critério desejado
- Se x é o número de sucessos na amostra, o estimador pontual da proporção populacional é:

$$\hat{p} = \frac{x}{n}$$

Estimadores pontuais para proporções

Inferência I

Felipe Figueiredo

Example

- População: fumantes no prédio
- Parâmetro: $p =$ proporção de fumantes no prédio
- Estimativa: $\hat{p} =$ proporção de fumantes na sala

Intervalos de confiança para proporções

- Podemos construir um intervalo de confiança de maneira análoga à usada para médias
- A margem de erro considera a proporção de sucessos $\hat{p}$ e a proporção de fracassos $\hat{q} = 1 - \hat{p}$

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- Essa aproximação é válida sempre que $n\hat{p} \geq 5$ e $n\hat{q} \geq 5$ (amostras grandes)
- O IC fica então $\hat{p} \pm E = (\hat{p} - E, \hat{p} + E)$

Exercício

Exercício

Num estudo para descrever o perfil dos pacientes adultos atendidos no ambulatório de um posto de saúde, uma amostra de 70 pacientes adultos foi selecionada ao acaso entre o total de pacientes atendidos no posto durante os últimos três anos, coletando-se dos prontuários desses pacientes dados relativos à idade, à escolaridade e a outros fatores de interesse.

Para a variável escolaridade, observou-se que 19 pacientes da amostra eram analfabetos.

Exercício

Exercício

1. Forneça uma estimativa pontual, um intervalo de 90% de confiança e um intervalo de 95% de confiança para proporção de analfabetos dentre os adultos atendidos neste ambulatório nos últimos três anos. Interprete e compare os intervalos de confiança.

Exercício

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad \hat{p} = \frac{19}{70} \approx 0.27$$

$$z_c(95\%) = 1.96 \qquad \hat{q} = 1 - 0.27 = 0.73$$

$$z_c(90\%) = 1.645 \qquad n = 70$$

Exercício

Solução

- IC de 90% (c=0.90)

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \sqrt{\frac{0.27 \times 0.73}{70}} \approx 0.09$$

$$IC_{0.90} = 0.27 \pm 0.09 = (0.18, 0.36)$$

- IC de 95% (c=0.95)

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96 \sqrt{\frac{0.27 \times 0.73}{70}} \approx 0.10$$

$$IC_{0.95} = 0.27 \pm 0.10 = (0.17, 0.37)$$

Tamanho da amostra (médias)

- Podemos aumentar a precisão do IC sem diminuir o nível de confiança
- Para isto, basta aumentar o tamanho da amostra
- Revirando a fórmula da margem de erro E, temos:

Tamanho da amostra (médias)

$$E = \frac{z_c \cdot s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z_c \cdot s}{E}$$

$$n = \left(\frac{z_c \cdot s}{E}\right)^2$$

Exercício

Exercício

Encontre o tamanho mínimo da amostra que dará uma margem de erro $E = 2$ ao nível de confiança $c = 0.95$ com desvio-padrão amostral $s = 6.1$

$$n \geq \left(\frac{z_c \cdot s}{E}\right)^2$$

Solução

$$n \geq \left(\frac{1.96 \times 6.1}{2}\right)^2 \approx 35.7$$

Portanto, n precisa ser no mínimo 36.

Recapitulando

- Quanto maior o nível de confiança (exigência), maior a amplitude do IC (menos precisão)
- Quanto maior o desvio-padrão (variabilidade) da amostra, maior a amplitude do IC (menos precisão)
- Quanto maior o tamanho da amostra (dados), menor a amplitude do IC (mais precisão)
- Dado um nível de confiança e uma margem de erro, podemos estimar o tamanho mínimo da amostra que gera este IC.