

1. Introdução à Inferência Estatística

Muito embora em algumas situações a amostra seja ela própria o objecto do estudo e da análise estatística, o que é mais comum é pretendermos obter informações acerca de uma população inteira, tal que, por motivos do seu tamanho ou outros, não é passível de estudo completo.

Até agora, o estudo debruçou-se:

- sobre a amostra – viu-se de que forma podemos organizar um conjunto de dados e tirar algumas conclusões preliminares (Estatística Descritiva)

- sobre a população – como caracterizar uma população (Teoria das Probabilidades, Distribuições de Variáveis discretas e de Variáveis contínuas, Algumas Distribuições Discretas / Contínuas Univariadas).

Ao longo deste capítulo e dos seguintes, iremos ver como é possível (ou não) transpor as conclusões tiradas anteriormente acerca de uma amostra para uma população inteira da qual ela foi retirada e também quantificar a confiança que podemos esperar dessas mesmas conclusões.

1.1 Parâmetros populacionais e variáveis amostrais

Um **parâmetro populacional** diz respeito a uma população concreta em estudo e portanto possui um valor único, logo, constante, ainda que desconhecido ¹.

Uma **variável amostral** possui um valor variável (varia consoante a amostra ou conjunto de dados seleccionados a partir da população) ainda que se conheça perfeitamente um dos seus possíveis valores, como é normalmente o caso.

Exemplo - A média das alturas de todos os portugueses residentes em Portugal Continental no dia 1 de Janeiro de 2000 possui um único valor, mas nós não o conhecemos. A média populacional é, assim um parâmetro populacional constante (único).

Por outro lado, é perfeitamente possível conhecer a média das alturas de 1000 portugueses seleccionados ao acaso na população anterior. No entanto, sabemos que essa média será quase sempre diferente da obtida num diferente conjunto de outras 1000 pessoas.

A média amostral é portanto uma variável – uma variável amostral.

Por terem definições e naturezas distintas, há necessidade de distinguir claramente entre parâmetros populacionais e variáveis amostrais em termos de nomenclatura. Na tabela 1, encontra-se resumida uma nomenclatura internacionalmente aceite (há outras) e que será utilizada nestas aulas, para alguns parâmetros populacionais mais importantes.

Acresce-se que :

A dimensão populacional será denotada por **M**.

A dimensão da amostra será denotada por **n**.

Nota : Apesar de respeitante à amostra, a dimensão amostral não é considerada uma variável, já que estando definida à partida não varia de possível amostra para possível amostra.

¹ Lembra-se que uma população só está correctamente definida quando indubitavelmente especificada a sua situação espaço-temporal.

Tabela 1 – Nomenclatura para alguns parâmetros populacionais e variáveis amostrais.

Nome comum	Parâmetro populacional		Variável amostral	
	nome	símbolo	nome	símbolo
<u>média</u>	média populacional, média, valor médio ou valor esperado	μ	média amostral	\bar{x}
<u>desvio padrão</u>	desvio padrão ou desvio padrão populacional	σ	desvio padrão amostral	s
<u>variância</u>	variância ou variância populacional	σ^2	variância amostral	s²
<u>proporção</u>	proporção ou proporção populacional	π	proporção amostral	f
<u>declive</u>	declive populacional	β	declive amostral	b
<u>ordenada na origem</u>	ord. origem populacional	α	ord. origem amostral	a
<u>coeficiente de correlação</u>	coef. corr. populacional	ρ	coef. corr. amostral	r

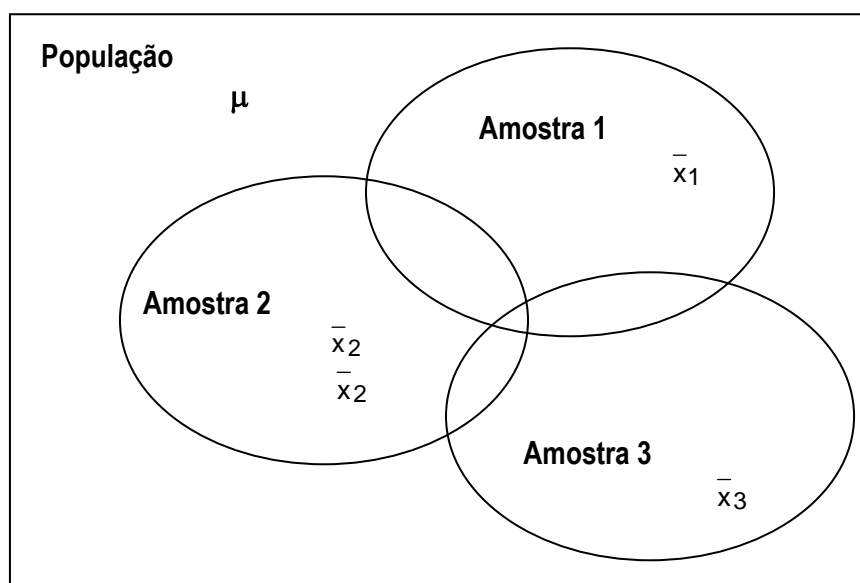


Figura 1 – População e amostra.

Ilustração da diferença entre os conceitos de média populacional e de média amostral. Enquanto que a média populacional tem um único valor, a média amostral varia de amostra para amostra.

2. Breves noções de Estimação Pontual

2.1 Conceitos básicos

Estimação – Processo que tem por objectivo quantificar um parâmetro populacional do qual não é possível conhecer com rigor o seu verdadeiro valor.

Estimação Pontual – Processo de estimação que encontra um valor único para o parâmetro populacional a estimar ².

Estimador – Variável amostral que permite estimar um parâmetro populacional; isto é, calcular um valor, não necessariamente o verdadeiro.

Estimativa – Estimador concretizado num valor.

Conhecem-se, através de vários métodos, os melhores estimadores pontuais de vários parâmetros populacionais ³. Alguns, a saber, estão indicados na tabela 3.2.

Tabela 2 – Melhores estimadores de alguns parâmetros populacionais.

<i>Parâmetro populacional</i>		<i>Melhor estimador</i>	
<i>nome</i>	<i>símbolo</i>	<i>nome</i>	<i>símbolo</i>
média populacional	μ	média amostral	\bar{x}
desvio padrão populacional	σ	desvio padrão amostral	s
variância populacional	σ^2	variância amostral	s²
proporção populacional	π	proporção amostral	f
declive populacional	β	declive amostral	b
ord.origem populacional	α	ord.origem amostral	a
coef. correlação populacional	ρ	coef. correlação amostral	r

Notação de estimadores – O estimador de um parâmetro pode ser representado, quer pela notação própria que já possua, quer pela notação do parâmetro populacional acrescida de um assento circunflexo (^).

Assim : $\hat{\mu} = \bar{x}$; $\hat{\sigma} = s$; etc...

² Surge por oposição à **Estimação Intervalar**, que situa o valor do parâmetro populacional num intervalo de valores.

³ Um bom estimador deve possuir várias propriedades (*centricidade, consistência eficiência,...*). Existem vários métodos que nos permitem encontrar bons estimadores para os mais variados parâmetros populacionais (*método dos momentos, método da máxima verosimilhança, método dos mínimos quadrados,...*). O estudo das propriedades dos estimadores bem como dos métodos referidos enocntram-se fora do âmbito do desta disciplina.

Note por exemplo que $\hat{\mu} \neq \mu$, pois, enquanto que μ é um parâmetro populacional, constante, $\hat{\mu}$ é na verdade uma variável que em geral adquirirá valores diferentes quando baseado em amostras diferentes.

A maior ou menor confiança que podemos depositar numa estimativa depende de :

- qualidade do estimador – os estimadores apresentados são, como já foi dito, os melhores estimadores para os parâmetros referidos
- tamanho da amostra – quanto maior a amostra, maior a confiança na estimativa
- método de amostragem utilizado – como já vimos, alguns métodos de amostragem são mais fiáveis do que outros
- incerteza na medição dos valores amostrais – tem a ver com exactidão e precisão do equipamento empregue, dos métodos de análise, da experiência dos operadores, etc.

Exemplo - Dada amostra de valores de certa variável quantitativa apresentou uma média de 4,3 e um desvio padrão de 2,1. Pretende-se estimar a média populacional e o desvio padrão populacional.

Parâmetros a estimar : μ e σ
 Estimadores seleccionados : $\hat{\mu} = \bar{x}$ e $\hat{\sigma} = s$
 Estimativas obtidas : $\hat{\mu} = 4,3$ e $\hat{\sigma} = 2,1$

Exemplo - Pretendendo-se determinar a constante de proporcionalidade k entre a absorvância e a concentração de uma determinada solução (NOTE : Lei de Beer : $A = k \cdot C$), recolheu-se um conjunto de dados bivariados (C, A) e determinou-se o declive de uma recta de ajuste, tendo-se encontrado o valor 2567. Qual será o verdadeiro valor da constante de proporcionalidade ?

Não podemos determinar mediante uma amostra de valores o verdadeiro valor da constante de proporcionalidade.

*No entanto, podemos ter uma estimativa : $\hat{k} = 2567$
 A confiança a depositar nesta estimativa depende do número de dados amostrais válidos e da incerteza que lhes possamos atribuir.*

2.2 Estimação pontual vs estimação intervalar

A Estimação Pontual apresenta em relação à Estimação Intervalar (que veremos no capítulo seguinte) a desvantagem de ser complexo quantificar a incerteza presente nas estimativas obtidas e assim conhecer o grau de confiança que podemos ter nos resultados obtidos.

Tem porém a vantagem de, uma vez conhecidos os estimadores, ser de extremamente simples e expedita aplicação, facilitando também a utilização dos resultados obtidos em cálculos subsequentes.

- FIM -