

Ficha — Amostragem e estimação: palpite informado sobre a população

11.º Ano MACS 4 páginas 12 exercícios

Objetivo: Compreender por que é necessária uma **amostra aleatória** para inferir sobre a população, distinguir **parâmetro e estimador**, reconhecer a **variabilidade amostral** (estimativas pontuais limitadas) e os processos de seleção **com e sem reposição**.

Revisão rápida

População	Amostra
Parâmetro: valor fixo desconhecido (μ , p , σ)	Estimador/estatística: valor calculado na amostra ($\bar{x}$, $\hat{p}$, s)
é o que se quer saber	é o que se tem para estimar

Porque aleatória: só com seleção aleatória se pode usar a **probabilidade para quantificar o erro** cometido ao inferir para a população — é o **primeiro passo** de uma inferência eficiente. Seleção por **conveniência** (quem está à mão) → viés sem medida (não há «erro com probabilidade»).

Variabilidade amostral: o valor do estimador **depende da amostra** → há **tantas estimativas diferentes quantas as amostras** da mesma dimensão — isso não é erro de cálculo, é a **própria aleatoriedade**. Consequência: a **estimativa pontual** (um número) é sempre **limitada** — nunca se sabe se a amostra foi «a boa» → daí a necessidade de **intervalos** (ficha 12).

Com vs sem reposição: *com* = cada escolha não altera as probabilidades seguintes (independentes) · *sem* = a população da urna encurta a cada escolha (dependentes — ficha 7). A dimensão da amostra é sempre **n** escolhas.

Exemplo do documento (população: 3 mulheres + 2 homens, amostras de **dimensão 2**): 10 amostras possíveis → distribuição da **proporção de homens** $\hat{p}$: 0 (3/10) · 0,5 (6/10) · 1 (1/10) → média da distribuição = **0,4** = **p** da população.

Estratégia: definir população e parâmetro → desenhar o **processo** de seleção (aleatório? com/sem reposição?) → enumerar/simular amostras → ver a **distribuição das estimativas** → concluir o que é fiável (e o que falta: o intervalo).

Exercícios

1. Completar: o parâmetro pertence à _____ e o estimador à _____ · $\bar{x}$ estima _____ · $\hat{p}$ estima _____ · sem uma seleção _____, não é possível quantificar o _____ com probabilidade.
2. Porque aleatória: um inquérito à entrada de um cinema pretende estimar a percentagem da cidade que vê filmes. (a) A amostra «só quem entra no cinema» é aleatória? Que viés introduz? (b) Desenha um processo aleatório melhor (sorteio telefónico? quota?).
3. O exemplo do documento: população {M, M, M, H, H} (3 mulheres, 2 homens), amostras de dimensão 2 sem reposição. (a) Enumera as 10 amostras. (b) Constrói a distribuição de $\hat{p}$ = % de homens (3/10, 6/10, 1/10). (c) Calcula a média da distribuição e compara com $p = 2/5 = 0,4$.

4. Com vs sem reposição: repete o ex. 3 com reposição (25 amostras ordenadas). (a) Obtém $P(\hat{p} = 0)$, $P(\hat{p} = 0,5)$, $P(\hat{p} = 1)$ (dica: 9/25, 12/25, 4/25). (b) A média continuou 0,4? (c) O que mudou na dispersão das estimativas (compara 3/10 com 9/25)?

5. Variabilidade em números: a população {2, 4, 6, 8, 10} tem $\mu = 6$. Três amostras de dimensão 2 dão $\bar{x} = 3$, $\bar{x} = 7$ e $\bar{x} = 9$. (a) Qual delas «erra» mais e qual «acerta»? (b) Podias escolher a amostra certa a priori? (c) O que isso diz sobre confiar num único número?

6. Erros: corrige (a) « $\bar{x}$ é o parâmetro da população» · (b) «o valor da amostra é o valor verdadeiro da população» · (c) «qualquer grupo de alunos serve como amostra».

7. Viés sem conserto: num fórum de leitura online se pergunta «quantos livros lês por mês?». (a) Que população «não é» a que se quer (amostra de auto-seleção)? (b) Porque é que nem aumentando o número de respostas o viés desaparece? (c) O que faltaria (processo aleatório + cobertura da população-alvo)?

8. Dimensão da amostra: com a população {M, M, M, H, H}, compara amostras de dimensão 1 ($\hat{p} \in \{0, 1\}$) e dimensão 2 ($\hat{p} \in \{0, 0,5, 1\}$). (a) Em que sentido a dimensão 2 «melhorou» a estimativa? (b) O que se espera para $n = 4$? (Antecipa a ficha 11: variabilidade menor com n maior.)

9. Contar é impossível: (a) Quantas amostras de dimensão 2 há em 5 pessoas? (Em 1000 pessoas?) (dica: combinações). (b) Porque é que «enumerar todas as amostras» deixa de ser viável e que ferramenta matemática resolve (probabilidade + TLC — ficha 11)?

10. Estimativa pontual numa sondagem: 52 % dos inquiridos apoiam X ($n = 800$). (a) O que esta estimativa não diz (o valor verdadeiro? o erro?). (b) Porque é que a notícia deveria dizer «52 % $\pm$ margem» (antecipa o IC — ficha 12)? (c) Que papel teve a seleção aleatória (ou não) no erro total?

11. Tecnologia: descreve como simular amostras numa folha de cálculo — (a) sortear valores (função aleatória) com e sem reposição, (b) calcular $\bar{x}$ ou $\bar{p}$ de cada amostra (100 amostras), (c) histogramar os resultados $\rightarrow$ o que o gráfico mostra (a distribuição de amostragem — ficha 11, AT 16).

12. Síntese: «Estimar é um palpite informado.» Explica (4–5 linhas) porquê a amostra aleatória torna o palpite mensurável (probabilidade do erro), porquê qualquer estimativa pontual é limitada (variabilidade amostral: tantas estimativas quantas amostras) e o que é preciso acrescentar para decidir com segurança — a incerteza declarada que os intervalos de confiança vão dar (ficha 12).

Soluções — Amostragem e estimação

11.º Ano

MACS

Resolução completa

1. população · amostra · μ · p · aleatória · erro.

2. (a) Não é aleatória: quem entra no cinema **já** é espectador — a amostra é circular (viés de seleção: a cidade real tem quem não vai ao cinema). (b) Processo aleatório: **sorteio** de moradores do arquivo telefónico (ou por quotas por bairro/idade) com contactos escolhidos **ao acaso** — assim, todos têm probabilidade conhecida de serem selecionados (permite quantificar o erro depois).

3. (a) 10 amostras (sem reposição): MM, MM, MM (3 — as mulheres distinguem-se por posição) · MH×6 (2H × 3M = 6 pares mistos) · HH×1. (b) $p^* = 0$ (0 homens) em $3/10 = 0,3$ · $p^* = 0,5$ (1 de 2) em $6/10 = 0,6$ · $p^* = 1$ (2 de 2) em $1/10 = 0,1 \rightarrow$ soma 1 ✓. (c) Média = $0(0,3) + 0,5(0,6) + 1(0,1) = 0,4$ — igual a $p = 2/5 = 0,4$: as estimativas oscilam, mas **em média acertam** (não-enviesado).

4. (a) Com reposição (25 pares ordenados): $P(p^*=0) = P(MM) = (3/5)^2 = 9/25$ · $P(p^*=0,5) = P(MH \text{ ou } HM) = 2 \times 3/5 \times 2/5 = 12/25$ · $P(p^*=1) = P(HH) = (2/5)^2 = 4/25 \rightarrow$ soma $25/25$ ✓. (b) Média = $0(9/25) + 0,5(12/25) + 1(4/25) = 6/25 + 4/25 = 10/25 = 0,4$ — sim, continua! (c) Dispersão mudou: os extremos **ganham peso** ($0,3 \rightarrow 0,36$ e $0,1 \rightarrow 0,16$) e o «0,5» caiu ($0,6 \rightarrow 0,48$) — a reposição dá mais liberdade às extremidades (as escolhas são independentes).

5. (a) «Erra» mais $\bar{x} = 3$ (distância $|3 - 6| = 3$) e $|9 - 6| = 3$ também; «acerta» o $\bar{x} = 7$ (distância 1). (b) **Não** — antes de sortear, qualquer amostra é igualmente possível; não se escolhe a certa. (c) Confiar num único número é arriscado: o mesmo processo dá 3, 7 ou 9 — **a estimativa pontual é instável**, por isso se declara a incerteza (e se pedem intervalos, ficha 12).

6. (a) **Falso** — $\bar{x}$ é **estatística** da amostra; o parâmetro μ é da população (grego vs latino, ficha 8). (b) **Falso** — o valor da amostra é uma **estimativa** (diferente de amostra para amostra — ex. 5). (c) **Falso** — tem de ser **aleatória** (com probabilidades conhecidas); «quem está à mão» introduz viés não medível (ex. 2 e 7).

7. (a) Quem responde é quem **já le muito** (fórum de leitura) $\rightarrow$ a amostra superestima a média da população. (b) Porque o viés é de **seleção**, não de tamanho: mais respostas dão **mais precisão à volta do valor errado** — o alvo continua deslocado. (c) Faltaria um **processo aleatório** sobre a população-alvo (cidadãos, com idades/locais sorteados) — daí a probabilidade poder medir o erro.

8. (a) Com $n = 1$ só há $\{0, 1\}$ (estimativas extremas: ou tudo ou nada); com $n = 2$ aparece o **0,5** e as estimativas ficam mais «próximas» de p — mais valores possíveis = maior resolução. (b) Para $n = 4$ esperam-se $p^* \in \{0, 0,25, 0,5, 0,75, 1\}$ — a distribuição concentra-se mais em torno de 0,4 (menos dispersão). É o embrião do **TLC**: variabilidade decrescente com n (ficha 11).

9. (a) Em 5 pessoas: $C(5, 2) = 10$ (ex. 3) · em 1000: $C(1000, 2) = 1000 \cdot 999 / 2 = 499\,500$ (e para $n = 30$, é um número com dezenas de dígitos). (b) Enumerar deixa de ser viável — e não é preciso: a **probabilidade** descreve a distribuição das estimativas **por fórmula**, e o **TLC** (ficha 11) diz-nos a forma dela sem contar nada (é isso que torna a inferência possível em populações grandes).

10. (a) Não diz que o valor verdadeiro seja 52 % (pode ser 49 %, 55 %...) nem **qual o erro** — a estimativa é um número sem margem. (b) Porque a notícia só é honesta com a **incerteza declarada**: «52 %, margem de ± 3 %» $\rightarrow$ intervalo onde p vive com confiança (ficha 12). (c) Se a seleção **não** foi aleatória (recursos online, não-respostas), há **viés adicional** que a margem não cobre — a margem só vale para erros de amostragem aleatória.

11. (a) Sortear com a função aleatória: =ALEATÓRIO.ENTRE(início; fim) (com reposição) ou embaralhar a lista e cortar n (sem reposição). (b) Repetir 100 vezes e registar $\bar{x}$ ou p^* de cada (coluna de resultados). (c) **Histogramar** essa coluna $\rightarrow$ vê-se a **distribuição de amostragem**: em torno de quê, quão larga, e como estreita com n maior — exatamente o que a ficha 11 formaliza (e o AT 16 simula).

12. A amostra **aleatória** dá ao palpite uma escala: como cada elemento tinha probabilidade conhecida de entrar, o **erro é um objeto probabilístico** (não uma opinião). Mas qualquer estimativa pontual é **limitada**: com a mesma população, outras amostras dão outros valores (ex. 5: 3, 7, 9) — «tantas estimativas quantas amostras». E é por isso que um número sozinho não chega para decidir: falta a **incerteza declarada** — e é ela que os **intervalos de confiança** (ficha 12) acrescentam à estimativa, dizendo com que confiança o parâmetro vive num intervalo e não num ponto.

Nota: quatro blocos — **parâmetro vs estimador** · **aleatoriedade e viés** · **variabilidade amostral com contas** · **do ponto ao intervalo**. Liga: fichas 7–9 (probabilidade) · ficha 11 (distribuição de amostragem e TLC) · ficha 12 (IC) · AT 16 (simulação) · resumo 3.