

Ficha 2 — Distribuição de amostragem e Teorema Limite Central

12.º Ano

Matemática A

4 páginas

12 exercícios

Objetivo: compreender a distribuição de amostragem de uma estatística, calcular o erro padrão da média, aplicar o Teorema Limite Central e interpretar a variabilidade das estimativas em função do tamanho da amostra.

Revisão rápida — como se comporta a média amostral

- **Distribuição de amostragem** de uma estatística é a distribuição (teórica) de todos os valores que essa estatística toma **em todas as amostras possíveis** do mesmo tamanho retiradas da população. Não se confunde com a distribuição da população.
- A **média** da distribuição de amostragem de $\bar{x}$ é igual à média populacional: $\mu_{\bar{x}} = \mu$.
- O **desvio padrão** da distribuição de amostragem de $\bar{x}$ chama-se **erro padrão** e vale $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, onde σ é o desvio padrão da população e n o tamanho da amostra.
- Consequência prática: para **quadruplicar** a precisão (reduzir o erro padrão a metade) é preciso multiplicar o tamanho da amostra por **4** — o erro padrão decresce com a raiz de n , não com n .
- **Teorema Limite Central (TLC)**: para amostras de tamanho n suficientemente grande, a distribuição de amostragem de $\bar{x}$ é **aproximadamente normal**, qualquer que seja a forma da distribuição da população. Escreve-se $\bar{x} \approx N(\mu, \sigma/\sqrt{n})$. Na prática usa-se $n \geq 30$.
- Se a **população já for normal**, então a distribuição de $\bar{x}$ é **exatamente normal** para qualquer n (não sendo preciso invocar o TLC).

Exercícios

1. Parâmetros da população: considere o lançamento de um dado equilibrado (valores 1 a 6). Determina a média μ e o desvio padrão σ da população, indicando o valor exato e arredondado a três casas decimais.

2. Erro padrão: para amostras de 4 lançamentos, calcula o erro padrão $\sigma/\sqrt{n}$ da média amostral (três casas decimais).

3. Mais observações: calcula o erro padrão para amostras de 25 lançamentos e compara-o com o obtido na questão 2, comentando a melhoria de precisão.

4. Quadruplicar a amostra: calcula o erro padrão para $n = 16$ e para $n = 64$ e mostra que quadruplicar o tamanho da amostra divide o erro padrão por 2.

5. Enunciar o TLC: enuncia o Teorema Limite Central e explica em que situação é necessário recorrer-lhe (e em que situação é dispensável).

6. População normal: se a população for normal com $\mu = 50$ e $\sigma = 8$, qual é a distribuição de amostragem de $\bar{x}$ para amostras de 10 elementos? Indica os parâmetros dessa distribuição.

7. Simulação — centro: numa simulação retiraram-se 20 000 amostras de 4 lançamentos e obteve-se uma média das médias de 3,4993. Que valor teórico se esperava e por que razão o valor simulado não coincide exatamente?

8. Simulação — dispersão: na mesma simulação ($n = 4$), o desvio padrão das 20 000 médias foi 0,8537. Compara este valor com o erro padrão calculado na questão 2 e explica o que a concordância confirma.

9. Probabilidade com o TLC: usando a distribuição de amostragem de $\bar{x}$ para $n = 25$ e a aproximação normal, calcula a probabilidade de a média de 25 lançamentos ser superior a 4. (Indica o valor de z e o resultado a quatro casas decimais.)

10. Porquê menos dispersão? Explica, com as tuas palavras, por que razão a média de 25 lançamentos é mais estável (varia menos) do que a média de 4 lançamentos.

11. Dispersão e enviesamento: explica a diferença entre reduzir o erro padrão e corrigir o enviesamento de uma amostra. Aumentar o tamanho da amostra resolve os dois problemas?

12. Aplicação: para estimar o gasto médio mensal dos clientes de uma loja, recolheu-se uma amostra de 100 clientes, obtendo-se $\bar{x} = 84$ €. Sabendo que $\sigma = 20$ €, descreve a distribuição de amostragem de $\bar{x}$ e calcula o erro padrão, indicando o que ele mede.

Soluções — Distribuição de amostragem e TLC

1. A média é $\mu = (1 + 2 + 3 + 4 + 5 + 6)/6 = 21/6 = 3,5$. Quanto à variância, $E(X^2) = (1 + 4 + 9 + 16 + 25 + 36)/6 = 91/6$, logo $\sigma^2 = 91/6 - 3,5^2 = 15,1667 - 12,25 = 35/12 \approx 2,9167$. Assim $\sigma = \sqrt{35/12} \approx 1,708$.
 2. $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1,707825/\sqrt{4} = 1,707825/2 \approx 0,854$.
 3. Para $n = 25$: $\sigma_{\bar{x}} = 1,707825/5 \approx 0,342$. O erro padrão passou de 0,854 para 0,342, ou seja ficou cerca de 2,5 vezes menor — com cinco vezes mais lançamentos, a média amostral torna-se bastante mais precisa.
 4. Para $n = 16$: $1,707825/4 \approx 0,427$. Para $n = 64$: $1,707825/8 \approx 0,213$. Ora $0,427/0,213 \approx 2$, exatamente: como o erro padrão vale $\sigma/\sqrt{n}$ e $\sqrt{64} = 2\sqrt{16}$, quadruplicar n divide a raiz por 2 e portanto **divide o erro padrão por 2**. É a razão pela qual ganhar precisão em estatística é «caro»: cada duplicação de precisão exige quadruplicar a amostra.
 5. **TLC**: para amostras de tamanho n suficientemente grande, a distribuição de amostragem da média $\bar{x}$ é aproximadamente normal, com média μ e desvio padrão $\sigma/\sqrt{n}$ — **independentemente da forma da distribuição da população**. É necessário recorrer-lhe quando a população **não é normal** (por exemplo, uma população claramente assimétrica) e se pretende usar a normalidade de $\bar{x}$; é **dispensável** quando a população já é normal, pois nesse caso $\bar{x}$ é exatamente normal para qualquer n .
 6. Sendo a população normal, $\bar{x}$ tem distribuição **exatamente normal** para qualquer n (não é preciso invocar o TLC). Os parâmetros são $\mu_{\bar{x}} = \mu = 50$ e $\sigma_{\bar{x}} = 8/\sqrt{10} \approx 2,530$. Escreve-se $\bar{x} \sim N(50; 2,530)$.
 7. O valor teórico esperado é $\mu = 3,5$, pois a média da distribuição de amostragem é igual à média da população (o estimador é centrado). O valor simulado (3,4993) não coincide exatamente porque uma simulação é uma **aproximação finita**: trabalha-se com 20 000 amostras, não com «todas» as amostras possíveis, e existe ainda o efeito do gerador de números pseudoaleatórios. A diferença (0,0007) é, no entanto, muito pequena — o que confirma o resultado teórico.
 8. O valor simulado (0,8537) é praticamente igual ao erro padrão teórico calculado na questão 2 (0,854). Esta concordância confirma experimentalmente a fórmula $\sigma_{\bar{x}} = \sigma/\sqrt{n}$: a dispersão das médias amostrais obtida por simulação reproduz a que a teoria prevê. É uma verificação independente do resultado — a teoria e a experiência dão o mesmo número.
 9. Com $n = 25$, a distribuição de $\bar{x}$ tem média 3,5 e desvio padrão 0,341565 (questão 3). Padronizando: $z = (4 - 3,5)/0,341565 \approx 1,4639$. Então $P(\bar{x} > 4) = 1 - \Phi(1,4639) \approx 0,0716$. Ou seja, em cerca de 7% das amostras de 25 lançamentos a média excederia 4 — um acontecimento pouco frequente, mas possível.
 10. A média de 25 lançamentos é mais estável porque, ao somar mais valores, as flutuações individuais **compensam-se**: um lançamento alto tende a ser equilibrado por outro baixo. Quanto maior o número de observações somadas, menor o efeito relativo de cada valor extremo, e por isso a média resultante varia menos de amostra para amostra. É exatamente isso que a fórmula $\sigma/\sqrt{n}$ traduz: o erro padrão diminui quando n aumenta.
 11. O **erro padrão** mede a dispersão das estimativas em torno do valor médio das estimativas — é a variabilidade devida ao acaso da amostragem, e diminui quando n aumenta. O **enviesamento** é um desvio **sistemático**: a estimativa erra sempre no mesmo sentido porque o processo de amostragem está mal concebido (por exemplo, uma amostra de conveniência). Aumentar n **reduz o erro padrão, mas não corrige o enviesamento**: uma amostra grande mas enviesada continua a dar resultados errados, apenas com aparência de maior precisão. Primeiro garante-se a representatividade; só depois se ganha precisão aumentando a amostra.
 12. A distribuição de amostragem de $\bar{x}$ tem média $\mu_{\bar{x}} = \mu$ (o gasto médio de todos os clientes) e desvio padrão $\sigma_{\bar{x}} = 20/\sqrt{100} = 2$ €. Pelo TLC, como $n = 100 \geq 30$, essa distribuição é aproximadamente normal: $\bar{x} \approx N(\mu, 2)$. O erro padrão de 2 € mede a **dispersão esperada das médias amostrais**: se se recolhessem muitas amostras de 100 clientes, as respetivas médias afastar-se-iam de μ , em média, cerca de 2 € — é a precisão típica da estimativa.
- Nota:** o resultado mais importante desta ficha é a forma do erro padrão — $\sigma/\sqrt{n}$. A raiz quadrada explica por que razão aumentar a amostra melhora a precisão cada vez mais devagar, e é por isso que nenhuma sondagem atinge precisão perfeita apenas por crescer.