

Resumo — Introdução à inferência estatística

12.º Ano

Matemática A

2 páginas

Síntese

Esta síntese reúne a introdução à inferência estatística do 12.º ano: raciocínio indutivo, vocabulário (população, amostra, parâmetro, estatística, estimador e estimativa), amostragem aleatória, distribuição de amostragem, erro padrão, Teorema Limite Central e intervalos de confiança para a média e para a proporção.

1. Raciocínio inferencial e vocabulário

- **Raciocínio indutivo:** parte de casos particulares (a amostra) para uma conclusão geral (a população). A conclusão é **plausível**, não certa — é o raciocínio próprio da estatística.
- **População:** conjunto total sobre o qual se quer concluir. **Amostra:** subconjunto efetivamente observado.
- **Parâmetro:** característica **fixa e desconhecida** da população (μ , p). **Estatística:** característica calculada na amostra ($\bar{x}$, $\hat{p}$) — varia de amostra para amostra.
- **Estimador:** a regra usada ($\bar{x}$ é um estimador de μ). **Estimativa:** o número obtido numa amostra concreta ($\bar{x} = 12,4$).
- **Amostra aleatória:** todos os elementos têm a mesma probabilidade de ser escolhidos. Uma amostra grande mas **enviesada** (de conveniência) é pior do que uma amostra pequena bem escolhida.

2. Distribuição de amostragem e erro padrão

- A **distribuição de amostragem** de uma estatística é a distribuição dos seus valores em todas as amostras possíveis do mesmo tamanho. Não se confunde com a distribuição da população.
- A média da distribuição de amostragem de $\bar{x}$ é μ (o estimador é centrado); o seu desvio padrão chama-se **erro padrão**: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.
- O erro padrão mede a **dispersão das estimativas**, não a dispersão dos indivíduos. Diminui com $\sqrt{n}$: **quadruplicar n divide o erro padrão por 2**.

3. Teorema Limite Central

- Para n suficientemente grande, a distribuição de amostragem de $\bar{x}$ é **aproximadamente normal**, qualquer que seja a forma da população: $\bar{x} \approx N(\mu, \sigma/\sqrt{n})$. Na prática, $n \geq 30$.
- Se a população **já for normal**, então $\bar{x}$ é **exatamente normal** para qualquer n — o TLC é dispensável.
- É este teorema que permite usar a normal padrão para construir intervalos de confiança e calcular probabilidades sobre $\bar{x}$.

4. Intervalo de confiança para o valor médio

- **Estimativa pontual:** $\bar{x}$. **Estimativa intervalar:** $\bar{x} \pm z_{1-\alpha/2} \times \sigma/\sqrt{n}$, onde $\Phi(z_{1-\alpha/2}) = 1 - \alpha/2$.
- Quantis a memorizar: **90% → 1,645; 95% → 1,96; 99% → 2,576**.
- Exemplo: com $\bar{x} = 12$, $\sigma = 3$ e $n = 100$, a margem a 95% é $1,96 \times 0,3 = 0,588$, logo o intervalo é **[11,412; 12,588]**.
- **Mais confiança ⇒ intervalo mais largo:** para 99% a margem sobe para 0,7728. Confiança e precisão estão em tensão.

5. Intervalo de confiança para a proporção

- Estimativa pontual: $\hat{p}$. Intervalo: $\hat{p} \pm z_{1-\alpha/2} \times \sqrt{\hat{p}(1-\hat{p})/n}$.
- **Condições:** amostra aleatória e $n\hat{p} \geq 5$ e $n(1-\hat{p}) \geq 5$.
- Exemplo: $\hat{p} = 0,62$ com $n = 1000$ dá erro padrão 0,015349 e margem 0,0301, logo **[0,5899; 0,6501]** — de 59,0% a 65,0%.
- **Margem de erro máxima** (pior caso $\hat{p} = 0,5$): $1,96 \times \sqrt{0,25/n}$ — anunciável **antes** de conhecer o resultado. Com $n = 1000$ vale $\pm 3,10\%$; com $n = 2000$, $\pm 2,19\%$.
- **Sondagens:** dois valores cujos intervalos se **sobrepõem** não permitem concluir qual é maior — uma diferença de 2 pontos percentuais com margem de 3% não é uma diferença.

6. Dimensionar a amostra

- Para uma margem de erro pretendida MA, resolve-se a desigualdade em ordem a n e arredonda-se **sempre por excesso**.
- Média:** $n \geq (z \sigma / MA)^2$. **Proporção:** $n \geq z^2 \times 0,25 / MA^2$ (com $z = 1,96$, $z^2 \approx 3,8416$).
- Exemplos: margem máxima de 5% $\Rightarrow n = 385$ (com 384 a margem seria 5,001%, acima do limite); margem de 3% $\Rightarrow n = 1068$.
- Ganhar precisão é caro: passar de 5% para 3% quase **triplica** a amostra.

7. Interpretação correta do nível de confiança

- O parâmetro μ (ou p) é um valor **fixo**, embora desconhecido; o **intervalo é que é aleatório**, pois depende da amostra.
- «95% de confiança» significa que o **método** produz intervalos que contêm o parâmetro em 95% dos casos, quando se repetisse a amostragem muitas vezes — não é a probabilidade de μ estar neste intervalo concreto.
- Uma simulação com 2000 intervalos ($n = 25$, $\sigma = 1$, $\mu = 0$) obteve **1891 intervalos contendo μ** — proporção **0,9455**, próxima de 0,95. Confirma o significado do nível de confiança.
- Falso:** «95% dos elementos da população pertencem ao intervalo» — o intervalo é sobre a **média**, não sobre os indivíduos.

8. Tabela-resumo

Grandeza	Fórmula	Observação
Erro padrão de $\bar{x}$	$\sigma/\sqrt{n}$	quadruplicar n divide por 2
IC para a média	$\bar{x} \pm z \times \sigma/\sqrt{n}$	$z = 1,96$ a 95%
Erro padrão de $\hat{p}$	$\sqrt{\hat{p}(1 - \hat{p})/n}$	exige $n \hat{p} \geq 5$ e $n(1 - \hat{p}) \geq 5$
IC para a proporção	$\hat{p} \pm z \times \sqrt{\hat{p}(1 - \hat{p})/n}$	$z = 1,96$ a 95%
Margem máxima (proporção)	$1,96 \times \sqrt{0,25/n}$	pior caso $\hat{p} = 0,5$
Amostra para a média	$n \geq (z \sigma / MA)^2$	arredondar por excesso
Amostra para a proporção	$n \geq 3,8416 \times 0,25 / MA^2$	arredondar por excesso

9. Erros frequentes a evitar

- Dizer que μ pertence ao intervalo com 95% de probabilidade (μ é fixo; o intervalo é aleatório).
- Confundir intervalo de confiança para a **média** com intervalo que contém 95% dos **indivíduos**.
- Comparar sondagens sem olhar para as margens de erro, ou concluir uma vantagem a partir de intervalos sobrepostos.
- Usar a fórmula da proporção sem verificar $n \hat{p} \geq 5$ e $n(1 - \hat{p}) \geq 5$.
- Arredondar a dimensão da amostra por defeito (falha a precisão pedida).
- Aumentar a amostra esperando corrigir um **enviesamento**: isso reduz o erro padrão, não o desvio sistemático.
- Esperar que duplicar a amostra reduza a margem para metade — a margem depende de $1/\sqrt{n}$.

Controlo rápido: com $n = 1000$ a margem máxima de uma sondagem é 3,10%; para a reduzir a 3% são precisos 1068 inquiridos; e uma simulação de 2000 intervalos a 95% ($n = 25$) conteve μ em 1891 casos (0,9455) — perto dos 95% esperados. Se te pedirem «95% de quê?», a resposta é: **95% dos intervalos construídos deste modo contêm o parâmetro**.