

Teste de Matemática — Introdução à inferência estatística (12.º Ano)

10 questões 20 valores 3 páginas

Instruções: responde a todas as questões, apresentando os cálculos e as justificações. Cada questão vale 2 valores. As soluções encontram-se na página 3.

Grupo I — Amostragem e estimação

1. (2 valores) Distingue parâmetro de estatística e apresenta um exemplo de cada, no contexto de um estudo sobre a altura dos alunos de uma escola.

2. (2 valores) Considera uma população com 5 elementos. Quantas amostras distintas de tamanho 2 se podem formar com reposição e sem reposição? Justifica.

3. (2 valores) Numa escola com 1000 alunos recolheu-se uma amostra aleatória de 50 alunos, obtendo-se um tempo médio de estudo de 12,4 horas por semana. Identifica, neste contexto, a população, a amostra, o parâmetro, a estatística, o estimador e a estimativa.

Grupo II — Distribuição de amostragem e Teorema Limite Central

4. (2 valores) O lançamento de um dado equilibrado tem valor médio $\mu = 3,5$ e desvio padrão $\sigma = \sqrt{35/12} \approx 1,708$. Calcula o erro padrão da média para amostras de 4 lançamentos e interpreta o valor obtido.

5. (2 valores) Enuncia o Teorema Limite Central e explica por que razão é um resultado central para a inferência estatística.

6. (2 valores) Com $\sigma = \sqrt{35/12} \approx 1,708$, calcula o erro padrão da média para $n = 25$ e para $n = 100$. Comenta a relação entre o tamanho da amostra e o erro padrão.

Grupo III — Intervalos de confiança

7. (2 valores) Numa amostra de 100 observações obteve-se $\bar{x} = 12$ e sabe-se que $\sigma = 3$. Constrói o intervalo de confiança a 95% para o valor médio μ e indica a margem de erro.

8. (2 valores) Numa sondagem com 1000 inquiridos, 62% declararam apoiar uma medida. Verifica as condições de aplicação da fórmula e constrói o intervalo de confiança a 95% para a proporção p .

9. (2 valores) Pretende-se uma sondagem com margem de erro máxima de 3 pontos percentuais, a 95% de confiança. Determina a dimensão mínima da amostra e verifica o resultado.

10. (2 valores) Explica o que significa, rigorosamente, «95% de confiança» e mostra por que razão é incorreto afirmar que «há 95% de probabilidade de μ pertencer a este intervalo».

Soluções — Introdução à inferência estatística

1. O **parâmetro** é uma característica **fixa e desconhecida** da população — aqui, a altura média de **todos** os alunos da escola (μ). A **estatística** é uma característica calculada na **amostra** e varia de amostra para amostra — a altura média medida num grupo de alunos escolhidos ao acaso ($\bar{x}$). (Aceita-se qualquer exemplo coerente, desde que distinga o valor populacional do valor amostral.)

2. **Com reposição:** cada escolha tem 5 opções e a ordem conta, logo $5^2 = 25$ amostras distintas. **Sem reposição:** não se repete o mesmo elemento, pelo que se conta o número de pares não ordenados: ${}^5C_2 = 5 \times 4/2 = 10$ amostras. A diferença explica por que razão a amostragem sem reposição fornece, para o mesmo n , mais informação sobre a população.

3. **População:** os 1000 alunos da escola. **Amostra:** os 50 alunos inquiridos. **Parâmetro:** o tempo médio de estudo de todos os 1000 alunos (μ , desconhecido). **Estatística:** o tempo médio de estudo dos 50 alunos da amostra ($\bar{x}$). **Estimador:** a média amostral $\bar{x}$, usada como regra para estimar μ . **Estimativa:** o valor concreto obtido, 12,4 horas por semana.

4. O erro padrão é $\sigma/\sqrt{n} = 1,708/\sqrt{4} = 1,708/2 \approx 0,854$. Significa que, se recolhêssemos muitas amostras de 4 lançamentos e calculássemos a média de cada uma, essas médias dispersar-se-iam em torno de 3,5 com um desvio padrão de cerca de 0,854 — bem menos do que o desvio padrão de 1,708 de um lançamento isolado, porque as médias de 4 lançamentos são menos variáveis do que lançamentos individuais.

5. O **Teorema Limite Central** afirma que, para n suficientemente grande (na prática $n \geq 30$), 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 central para a inferência porque permite usar a distribuição normal padrão para construir intervalos de confiança e calcular probabilidades sobre $\bar{x}$, mesmo quando nada se sabe sobre a distribuição da população original; sem ele, apenas se poderiam tirar conclusões para populações já normais.

6. Para $n = 25$: $1,708/\sqrt{25} = 1,708/5 \approx 0,342$. Para $n = 100$: $1,708/10 \approx 0,171$. O erro padrão **depende de $\sqrt{n}$** , não de n : quadruplicar a amostra de 25 para 100 (fator 4 em n) dividiu o erro padrão por 2 (fator $\sqrt{4} = 2$). Assim, ganhos de precisão exigem aumentos de amostra cada vez maiores — para reduzir o erro padrão a metade é preciso **quadruplicar** n .

7. A margem de erro é $1,96 \times \sigma/\sqrt{n} = 1,96 \times 3/\sqrt{100} = 1,96 \times 0,3 = 0,588$. O intervalo de confiança a 95% é $12 \pm 0,588$, isto é **[11,412; 12,588]**.

8. Condições: $n\hat{p} = 1000 \times 0,62 = 620 \geq 5$ e $n(1 - \hat{p}) = 1000 \times 0,38 = 380 \geq 5$ — estão reunidas, pelo que a fórmula é aplicável. O erro padrão é $\sqrt{(0,62 \times 0,38/1000)} = \sqrt{0,0002356} \approx 0,015349$ e a margem é $1,96 \times 0,015349 \approx 0,0301$. O intervalo é $0,62 \pm 0,0301 = \mathbf{[0,5899; 0,6501]}$, ou seja de cerca de 59,0% a 65,0%.

9. Exige-se $1,96 \times \sqrt{(0,25/n)} \leq 0,03$ (usa-se a margem máxima, por se desconhecerem os resultados). Então $\sqrt{(0,25/n)} \leq 0,03/1,96 = 0,015306$, donde $0,25/n \leq 0,00023428$ e $n \geq 1067,11$. Arredondando **por excesso**, **$n = 1068$** . Verificação: $n = 1068 \rightarrow \text{margem} \approx 2,999\% \checkmark$; $n = 1067 \rightarrow \text{margem} \approx 3,000\% \times$ (excede ligeiramente o limite) — confirma que não se pode arredondar por defeito.

10. «95% de confiança» significa que o **método** de construção do intervalo produz intervalos que contêm μ em **95% das repetições** da amostragem. É uma propriedade do procedimento, não do intervalo particular obtido. Afirmar «há 95% de probabilidade de μ pertencer a este intervalo» é incorreto porque μ é um valor **fixo** (embora desconhecido), não aleatório: depois de calculado, o intervalo concreto ou contém μ ou não contém, e a «probabilidade» de 95% refere-se à proporção de intervalos corretos numa amostragem repetida. Uma simulação de 2000 intervalos a 95% ($n = 25$) conteve μ em 1891 casos (proporção 0,9455) — próximo de 0,95, o que ilustra exatamente este significado.

Nota: a questão 10 é a mais exigente do teste — distingue quem sabe calcular um intervalo de quem compreende o que ele significa.