

Ficha 4 — Intervalo de confiança para a proporção

12.º Ano

Matemática A

4 páginas

12 exercícios

Objetivo: construir intervalos de confiança para uma proporção p , calcular a margem de erro de uma sondagem, dimensionar a amostra para a precisão pretendida e ler criticamente sondagens divulgadas na comunicação social.

Revisão rápida — intervalo de confiança para p

- Quando o parâmetro de interesse é uma **proporção** p (fração da população com uma dada característica), a estimativa pontual é a proporção amostral $\hat{p}$ = **número de casos favoráveis** / n .
- Intervalo de confiança a $100(1 - \alpha)\%$ para p :** $\hat{p} \pm z_{1-\alpha/2} \times \sqrt{\hat{p}(1 - \hat{p})/n}$, com os mesmos quantis da normal padrão (1,645 para 90%, **1,96** para 95%, 2,576 para 99%).
- Condições de aplicação:** a amostra tem de ser aleatória e suficientemente grande, verificando-se $n\hat{p} \geq 5$ e $n(1 - \hat{p}) \geq 5$. Só nestas condições a distribuição de amostragem de $\hat{p}$ é bem aproximada pela normal (Teorema Limite Central).
- Margem de erro máxima.** Como $\hat{p}(1 - \hat{p})$ é tanto maior quanto mais próximo de 0,5 estiver $\hat{p}$, o **pior caso** é $\hat{p} = 0,5$. Daí a fórmula usada nos estudos de opinião: **MA máxima** = $1,96 \times \sqrt{0,25/n}$ — válida **independentemente** do resultado observado.
- Regra prática:** como $1,96^2 \approx 3,84$, para uma margem máxima MA basta $n \geq 3,84 \times 0,25 / MA^2$, arredondando **sempre por excesso**.
- Leitura crítica:** uma sondagem com $n = 1000$ tem margem até $\pm 3,1$ pontos percentuais. Duas percentagens cujos intervalos **se sobrepõem** não permitem concluir qual é a maior — a diferença observada pode ser apenas erro de amostragem.

Exercícios

1. Numa sondagem com 1000 inquiridos, 62% declararam apoiar uma medida. Constrói o intervalo de confiança a 95% para a proporção de apoiantes e indica a margem de erro (em pontos percentuais).

2. Numa amostra de 400 pessoas, 200 apresentam a característica em estudo. Constrói o intervalo de confiança a 95% para p .

3. Numa amostra de 1000 inquiridos, 48% afirmaram ter intenção de votar num certo candidato. Constrói o intervalo de confiança a 95% e explica o que se pode (e o que não se pode) concluir sobre uma eventual maioria.

4. Com a mesma proporção de 62% mas apenas 100 inquiridos, constrói o intervalo a 95% e compara a sua largura com a da questão 1, comentando a perda de precisão.

5. Justifica a fórmula da margem de erro máxima de uma sondagem, $1,96 \times \sqrt{(0,25/n)}$, e calcula-a para $n = 500$, $n = 1000$ e $n = 2000$.

6. Uma empresa quer uma margem de erro máxima de 5 pontos percentuais, com 95% de confiança. Determina a dimensão mínima da amostra e verifica o resultado.

7. Repete a questão anterior para uma margem de erro máxima de 3 pontos percentuais e comenta quanto custa, em número de inquéritos, passar de 5% para 3%.

8. Verifica se a fórmula da proporção pode ser aplicada nos casos seguintes: (a) $n = 1000$ e $\hat{p} = 0,62$; (b) $n = 40$ e $\hat{p} = 0,05$; (c) $n = 200$ e $\hat{p} = 0,97$. Justifica em cada caso.

9. Um jornal anuncia: «A medida é apoiada por 62% dos portugueses, com uma margem de erro de 3%.» Explica por que razão a frase, tal como está escrita, é incorreta, e reescreve-a de forma correta.

10. Sondagem com dois candidatos: numa amostra de 1000 inquiridos, o candidato A obtém 43% e o candidato B 41%. Constrói os intervalos de confiança a 95% de ambos e conclui se é legítimo afirmar que A está à frente de B.

11. Justificação: explica por que razão a margem de erro de uma sondagem depende do resultado observado quando se usa $\sqrt{(\hat{p}(1-\hat{p}))/n}$, e não depende quando se usa a margem máxima $\sqrt{(0,25/n)}$.

12. Aplicação: para uma margem de erro máxima de 2 pontos percentuais (95%), mostra que são necessários cerca de 2401 inquiridos e explica por que razão duplicar a amostra de 1000 para 2000 não reduz a margem para metade.

Soluções — Intervalo de confiança para a proporção

1. Tem-se $\hat{p} = 0,62$ e $n = 1000$. O erro padrão é $\sqrt{(0,62 \times 0,38/1000)} = \sqrt{0,0002356} \approx 0,015349$. A margem de erro é $1,96 \times 0,015349 \approx 0,0301$, ou seja cerca de **3,0 pontos percentuais**. O intervalo é $0,62 \pm 0,0301$, isto é **[0,5899; 0,6501]** — aproximadamente de **59,0% a 65,0%**.

2. $\hat{p} = 200/400 = 0,5$. O erro padrão é $\sqrt{(0,5 \times 0,5/400)} = \sqrt{0,000625} = 0,025$ e a margem é $1,96 \times 0,025 = 0,049$. O intervalo a 95% é **[0,451; 0,549]**, ou seja de **45,1% a 54,9%**.

3. $\hat{p} = 0,48$, $n = 1000$: erro padrão $\sqrt{(0,48 \times 0,52/1000)} \approx 0,015799$ e margem $1,96 \times 0,015799 \approx 0,031$. O intervalo é **[0,449; 0,511]** — de **44,9% a 51,1%**. Conclui-se que 50% está **dentro** do intervalo: os dados são compatíveis tanto com uma minoria como com uma maioria (mesmo uma maioria muito ligeira). **Não** se pode afirmar que o candidato tem maioria nem que não tem — a sondagem não permite decidir.

4. Com $\hat{p} = 0,62$ e $n = 100$: erro padrão $\sqrt{(0,62 \times 0,38/100)} \approx 0,048539$ e margem $1,96 \times 0,048539 \approx 0,0951$. O intervalo é **[0,5249; 0,7151]** — de **52,5% a 71,5%**. A largura passou de cerca de 6,0 pontos percentuais (questão 1) para **19,0 pontos percentuais**: dez vezes menos inquéritos deram um intervalo mais de três vezes mais largo. É a raiz quadrada no erro padrão a explicar este comportamento.

5. Na expressão $\sqrt{(\hat{p}(1 - \hat{p})/n)}$, o produto $\hat{p}(1 - \hat{p})$ é máximo quando $\hat{p} = 0,5$, valendo 0,25; para qualquer outro valor é menor. Tomando o pior caso, a margem **nunca excede** $1,96 \times \sqrt{(0,25/n)}$, pelo que esta fórmula dá um limite superior válido **sem conhecer o resultado** da sondagem. Valores: $n = 500 \rightarrow 1,96 \times 0,022361 \approx 0,0438$ (**4,4%**); $n = 1000 \rightarrow 0,0310$ (**3,1%**); $n = 2000 \rightarrow 0,0219$ (**2,2%**).

6. Exige-se $1,96 \times \sqrt{(0,25/n)} \leq 0,05$, isto é $\sqrt{(0,25/n)} \leq 0,025510$, donde $0,25/n \leq 0,00065077$ e $n \geq 384,16$. Arredondando por excesso, **$n = 385$** . Verificação: com $n = 385$ a margem máxima é $1,96 \times \sqrt{(0,25/385)} \approx 0,04995\%$ ✓; com $n = 384$ seria $5,001\% \times$ — confirma que **não** se pode arredondar por defeito.

7. Para 3%: $n \geq 3,8416 \times 0,25/0,0009 \approx 1067,11$, logo **$n = 1068$** . Verificação: $n = 1068 \rightarrow \approx 2,999\% \times$ ✓; $n = 1067 \rightarrow \approx 3,000\% \times$ (fica acima do limite). Passar da margem de 5% para 3% exige subir de **385 para 1068 inquiridos** — quase o triplo da amostra para ganhar dois pontos percentuais de precisão. É a demonstração prática de que a precisão estatística se paga caro.

8. A condição é $n\hat{p} \geq 5$ e $n(1 - \hat{p}) \geq 5$. (a) $n\hat{p} = 620$ e $n(1 - \hat{p}) = 380$, ambos ≥ 5 : **pode aplicar-se**. (b) $n\hat{p} = 40 \times 0,05 = 2 < 5$: **não pode aplicar-se** (a aproximação normal não é adequada para acontecimentos raros com amostras pequenas). (c) $n(1 - \hat{p}) = 200 \times 0,03 = 6 \geq 5$ e $n\hat{p} = 194 \geq 5$: **pode aplicar-se**, embora com cautela por o valor estar muito próximo do limite.

9. A frase é incorreta porque apresenta os 62% como se fossem o valor exato da proporção populacional, quando são apenas a **estimativa** obtida numa amostra; além disso, «margem de erro de 3%» sem indicação do **nível de confiança** é ambíguo. Formulação correta: «**Com base numa amostra aleatória de 1000 inquiridos, a proporção de apoiantes foi estimada em 62%. O intervalo de confiança a 95% para essa proporção vai de cerca de 59% a 65% (± 3 pontos percentuais).**» Assim fica claro que 62% é uma estimativa e que existe um intervalo de valores plausíveis.

10. Para A: erro padrão $\sqrt{(0,43 \times 0,57/1000)} \approx 0,015656$ e margem $\approx 0,0307$, logo **[0,3993; 0,4607]** (de 39,9% a 46,1%). Para B: erro padrão $\sqrt{(0,41 \times 0,59/1000)} \approx 0,015553$ e margem $\approx 0,0305$, logo **[0,3795; 0,4405]** (de 38,0% a 44,0%). Os intervalos **sobrepõem-se largamente** (a zona comum vai de 39,9% a 44,0%), portanto **não é legítimo afirmar** que A está à frente de B: uma vantagem de apenas 2 pontos percentuais está dentro da margem de erro da sondagem.

11. Quando se usa $\sqrt{(\hat{p}(1 - \hat{p})/n)}$, a quantidade $\hat{p}(1 - \hat{p})$ **depende do valor observado**: se a característica for rara ($\hat{p}$ próximo de 0 ou de 1), esse produto é pequeno e o intervalo é mais estreito. Já a margem máxima $\sqrt{(0,25/n)}$ toma o valor **mais desfavorável possível** ($\hat{p} = 0,5$, onde $\hat{p}(1 - \hat{p}) = 0,25$), pelo que representa um limite superior que se pode anunciar **antes de conhecer os resultados** — que é precisamente o que os estudos de opinião fazem, ao publicar a margem de erro máxima prevista no plano amostral.

12. Para uma margem máxima de 2%: $n \geq 3,8416 \times 0,25/0,0004 = 2401$ (valor exato: com $n = 2401$ tem-se $1,96 \times 0,5/49 = 0,02$, exatamente no limite). Duplicar a amostra de 1000 para 2000 **não divide** a margem por 2 porque a margem é proporcional a **$1/\sqrt{n}$** , não a $1/n$: com $n = 1000$ a margem máxima é 3,10% e com $n = 2000$ é 2,19% (e não 1,55%). Para reduzir a margem a metade seria necessário multiplicar n por **4** (de 1000 para 4000), obtendo-se aí cerca de 1,55%. É a mesma «lei da raiz quadrada» que governa o erro padrão da média.

Nota: a mensagem central desta ficha está na questão 10 — comparar sondagens sem olhar para as margens de erro é o erro mais frequente na comunicação social. Uma diferença menor do que a margem de erro **não é uma diferença**, estatisticamente falando.