Comparação de proporções

Observação sobre notação

- Nesta aula:
- *p* indicará proporção
- *P* indicará o valor de *P* (nível descritivo)

Introdução

- Supondo que estejamos interessados em estimar a proporção de indivíduos de uma população, π, que possua um determinado atributo. Ao selecionar uma amostra de tamanho *n*, observamos *r* animais que apresentam o atributo. A estimativa para a proporção populacional é, então,

$$p = \frac{r}{n}$$

Propriedades de distribuição de proporções amostrais

- A distribuição (para proporção) é aproximadamente Normal se o tamanho da amostra é grande; a princípio, a distribuição de uma proporção segue a distribuição Binomial.
- Supondo distribuição Normal, o desvio padrão para uma proporção (*p*) pode ser estimado por

$$s_{prop} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p\,(1-p)}{n}}$$

Propriedades de uma distribuição amostral de proporção

- O intervalo de confiança (IC) de 95% para uma proporção é dado por (considerando aproximação Normal)

$$p \pm 1{,}96\, s_{prop} =$$

$$= \left(p - 1{,}96\sqrt{\frac{p\,(1-p)}{n}}\,, \quad p + 1{,}96\sqrt{\frac{p\,(1-p)}{n}} \right)$$

A interpretação para este IC é que há 95% de probabilidade de que a proporção verdadeira esteja contida no intervalo indicado acima.

Exemplo

- Amostras de sangue de 115 cabeças de gado foram analisadas através de um teste sorológico para *Leptospira* e, de acordo com o título, cada amostra foi classificada em positiva ou negativa. Na amostra, 36 animais apresentaram títulos positivos.

$$p = \frac{36}{115} = 0{,}313 \qquad s_{prop} = \sqrt{\frac{p\,(1-p)}{n}} = \sqrt{\frac{0{,}313\,(1-0{,}313)}{115}} = 0{,}043$$

Intervalo de confiança de 95%:

$$p \pm 1{,}96\, s_{prop} =$$

$$= \left(0{,}313 - 1{,}96 \times 0{,}043\,, \quad 0{,}313 + 1{,}96 \times 0{,}043\right)$$

$$= (0{,}228;\ 0{,}398)$$

Há 95% de probabilidade de que a proporção verdadeira esteja entre 0,23 e 0,40.

Comparação de uma proporção

- Testando hipóteses sobre uma proporção simples
- Exemplo: Deseja-se saber se a proporção de fêmeas de coelhos selvagens é de 50%, ou seja, se há uma razão 1:1 entre machos e fêmeas desta espécie.
- 1) Hipóteses do teste: $H_0 : \text{p} = 0{,}5$

  $H_1 : \text{p} \neq 0{,}5$

- 2) Dados obtidos:
  - Em 297 nascimentos, 167 foram de fêmeas
  - Proporção observada de fêmeas:

$$\text{p} = \frac{167}{297} = 0{,}562$$

- 3) Resultados do Minitab

Test and CI for One Proportion

Test of p = 0,5 vs p not = 0,5

Sample	X	N	Sample p	95,0% CI	Z-Value	P-Value
1	167	297	0,562290	(0,505868; 0,618711)	2,15	0,032

- 4) Como o valor de P é 0,032, há evidência para se rejeitar a hipótese nula, para um nível de significância de 5%.
- 5) IC 95% : (0,51 ; 0,62)
  - O valor 0,50 não está contido no IC 95%. No entanto, devemos ser cautelosos ao concluir que possa haver algum fator afetando a razão entre machos e fêmeas, uma vez que o valor de P do teste está próximo de 5% e a proporção de fêmeas é de 56,2 %, excedendo por pouco 50%.
- Observação: Não confundir a proporção *p* da amostra com o valor de P do teste estatístico. São diferentes!

Comparação de 2 proporções (Amostras independentes)

- Podemos testar a hipótese de que 2 proporções populacionais são iguais de 2 modos:
  - teste χ^2
  - usando a aproximação Normal para distribuição Binomial e fazendo a comparação das duas proporções

  Os dois testes produzem resultados idênticos

Comparação de 2 proporções em uma tabela 2x2

- Raciocínio:
  - Se não há associação entre o resultado e o grupo, então seria de se esperar que as proporções de sucesso sejam as mesmas nos dois grupos

---

Comparação de 2 proporções (Intervalo de confiança)

- Intervalo de confiança para a diferença entre duas proporções:

$$(p_1 - p_2) \pm 1{,}96\sqrt{\frac{p_1\,(1-p_1)}{n_1} + \frac{p_2\,(1-p_2)}{n_2}}$$

Comparação de 2 proporções (Amostras independentes)

- **Exemplo:** Pesquisadores decidiram avaliar se a proporção de cães machos é idêntica em cães domiciliados e não-domiciliados. Fizeram um levantamento em um determinado município, e observaram que, dos 510 cães domiciliados amostrados, 301 eram machos, e, dentre os 230 não-domiciliados recolhidos, 97 eram machos. Pergunta-se: há diferença estatística entre as duas proporções?

- 1) Estabelecer as hipóteses do teste:
  - Hipótese nula: Proporção de machos é igual em cães domiciliados e não-domiciliados.
  - Hipótese alternativa: Proporção de machos é diferente nos dois grupos.
- 2) Dados obtidos:
  - Domiciliados: p_1 = 301 / 510 = 0,59 ou 59%
  - Não-domiciliados: p_2 = 97 / 230 = 0,42 ou 42%

- 3) Resultados do Minitab

Test and CI for Two Proportions

Sample	X	N	Sample p
1	301	510	0,590196
2	97	230	0,421739

Estimate for p(1) - p(2): 0,168457

95% CI for p(1) - p(2): (0,0916781; 0,245236)

Test for p(1) - p(2) = 0 (vs not = 0): Z = 4,30 P-Value = 0,000

- 4) Decidir se rejeita ou não a hipótese nula:

  Como o valor de P < 0,001, inferior a 5%, há evidências para se rejeitar a hipótese nula. Ou seja, observamos uma diferença estatística significativa entre as proporções de machos nos dois grupos.
- 5) IC 95% para a diferença entre as proporções nos dois grupos: (0,092 ; 0,245).

  O valor 0 (zero), que indicaria igualdade entre as duas proporções, não está incluso no IC 95%, levando-nos à mesma conclusão do item anterior.

Teste χ^2

valores observados

	Machos	Fêmeas	Total
Domic.	301	209	510
Não domic.	97	133	230
Total	398	342	740

valores esperados

	Machos	Fêmeas	Total
Domic.	274	236	510
Não domic.	124	106	230
Total	398	342	740

χ^2=18,10 p<0,001

Exemplo (aula de Associação)

- Camundongo não obeso diabético (NOD) desenvolve diabetes autoimune que é usada como modelo para o diabetes juvenil humano insulino-dependente. Nos camundongos da colônia de Hawkins, a incidência para machos e fêmeas era de 24% e 73%, respectivamente. Hawkins investigou a causa desta diferença entre os sexos avaliando o efeito da castração precoce na incidência de diabetes em camundongos NOD machos.

Exemplo

- De 100 camundongos NOD machos, 50 foram castrados um dia após o nascimento e outros 50 foram submetidos a uma falsa cirurgia. Os camundongos foram mantidos por 140 dias, e amostras de sangue foram colhidas a cada duas semanas a partir do 42 dia. Diabetes foi diagnosticada como estando presente quando três amostras consecutivas apresentaram níveis de glicose superiores a 200mg/dL. O experimento permitiu determinar que a castração provocou um aumento na incidência de diabetes (52%) quando comparado com animais não castrados (24%) no dia 112.

Teste χ^2

valores observados

	Castrados	Controle	Total
Com diabetes	26	12	38
Sem diabetes	24	38	62
Total	50	50	100

valores esperados

	Castrados	Controle	Total
Com diabetes	19	19	38
Sem diabetes	31	31	62
Total	50	50	100

χ^2=8,319 p=0,004

Teste de comparação de duas proporções

Hipóteses do teste:

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

- Proporção de castrados com diabetes (p_1): $p_1 = \frac{26}{50} = 0{,}52$
- Proporção de animais submetidos à falsa cirurgia com diabetes (p_2): $p_2 = \frac{12}{50} = 0{,}24$

Teste de comparação de duas proporções

Resultados do Minitab:

Test and CI for Two Proportions

Sample	X	N	Sample p
1	26	50	0.520000
2	12	50	0.240000

Estimate for p(1) - p(2): 0.28

95% CI for p(1) - p(2): (0.0978182, 0.462182)

Test for p(1) - p(2) = 0 (vs not = 0): Z = 3.01 P-Value = 0.003

Para um nível de significância de 5%, rejeitamos a hipótese nula de proporções iguais e dizemos que foi observada uma diferença estatística significativa entre as proporções (P=0,003). Essa decisão é idêntica à tomada ao se utilizar o teste de χ^2.