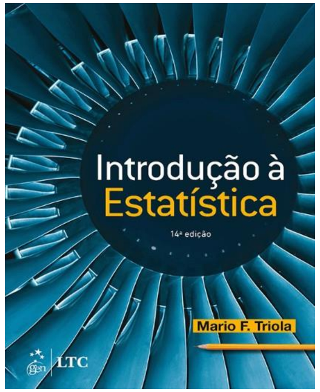
Aula 10 – Modelos Lineares Generalizados

Teste de Qui-quadrado

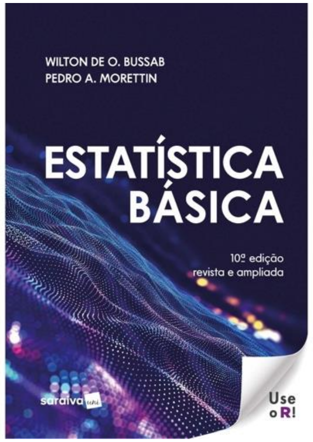
Ao final da aula você será capaz de:

- Identificar quando o teste de qui-quadrado é apropriado (dados categóricos).
- Diferenciar os dois principais tipos de teste de qui-quadrado: Aderência (Goodness-of-Fit) e Independência/Associação.
- Formular hipóteses nula (H_0) e alternativa (H_1) para ambos os testes.
- Calcular as frequências esperadas e a estatística χ^2 para problemas biológicos simples.
- Interpretar o resultado do teste com base no valor de *P* e nos graus de liberdade para tirar conclusões biológicas.

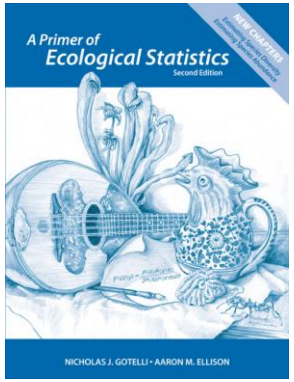
Cap 12



Cap 11

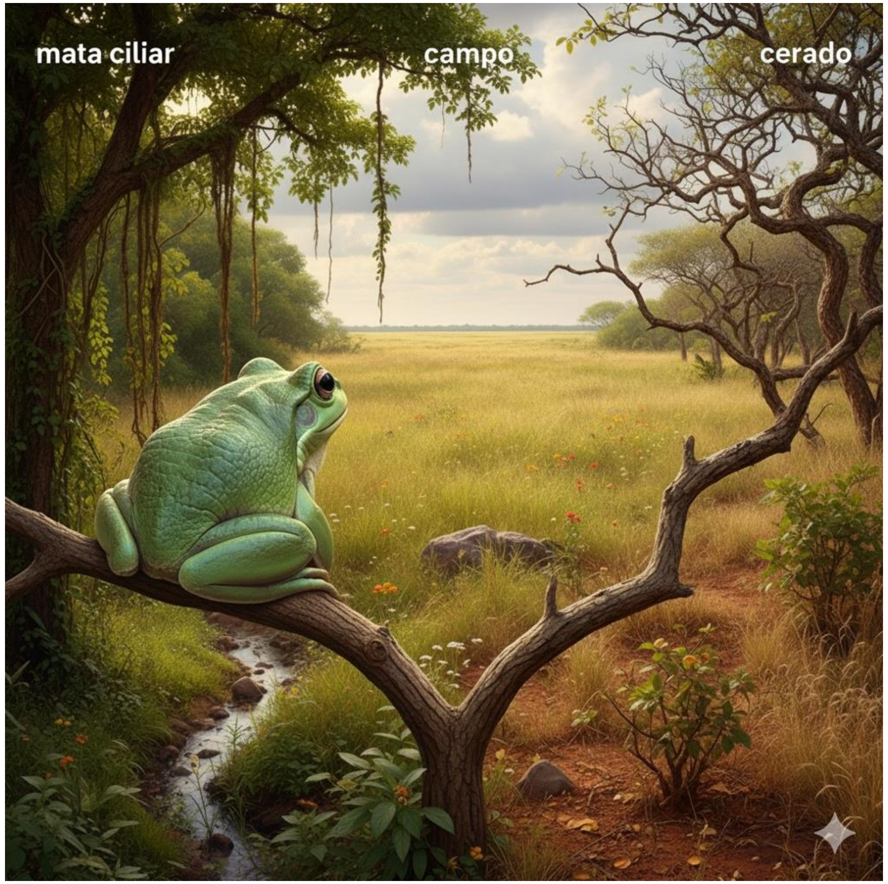


Cap 14



Cap 11

Cenário biológico



- Imaginem que vocês são biólogos de campo estudando a **preferência de habitat** de uma espécie de anfíbio em uma reserva com **três tipos de vegetação**: mata ciliar, campo e cerrado.
- Vocês **contaram quantos indivíduos foram encontrados em cada ambiente**.
- A pergunta é: essa espécie se distribui aleatoriamente pelo espaço ou **ela prefere um tipo de vegetação**? Como podemos testar isso com confiança?

Papel do Teste de qui-quadrado

- Muitos dados em biologia não são contínuos (como altura ou peso), mas sim categóricos ou qualitativos. Os dados do nosso exemplo são as contagens de animais em diferentes categorias de habitat.
- Teste de Qui-Quadrado (χ^2) é usado para analisar esse tipo de dado. O teste compara o que você observou no campo com o que você esperaria se não houvesse nenhum padrão.
  - Teste de Aderência (Goodness-of-Fit): Os nossos dados se ajustam a uma teoria?
  - Teste de Independência: Duas variáveis categóricas estão associadas?

Teste de Aderência (Goodness-of-Fit):

- **Conceito:** O teste de aderência é usado para verificar se a distribuição de frequência de uma variável categórica em uma amostra se ajusta a uma distribuição teórica ou esperada.
- **Frequências Observadas (O):** O que você realmente contou no seu estudo. São os seus dados brutos.
- **Frequências Esperadas (E):** O que você esperaria contar se sua hipótese nula fosse verdadeira.
- **Hipótese Nula (H_0):** Não há diferença significativa entre as frequências observadas e as esperadas. No nosso exemplo do anfíbio, H_0 seria: "A espécie não tem preferência de habitat e se distribui de forma uniforme entre os três ambientes".

Anatomia do qui-qudrado

- Cálculo das Frequências Esperadas (para distribuição uniforme):

$$E = \frac{Tamanho\ total\ da\ amostra\ (n)}{Número\ de\ categorias\ (k)}$$

- Fórmula do qui-quadrado

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Componentes da fórmula

- $(O-E)$: Mede a diferença (desvio) para cada categoria.
- $(O-E)^2$: Eleva ao quadrado para que os desvios não se anulem e para dar mais peso a grandes desvios.
- $(\ldots)^2/E$: Padroniza o desvio em relação ao que era esperado. Uma diferença de 10 é mais importante se você esperava 5 do que se esperava 100.
- $\sum$: Soma os desvios padronizados de todas as categorias para obter uma medida total de "desajuste".

Interpretação

- Um valor de χ^2 próximo de zero significa que seus dados observados se ajustam muito bem ao esperado (você provavelmente não rejeitará H_0).
- Um valor grande de χ^2 significa um grande desajuste (você provavelmente rejeitará H_0).

Exemplo prático: Genética Mendeliana

- Um cruzamento duplos heterozigotos de ervilhas (amarelas/lisas com verdes/rugosas) deve, teoricamente, produzir uma prole na proporção fenotípica de 9:3:3:1 (amarela/lisa : amarela/rugosa : verde/lisa : verde/rugosa).
- Um botânico realizou o cruzamento e obteve 500 sementes com os seguintes resultados:
  - 290 amarelas/lisas,
  - 88 amarelas/rugosas,
  - 95 verdes/lisas
  - 27 verdes/rugosas.
- H_0: Os resultados observados se

ajustam à proporção esperada de 9:3:3:1

Cálculo passo-a-passo

- **Calcular Frequências Esperadas (E)**:
  - Total da prole = 500. Proporção total = 9+3+3+1 = 16.
  - E (Amarela/Lisa) = (9/16)×500=281.25
  - E (Amarela/Rugosa) = (3/16)×500=93.75
  - E (Verde/Lisa) = (3/16)×500=93.75
  - E (Verde/Rugosa) = (1/16)×500=31.25

Calcular o χ^2:

- $\chi^2 = \frac{(290-281.25)^2}{281.25} + \frac{(88-93.75)^2}{93.75} + \frac{(95-93.75)^2}{93.75} + \frac{(27-31.25)^2}{31.25}$
- $\chi^2 = 0.27 + 0.35 + 0.02 + 0.58 = 1.22$

  - Graus de liberdade (gl): *k* – 1 = 4 – 1 = 3

Decisão: Método do valor crítico

- Compare o χ^2 calculado (1.22) com o valor crítico da tabela para gl=3 e α=0.05 (que é 7.81).
- Como 1.22 < 7.81, não rejeitamos H_0.
- "Os dados são consistentes com a teoria de Mendel."

Tabela A.2 Valores de χ^2 segundo os graus de liberdade e o valor de α

Graus de liberdade	Valor de α		
	10%	5%	1%
1	2,71	3,84	6,64
2	4,60	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,72
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18			

Sonia Vieira

A Ecologia das Balas (15 min)

- Formem 5 grupos
- Cenário: "A empresa Mars, Inc. afirma que a distribuição de cores dos M&M's em um pacote é: 24% azul, 20% laranja, 16% verde, 14% amarelo, 13% vermelho e 13% marrom. Um biólogo desconfiado comprou um pacote grande com 150 balas e encontrou as seguintes contagens: (forneça uma tabela com contagens observadas, por exemplo: 30 azuis, 25 laranjas, 28 verdes, 20 amarelos, 25 vermelhos, 22 marrons)."

A Ecologia das Balas (15 min)

- Tarefa do Grupo:
  - Formulem a hipótese nula (H_0).
  - Calculem as frequências esperadas (E) para cada cor em uma amostra de 150 balas.
  - Calculem o valor da estatística χ2
  - Determinem os graus de liberdade.
  - Perguntem para mim qual é o valor crítico
  - Tomem uma decisão: os dados dão motivos para duvidar da afirmação da empresa?
- Debriefing (5 min): Um ou dois grupos que compartilham resultados e conclusão.

Teste de Associação

- Usamos este teste quando temos duas variáveis categóricas para cada indivíduo da amostra e queremos saber se elas são associadas ou independentes.
- **Tabelas de Contingência**: Os dados são organizados em uma tabela de dupla entrada, também chamada de tabela de contingência. As linhas representam as categorias de uma variável e as colunas as da outra.
- **Hipótese Nula (H_0):** As duas variáveis são independentes. Não existe associação entre elas

Cálculo das frequências esperadas

- A lógica aqui é um pouco diferente. Se as variáveis são independentes, a frequência em cada célula deve ser proporcional aos totais de sua linha e coluna

$$E = \frac{(Total\ da\ linha) \times\ (Total\ da\ Coluna)}{Total\ geral}$$

Tabela de contingência

TABLE 11.2 **Two-way contingency table summarizing the relationship between protection and population status**

Population status	Protection status: Not protected	Protection status: Protected	*Row total*
Declining	$Y_{1,1} = 18$	$Y_{1,2} = 8$	$\sum_{j=1}^{m} Y_{1,j} = 26$
Stable or increasing	$Y_{2,1} = 15$	$Y_{2,2} = 32$	$\sum_{j=1}^{m} Y_{2,j} = 47$
Column total	$\sum_{i=1}^{n} Y_{i,1} = 33$	$\sum_{i=1}^{n} Y_{i,2} = 40$	$\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{i,j} = 73$

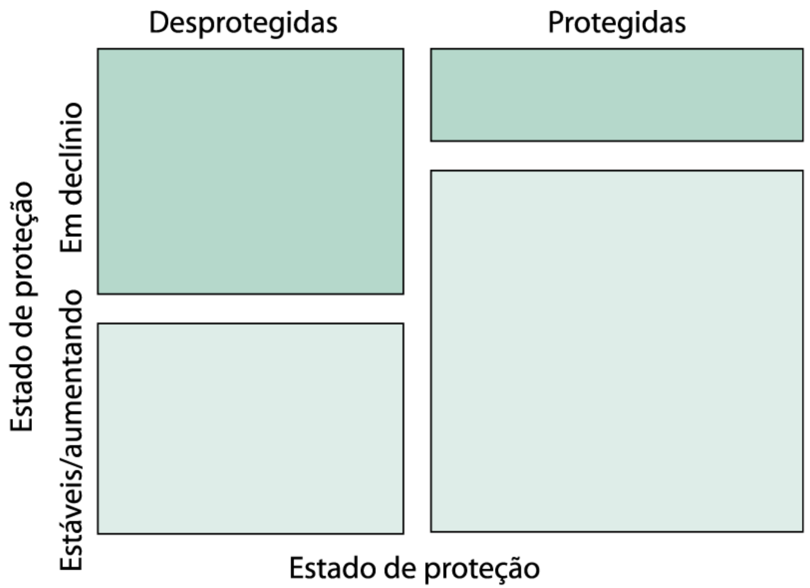


Figura 11.1 Ilustração de um gráfico mosaico. Esse gráfico retrata a relação entre o estado das populações de plantas raras e quando as terras em que elas ocorrem são protegidas ou não (Farnsworth, 2004). As frequências nas células (da Tabela 11.2) são representadas como "retângulos", cujos tamanhos são proporcionais à frequência relativa no conjunto de dados. A largura das colunas é proporcional ao total das colunas. Por exemplo, como 40 das 73 populações são protegidas, a coluna da proteção ocupa 55% (ou 40/73) da largura do gráfico. A altura de cada retângulo é proporcional à frequência da célula. Deste modo, 32 das 40 populações protegidas estão estáveis ou aumentando, portanto seu retângulo ocupa 32/40, ou 80%, da coluna da direita.

Gotelli & Ellison

Exemplo: Vacinas e autismo

TABLE 11-1 Results From a Study of a Link Between the MMR Vaccine and Autism

	Unvaccinated	Vaccinated	Total
Autism	25	64	89
No Autism	362	1427	1789
Total	387	1491	1878

H_0: A ocorrência de autismo é independente do status de vacinação com a vacina MMR

Cálculo das frequências esperadas (*E*)

- E (Autismo, Não Vacinado) = (89×387)/1878=18,34

- E (Autismo, Vacinado) = (89×1491)/1878=70,66

- Os totais de linha e coluna permanecem os mesmos
- $\chi^2 = 3.198$
- Graus de liberdade: gl = (nº de linhas – 1) x (nº de colunas – 1) = (2 -1) x (2 - 1)=1

Conclusão

- O valor crítico para gl=1 e α=0.05 é 3.841. Como nosso χ2 calculado (3.198) é menor que o crítico, não rejeitamos a H_0.
  - "Com base nestes dados, não há evidência estatística de uma associação entre a vacina MMR e o autismo”

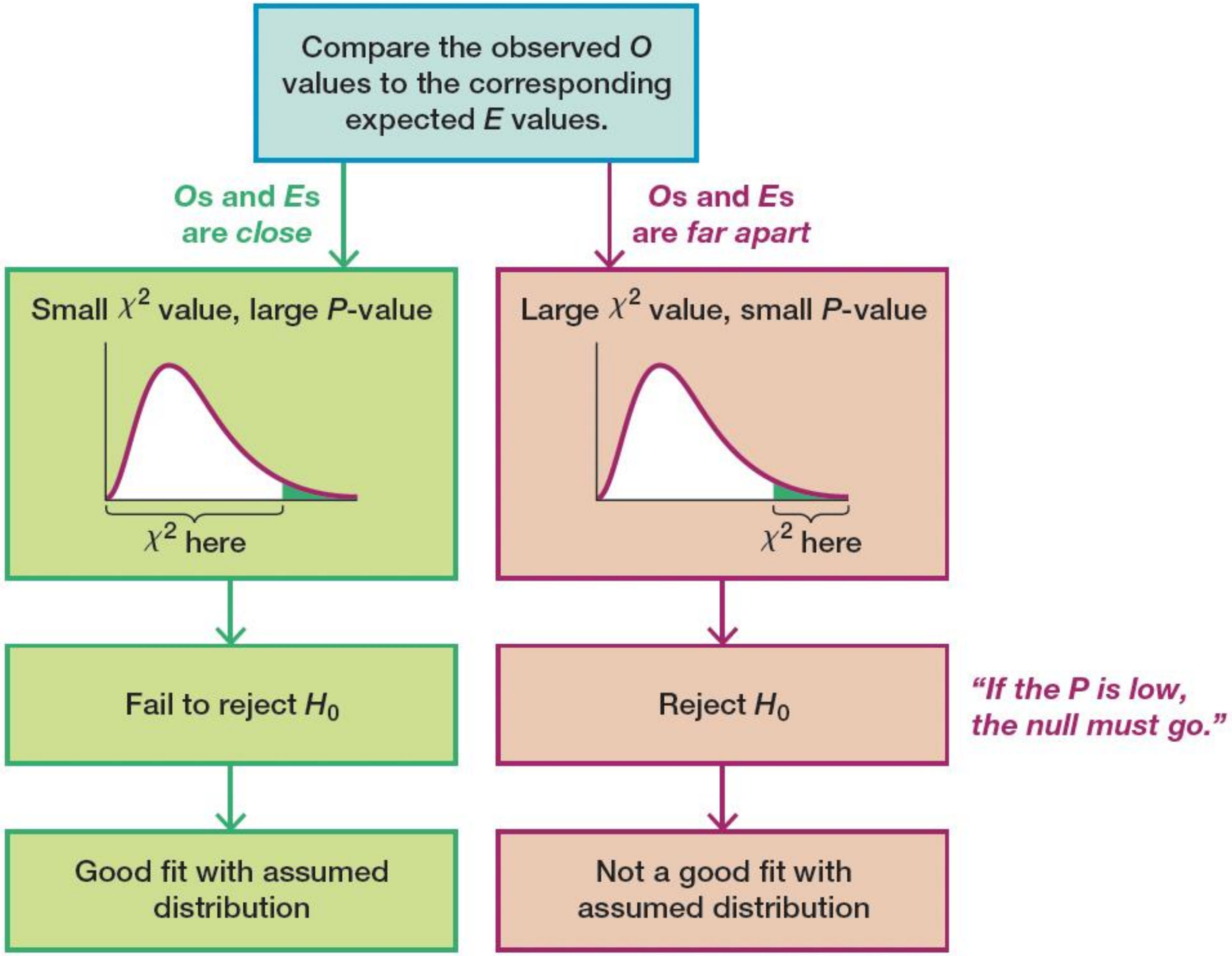


FIGURE 11-1 Relationships Among the χ^2 Test Statistic, P-Value, and Goodness-of-Fit

Resumo e Principais Cuidados

- **Aderência**: Uma variável, compara com uma teoria/distribuição conhecida.
- **Independência**: Duas variáveis, verifica se há associação entre elas.
- Restrições e pressupostos do teste:
  - Os dados devem ser contagens (frequências), não porcentagens ou proporções.
  - As observações devem ser independentes (a escolha de um indivíduo não influencia a de outro).
  - As frequências esperadas (E) não devem ser muito pequenas. Uma regra comum é que E deve ser
  - ≥5 para a maioria das células. Se isso não for atendido, outros testes como o Teste Exato de Fisher são mais apropriados.