

# Estatística Básica

Aula 13 (???)

Análise de dados categóricos

Pavel Dodonov

[pdodonov@gmail.com](mailto:pdodonov@gmail.com)

[anotherrecoblog.wordpress.com](http://anotherrecoblog.wordpress.com)

# Tipos de variáveis operacionais

Categórica  
(qualitativa)

Numérica  
(quantitativa)

Ordinal  
(semi-quantitativa)

# Tipos de variáveis operacionais

Categórica  
(qualitativa)

Numérica  
(quantitativa)

Ordinal  
(semi-quantitativa)

# Diferentes situações

Variáveis  
explanatórias  
categóricas

Variáveis  
explanatórias  
contínuas

Variável-  
resposta  
categórica

Variáveis explanatórias  
contínuas e categóricas


# Diferentes situações

Variáveis  
explanatórias  
categóricas

Tabela de contingência  
Qui-quadrado  
Teste exato de Fisher

Variáveis  
explanatórias  
contínuas

Variável-  
resposta  
categórica



Variáveis explanatórias  
contínuas e categóricas

# Diferentes situações

Variáveis  
explanatórias  
categóricas

**Regressão logística**  
**Regressão multinomial**  
**GLM – distribuição**  
**binomial**

Variáveis  
explanatórias  
contínuas

Variável-  
resposta  
categórica



Variáveis explanatórias  
contínuas e categóricas

# Diferentes situações

Variáveis  
explanatórias  
categóricas

Variáveis  
explanatórias  
contínuas

Variável-  
resposta  
categórica



**GLM – distribuição binomial**

Variáveis explanatórias  
contínuas e categóricas

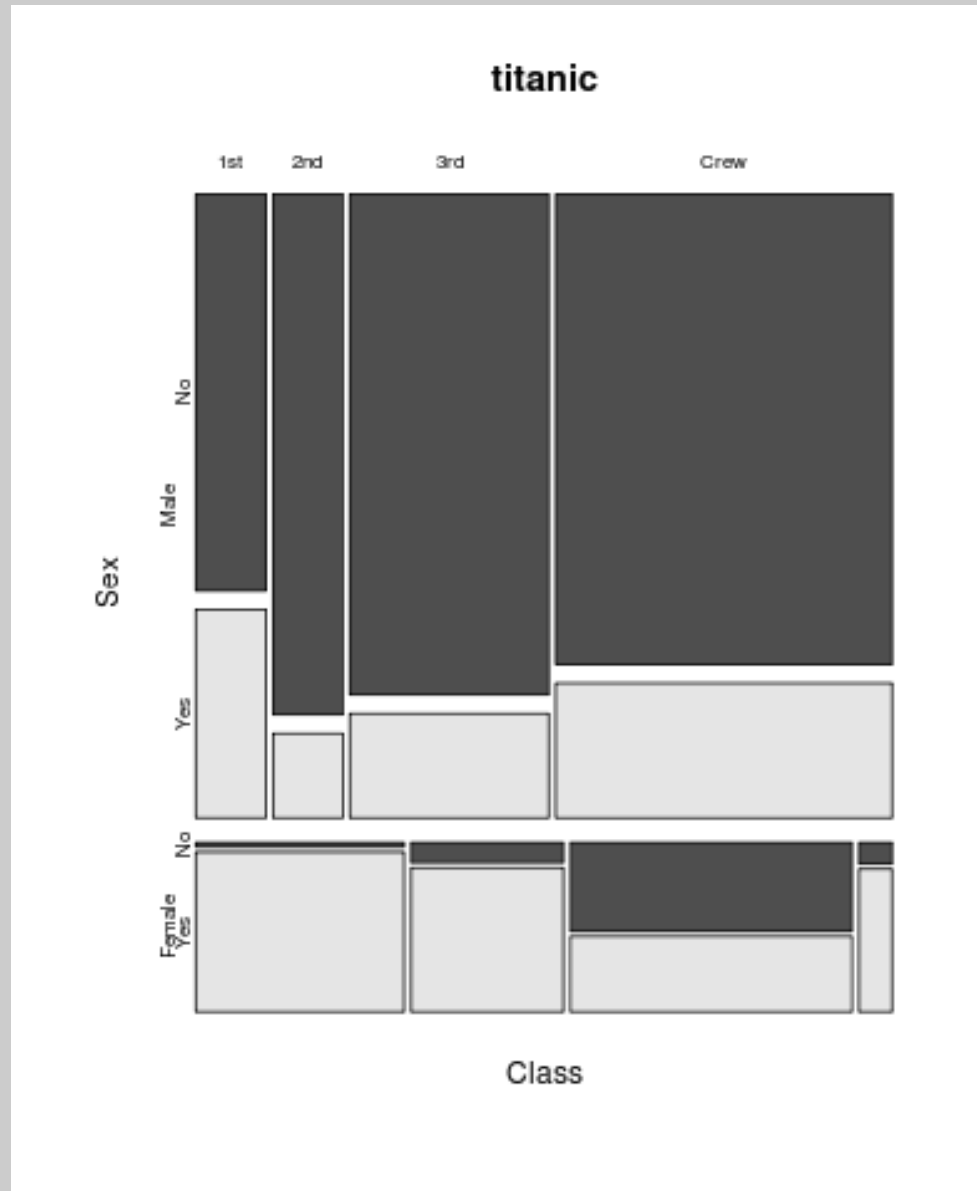
# Categórica VS Categórica

Exemplo: Sobreviventes do Titanic

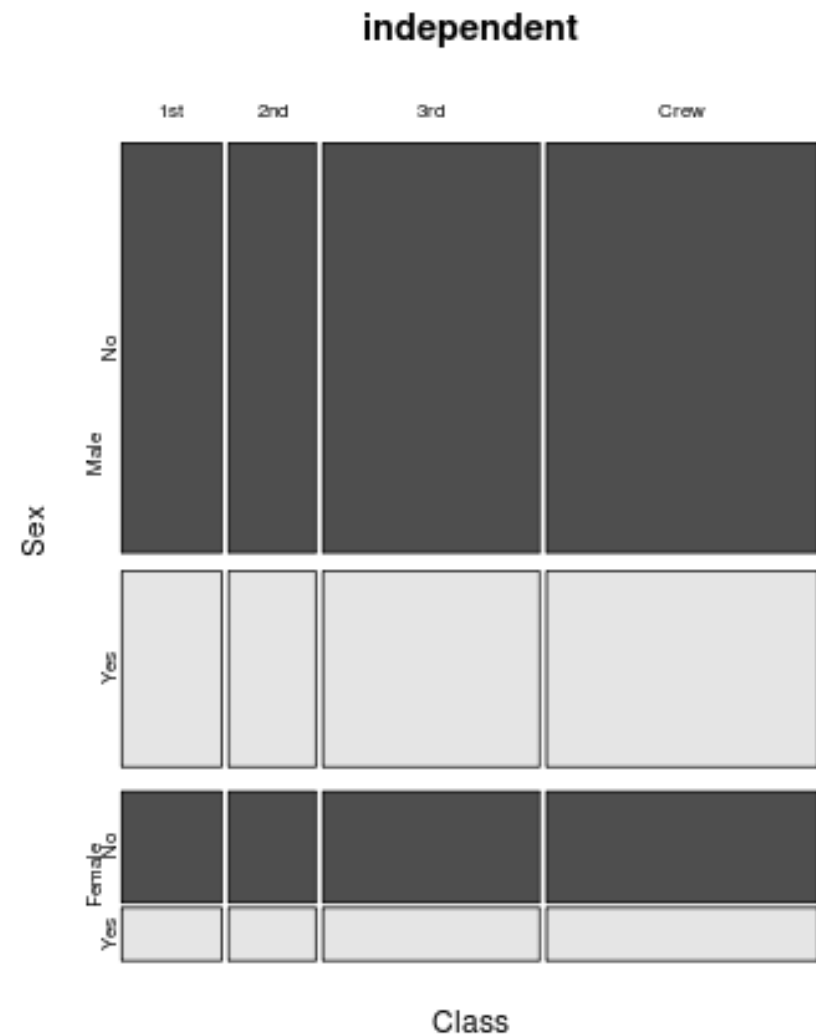
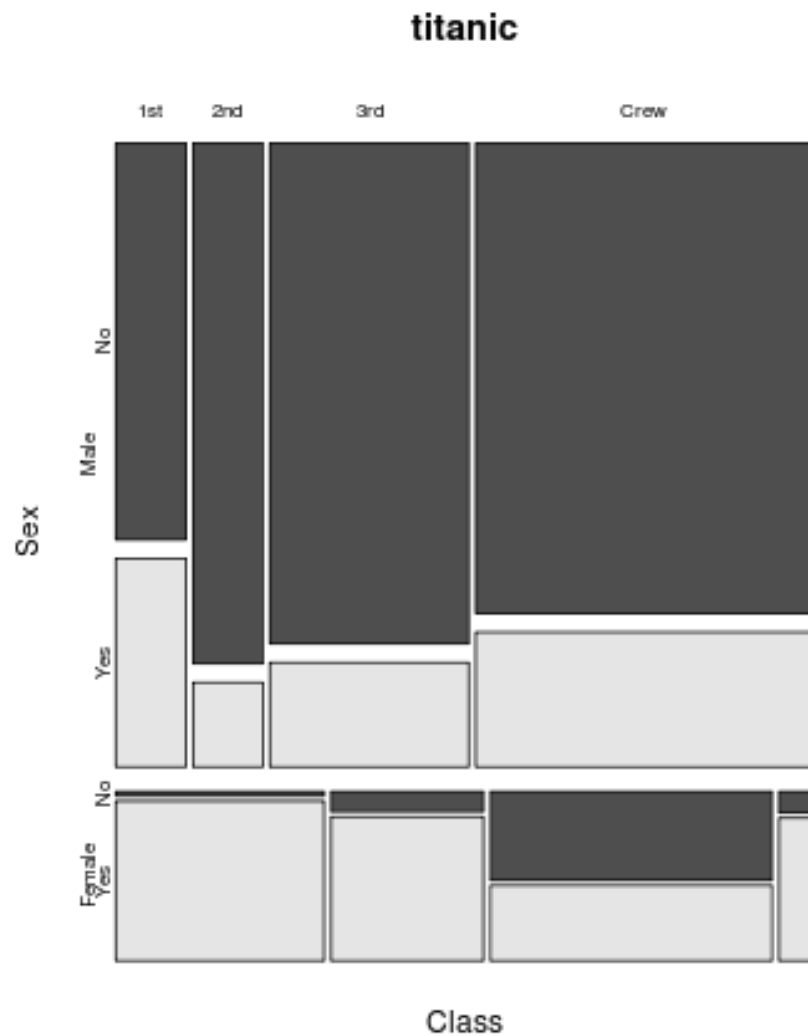
Gênero	Sobreviveu?	1a Classe	2a Classe	3a Classe	Tripulação
Masculino	Não	118	154	422	670
	Sim	62	25	88	192
Feminino	Não	4	13	106	3
	Sim	141	93	90	20



# Representação gráfica



# Comparando com mortalidade homogenea

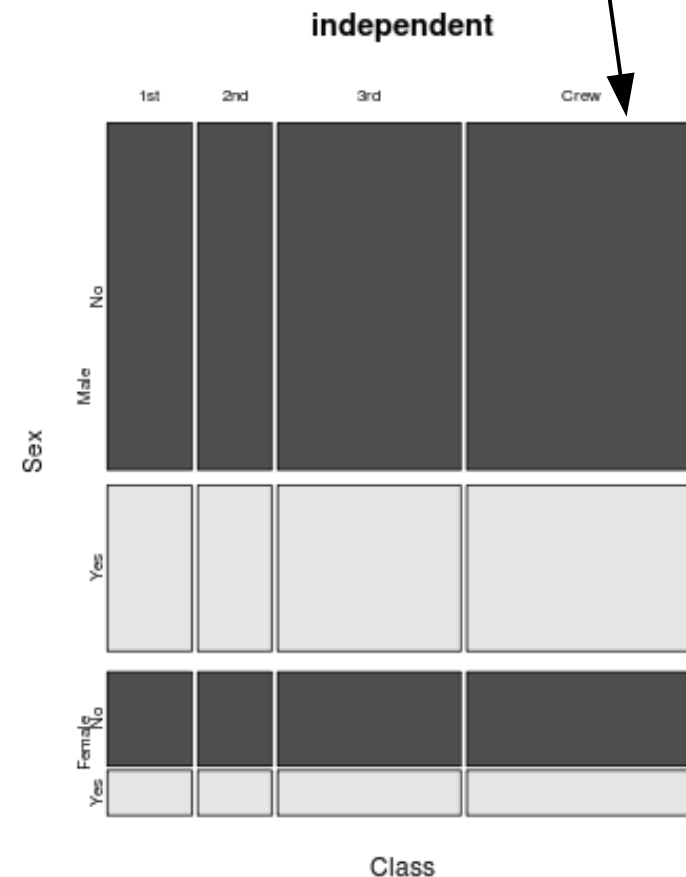
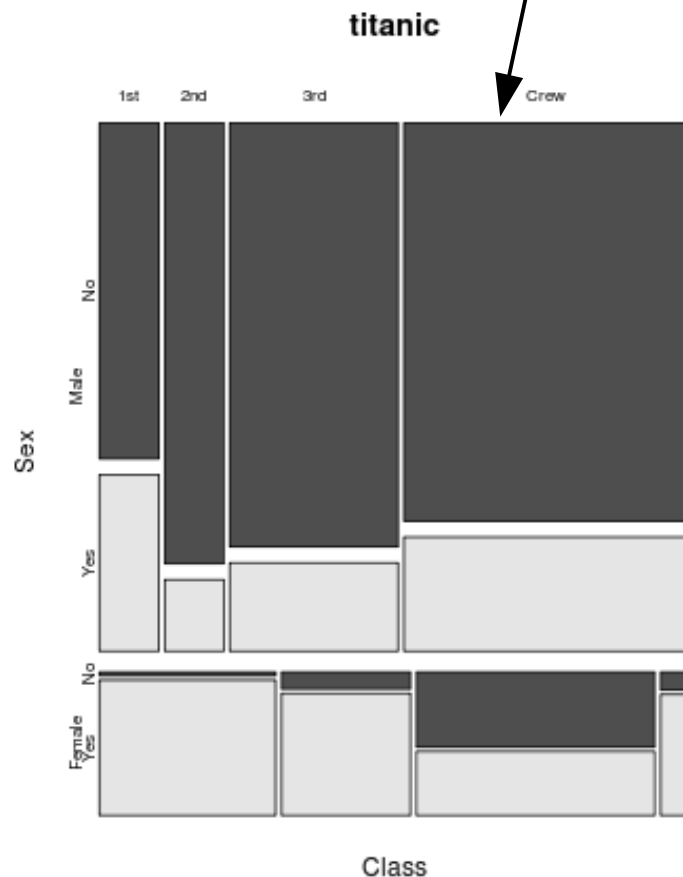


# Comparando...

- Parece que há uma **associação** entre gênero e classe com mortalidade.
  - **Associação**: análogo à **correlação**, mas para variáveis **categóricas**.

# Qui-quadrado

- Compara o **observado** com o **esperado** (sob a hipótese nula de não haver associação)



# Qui-quadrado

- Compara o **observado** com o **esperado** (sob a hipótese nula de não haver associação)

$$\chi^2 = \sum \frac{(\textit{Observado} - \textit{Esperado})^2}{\textit{Esperado}}$$

# Teste exato de Fisher

- Compara o **observado** com **todas as distribuições** possíveis das **amostras** entre os **grupos**



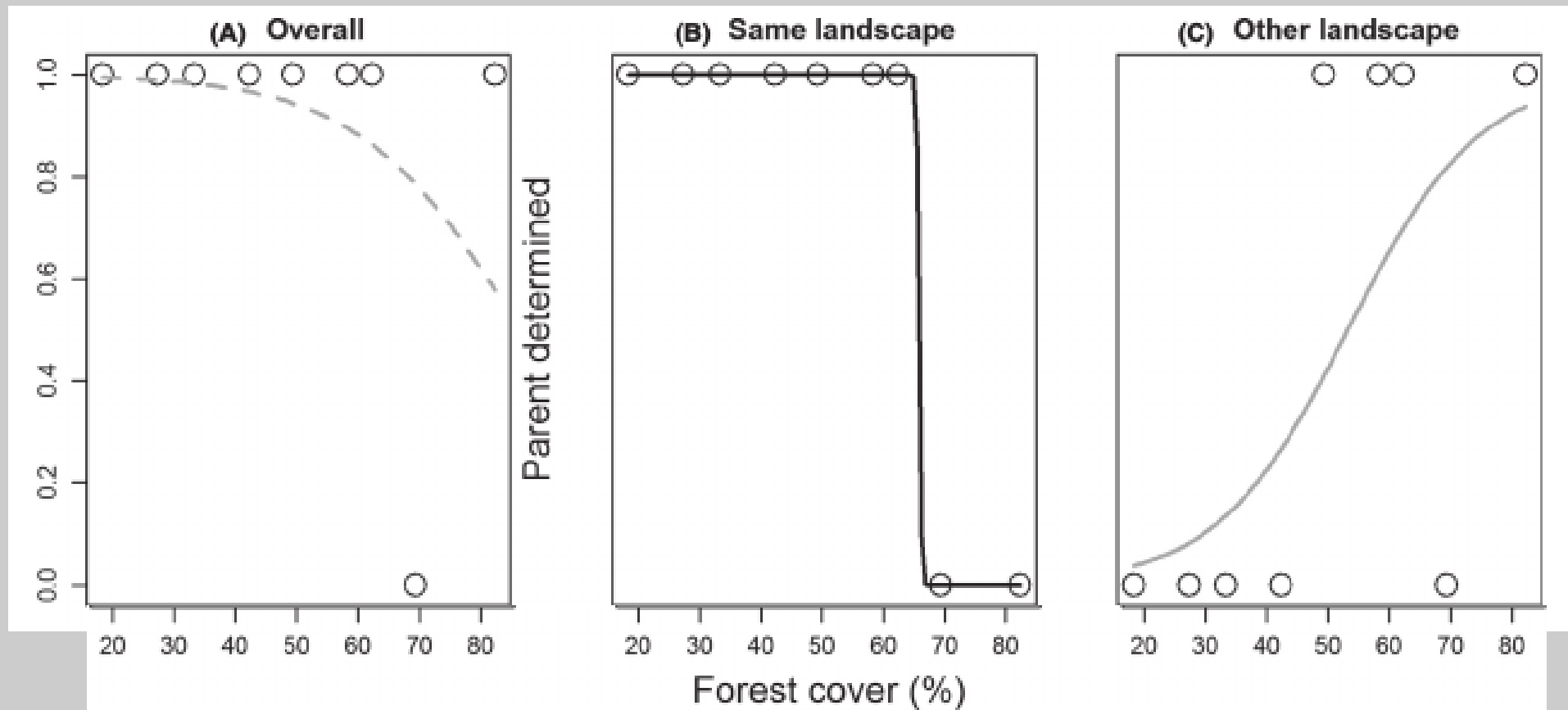
# Teste exato de Fisher

- Compara o **observado** com **todas as distribuições** possíveis das **amostras** entre os **grupos**
  - É **exato** porque o P-valor é calculado exatamente, não aproximado com base em uma distribuição assintótica.



# Categórica (binária) VS contínua

Exemplo: Desmatamento VS genética de palmito

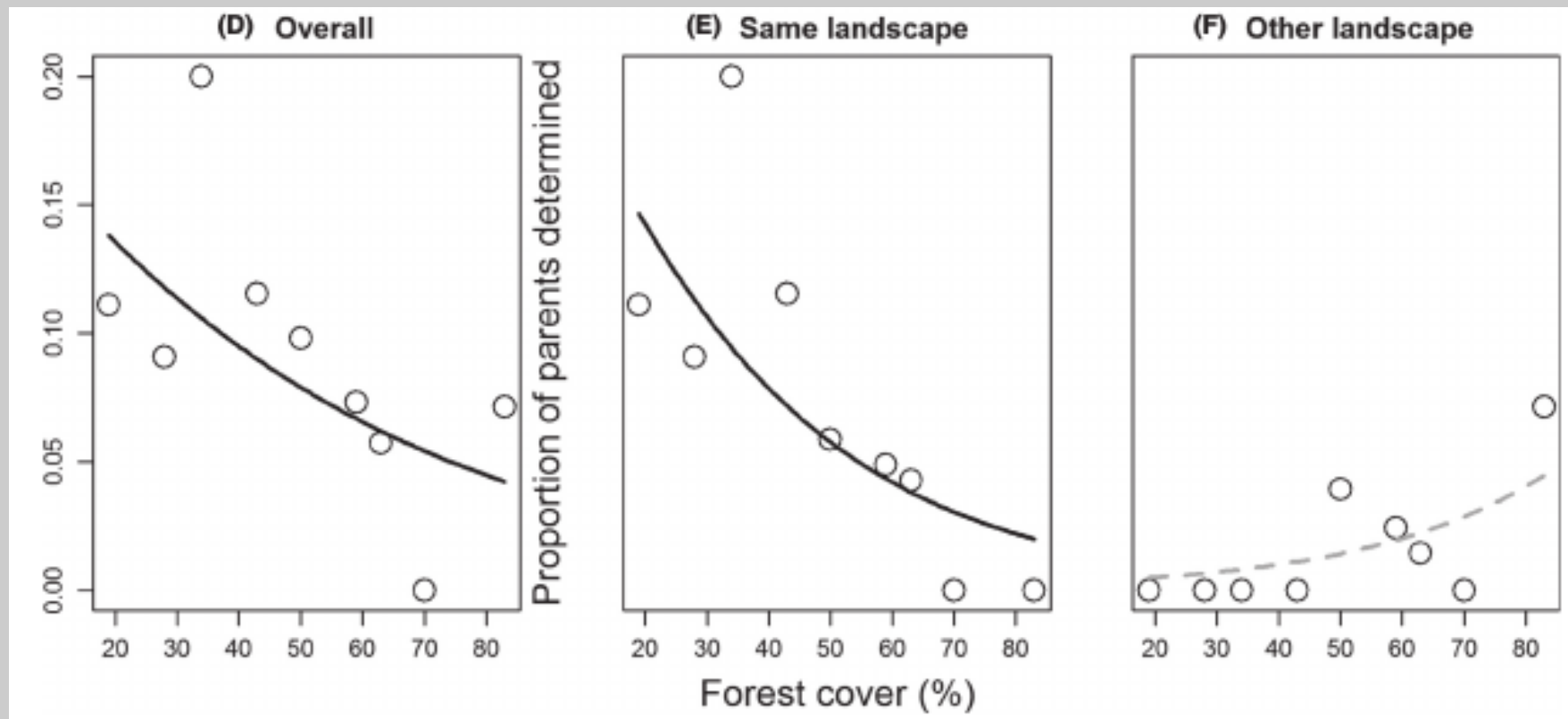


Presença/ausência de plântulas com adulto encontrado em algum lugar, na mesma paisagem ou em outra paisagem. Com maior cobertura florestal, há maior possibilidade do parental estar longe.



# Proporção VS contínua

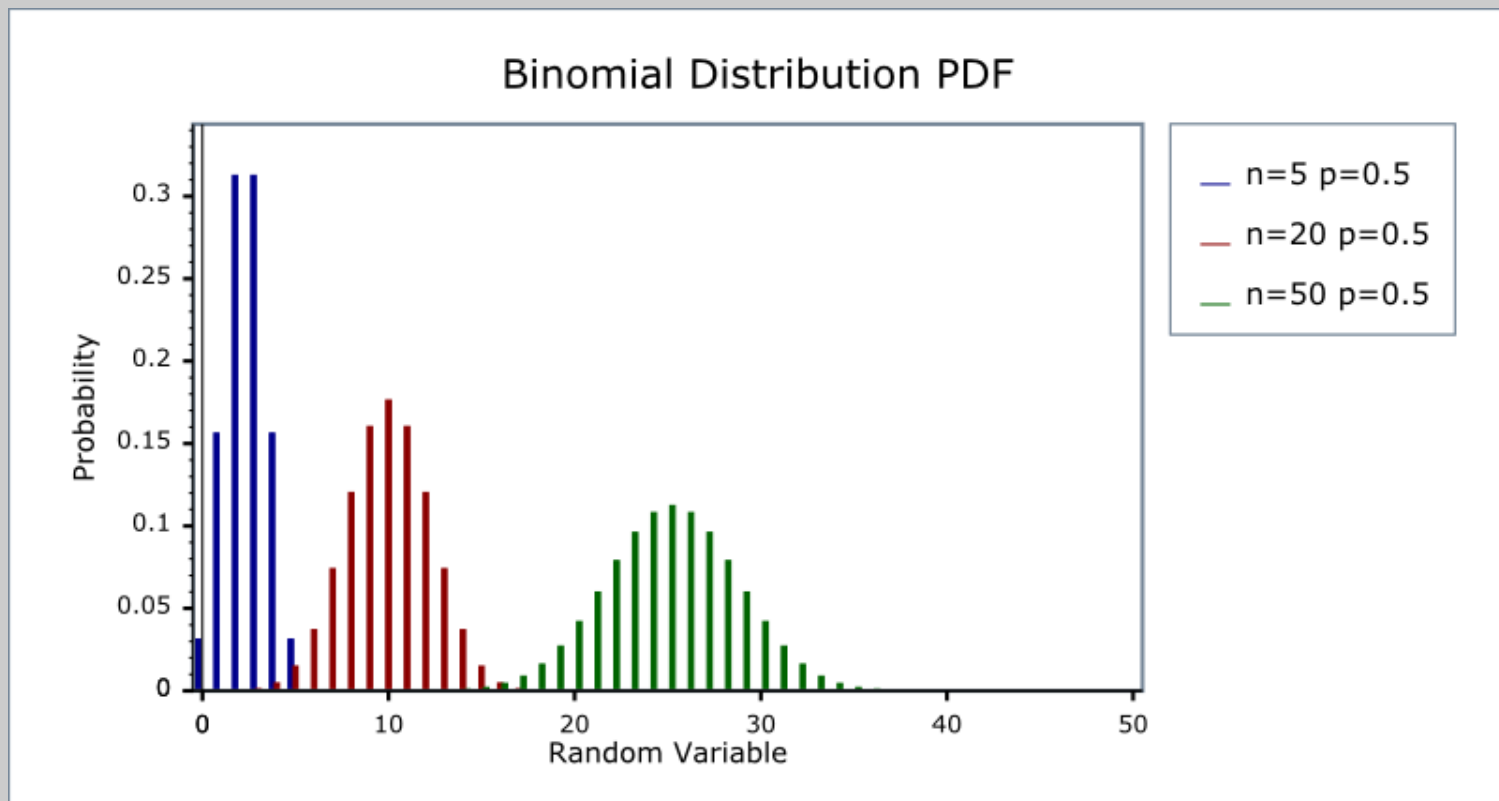
Exemplo: Desmatamento VS genética de palmito



Proporção de plântulas com adulto encontrado em algum lugar, na mesma paisagem ou em outra paisagem. Com maior cobertura florestal, há maior possibilidade do parental estar longe.

# Distribuição binomial

- Proporção de sucessos em um certo número de testes
  - Presença/ausência: sucesso ou falha em um único teste.



# Diferença importante

	Y contínuo, X categórico	Y categórico, X contínuo
Exemplo de pergunta	O valor médio (ou outro parâmetro) de Y difere entre níveis de X?	O X influencia a probabilidade de eu obter um sucesso (ou um certo nível) em Y?
Exemplo de gráfico	<i>Jitter plot</i> , <i>box plot</i> , gráfico de barras	Gráfico de dispersão (possivelmente com <i>jitter</i> no Y)
Exemplo de análise	Teste t de Student, ANOVA, Kruskal-Wallis, Kolmogorov-Smirnov	GLM com distribuição binomial