

Introdução à Regressão Linear Simples

Fernando Bispo, Jeff Caponero

Sumário

Introdução	2
Objetivo	2
Análise dos dados	2
Histograma	4
Dispersão	5

Table 1: Medidas Resumo para o sexo feminino.

	Min	Q1	Med	Média	Q3	Max	D. Padrão	CV
Comprimento Total	75,0	85	88,5	87,91	90,5	96,5	4,18	0,05
Largura Crânio	51,5	55	56,4	56,59	57,7	67,7	2,57	0,05

Table 2: Medidas Resumo para o sexo masculino

	Min	Q1	Med	Média	Q3	Max	D. Padrão	CV
Comprimento Total	76	84,0	86,0	86,51	89,5	96,0	4,34	0,05
Largura Crânio	50	54,8	56,3	57,09	59,0	68,6	3,45	0,06

Introdução

O presente estudo visa a aplicação das técnicas até o presente momento adquiridas para a implementação da Regressão Linear Simples nos dados contendo medidas morfológicas de 104 gambás das montanhas, que foram capturados através de armadilhas em sete localizações na Inglaterra. As variáveis contidas nesse arquivo são:

- sexo (**sex**);
- largura do crânio (**skullw**);
- comprimento total (**totlngth**).

Objetivo

O objetivo dessa análise responderá as seguintes questões:

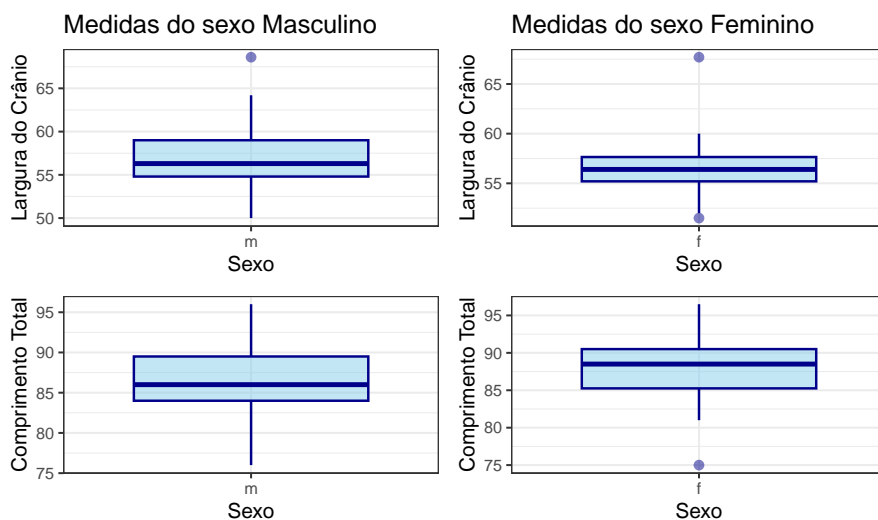
Análise dos dados

Composto por 3 características (variáveis) morfológicas dos gambás, duas destas são classificadas como aritméticas contínuas, sendo estas a largura do crânio (**skullw**) e o comprimento total (**totlngth**) e uma variável classificada como categórica ordinal, sendo esta sexo(**sex**).

Abaixo seguem as tabelas com as principais medidas resumo por sexo dos gambás.

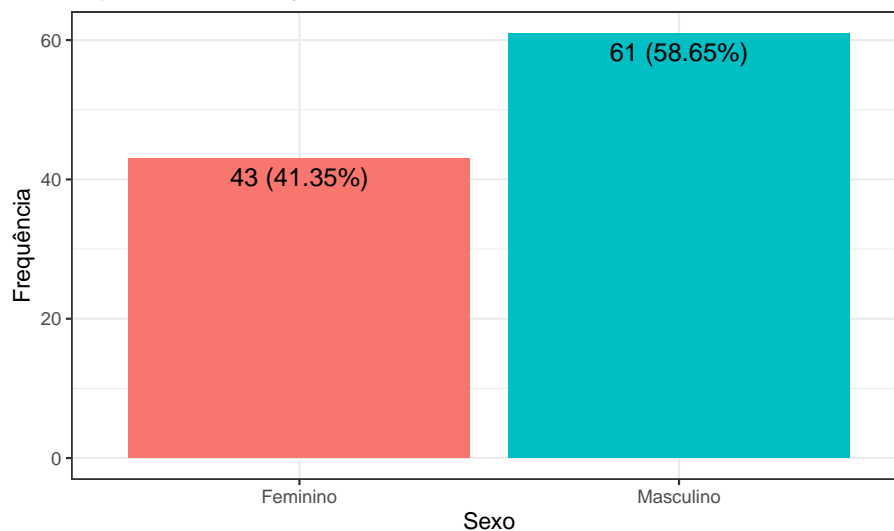
A fim de se possibilitar uma análise mais visual, as Figuras 1 e 2 trazem os gráficos de caixa (*BoxPlot*) com as características morfológicas dos gambas separadas por sexo.

Figura 1: Comparação das medidas morfológicas por sexo



A Figura 2 retrata a distribuição do sexo, possibilitando identificar o quantitativo de informações por sexo.

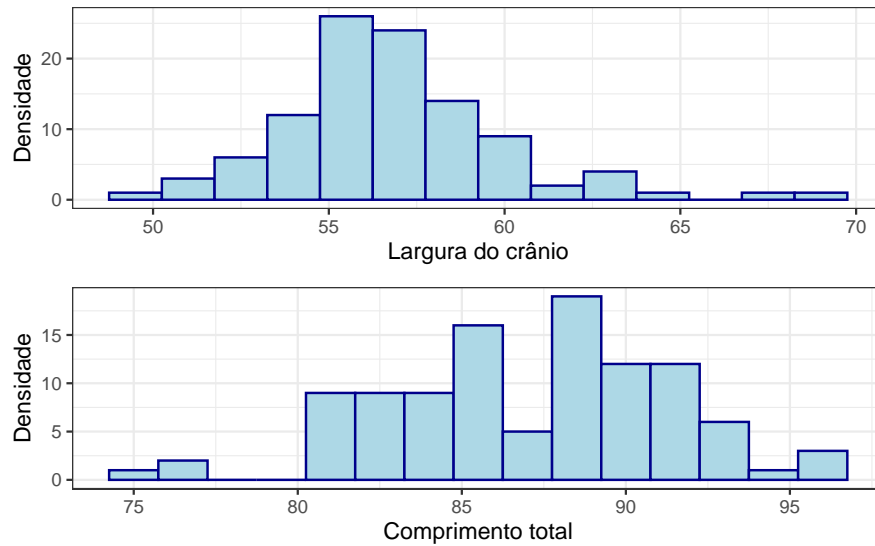
Figura 2: Distribuição da variável sexo



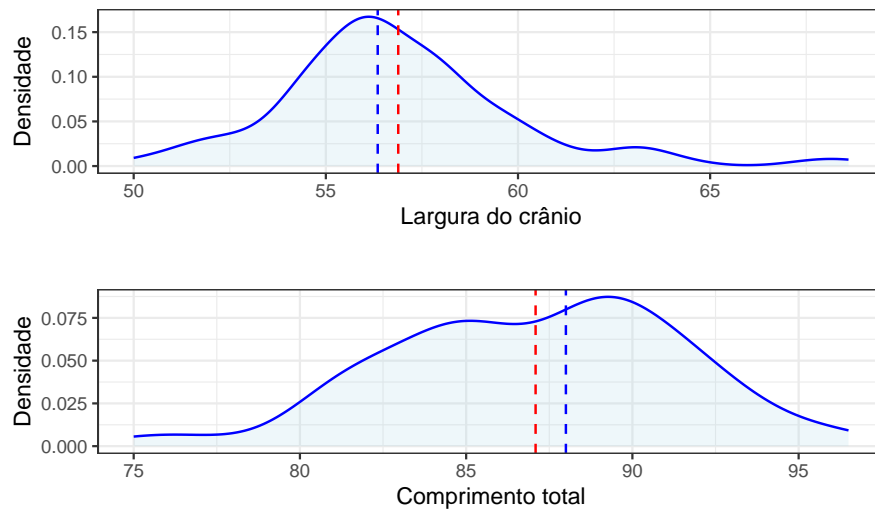
É possível identificar pontos atípicos (*outliers*) presentes nos gráficos presentes

na Figura 2.

Histograma



A fim de se identificar com maior clareza a distribuição dos dados das variáveis constantes nos histogramas seguem gráficos de densidade juntamente com as medias e medianas sobrepostas.



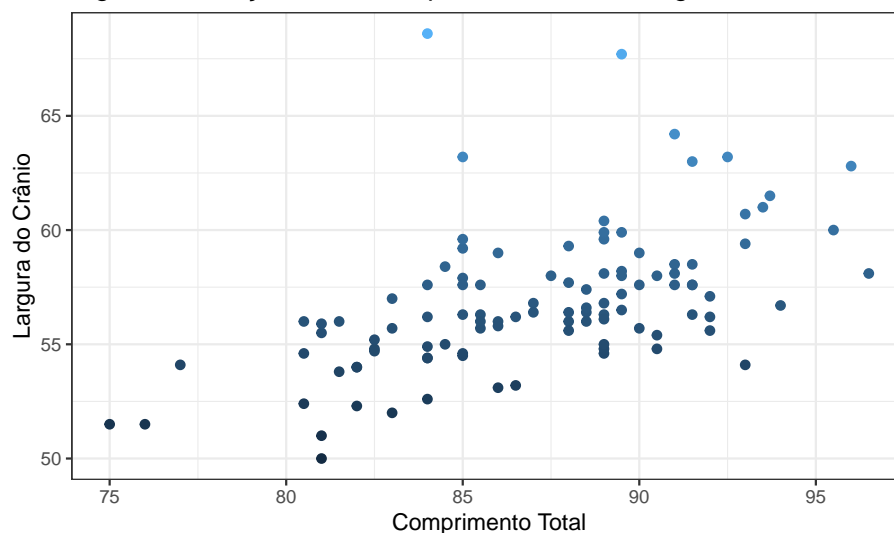
Dispersão

Com o intuito de comparar a relação entre as variáveis largura do crânio (**skullw**) e comprimento total (**totlngth**), foi construída a Figura 5.

É possível identificar uma relação linear positiva entre as características em análise, ao se calcular o coeficiente de correlação de Pearson $\hat{\rho}$ estimado obteve-se o valor de 0.5264 caracterizando uma relação **regular**.

Para testar a significância da correlação linear ao nível de significância de 5% foi realizado o teste de hipótese para correlação linear, em que a hipótese nula foi rejeitada, podendo assumir com 95% de confiança que há correlação linear entre as variáveis estudadas.

Figura 5: Relação entre Comprimento Total e Largura do Crânio



Através da Figura 5 é possível identificar com mais clareza pontos atípicos na relação entre as variáveis morfológicas, a fim de identificar o quão influente são estes pontos no comportamento geral dos dados, a Figura 6 foi construída sem estes pontos.