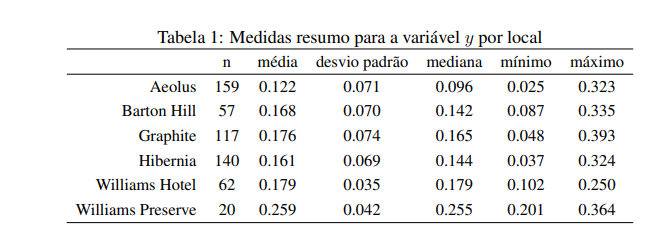
Exercício 1

(a) Classifique as variáveis y e local. Obtenha a média e os valores mínimo e máximo de y.



*Classificação:*

*Variável y (proporção de gordura corporal: quantitativa discreta*

*Local: qualitativa nominal*

*Média de y: 0.1591045*

*Valor máximo: 0.393*

*Valor mínimo: 0.025*

(b) Utilizando o coeficiente de variação, compare os grupos de morcegos dos diferentes locais quanto à variabilidade da variável y.

Coeficientes por grupo:

{'Aeolus': 0.582,

'Barton Hill': 0.417,

'Graphite': 0.42,

'Hibernia': 0.429,

'Williams Hotel': 0.196,

'Williams Preserve': 0.162}

Podemos observar que os grupos de morcegos variam conforme a variável y nesta ordem (de local que menos varia para o local que mais varia): Williams Preserve (16,2%), Williams Hotel (19,6%), Barton Hill (41,7%), Graphite (42,0%), Hibernia (42,9%), Aeolus (58,2%).

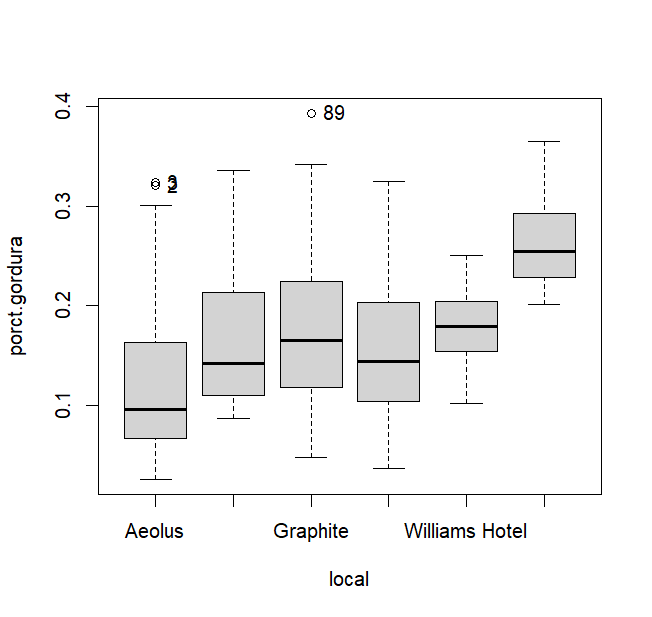
(c) Os valores do primeiro e terceiro quartil de y para os morcegos da região Aeolus são, respectivamente, 0.067 e 0.164. Com essas informações adicionais é possível construir o boxplot para a proporção de gordura corporal dos morcegos amostrados nessa região? Se sim, obtenha o boxplot. Se não, justifique.

𝐿𝑆 = 𝑄3 + 1,5(𝑄3 − 𝑄1)

𝐿𝐼 = 𝑄1 − 1,5(𝑄3 − 𝑄1)

São necessários para a construção de um boxplot os valores: do primeiro quartil, do segundo quartil, da mediana, dos valores máximos e mínimos (para representá-los como outliers ou máximos e mínimos típicos) e os máximos e mínimos típicos caso estes não sejam outliers. Como não possuímos as informações de máximos e mínimos típicos quando estes não são outliers, é possível construir um boxplot, porém ele ficará incompleto, uma vez que não contará com a linha do “máximo típico”, uma vez que só possuímos o valor máximo (que é superior ao limite superior)

(d) Obtenha os boxplots da variável y para os diferentes locais (utilize o R e o conjunto de dados disponível no arquivo Morcegos\_Cheng.). Comente.



Exercício 2

A Figura 1 apresenta o histograma da variável y.

(a) Obtenha os valores aproximados para a mediana e o terceiro quartil de y com base no histograma. Compare esses valores com os exatos, obtidos através do software R.

Considerando que a frequência relativa é igual a h/base, a tabela abaixo e o fato de que a mediana corresponde ao valor na posição 50% e o terceiro quartil o valor que deixa 75% dos valores abaixo dele, podemos assumir 0,15 e

| Cálculo da fr | Valor da fr | % acumulada |
| --- | --- | --- |
| 0,685\*0,05 | 3,43% | 3,43% |
| 4,631\*0,10 | 46,31% | 3,43%+46,31% = 49,74% |
| 4,252\*0,05 | 21,26% | 49,74% + 21,26% = 71% |
| 2,45\*(0,3-0,2) | 24,50% | 71% + 24,50% = 95,5% |
| 0,829\*(0,35-0,3) | 4,14% | 95,5% + 4,14% = 99,64% |
| 0,072\*(0,4-0,35) | 0,36% | 99,64% + 0,36% = 100% |

49,74% é bem próximo de 50%. Portanto, 0,15 é bem próximo da mediana aproximada

75% - 71% = 4%.

4% representa 16.33% de 24,50%.

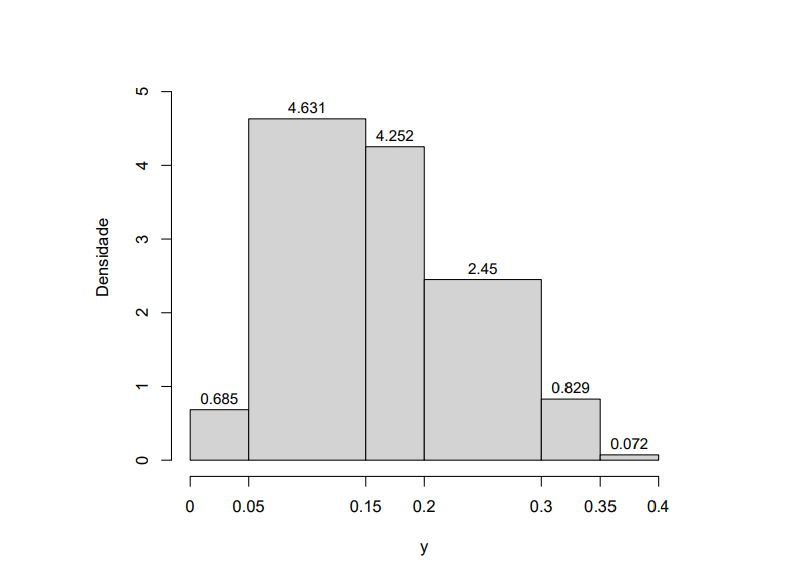
16.33% de (0,3-0,2) = 0.0163.

0,2 + 0.0163 = 0.216 =

Q3 aproximado

Mediana real = 0.1512305

Q3 real = 0.2071649



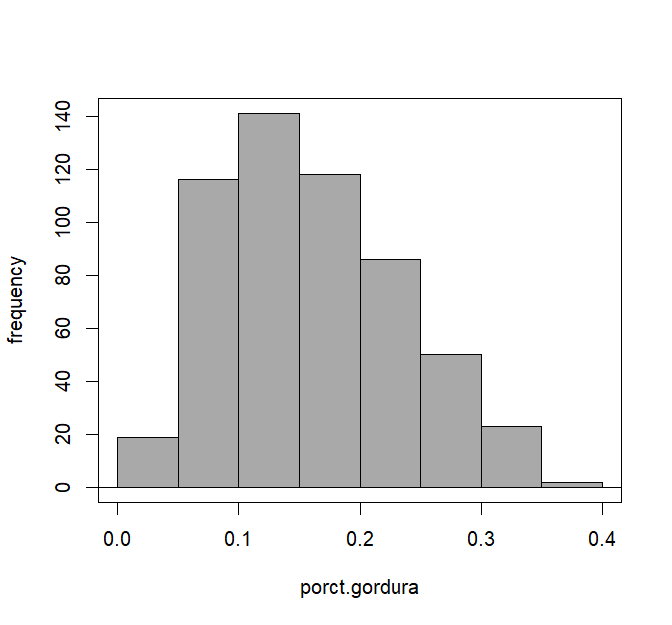
(b) Apresente a tabela de frequências que originou o histograma da Figura 1. Inclua as frequências relativas e absolutas para cada classe.

| Frequência (n) | Cálculo da fr | Valor da fr | % acumulada |
| --- | --- | --- | --- |
| 19 | 0,685\*0,05 | 3,43% | 3,43% |
| 257 | 4,631\*0,10 | 46,31% | 3,43%+46,31% = 49,74% |
| 118 | 4,252\*0,05 | 21,26% | 49,74% + 21,26% = 71% |
| 136 | 2,45\*(0,3-0,2) | 24,50% | 71% + 24,50% = 95,5% |
| 23 | 0,829\*(0,35-0,3) | 4,14% | 95,5% + 4,14% = 99,64% |
| 2 | 0,072\*(0,4-0,35) | 0,36% | 99,64% + 0,36% = 100% |

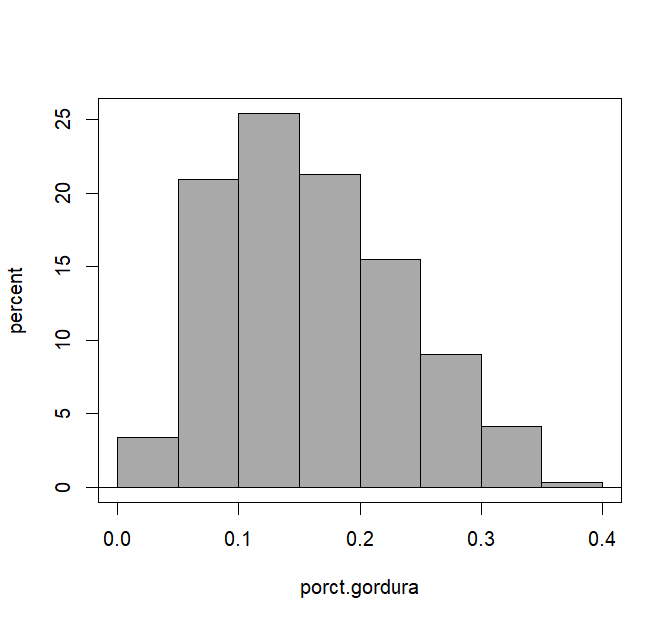
(c) Utilizando o software R, obtenha um histograma com classes iguais e o boxplot da variável y

Não sabemos qual variável deve ser adotada no eixo y, então anexamos todas:

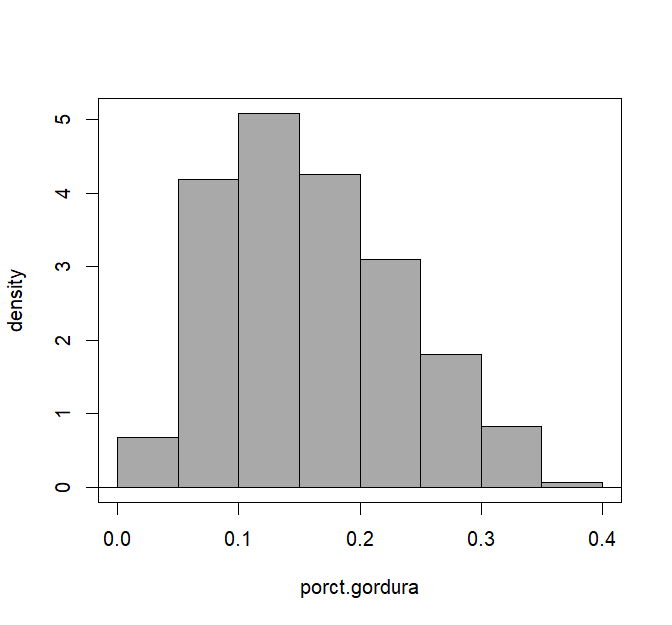
Frequência



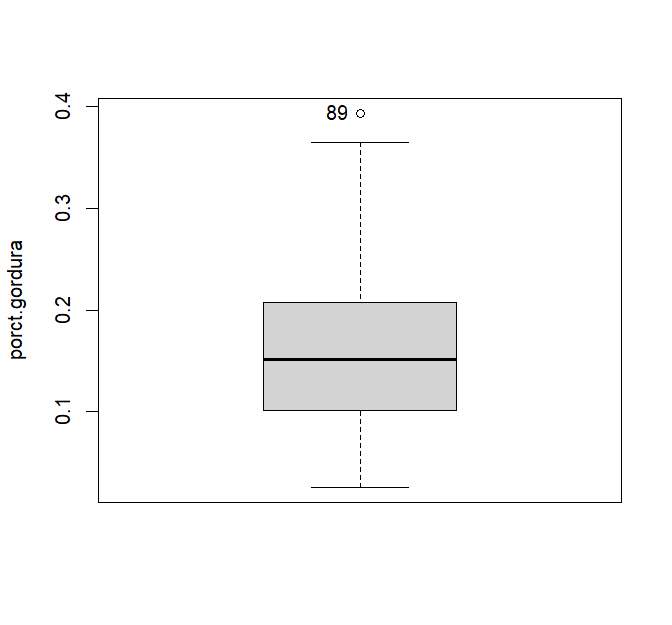
Percentual



Densidade

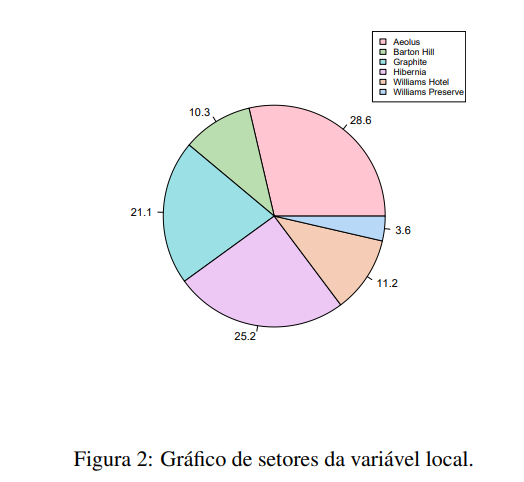


Boxplot variável y



Exercício 3

O gráfico de setores para a variável local é apresentado na Figura 2.

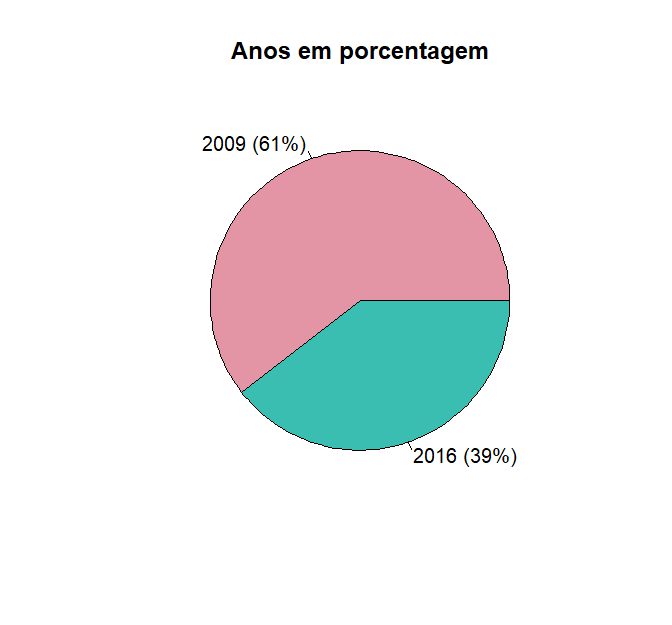


(a) Observando o gráfico da Figura 2, o que podemos dizer a respeito da variável local?

De acordo com o gráfico, podemos observar que as frequências relativas da variável dias em relação à variável local variam muito. Portanto, podemos inferir que a variável local pode ser uma variável que influencia o tempo de hibernação dos morcegos.

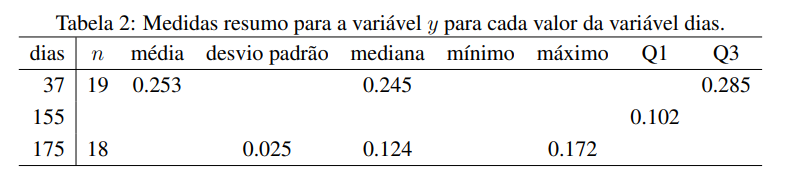
(b) Qual outro gráfico poderia ser utilizado para essa variável? Desenhe, manualmente, esse gráfico.

(c) Utilizando o R, obtenha um gráfico de setores para a variável ano.



Exercício 4

Nesta questão, vamos estudar a relação entre as variáveis y e dias, considerando os dados dos morcegos amostrados na região Barton Hill. A tabela a seguir apresenta algumas medidas descritivas da variável y por dias.



(a) Utilizando o R, complete a Tabela 2 com os valores faltantes.

| dias | n | média | dv | mediana | minimo | maximo | qi | q3 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 37 | 19 | 0.253 | 0.0478 | 0.245 | 0.152145 | 0.335481 | 0.217 | 0.289 |
| 155 | 20 | 0.125 | 0.0288 | 0.116 | 0.086570 | 0.179070 | 0.103 | 0.180 |
| 175 | 57 | 0.088 | 0.035 | 0.082 | 0.038252 | 0.171886 | 0.064 | 0.110 |

“Dados”

“Modificações de variáveis no conjunto dos dados”

“Converter variável numérica para fator”

“Estatísticas”

“Resumos”

“Resumos numéricos”

(b) Sem utilizar ferramentas computacionais, desenhe, em um mesmo gráfico, os boxplots da variável y para cada valor da variável dias. Comente sobre a relação entre as variáveis y e dias. Considerando o contexto do problema, essa relação é esperada?