



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE  
COIMBRA

# Relatório Trabalho Prático TP1

## Teoria de Informação - 2023/24

PL7



**João Vaz - 2022231087**  
**Francisco Pereira - 2022217071**  
**Francisco Caetano - 2022217054**

Pela visualização dos gráficos relativos ao MPG e os restantes aspetos conseguimos retirar algumas conclusões:

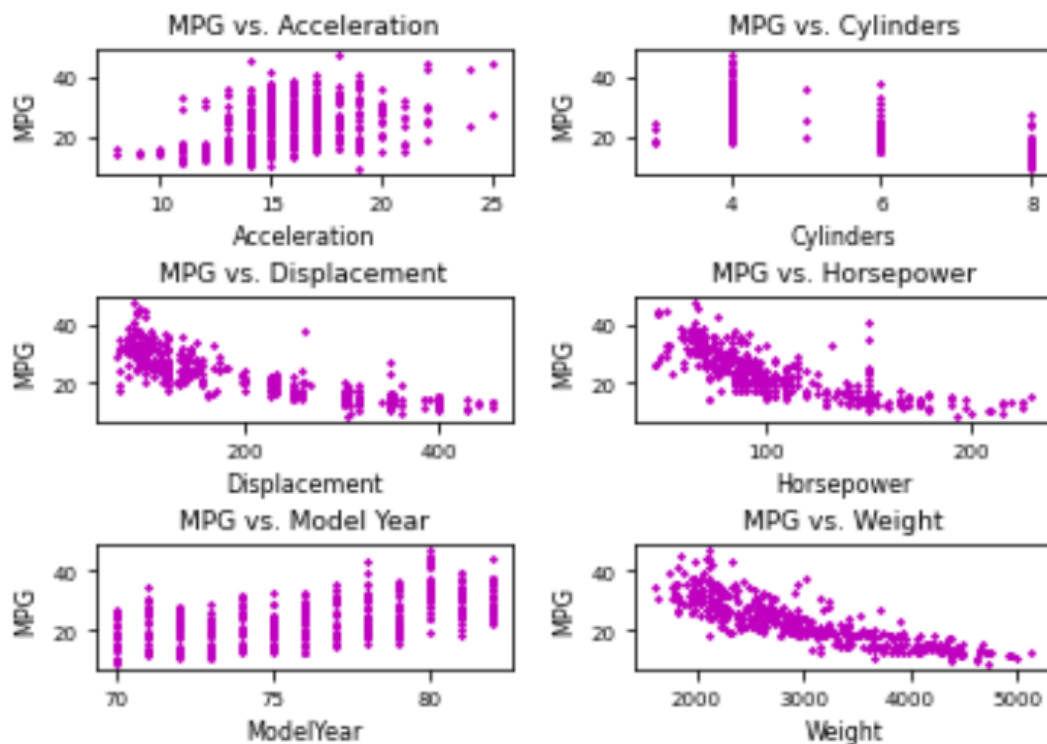


Figure 1

Entre os aspetos "Acceleration" e "ModelYear" é possível observar que à medida que estas aumentam o "MPG" também aumenta, ou seja, admitem uma relação de proporcionalidade direta entre si.

No entanto, com os restantes, "Displacement", "Cylinders", "Horsepower" e "Weight" verifica-se uma relação de proporcionalidade inversa, ou seja, quando estas aumentam o MPG diminui.

De seguida, realizou-se o cálculo do valor médio (teórico) de bits por símbolo, ou seja, o cálculo da entropia para cada símbolo de cada variável e bem como a total.

Resultados:

```
Acceleration: 3.50 bits por símbolo
Cylinders: 1.59 bits por símbolo
Displacement: 5.73 bits por símbolo
Horsepower: 5.84 bits por símbolo
ModelYear: 3.69 bits por símbolo
Weight: 8.39 bits por símbolo
MPG: 4.84 bits por símbolo
Dados Totais: 7.21 bits por símbolo
```

Perante este resultados, denotamos que existem diferentes valores para cada uma das variáveis, devido à variabilidade ou não de valores que existem.

Neste caso, verificamos que o "Weight", "Horsepower" e "Displacement" têm uma alta variedade de valores e por isso requerem mais bits para os representar, logo a entropia é mais elevada. Enquanto, por exemplo, a "Acceleration", apresenta uma entropia menor, pelo défice de variabilidade de valores.

Nos "Dados Totais", é apresentado o valor da entropia geral do conjunto de dados.

Os valores obtidos na codificação de Huffman, número médio de bits por símbolo para cada variável são bastante similares aos valores obtidos no ponto 7 ( bits por símbolo).

De facto, isto acontece pelo facto de os dois conceitos se relacionarem. A codificação de Huffman é ótima em termos de eficiência para conjuntos de símbolos cujas probabilidades sejam diferentes.

No entanto, o número médio de bits por símbolo não pode ultrapassar o valor da entropia (valor teórico, calculados no ponto 7) +1 e isso verifica-se nos nossos resultados.

A redução da variância dos comprimentos, especialmente em contextos estatísticos e de análise de dados, é algo bastante importante e tido muitas vezes como uma meta.

Existem diversas formas e ações que podemos tomar caso o nosso objetivo seja essa redução: certificarmos-nos de que os dados coletados sejam precisos e consistentes; erros de medição e coleta podem levar a uma variância maior e caso os dados sejam obtidos por meio de uma amostra, é importante que essa amostra represente de forma fidedigna a população ou o objeto que estamos a analisar.

Em muitos casos, o aumento do tamanho dessa amostra também reduz a variância dos comprimentos. Para além disso, em análises estatísticas, normalizar os dados, como por exemplo, subtrair a média e dividir pelo desvio padrão, pode reduzir a variância dos dados.

```
Acceleration: 3.50 bits por símbolo
Cylinders: 1.59 bits por símbolo
Displacement: 5.73 bits por símbolo
Horsepower: 5.84 bits por símbolo
ModelYear: 3.69 bits por símbolo
Weight: 8.39 bits por símbolo
MPG: 4.84 bits por símbolo
Dados Totais: 7.21 bits por símbolo
-----
Número médio de bits por símbolo da Acceleration (Huffman): 3.54
Variância dos comprimentos (Huffman): 0.81
Número médio de bits por símbolo da Cylinders (Huffman): 1.73
Variância dos comprimentos (Huffman): 0.71
Número médio de bits por símbolo da Displacement (Huffman): 5.76
Variância dos comprimentos (Huffman): 1.76
Número médio de bits por símbolo da Horsepower (Huffman): 5.87
Variância dos comprimentos (Huffman): 2.19
Número médio de bits por símbolo da ModelYear (Huffman): 3.73
Variância dos comprimentos (Huffman): 0.20
Número médio de bits por símbolo da Weight (Huffman): 8.46
Variância dos comprimentos (Huffman): 0.44
Número médio de bits por símbolo da MPG (Huffman): 4.87
Variância dos comprimentos (Huffman): 0.88
```

Com base nos valores dos coeficientes de correlação (correlação de Pearson) e das informações mútuas entre MPG e outras variáveis, podemos estabelecer as seguintes relações:

MPG vs. Acceleration: A correlação de Pearson é positiva mas não muito próximo de 1, o que indica uma relação positiva fraca com o MPG. Além disso, a informação mútua é relativamente alta, sendo a sua relação mais complexa do que uma correção linear, existindo dependências não lineares significativas.

MPG vs. Cylinders: A correlação de Pearson é negativa próxima de -1, o que indica uma relação negativa forte com o MPG. Além disso, a informação mútua é relativamente alta, sugerindo que a relação não é estritamente linear.

MPG vs. Displacement: A correlação de Pearson é negativa próxima de -1, o que indica correlação negativa forte com o MPG. A informação mútua é relativamente alta, sugerindo que não se limita a uma correlação linear

MPG vs. Horsepower: A correlação de Pearson é negativa próxima de -1, o que indica uma correlação negativa forte com a variável MPG. A informação mútua é relativamente alta, sugerindo que não se trata de uma relação linear.

MPG vs. ModelYear: A correlação de Pearson é positiva mas não muito próxima de 1, o que indica uma relação positiva fraca com o MPG. Além disso, a informação mútua é relativamente alta, sugerindo que não se trata de uma relação estritamente linear.

MPG vs. Weight: A correlação de Pearson é negativa muito próxima de -1, o que indica uma correlação negativa forte com o MPG. A informação mútua é relativamente, sugerindo uma relação mais complexa e não só estritamente linear.

```
Coeficiente de correlação (MPG / Acceleration): 0.4135853380757749
Coeficiente de correlação (MPG / Cylinders): -0.7760589899625314
Coeficiente de correlação (MPG / Displacement): -0.8047025675071012
Coeficiente de correlação (MPG / Horsepower): -0.7551351418256331
Coeficiente de correlação (MPG / ModelYear): 0.5872638852454327
Coeficiente de correlação (MPG / Weight): -0.8321486403128853
-----
Informação mútua entre MPG e Acceleration : 0.8720358370364727
Informação mútua entre MPG e Cylinders : 0.962178641086926
Informação mútua entre MPG e Displacement : 2.646895257681372
Informação mútua entre MPG e Horsepower : 2.56845025849273
Informação mútua entre MPG e ModelYear : 1.0294236623155246
Informação mútua entre MPG e Weight : 4.610309603870352
-----
```

Em resumo, ambas as medidas são consistentes na descrição das relações entre MPG e as variáveis, fornecendo informações complementares sobre a dependência estatística.

Relativamente, aos valores estimados dos MPG temos:

```
MAE para MPG_estimado: 2.5513513513513515
MAE para MPG_estimado_sem _aceleracao(menor MI): 3.2226044226044226
MAE para MPG_estimado_sem _peso(maior MI): 17.42850122850123
```

Os valores estimados não são exatamente iguais aos reais, sendo que o MAE é igual a 2.55.

Os valores estimados sem o termo da aceleração(menor MI) ainda se afastam mais dos reais sendo o seu MAE igual a 3.22

Os valores sem o termo do weight(maior MI) afastam-se completamente dos reais visto e o seu MAE é igual a 17.43.