Questão 1.

	Estimativa	Erro padrão	Valor t	Pr(>|t|)
Intercepto	2,939	0,3119	9,42	< 0,0001
TV	0,046	0,0014	32,81	< 0,0001
rádio	0,189	0,0086	21,89	<0,0001
jornal	-0,001	0,0059	-0,18	0,8599

- Testando β_0 : Temos $H_0 : \beta_0 = 0$ ou $H_1 : \beta_0 \neq 0$, estamos testando se o modelo ajustado possui ou não o intercepto. Vemos que existe evidência para a rejeição de H_0, indicando assim que o intercepto é significativo para o modelo. Podemos concluir que mesmo sem investir em publicidade nos três veículos de comunicação ainda se tem um número médio de vendas base.
- Testando β_1 : Temos $H_0 : \beta_1 = 0$ ou $H_1 : \beta_1 \neq 0$, estamos testando se a variável TV é significativa para modelo. Vemos que existe evidência para a rejeição de H_0, indicando assim que a variável TV é significativa para o modelo.
- Testando β_2 : Temos $H_0 : \beta_2 = 0$ ou $H_1 : \beta_2 \neq 0$, estamos testando se a variável rádio é significativa para modelo. Vemos que existe evidência para a rejeição de H_0, indicando assim que a variável rádio é significativa para o modelo.
- Testando β_3 : Temos $H_0 : \beta_3 = 0$ ou $H_1 : \beta_3 \neq 0$, estamos testando se a variável jornal é significativa para modelo. Vemos que existe evidência para a não rejeição de H_0, indicando assim que a variável jornal não é significativa para o modelo. Portanto, a publicidade em jornal não é significativa para as vendas.

Questão 2.

a) Para o modelo $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 GPA + \widehat{\beta}_2 IQ + \widehat{\beta}_3\text{Gênero} + \widehat{\beta}_4 GPA \times IQ + \widehat{\beta}_5 GPA \times \text{Gênero}$ temos que o item correto é,

Para um valor de IQ e GPA, homens ganham em média mais do que as mulheres, desde que GPA seja grande o suficiente.

Para verificar isso, considere que para homens,

$$\begin{aligned}\widehat{Y}_H &= \widehat{\beta}_0 + \widehat{\beta}_1 GPA + \widehat{\beta}_2 IQ + \widehat{\beta}_4 GPA \times IQ \\ &= 55 + 20GPA + 0,07IQ + 0,01GPA \times IQ\end{aligned}$$

e para as mulheres o modelo é

$$\begin{aligned}\widehat{Y}_M &= \widehat{\beta}_0 + \widehat{\beta}_1 GPA + \widehat{\beta}_2 IQ + \widehat{\beta}_3 + \widehat{\beta}_4 GPA \times IQ + \widehat{\beta}_5 GPA \times 1 \\ &= \underbrace{\widehat{\beta}_0 + \widehat{\beta}_1 GPA + \widehat{\beta}_2 IQ + \widehat{\beta}_4 GPA \times IQ}_{\widehat{Y}_H} + \widehat{\beta}_3 + \widehat{\beta}_5 GPA \times 1 \\ &= \widehat{Y}_H + 35 - 10GPA \times 1\end{aligned}$$

Assim, para valores fixados de IQ e GPA, notamos que para um GPA(maior que 3,5) muito grande os homens em média ganham mais do que as mulheres.

b) O valor predito para uma mulher com IQ =110 e GPA = 4 é 137,1 mil dólares.

c) Falso, para verificar se existe interação ou não entre IQ e GPA, é necessário testar $H_0 : \widehat{\beta}_4 = 0$ e não rejeitar H_0.

Questão 3.

Temos que $\hat{\beta} = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ e

$$\begin{aligned} y_N &= \hat{\beta} x_N \\ &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x_N \\ &= \sum_{i=1}^{n} \frac{x_i x_N}{\sum_{i=1}^{n} x_i^2} y_i \\ &= \sum_{i=1}^{n} a_i y_i \end{aligned}$$

em que $a_i = \frac{x_i x_N}{\sum_{i=1}^{n} x_i^2}$

Questão 4. O modelo linear simples ajustado é $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$, com $\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x}$ e $\widehat{\beta} = \frac{S_{xy}}{S_{xx}}$. Assim, para saber se um ponto (z, w) está na reta, basta verificar se $f(z) = \alpha + \beta z = w$, ou seja,

$$\begin{aligned} \widehat{y}_i(\bar{x}) &= \widehat{\alpha} + \widehat{\beta}\bar{x} \\ &= \bar{y} - \widehat{\beta}\bar{x} + \widehat{\beta}\bar{x} \\ &= \bar{y}. \end{aligned}$$

Portanto, o ponto $(\bar{x}, \bar{y})$ pertence a reta.

Questão 5. Considerando $\bar{x} = \bar{y} = 0$ teremos que o coeficiente de regressão linear é dado por $r = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2)}}$ e

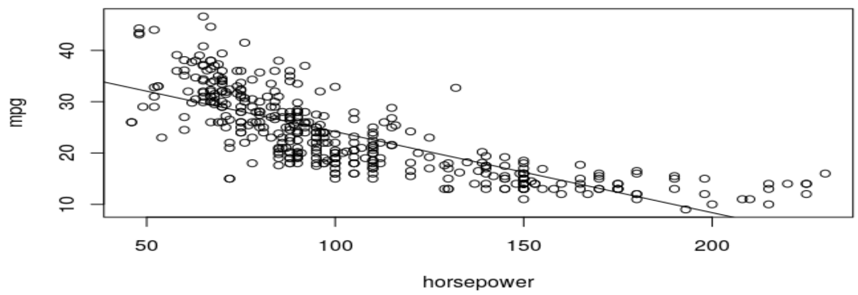
$$r^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2)} = \frac{S_{xy}^2}{S_{xx} S_{yy}} = R^2.$$

Questão 6.

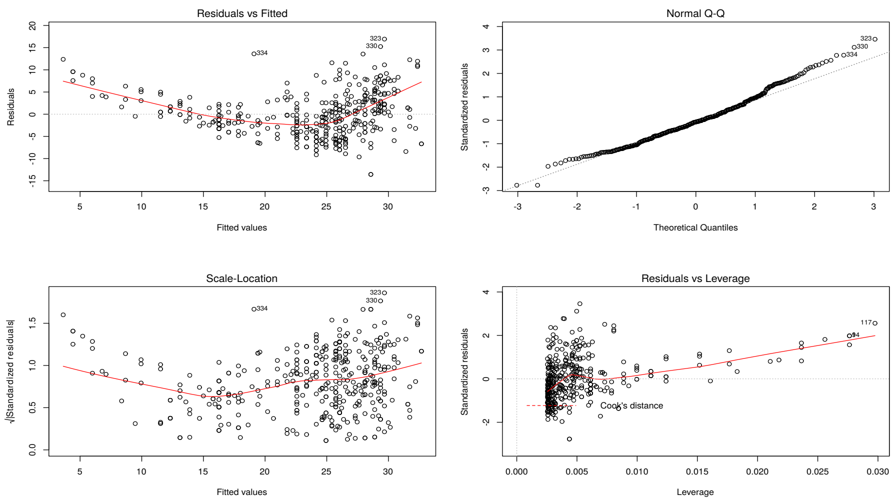
a) Ajustando o modelo $mpg \sim horsepower$ vemos que existe relação entre essas duas variáveis e que a força dessa relação é negativa e igual a -77,8%. Com o modelo ajustado, a predição para mpg quando $horsepower = 98$ é 24,47 e seu intervalo predito de confiança é [23,97 ; 24,96].

	Estimativa	Erro padrão	Valor t	Pr(>|t|)
(Intercept)	39,9359	0,7175	55,66	$< 0,001$
horsepower	-0,1578	0,0064	-24,49	$< 0,001$

b) Apresentamos o gráfico de dispersão com a reta estimada.



c) O gráfico de diagnóstico dos resíduos.

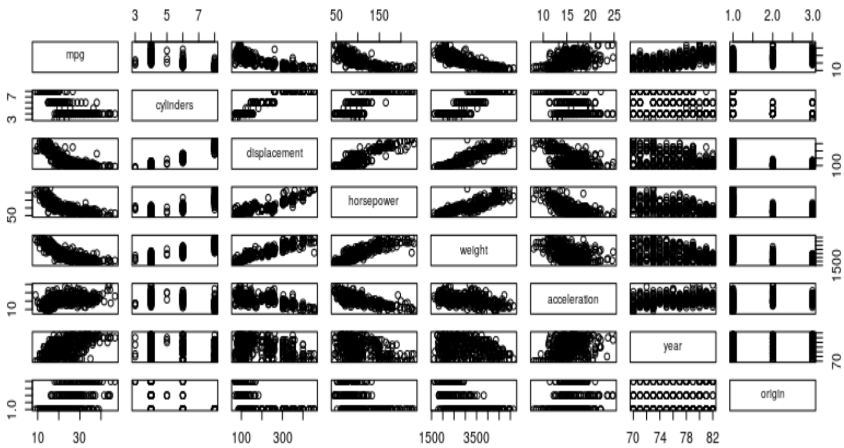


Nos gráficos de diagnóstico vemos indicativos contra as suposições de linearidade, normalidade e homoscedasticidade dos resíduos. Olhando o gráfico *Residuals vs Leverage* notamos a presença de possíveis pontos de influência.

```
require(ISLR)
attach(Auto)
cor(mpg,horsepower)
fit <- lm(mpg ~ horsepower, data = Auto)
summary(fit)
predict(fit, data.frame( horsepower = 98), level = 0.95,interval ="confidence")
plot(horsepower, mpg ); abline(fit)
par(mfrow = c(2,2))
plot(fit)
```

Questão 7.

a) A matriz scatterplot dos dados Auto sem a variável names.



b) Matriz de correlações

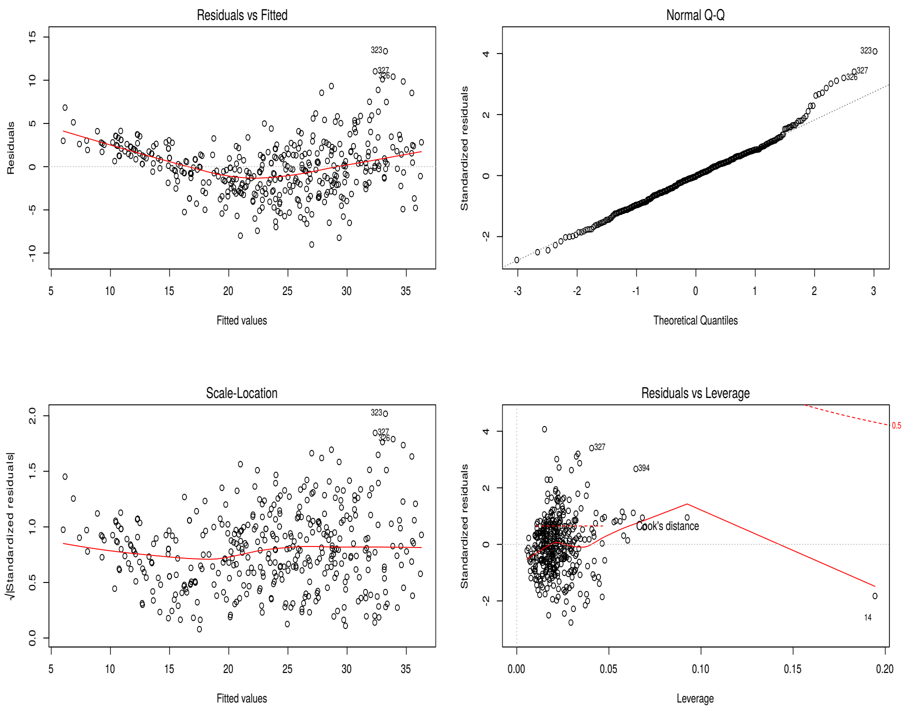
	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
mpg	1.00	-0.78	-0.81	-0.78	-0.83	0.42	0.58	0.57
cylinders	-0.78	1.00	0.95	0.84	0.90	-0.50	-0.35	-0.57
displacement	-0.81	0.95	1.00	0.90	0.93	-0.54	-0.37	-0.61
horsepower	-0.78	0.84	0.90	1.00	0.86	-0.69	-0.42	-0.46
weight	-0.83	0.90	0.93	0.86	1.00	-0.42	-0.31	-0.59
acceleration	0.42	-0.50	-0.54	-0.69	-0.42	1.00	0.29	0.21
year	0.58	-0.35	-0.37	-0.42	-0.31	0.29	1.00	0.18
origin	0.57	-0.57	-0.61	-0.46	-0.59	0.21	0.18	1.00

c) Modelo ajustado:

	Estimativa	Erro padrão	Valor t	Pr(>|t|)
(Intercept)	-17,95	4,68	-3,84	$< 0,001$
cylinders	-0,49	0,32	-1,52	0,13
displacement	0,02	0,01	3,13	$< 0,001$
horsepower	-0,02	0,01	-1,33	0,19
weight	-0,01	0,00	-10,24	$< 0,001$
acceleration	0,08	0,10	0,81	0,42
year	0,78	0,05	15,01	$< 0,001$
factor(origin)2	2,63	0,57	4,64	$< 0,001$
factor(origin)3	2,85	0,55	5,16	$< 0,001$

No modelo ajustado notamos que as variáveis $cylinders, horsepower$ e $acceleration$ são não significativas para o modelo ao nível de significância de 10%. Notamos também que o modelo explica 82% da variabilidade dos dados.

d)



Nos gráficos de diagnóstico vemos indicativos contra as suposições de linearidade e normalidade. No entanto, podemos considerar a suposição de homoscedasticidade. Olhando o gráfico *Residuals vs Leverage* notamos a presença de possíveis pontos de influência.

e) As interações entre cylinders e acceleration, displacement e weight, displacement e year, horsepower e year, acceleration e year, acceleration e factor(origin)2, acceleration e factor(origin)3, year e factor(origin)2, year e factor(origin)3 foram significativas para o modelo ao nível de significância de 10%.

	Estimativa	Erro padrão	Valor t	Pr(>|t|)
(Intercept)	44,0119	51,4674	0,86	0,3930
cylinders	3,3016	8,1865	0,40	0,6870
displacement	-0,3529	0,1974	-1,79	0,0746
horsepower	0,5312	0,3390	1,57	0,1180
weight	-0,0033	0,0182	-0,18	0,8580
acceleration	-6,0483	2,1466	-2,82	0,0051
year	0,4833	0,5923	0,82	0,4151
factor(origin)2	-35,1651	12,6020	-2,79	0,0055
factor(origin)3	-37,6464	14,2613	-2,64	0,0087
cylinders:displacement	-0,0063	0,0071	-0,89	0,3747
cylinders:horsepower	0,0145	0,0246	0,59	0,5551
cylinders:weight	0,0006	0,0009	0,63	0,5287
cylinders:acceleration	0,3658	0,1671	2,19	0,0293
cylinders:year	-0,1447	0,0965	-1,50	0,1348
cylinders:factor(origin)2	-0,7210	1,0883	-0,66	0,5081
cylinders:factor(origin)3	1,2256	1,0070	1,22	0,2244
displacement:horsepower	-0,0001	0,0003	-0,19	0,8502
displacement:weight	0,0000	0,0000	1,83	0,0684
displacement:acceleration	-0,0025	0,0034	-0,76	0,4484
displacement:year	0,0045	0,0024	1,86	0,0638
displacement:factor(origin)2	-0,0336	0,0422	-0,80	0,4259
displacement:factor(origin)3	0,0538	0,0414	1,30	0,1955
horsepower:weight	-0,0000	0,0000	-1,15	0,2497
horsepower:acceleration	-0,0034	0,0039	-0,88	0,3821
horsepower:year	-0,0064	0,0039	-1,65	0,0995
horsepower:factor(origin)2	-0,0049	0,0506	-0,10	0,9234
horsepower:factor(origin)3	0,0229	0,0625	0,37	0,7145
weight:acceleration	-0,0001	0,0002	-0,29	0,7741
weight:year	-0,0001	0,0002	-0,37	0,7122
weight:factor(origin)2	0,0023	0,0027	0,85	0,3970
weight:factor(origin)3	-0,0045	0,0035	-1,29	0,1971
acceleration:year	0,0614	0,0255	2,41	0,0164
acceleration:factor(origin)2	0,9234	0,2641	3,50	0,0005
acceleration:factor(origin)3	0,7159	0,3258	2,20	0,0286
year:factor(origin)2	0,2932	0,1444	2,03	0,0430
year:factor(origin)3	0,3139	0,1483	2,12	0,0350

```
#a
plot(Auto[,-9])
#b
cor(Auto[,-9])
#c
Auto$origin <- factor(origin)
fit1 <- lm(mpg~.,data = Auto[,-9])
summary(fit1)
# d
par(mfrow = c(2,2))
plot(fit1)

#e
fit1 <- lm(mpg~ .*.,data = Auto[,-9])
summary(fite)
```