1. Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

	TV	radio	newspaper	sales
TV	1.0000	0.0548	0.0567	0.7822
radio		1.0000	0.3541	0.5762
newspaper			1.0000	0.2283
sales				1.0000

1. TV Advertising:
$H_0 : \beta_{TV} = 0$: TV advertising has no effect on sales.
The p-value for TV advertising is less than 0.0001. Since this is much less than 0.05, we reject the null hypothesis and conclude that TV advertising has a significant effect on sales.

2. Radio Advertising:
$H_0 : \beta_{radio} = 0$Radio advertising has no effect on sales.
- The p-value for radio advertising is less than 0.0001. Since this is much less than 0.05, we reject the null hypothesis and conclude that radio advertising has a significant effect on sales.

3. Newspaper Advertising:
$H_0 : \beta_{newspaper} = 0$: Newspaper advertising has no effect on sales.
The p-value for newspaper advertising is 0.8599. Since this is much greater than 0.05, we fail to reject the null hypothesis and conclude that newspaper advertising does not have a significant effect on sales.

Conclusion
Based on the p-values in Table 3.4, we can draw the following conclusions that:
TV advertising and radio advertising both have a significant positive effect on sales, showing that the increase in TV and radio advertising budgets are associated with increased sales. While Newspaper advertising does not have a significant effect on sales, suggesting that changes in the newspaper advertising budget do not significantly impact sales. These results can allow the advertising company to focus their budget more on TV and Radio advertising compared to newspaper advertising.

3. Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = Interaction

between GPA and IQ, and X5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get β_0 = 50, β_1 = 20, β_2 = 0.07, β_3 = 35, β_4 = 0.01, β_5 = −10.

(a) Which answer is correct, and why?

i. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.
ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.
iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.
iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

Given the coefficients:

- β_0=50
- β_1=20 (for GPA)
- β_2=0.07 (for IQ)
- β_3=35 (for Level)
- β_4=0.01 (for interaction between GPA and IQ)
- β_5=−10 (for interaction between GPA and Level)

An equation can look like:

$Salary = 50 + 20 \times GPA + 0.07 \times IQ + 35 \times Level + 0.01 \times (GPA \times IQ) - 10 \times (GPA \times Level)$

For high school graduates (Level = 0)

$Salary = 50 + 20 \times GPA + 0.07 \times IQ + 0.01 \times (GPA \times IQ)$

For College graduates(Level = 1)

$Salary = 50 + 20 \times GPA + 0.07 \times IQ + 35 + 0.01 \times (GPA \times IQ) - 10 \times GPA$

This results in the equation:

$Difference = 35 - 10 \times GPA$

This means that when GPA is low, the equation above is positive, meaning that college graduates will earn more money. However when GPA is high enough so that the equation is negative high school graduates could earn more. Therefore for a fixed value of I and GPA, college graduates earn more on average given that the GPA is high enough.

(b) Predict the salary of a college graduate with an IQ of 110 and a GPA of 4.0.

$Salary = 50 + 20\times 4.0 + 0.07\times 110 + 35 + 0.01\times(4.0 \times 110) - 10\times 4.0 = 137.1$

This means that the predicted salary is $137,100

(c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

While the coefficient for the GPA/IQ interaction is small at 0.01 the significance of this interaction can be determined by more than just the magnitude of the coefficient, but also the significance of the statistics which is typically tested by the p-values. Without the p-values we are not able to define the significance with complete confidence. However based on the magnitude alone there is a minor practical impact. Therefore the statement above can be considered true. The small coefficient shows a minimal effect on the interaction.

8. This question involves the use of simple linear regression on the Auto data set.

(a) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output.

For example:

i. Is there a relationship between the predictor and the response?

There is a relationship between the predictor, horsepower, and the response, mpg. This is due to the greater than 1 F-statistic and a close to zero p-value of the F-statistic of < 2.2e-16. Thus we can reject the null hypothesis.

ii. How strong is the relationship between the predictor and the response?

The mean of mpg is 23.45 and the RSE of the linearModel was 4.9 which means the percent error is 20.9%. The R^2 of the linearModel was 0.6 which means 60.5% of the variance in mpg is explained by horsepower.

iii. Is the relationship between the predictor and the response positive or negative?

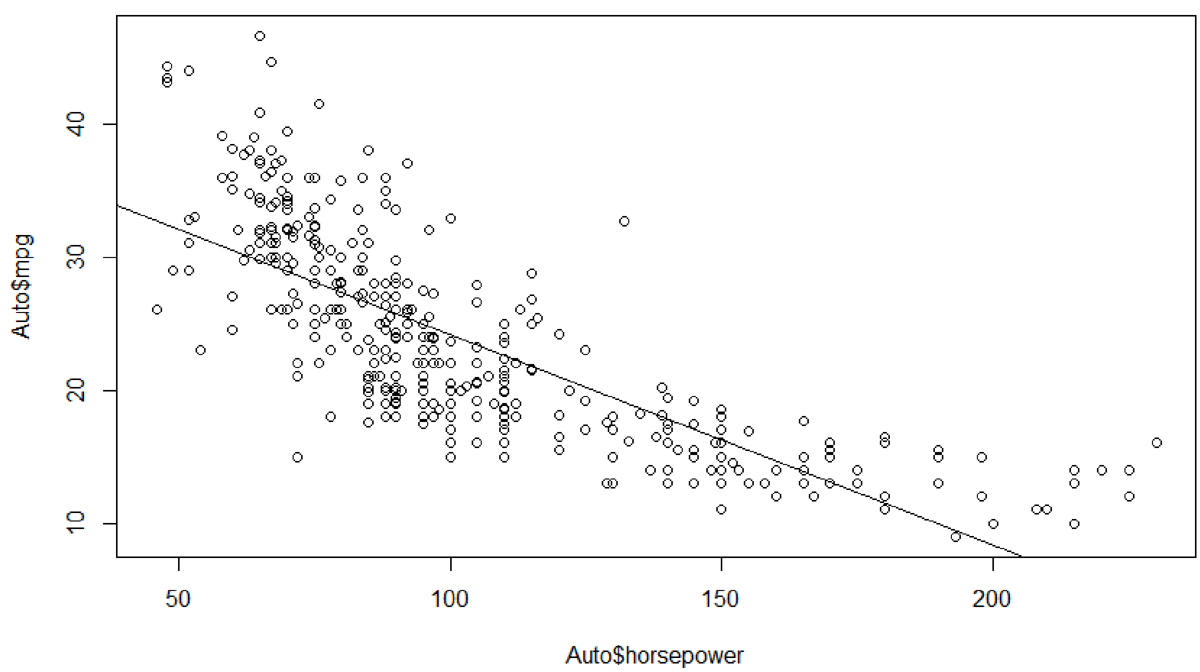
The relationship between the predictor and the response is negative.

iv. What is the predicted mpg associated with a horsepower of 98? What are the associated 95 % confidence and prediction intervals?

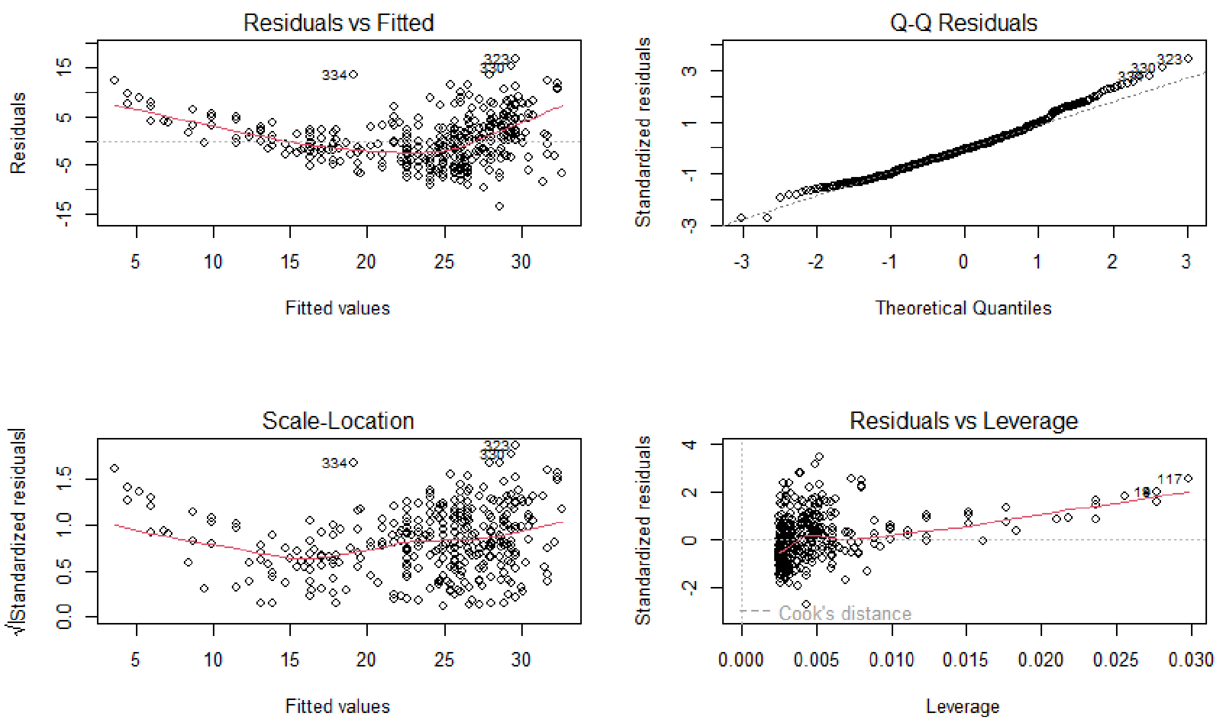
	fit	lwr	upr
1	24.46708	23.97308	24.96108

	fit	lwr	upr
1	24.46708	14.8094	34.12476

(b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.



(c) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.



The residuals of the plot show a non-linear relationship.