1.

Now we look at this table. The table shows the residual standard error , R^2, and F-statistic about the least square model for the regression of the sales vs medias like TV, newspaper and radio in the Advertising data.

I will analysis the F-statistic first. In the table, F-statistic is 570 and since it is far larger than 1. It shows enough evidence to reject null hypothesis. This means that at least one advertising media is related to sales.

2. mixed selection

In advertising example, the p-value increase as new predictors are added to the model. Newspaper is an example. In simple linear regression, the p-value shows that there is a relationship between sales and newspaper, but in multiple linear regression, the p-value of the newspaper increase to 0.8599 from less than 0.0001.

Therefore, in the mixed selection method, I will remove the newspaper because it’s p-value rises above a certain threshold. After removing, I will do another p-value check to see if the model have a sufficiently low p-value, and all variables outside the model would have a large p-value if added to the model.

3.

(new update to R^2) because R^2 is to describe the % of the variation is not described by the regression line. I can do 1 for 100% minus the percentage of the unexplained variation or error variation to the total variation

In the table, I see that the R^2 of the advertising data with 3 medias is 0.8972. Now according to the book, the R^2 value for only TV and radio is 0.89719. This brings a question: Why does the R^2 value decrease when I remove the newspaper from the model because of its large p-value?

It turns out that R^2 will always increase when more variables are added to the model. The fact is backed up because adding another variable to the least squares equation must allow the model to fit the training data more accurately

Also by noticing only a tiny increase in R^2 when newspaper is added shows that newspaper can be dropped from the model because it provide no real improvement in the model fit to the training data. If newspaper is added to the model, it might result in poor results due to overfitting.

Now in the table below, we can see that model that only contain TV had an R^2 of 0.61. However, adding radio to the model, the R^2 is 0.89719 which is a substantial improvement in R^2. This means that the model explain a much larger portion of the training data when radio is added to the TV model, which should predict sales better.

I can also further qualify this by looking at the p-value. According to the book, the model only contain TV and radio has the RSE of 1.681 and the model also containing newspaper has an RSE of 1.686. In contrast the RSE of the model only containing TV is 3.26. This shows that containing the model on radio with TV reduce the error and can predict sales more accurately than just using TV.

In addition to just looking at the numerical statistic R^2 and RSE, it is useful to plot the data because the graphic summaries can reveal problems with the model that is not visible from the numerical statistic. The graph below shows the linear regression plane that fit to sales suing TV and radio. The red point is the coordinate of the sales based on the budget spend on TV and radio.

Looking at the graph, I notice that the diagonal of the plane have many positive residual while negative residual tend to lie away from the diagonal where the budgets are more lopsided. This shows issue of overestimate sales for budgets that is lopsided to either TV or radio, while underestimating sales for budgets that was split between the two media.

This might show that there is a non-linear pattern that this linear model cannot predict accurately. Also noticing the positive residual, the model suggest a synergy effect between the TV and radio where when budget is spend equally between the TV and radio advertising increase the sales greatly than lopsided budget use.

4.

To determine the accuracy of the model. I use the model from question 3 above for $100,000 spend on TV and $20,000 spend on radio as an example. The 95% confidence interval is [10,985, 11,528] and the 95% prediction interval is [7,930, 14,580].

The intervals need to calculate in a few steps: first calculate the data matrix: follow the formula for calculating the data matrix. The data matrix should be shown like a reflection on the diagonal line matrix then invert the data matrix. To do this, I will use software or put an identity matrix beside the data matrix and row operate the matrix into reduced echelon form.

Then I will pick up the matrix outside of the reduce echelon rows and columns. Next I will calculate the Mean square error then multiply it to the inverted matrix. This would get the Covariance matrix of the model. Now start to plugin the x value of the prediction. For example, in the advertisement example, $100,000 for TV and $20,000 for radio.