1. **Exercises**

Part 1.1 Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point ($\bar{x}$,$\bar{y}$).

Answer:

Least Squares Line Equation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
$$= \hat{\beta}_0 + \hat{\beta}_1 x - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}$$
$$= \hat{\beta}_0 + \hat{\beta}_1 (x - \bar{x}) + \hat{\beta}_1 \bar{x}$$

We know that,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Substituting $\hat{\beta}_0$ in regression equation, we get:

$$\hat{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 (x - \bar{x}) + \hat{\beta}_1 \bar{x}$$
$$\hat{y} = \bar{y} - \hat{\beta}_1 (x - \bar{x})$$

If we substitute $x = \bar{x}$, we get:

$$\hat{y} = \bar{y} - \hat{\beta}_1 (\bar{x} - \bar{x})$$
$$\hat{y} = \bar{y}$$

Hence, we can say that least squares line pass through the $\bar{x}$ & $\bar{y}$. That means the

Part 1.2
Part 1.2-1
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.
Answer:

By seeing the p-values of the predictor's TV, radio, and newspaper. The TV and Radio have less p-value than Alpha (α) and the newspaper has a higher Alpha(α). Basically, the Alpha(α) value is 0.05. If the P-value
H_0 (Null Hypothesis): No relationship between predictors and response variables

H_1 (Alternative Hypothesis): There's a relationship between predictors and response variable

α: P (H_0 Rejected | H_0 is true) - 0.05 or 5%

P value<=α then the result does not support the hypothesis H0
P value > α then the result support the hypothesis H0

There is a higher correlation between the newspaper and radio, it has a 0.35 correlation this tells us the increase in newspaper sales also impacts/increases the sales of the radio.
The newspaper has a negative coefficient with a high p-value(0.86) than Alpha(α)(0.05), and the TV and radio have less P-Value than Alpha(α). which indicates that the results should Support the null hypothesis(H0). So there is no relationship between response and predictors.

Part 1.2
Part 1.2-3
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 $\beta\hat{}0 = 50$, $\beta\hat{}1 = 20$, $\beta\hat{}2 = 0.07$, $\beta\hat{}3 = 35$, $\beta\hat{}4 = 0.01$, $\beta\hat{}5 = -10$.
(**1.2-3-a**) Which answer is correct, and why?

Answer: The option C is correct.

Here, X1 = GPA, X2 = IQ, X3 = Level, X4= X1*X2, X5=X2*X3.

$\hat{\beta}_0$=50, $\hat{\beta}_1$=20,$\hat{\beta}_2$=0.07,$\hat{\beta}_3$=35,$\hat{\beta}_4$=0.01,$\hat{\beta}_5$=-10

Equation

$Y=\hat{\beta}_0+, \hat{\beta}_1 X1+\hat{\beta}_2\ X2+\hat{\beta}_3\ X3+\hat{\beta}_4\ (X1*X2) +\hat{\beta}_5(X1*X3)$

Y=50+20(X1) + 0.07(X2) +35(X3) + 0.01(X1*X2)-10(X1*X3)

Y=50+20(X1) + 0.07(X2) + 0.01(X1*X2) + X3(35-10(X1))

By taking the fixed values of IQ and the GPA, the Level is high school, take X3 = '0' and the High GPA of college students around X1=4. Here, the 35-10(X1) is multiplied with X3 is equal to X3(35-10(X1)) => 0(35-10(4))=>0. If you see the other option(iv), here the X3 value is '1' and the high GPA of college student X1=4, like 1(35-10(4)) will result in negative value '-5'. The option D with high GPA is reducing the total income. So, the option C is best.

1.2-3-b) Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

Answer:

Lets take,

X1=GPA=4

X2=IQ=110

X3=1(college grad)

Equation,

Y=50+20(X1) + 0.07(X2) + 0.01(X1*X2) + X3(35-10(X1))

Y= 50+20(4)+0.07(110)+0.01(4*110)+1(35-10(4))

Y=50+80+7.7+4.4 - 5

Y=137.1

The 137,100$ is the salary of a college graduate for given IQ and GPA.

1.2-3-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.

Answer:

False.

We have some few data missing. We have not taken the standard deviation error value of GPA/IQ Interaction variable. Also, the range of units of GPA,IQ and sales vary a lot as sales is measured in 1000's.So, we cannot conclude if small GPA/IQ coefficient means vary little interaction effect.Ex GPA scale is from 1.0 to 4.0 while we don't know exactly about the range of IQ. And also, we don't have standard error value of GPA/IQ interaction variable.

Part 1.2-4-a

Suppose that the true relationship between X and Y is linear, i.e. $Y = \beta 0 + \beta 1X + \epsilon$. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

Answer:

Yes, we would expect the RSS(Residual sum of squares) for cubic regression is less than linear regression. Using the linear regression is also a best option, when we have linearly separable data. If we take linear separable data, then we can draw regression line using linear and as well as with cubic regression. The cubic regression has the flexibility to pass through many values as possible as linear regression. If it has the more strong linearity then there will be no different using linear or cubic regression.

2. Programming

Part 2.1-A-i

We are not using the vehicle name as predictor for mpg (response). The car names are not relevant to the prediction problem and they are not reasonable to take, to predict our response (mpg) because it's a qualitative data.

Part 2.1-A-ii

We can infer from the summary that the greatest residual value is 12.99.

R^2 and Adjusted R-sq:

With an R2 value of 0.81, our model can account for 81% of the variation in the response variable around its mean. The value of 0.81 for R-squared and Adjusted R-squared is the same. R-square takes into account all independent variables that have an impact on the model's outcomes, whereas Adjusted R-squared only takes into account independent variables that have an impact on the dependent variable.

RMSE:

The residuals around the line of best fit had a standard deviation of 3.37, as shown by the RMSE that we obtained. The better match is shown by lower values. However, the RMSE value of 3.37 is small given that our Response values (mpg) range from 9 to 46.

RSE:

A regression model's residual standard error how well it fits the data. This model's RSE, which was trained using every predictor, is 3.41. This shows that the model has an average error of 3.41 in predicting the target variable.

Part 2.1-A-iv

Gaussian Distribution:

Yes, the histogram follows the Gaussian distribution. It is also known as normal distribution,the data is symmetric to the mean.the appearance of the data is near to the mean than the data far from

he mean.Not all symmetrical distributions are normal, but all normal distributions are symmetrical. Here in above image we can see that, we had high frequency at '0'(mean) than far the mean.

Part 2.1-B-i

To narrow down the features, i had picked 3 features having less P-value among all the features in the data set, and now i created a different model using subset function.

Part 2.1-B-ii

We can infer from the summary that the greatest residual value is 12.99.

R^2 and Adj-R sq:

The model here is created with only three features,With an R2 value of 0.81, our model can account for 81% of the variation in the response variable around its mean. The value of 0.81 for R-squared and Adjusted R-squared is the same. R-square takes into account all independent variables that have an impact on the model's outcomes, whereas Adjusted R-squared only takes into account independent variables that have an impact on the dependent variable.This got the same values when we created model with all the features

RMSE:

This also the same with the model we created with the all features.The residuals around the line of best fit had a standard deviation of 3.37, as shown by the RMSE that we obtained. The better match is shown by lower values. However, the RMSE value of 3.37 is small given that our Response values (mpg) range from 9 to 46.

RSE:

This is different value compared with previous model in 'a' we got 3.41 in model, here it is 3.39A regression model's residual standard error how well it fits the data. This model's RSE, which was trained using every predictor, is 3.39. This shows that the model has an average error of 3.39 in predicting the target variable.

Part 2.1-B-iv

Gaussian Distribution: It is also known as normal distribution,the data is symmetric to the mean.the appearance of the data is near to the mean than the data far from he mean.Not all symmetrical distributions are normal, but all normal distributions are symmetrical.

The data in the histogram above is centered around 0, so it follows the normal distribution; however, the residuals on the negative scale are more left-skewed. Since the model may have predicted a value that is lower than the actual value, it is not optimally fitted. The model is more biased toward the histogram's negative scale, we can also say. Due to the histogram's left-skewed distribution, the majority of the model's predicted values fall above the fitted line, or higher than the actual values. So, based on this, we can conclude that our model may, on average, overestimate the independent value, or mpg.

Part 2.1-b-v

Comparing models from (a) and (b)

The both models had performed well on the training data. The R-square and the Adjusted R-square for both the models are same(0.81).

The relative stranded error for the both models are different for the first model in (a) is 3.41 and for the model in (b) is 3.39, here the model in (b) is better because its having the low value than model in (a), the closer it gets to '0' the good the model.

The Root Mean Square Error(RMSE) is a measure that shows how the model regression line is fitted to data points.

The model(a) is trained on the all the features and model(b) is just trained on 3 features,But the model(b) had better performance among the both

Part 2.1-f

Prediction interval is a range,where the newly predicted values fall in that range. The prediction interval is made based on observation of previous data, and the prediction interval will also captures the uncertainty of the single value. The prediction interval is more wider than the confidence interval.

Confidence interval: In here we will calculate the upper and the lower bound values of the interval, and our predicted values will fall under the interval.we can run our program many times and the values fall under the same interval based on confidence percentage. The percentage we took here is 95% that is confidence level, that the values we predicted weill fall uder the lower and upper bounds of the interval.

Part 2.1-f-i

The Prediction interval has the interval range, where the all values fall in the interval in (e), in the confidence interval has only 8 matches which fall in the boundary of the confidence interval.where the range of prediction intervcal is much wider than the confidence interval.

Part 2.1-f-ii

By looking at 'd' and 'e'. we can tell that 'd' confidence interval has the small interval range and the 'e' has the large range of interval, which tells the fitting range. in 'd' we got only 8 values fall in that interval and in 'e', we got all the values'20' fall in the interval.This is because prediction interval must account for uncertainty of estimation population mean as well as random variation of individual