# DATA303 Assignment2

## Question1

Text

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

## Question2

When n = 2:

Graphical user interface, chart, scatter chart

Description automatically generated

When n = 4:

Graphical user interface, chart, line chart

Description automatically generated

When n = 6:

Chart, line chart

Description automatically generated

This is a combined graph:

Graphical user interface, chart

Description automatically generated

When n = 2, the model is a curved line, it does not fit all the data well.

As the size of the polynomial increases, when n = 6, although it fits all the data, even including the top one, the gap between each data is too much and it is expensive to fit.

N = 4 is the best fitted model since the gap around the top point is large but the gaps between other data points are not too much. And the curve almost fits all the data.

## Question3

Chart, histogram

Description automatically generated

As ρ increases, the line becomes flatter and flatter.

When ρ = 0, although the model fits all the data points (even for the top point), the gap between the data is too much and the amplitude is too large. It is very expensive to fit all the data.

When ρ = 4, the curve is too flat and cannot reach the highest point.

For the model when ρ = 0.25, we can see it sorts of fit the top point. But comparing to ρ = 0, we can see the gap between the data point is not that wide but for the gap around the highest point, it is still wider than the other gaps. Overall, we can say it is a good model.

For ρ = 1, we can also say it is a good line, but the line is flatter than the line of ρ = 0.25, so it is really depending on how much we care about the highest point. If we care it more, then ρ = 0.25 is a better fit, since the amplitude is wider which is closer to the top point. But if we don’t care the highest point that much, then the model of ρ = 1 fits better since it is closer to the points below but also have more amplitude with the highest point.

## Question4

a)

Text, letter

Description automatically generated

b）

Graphical user interface, text

Description automatically generated

Graphical user interface, chart, line chart

Description automatically generated

If we don’t center the data, we can see when ρ = 10 ^ 10 -5, the model fits the data well but there are a lot of unnecessary swings.

When ρ = 10 ^ 10 -4, 10 ^ 10 -3, 10 ^ 10 -2 and 10 ^ 10 -1. These four lines are really close to each other and they are all good models, we can observe that it almost fits the data. Although they did not fit the data that well, there are no unnecessary swings.

After that, when ρ increases (ρ = 1 and 10), these two models do not fit the data at all and getting close to the origin. They go down to 0.

C)

Graphical user interface, chart

Description automatically generated

If we center the data, we can see there are still unnecessary swings around the point when ρ is 10 \* 10 ^ -5. As ρ increases, when ρ = 10 ^ 10 -4, 10 ^ 10 -3, 10 ^ 10 -2 and 10 ^ 10 -1, the next four lines are still good, although they did not fit the data so well , but still in a reasonable fit, without unnecessary swings around the data.

But as ρ increases to 1 and 10, the lines are not approaching to 0, as the data is centered. We add mean back to the model. The two lines below is getting back to the mean of data. The models are kind of fitting the model. Although it is not that fitted since there are only 3 data point. But if the data set is big enough. It will become more useful since they are closer to the mean and will give a more reasonable fit.

d)

i)

Graphical user interface, chart, line chart

Description automatically generated

The length of x decreases significantly and then slowly decay.

For the residuals, at the beginning it starts with almost 0,

As ρ increases, the x gets smaller, and residuals gets bigger.

ii)

Graphical user interface

Description automatically generated with medium confidence

There is a turning point.

We can change x2(Y shown in the graph) and x3(Z shown in the graph) to a much smaller value easily and make the sum of x2 and x3 as small as possible. The reason we do that is because if we change x2 and x3 a lot, there will be only a really small change of product Ax for eliminating the bad behavior. That’s why we have a knee (turning point) there. After they reach the knee, most of them are closer to 0.

After that turning point, the x1 starts decreasing too.

iii)

Graphical user interface

Description automatically generated

When the ρ value is from 10 ^-5 to 10^-4, the line decreases significantly, that is the x path with lines with unnecessary swings.

When the ρ value is from 10 ^-4 and 10 -1, the line is an almost straight flat line. This part is related to the x path that four models almost fit the data point but without swings.

When the ρ value is from 1 to 10, the line is decreasing again, this part is related to the x path that two lines are much lower than the data point and getting closer to 0.