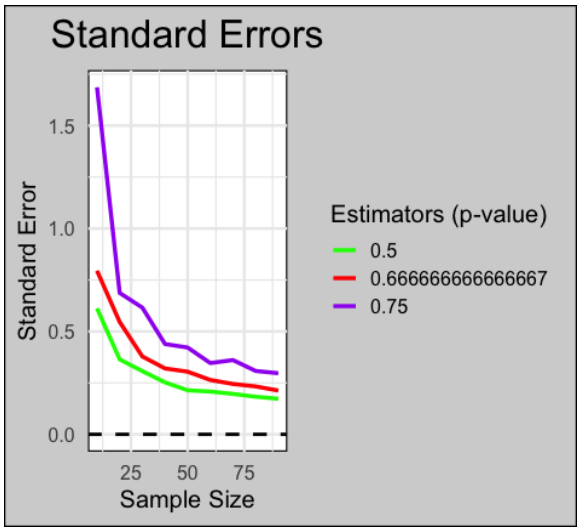
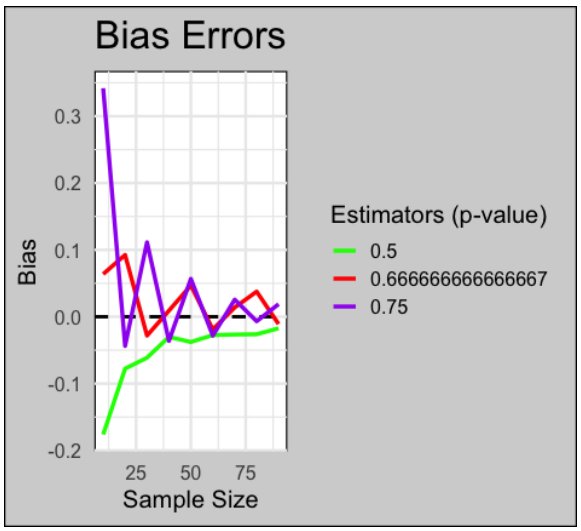
Statistical Modelling and Inference
Assessed Coursework Assignment 1

1(a). i.



The plot demonstrates that as the sample size increases standard error for the estimators (p-value = c(0.5, 0.667, 0.75)) decreases. However, the estimator with p-value equal to 0.5 might be unbiased because the plot demonstrates that with an increasing sample size compared to other estimators with p-value 0.667 and 0.75, it has the lowest standard error.

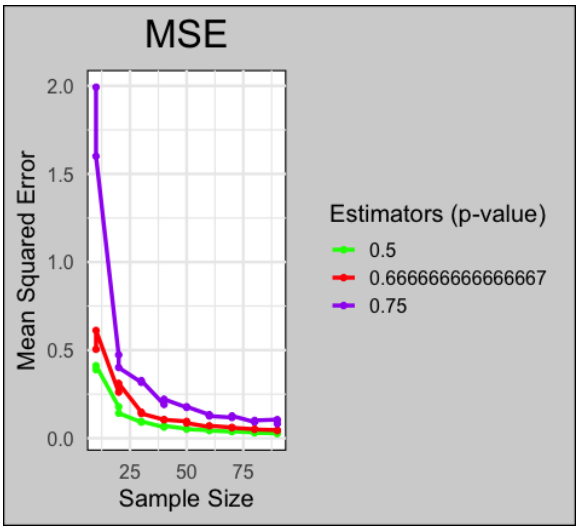


The plot demonstrates that as the sample size increases bias error for the estimators (p-value = c(0.5, 0.667, 0.75)) descreases. However, the estimator with p-value equal to 0.5 might be unbiased because the plot demonstrates that with an increasing sample size compared to other estimators with p-value 0.667 and 0.75, it has the lowest bias error.

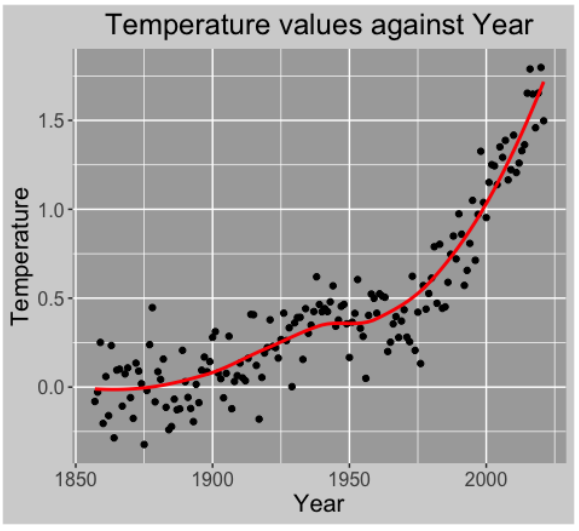
The observations have shown that the estimator with p-value equals to 0.5 is most likely unbiased compared to other estimators. Both Standard Error and Bias Error decreases as the sample size increases. The estimator with the p-value equals to 0.5 is the most precise measure because it provides the lowest standard and bias errors.

1 (b).

1 (a). ii.

	The plot demonstrates that MSE decreases for all estimators (p-value = c(0.5, 0.667, 0.75)) with an increasing sample size. It also shows that all the estimators are consistent, which means with an increasing sample size the parameter's accuracy is increasing as well. However, among all the estimators, the one with p-value equals to 0.5 is more consistent and shows lowest MSE.

2(a).

	The scatter plot demonstrates temperature values against the year. The smoothing line shows the overall trend. Temperature increased over the years, accelerating significantly after 1975. Over the years 1857 to 1899 temperature has a zero mean due to normalisation.

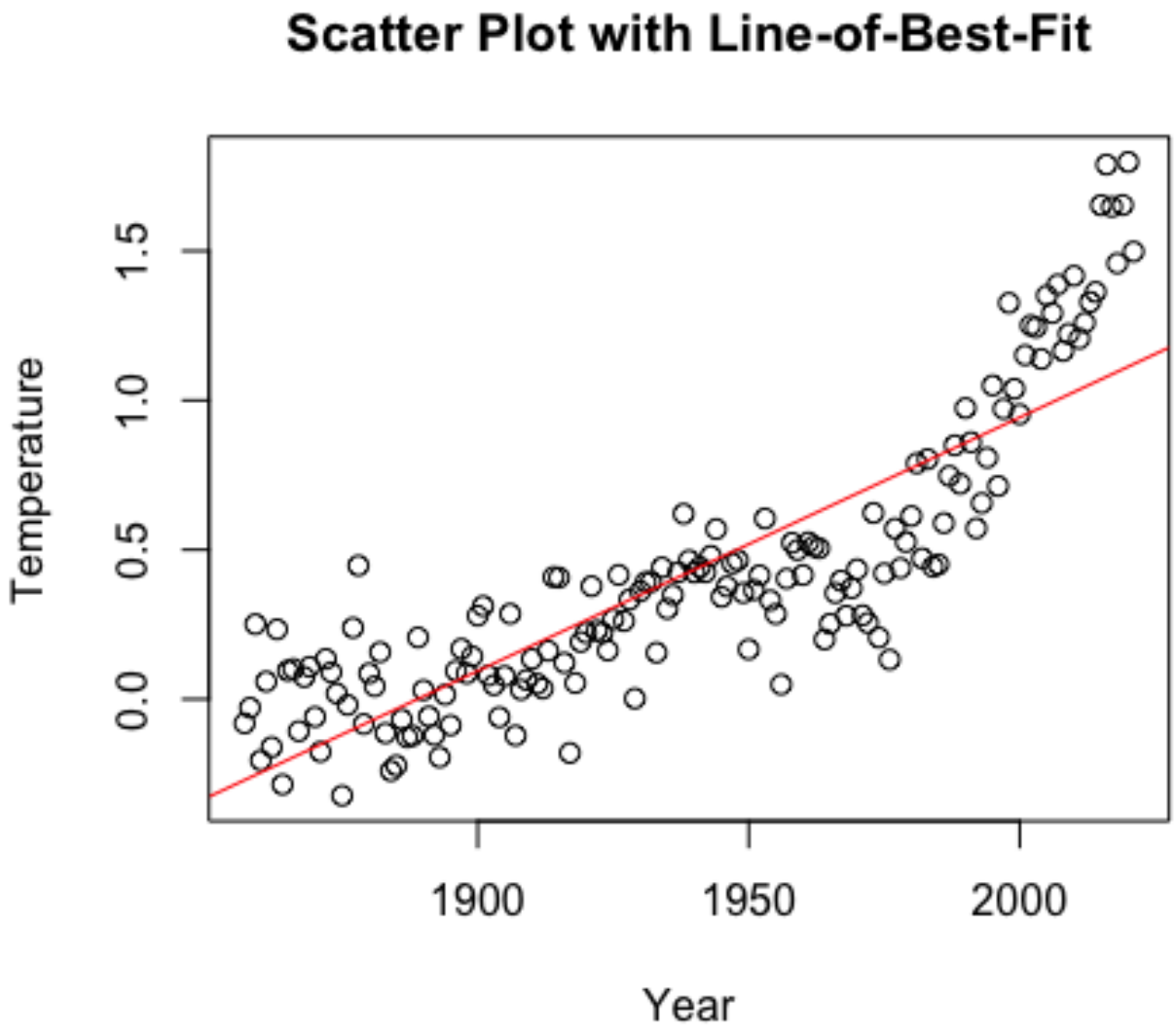
2(b).

The estimation showed that there is 0.008485213 change of temperature per year. The 95% confidence interval is 0.007693459 to 0.009276968 which

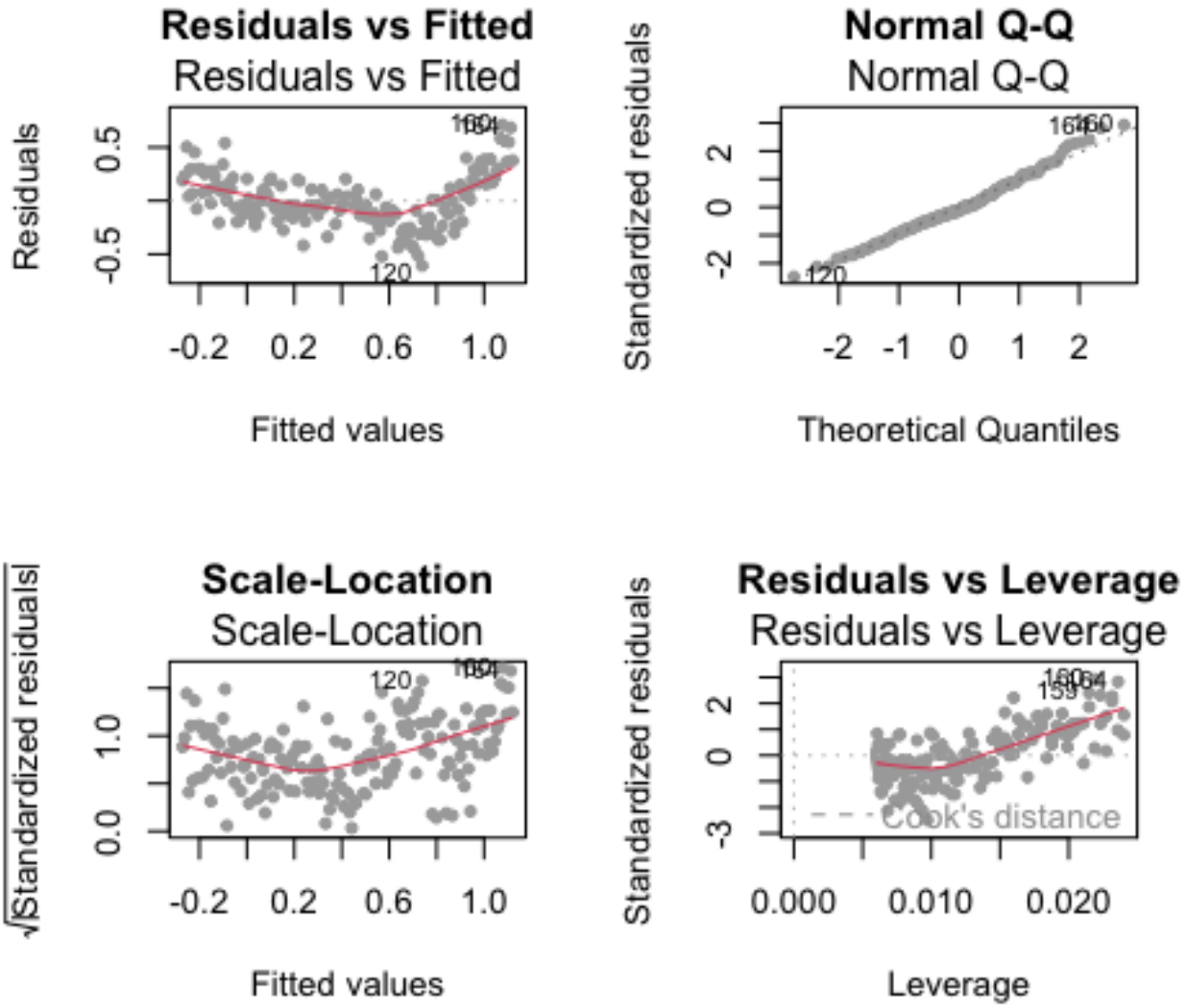
means that there is 95% confidence that actual temperature change per year is in this interval.

Null Hypothesis H0: The rate of change is 0, which means no change.
Alternative Hypothesis H1: The rate of change is not 0, which means there is a change.
The p-value is 1.265859 * 10ˆ-48 which is much less than 0.01. Therefore Null Hypothesis H0 is rejected as the rate of change is not zero.

2(c).



The Scatter Plot with Line-Best-of-Fit demonstrates the relationship between Year and temperature. It shows a steady increasing trend, as the year increases, the temperature increases as well. However, it can be seen that at the latest years after 2000 the distance between the line and variables is increasing. Therefore it can be assumed that the spread of the residuals may change and the independent variables change.



- The Residuals vs Fitted model shows predictions of the model and compares it to the actual data. A perfect model should demonstrate a normal spread out of variables, however this model shows that the spread is high.
- The Normal Q-Q model shows whether the residuals are normally distributed. The points should lie perfectly in a line, however this model does not demonstrate ideal normal distribution.

- The Scale-Location model is similar to The Residuals vs Fitted model but it mostly indicates consistency of the spread. The model demonstrates increase in spread across variables thus the spread is inconsistent.
- The Residuals vs Leverage model shows points that are the most significant in the model. This model does not have any significant points that stands out comparing to others.

2 (d). The study in comparison between polynomial regression model and linear regression model has shown that polynomial regression model fits significantly better at the 5% level of significance. The reason is the trend between year and temperature is not perfectly linear therefore it does not represent the trend perfectly.

2 (e). The polynomial regression model has demonstrated that mean temperature in 2040 is 3.25245 celsius. Also, the 95% confidence interval is between 2.744158 and 3.762742 celsius, which means we can be 95% confident that the temperature will be within the interval.