Math IA1 - Draft

Formulate

Introduction

Keeping a business profitable requires, depending on the type of business, the owner to have great foresight into what the consumers want, and how much to sell it for depending on the demand. The difficult part of this is knowing how much demand there will be. This is why approximating the future is a valuable skill in the business world, this can be achieved relatively simply through mathematical means, and you can end up with an accurate mathematical representation of the future in regards to a businesses sales. In this Problem-Solving and modelling task, a method to create 2 mathematical models will be found. Both of which will attempt to forecast the profit for the client, given the clients data on the past year of sales. One model will approximate the **annual sales figures through to 2027**, and the other **monthly sales figures for 2024**.

Translation

Line of best fit:

The simplest and most effective method of creating these models, is by using linear regression, which attempts to model the relationship between two variables by fitting a linear equation to the observed data (*Linear Regression*, 2024). This line has an equation of form $y = b + ax$, where y would be the sales figures, and x would time.

$$sales\ figures = b + a * time$$

In order to create this *line of best fit*, we can use the **Least-Squares Regression** method to find the values of b and a in the *line of best fit* equation.

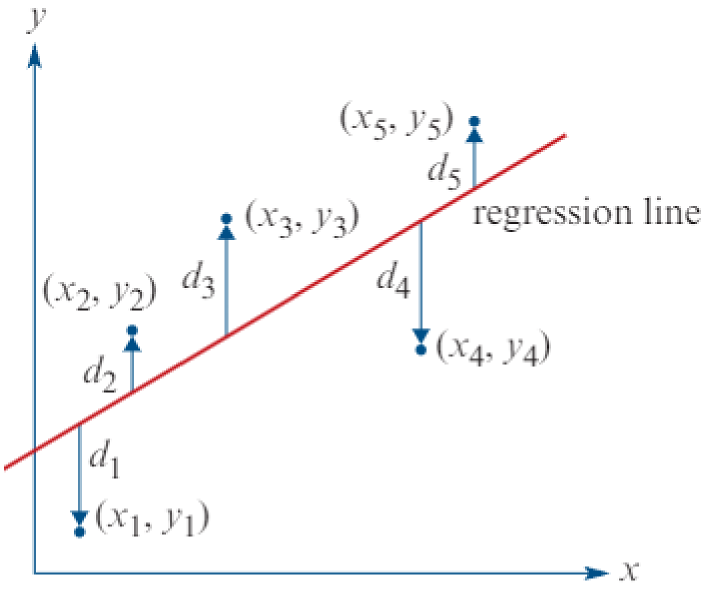

Figure 1, a graph showing a regression line

The line in this graph is a regression line, but isn't the *line of best fit*, the blue lines shown indicate the distance from each point to the regression line. This distance is called a **residual**. Squaring these residuals, and finding the sum of them finds **the Sum of Squared Residuals** The **Least-Squares Regression** method aims to minimize the sum of squared residuals. The reason the residuals are squared, is if the distance is negative, it will become positive, and if it is positive, it will still be positive. However this means that any outliers in the data affect the line of best fit more than data close to the ==average==. The equation of **the least squares regression line** can be found using the following formulae: $$b=\frac{rs_y}{s_x}$$ $$a=\bar{y}-b\bar{x}$$ where: - $b=slope$ - $a=intercept$ - $r=correlation \space coefficient$ - $s_x=standard \space deviation_x$ - $s_y=standard \space deviation_y$ - $\bar{x}=mean \space of \space x$ - $\bar{y}=mean \space of \space y$ %%Should show how to find r, sx, and sy?%% Now that the a and b values can be found, the line of best fit can be made using the equation $y=b+ax$.

Seasonalized & Deseasonalised data:

Whilst the line of best fit is accurate on most data, there is a characteristic of some time-series plots that is seasonalized. This means that the data regularly undergoes predictable changes every year (for example, could be any time period). Due to this characteristic, the line of best fit doesn't accurately represent the data. Often in business, sales vary due to seasons, for a swimsuit business, sales will be higher during summer, and this *seasonalisation* can be removed through the process of **deseasonalisation**. This process can reveal the underlying trend occurring in the plot, if the overall sales are increasing or decreasing. This in combination with the line of best fit is a great model to help analyze business sales, and help give advice to actions to perform business wise in the future.

Figure 2, a graph showing actual data vs deseasonalised data

The process of deseasonalisation can be completed by using the following formula, and repeating it for all data points in the dataset:

$$deseasonalised\ figure = \frac{actual\ figure}{SI}$$

The SI can be found using the formula:

$$SI = \frac{value\ for\ anum}{anum\ average}$$

$$anum\ average = \frac{anum_1 + anum_2 + \ldots + anum_n}{n}$$

Here is an example of finding the SI for a dataset:

Year	Summer	Autumn	Winter	Spring
1	920	1085	1241	446

$$anum\ average = \frac{920 + 1085 + 1241 + 446}{4} = 923$$

$$SI_{summer} = \frac{920}{923} = 0.997$$

$$SI_{autumn} = \frac{1085}{923} = 1.176$$

$$\dots$$

Performing this for the rest of anums (seasons) will get you:

Year	Summer	Autumn	Winter	Spring
1	0.997	1.176	1.345	0.483

Assumptions

- **Accurate Data**, it is assumed that the provided data is accurate, this directly affects the accuracy of the forecast.
- **Tax is included**, it is assumed that the sales data provided includes GST taxes. This assumption is important due to the fact that if it wasn't assumed, the forecast would be less than or greater than reality.
- **Tax variations**, it is assumed that tax remains the same, or at least on the same growth rate as included in the dataset. If not, the sales could decrease over time, but it is just tax going up.
- **Global Events**, it is assumed that no major global events will occur, that will directly affect the economy of the business location/nation, for example, COVID-19. This
- **Australian Dollar**, it is assumed that the money used in the dataset is Australian Dollar ($),

Observations:

Bibliography

Linear Regression. (2024). Yale.edu. http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm