Best Linear Unbiased Estimators (BLUE)

At the end of the last section, we noted that the **best linear regression model** is that which **minimizes the error term**

To do this, we can use the **OLS (ordinary least squares) method**

OLS

The OLS method **squares the value of each individual error term** (the **vertical distance** between each data point and the estimated best fit line, ε_i)

This is so that the error terms are **all positive values** and hence **sign equivalent + comparable for summation**

Once all ε_i are of the same format, the OLS method **sums the error terms** before selecting the model with the **lowest amount of predicted error** using **β and α estimates (α and β)**
Mathematically, this is shown as:

$$Min_{\underline{\alpha},\beta} \sum \varepsilon_i{}^2$$

Notation

Underlined ($\underline{\alpha}$) = an estimate
Bar lower case ($\overline{y}$) = mean
Bar upper case ($\overline{Y}$) = sample mean
Subscript i (x_i)= an individual value

Forming a Regression Line

Say we obtained a **sample of the different number of sales of trainers** at different prices:

Price (x)	2	3	5	6	7	8	9	10	13
Sales (y)	16	15	10	8	6	7	5	4	1

A known characteristic of any regression line is that it must go through the **average point of all the data points :**

$$Average\ data\ point\ =\ (\overline{X},\overline{Y})\ =\ (7{,}8)$$

Now we can define the regression line **in terms of x, y and m** using $\mathbf{x_1 = 7}$ **and** $\mathbf{y_1 = 8}$ as follows:

$$\boldsymbol{y - y_1 = m(x - x_1)}$$

$$y = mx\ - 7m + 8$$

We can now **find the deviations between the data points we have and this equation** to find the error terms:

x	y values on regression line	actual y value in sample	vertical deviation (ε_i)
2	$-5m + 8$	16	$5m + 8$
3	$-4m + 8$	15	$4m + 7$
5	$-2m + 8$	10	$2m + 2$
6	$-m + 8$	8	m
7	8	6	-2
8	$m + 8$	7	$-m - 1$
9	$2m + 8$	5	$-2m - 3$
10	$3m + 8$	4	$-3m - 4$
13	$6m + 8$	1	$-6m - 7$

Note that here, $\sum \varepsilon_i = 0$ since the **negatives and positives cancel each other out** as there will always be the **same deviations below the line and above**

Hence instead we **square each error term before summation**:

$$\sum \varepsilon_i{}^2 = (5m+8)^2 + (4m+7)^2 + (2m+2)^2 + m^2 + (-2)^2 + (-m-1)^2 + (-2m-3)^2 + (-3m-4)^2 + (-6m-7)^2 = 96m^2 + 266m + 196 = r$$

Since we want to **minimise this summed squared error term**, we can use **simple differentiation** to find m:

$$\frac{dr}{dm} = 192m + 266 = 0$$

$$m = -1.3854\ldots = \beta$$

Plug m into original regression equation to find **final regression line for data**:

$$y = mx - 7m + 8$$

$$y = -1.3854x + 17.6978$$

Where:

$$\beta = -1.3854$$

$$\alpha = 17.6978$$

With a perfect OLS model, the **predicted individual values will always equal the actual individual values** (i.e. $\underline{y_i} = y_i$)

> However this is **never possible due to variation** in the data points, causing individual data points to **differ from mean average**, $\overline{y}$, calculated through the **TSS**

Total Sum of Squares (TSS)

The total sum of squares (TSS) calculates the **total variation between the actual data points and the actual mean average**

It does this by **summing the squares of the differences between each actual data point (y_i) and the actual mean average ($\overline{y}$)**, shown mathematically as:

$$\text{TSS} = \sum (y_i - \overline{y})^2$$

Explained Sum of Squares (ESS)

The explained sum of squares (ESS) gives the **variation between the predictions made by the OLS model and the actual mean average**

It does this by **summing the squares of the differences between each estimated data point ($\underline{y_i}$) and the actual mean average ($\overline{y}$)**, shown mathematically as:

$$ESS = \sum (\underline{y_i} - \overline{y})^2$$

Residual Sum of Squares

Residual sum of squares (RSS) gives the **variation between the actual data and the predicted data by the OLS**

It does this by **summing the squares of the differences between each actual data point (y_i) and each estimated data point ($\underline{y_i}$), which is equal to the summation of the error terms squared (e_i),** shown mathematically as:

$$RSS = \sum(y_i - \underline{y_i})^2 = \sum {\varepsilon_i}^2$$

Coefficient of Determination

The coefficient of determination, symbolised by **R^2**, shows the **accuracy of the model (the % of population data that is explained by the regression line)**

Ranges from 0 (**completely inaccurate)** to 1 **(perfect model)**

The value of R^2 can be found using either of the following two formulas:

$$R^2 = \frac{ESS}{TSS}$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

A superior model should have a **low RSS and high R^2**

Gauss-Markov Theorem

A model should be **efficient, consistent, unbiased, and sufficient**

The **Gauss-Markov theorem gives the conditions** for the OLS model to have these **four qualities, and hence be the best linear unbiased estimator**

For an **unbiased estimator**, **two conditions** are required:

1. **The x variables are fixed upon repeated sampling**
    - This means that the **causation in the model must always go in one direction** (from the x variable to the y variable), and so no matter how many times the model is run, **changes to the y variable do not change the fixed x variable**
    - I.e. **change in sales do not determine price in model**
2. **Expected individual error is zero**
    - All error must **average to an expected error of 0 ($\mathbf{E}(e_i) = 0$)**
    - If this isn't the case, the variations of error terms are **predictable** and hence must be accounted for and **incorporated into a better model**

From here, all that is left is for the model to be the **best and most efficient model** with the **lowest variance**, dependent upon **two further conditions:**

3. **Homoscedasticity or constant variance (i.e. $\mathbf{E}({\varepsilon_i}^2) = \sigma^2$)**
    - Homoscedasticity means all error terms have the **same variance throughout the sample**
    - This is the opposite to **heteroscedasticity**, where there is **changing variance of error terms throughout the sample**, which means that the variance of the error terms **depends on the value of x**, suggesting model needs to be **changed** to take account of **additional errors that occur at certain x values**
    - **I.e. Sales should vary the same in the dataset irrespective of the price**

4. **No autocorrelation**
    - This means **errors should remain unrelated to each other** with **no covariance**
    - If error was correlated, knowledge of one error term would **allow other error terms to be predicted** and hence **incorporated into a better model**