

Java doc(word edition)

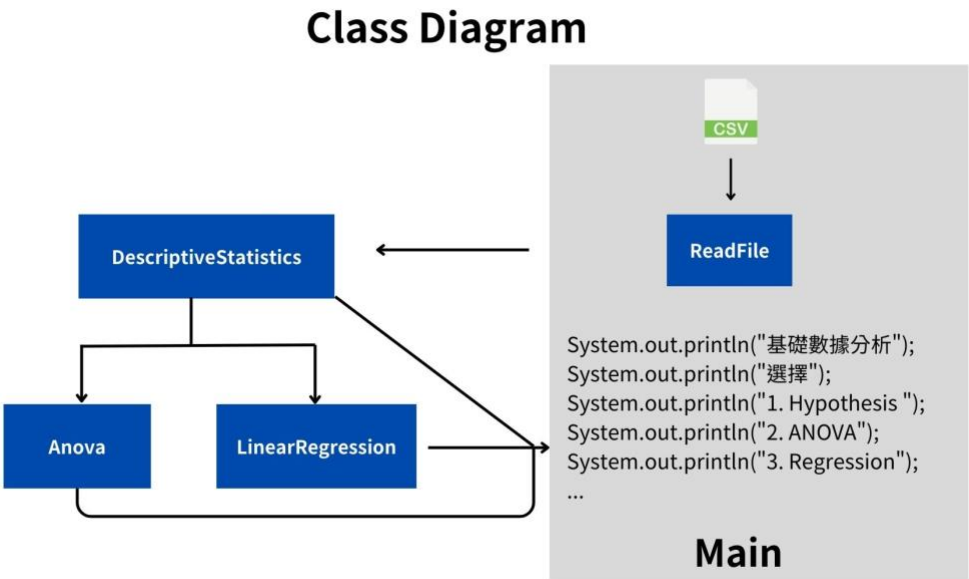
說明與動機

這是來自 Java Project 的說明文件，主要是為了方便使用者了解這個專案的功能，以及如何使用這個專案。

About class

這個專案有 6 個 class，分別是 ReadFile、DescriptiveStatistics、LinearRegression、Main、Anova、HypothesisTest。

相信以下圖表勝於千言萬語，以下是這個專案的類別圖。



類別圖

DescriptiveStatistics 類別說明

類別名稱：DescriptiveStatistics

類別說明： 這個類別提供了計算數據集描述性統計的方法，包括平均值、中位數、標準偏差等。

Constructors

建構函數	描述
DescriptiveStatistics(double[] data, String name)	用指定的數據陣列和數據集名稱初始

建構函數	描述
化 DescriptiveStatistics 實例。	

Methods

方法名	返回類型	描述
getData()	double[]	返回當前數據集。
getName()	String	返回數據集的名稱。
setData(double[] data)	void	設置新的數據集。
setName(String name)	void	設置數據集的新名稱。
mean()	double	計算數據集的平均值。
median()	double	計算數據集的中位數。
standardDeviation()	double	計算數據集的標準偏差。
sampleSize()	int	返回數據集的樣本大小。
populationVariance()	double	計算數據集的總體方差。
populationStandardDeviation()	double	計算數據集的總體標準偏差。
summary()	String	提供數據集的描述性統計摘要。

LinearRegression 類別說明

類別名稱： LinearRegression

繼承： DescriptiveStatistics

類別說明： 這個類別繼承自 DescriptiveStatistics，提供線性回歸分析的功能。它可以計算線性回歸模型的斜率和截距，並使用模型進行預測。

Constructors

建構函數	描述
LinearRegression(double[] xData, double[] yData, String name)	使用自變量和因變量的數據，以及數據集的名稱來構造 LinearRegression 物件。

Methods

方法名	返回類型	描述
calculateSlope()	double	計算回歸線的斜率（beta）。
calculateIntercept()	double	計算回歸線的截距（alpha）。
predict(double x)	double	使用線性回歸模型預測給定 x 值的 y 值。
summary()	String	提供線性回歸模型的摘要。
explain()	String	提供線性回歸的基本概念和模型解釋。

方法名	返回類型	描述
<code>description()</code>	<code>String</code>	提供簡單線性回歸的詳細說明和公式。

Anova 類別說明

類別名稱：Anova

繼承：DescriptiveStatistics

類別說明：這個類別提供了執行單因素方差分析（ANOVA）的方法。用於分析多組數據集之間的均值是否存在顯著差異。

Constructors

建構函數	描述
<code>Anova(double[][] groups, String name)</code>	使用多組數據構造 Anova 對象，每個子數組代表一組數據。

Methods

方法名	返回類型	描述
<code>flatten(double[][] arrays)</code>	<code>double[]</code>	將二維數組展平為一維數組。
<code>overallMean()</code>	<code>double</code>	計算總體均值。
<code>totalSumOfSquares()</code>	<code>double</code>	計算總體平方和（SST）。
<code>betweenGroupSumOfSquares()</code>	<code>double</code>	計算組間平方和（SSB）。
<code>withinGroupSumOfSquares()</code>	<code>double</code>	計算組內平方和（SSW）。
<code>calculateFValue()</code>	<code>double</code>	計算 ANOVA 的 F 值。
<code>summary()</code>	<code>String</code>	提供 ANOVA 分析的摘要。
<code>explain()</code>	<code>String</code>	提供 ANOVA 分析的基本概念和使用方法的解釋。
<code>description()</code>	<code>String</code>	提供 ANOVA 的詳細描述和公式。

HypothesisTest 類別說明

類別名稱：HypothesisTest

類別說明：這個類別用於進行假設檢定。它可以根據給定的數據集和假設值，執行假設檢定並計算相應的 p 值，以判斷是否拒絕虛無假設。

Constructors

建構函數	描述
<code>HypothesisTest(ArrayList<Double> data, double[] data_array, String name)</code>	使用數據集、數據數組和名稱來構造 HypothesisTest 物

建構函數	描述
------	----

Methods

方法名	返回類型	描述
<code>getAlpha()</code>	<code>double</code>	獲取顯著性水平（alpha）。
<code>getAvg()</code>	<code>double</code>	獲取平均值。
<code>getVar()</code>	<code>double</code>	獲取變異數。
<code>getXbar()</code>	<code>double</code>	獲取樣本平均值。
<code>getNum()</code>	<code>int</code>	獲取數據數量。
<code>setAlpha(double alpha)</code>	<code>void</code>	設置顯著性水平（alpha）。
<code>setAvg(double avg)</code>	<code>void</code>	設置平均值。
<code>setVar(double var)</code>	<code>void</code>	設置變異數。
<code>setNum(int num)</code>	<code>void</code>	設置數據數量。
<code>setXbar(double xbar)</code>	<code>void</code>	設置樣本平均值。
<code>setNullHypo(double hypo, String direct)</code>	<code>void</code>	設置虛無假設和對立假設。
<code>calculatePValue()</code>	<code>void</code>	計算假設檢定的 p 值。
<code>tCalculatePValue()</code>	<code>double</code>	計算 t-分佈的 p 值。
<code>analysis(double t)</code>	<code>void</code>	根據計算得到的 t-值或 z-值分析結果。
<code>tAnalysis(double t)</code>	<code>void</code>	根據 t-值和測試方向執行 t-分佈假設檢定。

Main 類別說明

1. 首先，先建立一個 `ArrayList`，用來存放 `DescriptiveStatistics` 物件，因為有可能會使用到兩種以上的資料集，所以用 `ArrayList` 來存放。
2. 讓使用者選擇數據來源，有三種選擇，分別是 CSV、手動輸入、使用範例。
 - CSV：使用者輸入 CSV 檔的路徑，程式會自動讀取 CSV 檔的數據，並將數據存放到 `ArrayList` 中。
 - 手動輸入：使用者輸入數據欄位名、數據個數、數據，程式會自動將數據存放到 `ArrayList` 中。
 - 使用範例：程式會自動將範例數據存放到 `ArrayList` 中。

3.接著，將 `ArrayList` 中的數據進行描述性統計分析，包括平均值、中位數、標準偏差等。(最基礎的 `DescriptiveStatistics` class)

4.接著，讓使用者選擇要進行的分析，有兩種進階選擇，分別是 ANOVA、Regression

- ANOVA：讓使用者選擇要進行 ANOVA 的資料，並計算 ANOVA 的 F 值。
 1. 先用 `description()`說明 ANOVA 的概念和公式
 2. 接著，列出所有的資料集，讓使用者選擇要進行 ANOVA 的資料集，並計算 ANOVA 的 F 值。
 3. 再來，用 `summary()`提供 ANOVA 分析的摘要。
 4. 最後，用 `explain()`提供 ANOVA 分析的基本概念和使用方法的解釋。
- Regression：讓使用者選擇要進行 Regression 的資料，並計算 Regression 的斜率和截距。
 1. 先用 `description()`說明 Regression 的概念和公式
 2. 接著，列出所有的資料集，讓使用者選擇要進行 Regression 的資料集，並計算 Regression 的斜率和截距。
 3. 再來，用 `summary()`提供 Regression 分析的摘要。
 4. 用 `explain()`提供 Regression 分析的基本概念和使用方法的解釋。
 5. 最後，使用 `predict()`來預測給定 x 值的 y 值。
- HypothesisTest：讓使用者選擇要進行 HypothesisTest 的資料，並計算 HypothesisTest 的 t 值和 p 值。

```
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Scanner;
import java.io.IOException;

public class Main {
    public static void main(String[] args) {

        //有可能會使用到兩種以上的資料集
        //所以用ArrayList 來存放

        Scanner sc = new Scanner(System.in);
```

```

ArrayList<DescriptiveStatistics> statsList = new ArrayList<>();

System.out.println("請選擇數據來源 ? 1:CSV 2:手動輸入 3.使用範例")
;

int choice1 = sc.nextInt();

switch(choice1){
    case 1:
        System.out.println("輸入 csv 檔: ");
        String filePath = sc.next();
        ReadFile.readDataToStatsList(filePath, statsList);
        break;
    case 2:
        while (true) {

            System.out.println("數據欄位名: ");
            String name = sc.next();

            System.out.println("請問有幾個數據: ");
            int n = sc.nextInt();
            double[] data = new double[n];

            System.out.println("輸入數據: ");
            for (int i = 0; i < n; i++) {
                data[i] = sc.nextDouble();
            }

            DescriptiveStatistics stats = new DescriptiveStatistics
tics(data, name);
            statsList.add(stats);

            System.out.println("Do you want to enter another se
t of data? (Y/N)");
            String answer = sc.next();

            if (answer.equalsIgnoreCase("N")) {
                break;
            }
        }
        break;
    case 3:
        System.out.println("為您提供以下示範");
        double[] data1 = {1, 2, 3, 4, 5};
        double[] data2 = {2, 5, 6, 7, 8};
        double[] data3 = {3, 4, 5, 6, 7};
        statsList.add(new DescriptiveStatistics(data1, "data1"));
        statsList.add(new DescriptiveStatistics(data2, "data2"));
        statsList.add(new DescriptiveStatistics(data3, "data3"));

```

```

        System.out.println("data1: " + Arrays.toString(data1));
        System.out.println("data2: " + Arrays.toString(data2));
        System.out.println("data3: " + Arrays.toString(data3));
        System.out.println("");
        System.out.println("-----");
        System.out.println("");
        break;
    }

    System.out.println("基礎數據分析");
    System.out.println(statsList.get(0).description());
    for (DescriptiveStatistics stats : statsList) {
        System.out.println(stats.summary());
        System.out.println();
    }

    System.out.println(statsList.get(0).explain());
    System.out.println("-----");
    System.out.println("");

    System.out.println("選擇要進行的分析：");
    System.out.println("1. HypothesisTest");
    System.out.println("2. ANOVA");
    System.out.println("3. Regression");
    int choice = sc.nextInt();
    System.out.println("");

    switch(choice){
        case 1:
            hypothesisTest test = new hypothesisTest(null, null, "name");

            System.out.println("歡迎使用「假設檢定」功能，請問有需要為您進行假設檢定之概念講解嗎？(輸入 y/n)");

            if (sc.next().equals("y")) {
                test.instruction();
            }

            System.out.println("第一個步驟為「確立虛無假設」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");

            if (sc.next().equals("y")) {
                test.instruction1();
            }

            System.out.println("\n 題目是否有提供敘述統計量？(輸入 y/n)");

            String userInput = sc.next();

```

```

        if (userInput.equals("y")) {

            System.out.println("請依序輸入「虛無假設值」以及選擇「左尾、右尾、雙尾」");
            test.setNullHypo(sc.nextDouble(), sc.next());

            System.out.println("您想進行 z 檢定還是 t 檢定 (a.z 檢定 b.t 檢定(僅提供 29 筆以下之數據使用), 輸入 a/b)");

            String testChoose = sc.next();
            double t = 0;

            if (testChoose.equals("a")) {

                System.out.println("請依序輸入「樣本平均數」、「母體變異數」、「信心水準」、「母體數量」");
                test.setXbar(sc.nextDouble());
                test.setVar(sc.nextDouble());
                test.setAlpha(Math.round(100 * (1 - sc.nextDouble())) / 100);

                test.setNum(sc.nextInt());

                System.out.println("第二個步驟為「計算檢定統計量」, 請問有需要為您進行該步驟之概念講解嗎? (輸入 y/n)");

                if (sc.next().equals("y")) {
                    test.instruction2();
                }
                test.calculatePValue();

                System.out.println("第三個步驟為「針對結果進行分析」, 請問有需要為您進行該步驟之概念講解嗎? (輸入 y/n)");

                if (sc.next().equals("y")) {
                    test.instruction3();
                }
                test.analysis(t);

            } else if (testChoose.equals("b")) {

                System.out.println("請依序輸入「樣本平均數」、「樣本變異數」、「信心水準」、「母體數量」");
                test.setXbar(sc.nextDouble());
                test.setVar(sc.nextDouble());
                test.setAlpha(Math.round(100 * (1 - sc.nextDouble())) / 100);

```



```

        test.setNum(sc.nextInt());

        System.out.println("第二個步驟為「計算檢定統計量」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");

        if (sc.next().equals("y")) {
            test.instruction2();
        }
        t = test.tCalculatePValue();
        System.out.println("可得此情形下，檢定統計量為" + t + "，自由度為" + (test.getNum() - 1));

        System.out.println("第三個步驟為「針對結果進行分析」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");

        if (sc.next().equals("y")) {
            test.instruction3();
        }
        test.tAnalysis(t);
    }
    //以下是要自己手動輸入資料的程式碼
} else if (userInput.equals("n")) {

    ArrayList<Double> dataList = new ArrayList<>();

    while (true) {
        System.out.print("輸入資料（輸入 'q' 結束）：");

        String userInputnum = sc.next();

        if (userInputnum.equalsIgnoreCase("q")) {
            break; // 用戶輸入 'q'，跳出迴圈
        }

        try {
            // 嘗試將輸入轉換為 double
            double number = Double.parseDouble(userInputnum);

            dataList.add(number);
        } catch (NumberFormatException e) {
            System.out.println("請輸入有效的數字或 'q' 來結束。");
        }
    }

    hypothesisTest test1 = new hypothesisTest(dataList, null, "name");

```

```

        System.out.println("請依序輸入「虛無假設值」以及「左尾、右尾、雙尾」");
        test1.setNullHypo(sc.nextDouble(), sc.next());

        System.out.println("您想進行 z 檢定還是 t 檢定 (a.z 檢定 b.t 檢定(僅提供 29 筆以下之數據使用)，輸入 a/b");

        String testChoose = sc.next();
        double t = 0;

        if (testChoose.equals("a")) {

            System.out.println("請輸入信心水準");
            test1.setAvg(test1.mean());
            test1.setVar(test1.populationVariance());
            test1.setAlpha(Math.round(100 * (1 - sc.nextDouble())) / 100);

            System.out.println("第二個步驟為「計算檢定統計量」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");
            if (sc.next().equals("y")) {
                test1.instruction2();
            }
            test1.calculatePValue();

            System.out.println("第三個步驟為「針對結果進行分析」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");
            if (sc.next().equals("y")) {
                test1.instruction3();
            }
            test1.analysis(t);

        } else if (testChoose.equals("b")) {

            System.out.println("請輸入信心水準");
            test1.setAvg(test1.mean() * dataList.size() / (dataList.size() - 1));
            test1.setVar(test1.populationVariance() * dataList.size() / (dataList.size() - 1));
            test1.setNum(dataList.size());
            test1.setAlpha(Math.round(100 * (1 - sc.nextDouble())) / 100);

            System.out.println("第二個步驟為「計算檢定統計量」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");
            if (sc.next().equals("y")) {
                test1.instruction2();
            }
        }
    }
}

```

```

        double t1 = test1.tCalculatePValue();
        System.out.println("可得此情形下，檢定統計量
為" + t1 + "，自由度為" + test.getNum());

        System.out.println("第三個步驟為「針對結果進
行分析」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)");
        if (sc.next().equals("y")) {
            test1.instruction3();
        }
        test1.tAnalysis(t1);
    }

    }

    System.out.println("");
    break;
case 2:
    //ANOVA-1 解釋
    System.out.println("要聽一下 Anova 的概念嗎？(Y/N)");
    String answera1 = sc.next();
    while(answera1.equalsIgnoreCase("Y")){
        System.out.println("概念講解");
        double[][] dataForAnova = {{1, 2, 3, 4, 5}, {2, 5, 6,
7, 8}};

        Anova anova = new Anova(dataForAnova, "ANOVA Test");
        System.out.println(anova.description());
        System.out.println("要在聽一次嗎？(Y/N)");
        String answera2 = sc.next();
        if (answera2.equalsIgnoreCase("N")) {
            System.out.println("好的，那我們進入實戰環節！");
            System.out.println("");
            break;
        }
    }

    //ANOVA-2 實戰
    System.out.println("請選擇要做 ANOVA 的資料(輸入對應數字，輸
入-1 結束)：");
    for (int i = 0; i < statsList.size(); i++) {
        System.out.println((i + 1) + ". " + statsList.get(i).
getName());
    }

    ArrayList<Integer> indexList = new ArrayList<>();
    while(true) {
        int index = sc.nextInt() - 1;
        if (index == -2) {

```

```

        break;
    } else if (index >= 0 && index < statsList.size()) {
        indexList.add(index);
        System.out.println("已選擇 " + statsList.get(index).getName());
    } else {
        System.out.println("無效的索引，請重新輸入");
    }
}

double[][] dataForAnova = new double[indexList.size()][];
for (int i = 0; i < indexList.size(); i++) {
    DescriptiveStatistics stats = statsList.get(indexList.get(i));
    dataForAnova[i] = stats.getData();
}

if (dataForAnova.length > 1) {
    Anova anova = new Anova(dataForAnova, "ANOVA Test");
    System.out.println(anova.summary());
    System.out.println("F 值: " + anova.calculateFValue());
    System.out.println(anova.explain());
} else {
    System.out.println("至少需要選擇兩組數據進行 ANOVA 分析");
}

break;
case 3:
    //提供線性回歸的解釋
    System.out.println("要聽一下線性回歸的概念嗎？(Y/N)");
    String answer = sc.next();
    while(answer.equalsIgnoreCase("Y")){
        System.out.println("概念講解");
        double[] xData = {1, 2, 3, 4, 5};
        double[] yData = {2, 5, 6, 7, 8};
        LinearRegression reg = new LinearRegression(xData, yData, "線性回歸示範");
        System.out.println(reg.description());
        System.out.println("要在聽一次嗎？(Y/N)");
        String answer1 = sc.next();
        if (answer1.equalsIgnoreCase("N")) {
            System.out.println("好的，那我們進入實戰環節！");
            System.out.println("");
            break;
        }
    }
}
for (int i = 0; i < statsList.size(); i++) {
    System.out.println((i + 1) + ". " + statsList.get(i).

```

```
getName());  
    }  
    System.out.println("請選擇要做迴歸分析的因變數：");  
    int index1 = sc.nextInt() - 1;  
    System.out.println("請選擇要做迴歸分析的自變數：");  
    int index2 = sc.nextInt() - 1;  
    String regName = statsList.get(index1).getName() + " vs.  
" + statsList.get(index2).getName();  
    LinearRegression reg = new LinearRegression(statsList.get  
(index1).getData(), statsList.get(index2).getData(), regName);  
  
    System.out.println("");  
    System.out.println("迴歸分析結果：");  
    System.out.println(reg.summary());  
    System.out.println(reg.explain());  
    System.out.println("-----");  
    System.out.println("");  
  
    System.out.println("請問要預測 y(應變數) 值嗎？(Y/N)");  
    String answer12 = sc.next();  
    while(answer12.equalsIgnoreCase("Y")){  
        System.out.println("請輸入 y 值：");  
        double x = sc.nextDouble();  
        System.out.println("預測的 x 值為：" + reg.predict(x));  
        System.out.println("要在預測一次嗎？(Y/N)");  
        String answer13 = sc.next();  
        if (answer13.equalsIgnoreCase("N")) {  
            System.out.println("好的，再見！");  
            break;  
        }  
    }  
}  
}  
}sc.close();  
}
```

Example

1:使用範例+回歸分析

請選擇數據來源 ? 1:CSV 2:手動輸入 3.使用範例

3

為您提供以下示範

data1: [1.0, 2.0, 3.0, 4.0, 5.0]

data2: [2.0, 5.0, 6.0, 7.0, 8.0]

data3: [3.0, 4.0, 5.0, 6.0, 7.0]

基礎數據分析

以下將把每個數據集的數據進行描述性統計分析，包括平均值、中位數、標準(偏)差、樣本大小、母體方差和母體標準偏差。

數據名稱: data1

平均值: 3.0

中位數: 3.0

標準偏差: 1.4142135623730951

樣本大小: 5

母體方差: 2.0

母體標準偏差: 1.4142135623730951

數據名稱: data2

平均值: 5.6

中位數: 6.0

標準偏差: 2.0591260281974

樣本大小: 5

母體方差: 4.24

母體標準偏差: 2.0591260281974

數據名稱: data3

平均值: 5.0

中位數: 5.0

標準偏差: 1.4142135623730951

樣本大小: 5

母體方差: 2.0

母體標準偏差: 1.4142135623730951

以下為您講解公式

平均值 = $\sum x / n$

中位數 = $(x[n/2] + x[n/2+1]) / 2$

標準偏差 = $\sqrt{(\sum (x - x \text{ 平均值})^2 / n)}$

母體方差 = $\sum (x - x \text{ 平均值})^2 / n$

母體標準偏差 = $\sqrt{(\sum (x - x \text{ 平均值})^2 / n)}$

選擇要進行的分析：

1. Hypothesis Testing
2. ANOVA
3. Regression

3

要聽一下線性回歸的概念嗎？(Y/N)

Y

概念講解

以下是簡單線性回歸的概念

=> involves one independent variable and one dependent variable.

Suppose: $y = B_0 + B_1x + \varepsilon$, then $E(y) = B_0 + B_1x$, $\varepsilon \sim NID(0, \sigma^2)$

Sampling and Fitted: $\hat{y} = b_0 + b_1x$

Estimated: $\hat{y} = b_0 + b_1x \rightarrow E(y) = B_0 + B_1x$

$b_0 \rightarrow B_0$; $b_1 \rightarrow B_1$

(residual) $e = y - \hat{y} \rightarrow \varepsilon = y - E(y)$

by $\min \Sigma(y - \hat{y})^2 = \min \Sigma(y - b_0 + b_1x)$

\wedge^2

由 $\partial SSE / \partial b_0 = 0$ and $\partial SSE / \partial b_1 = 0$

得 $b_1 = \Sigma(x - \bar{x})^2 (y - \bar{y})^2 / \Sigma(x - \bar{x})$

\wedge^2

$b_0 = \bar{y} - b_1\bar{x}$

要在聽一次嗎？(Y/N)

N

好的，那我們進入實戰環節！

1. data1
2. data2
3. data3

請選擇要做迴歸分析的因變數：

1

請選擇要做迴歸分析的自變數：

2

迴歸分析結果：

線性迴歸模型 - data1 vs. data2

斜率 (beta0): 1.4

截距 (beta1): 1.4000000000000004

本數據的迴歸模型是: $\hat{y} = 1.4x + 1.4000000000000004$

Statistical (True) Model is: $y = f(x) + \varepsilon$

其中: y = 應變數 ; x = 自變數

Statistical (True) Model \rightarrow Fitted model, eg: $\hat{y} = b_1x + b_0$

利用 data 來 尋找 Y 和 X 的關係 (不一定是因果關係)

請問要預測 **y**(應變數) 值嗎？(Y/N)

Y

請輸入 **y** 值：

100

預測的 **x** 值為：**141.4**

要在預測一次嗎？(Y/N)

N

好的，再見！

2.使用 csv 檔+ANOVA

請選擇數據來源？ **1**:CSV **2**:手動輸入 **3**.使用範例

1

輸入 csv 檔：

datak.csv

基礎數據分析

以下將把每個數據集的數據進行描述性統計分析，包括平均值、中位數、標準(偏)差、樣本大小、母體方差和母體標準偏差。

數據名稱：**score**

平均值：**79.21241736022337**

中位數：**79.37917081784681**

標準偏差：**11.795080357828915**

樣本大小：**100**

母體方差：**139.1239206476415**

母體標準偏差：**11.795080357828915**

數據名稱：**course_length**

平均值：**17.5**

中位數：**17.5**

標準偏差：**1.118033988749895**

樣本大小：**100**

母體方差：**1.25**

母體標準偏差：**1.118033988749895**

以下為您講解公式

平均值 = $\Sigma x / n$

中位數 = $(x[n/2] + x[n/2+1]) / 2$

標準偏差 = $\sqrt{(\Sigma(x - x \text{ 平均值})^2 / n)}$

母體方差 = $\Sigma(x - x \text{ 平均值})^2 / n$

母體標準偏差 = $\sqrt{(\Sigma(x - x \text{ 平均值})^2 / n)}$

選擇要進行的分析：

1. 離開

- 2. ANOVA
- 3. Regression
- 2

要聽一下 Anova 的概念嗎？(Y/N)

Y

概念講解

ANOVA 通過計算 F 值來測試組間差異的顯著性。F 值是組間均方 (MSB) 和組內均方 (MSW) 的比率，其中 $MSB = \text{組間平方和(SSB)} / \text{組間自由度(dfBetween)}$ ， $MSW = \text{組內平方和(SSW)} / \text{組內自由度(dfWithin)}$ 。高 F 值通常表明組間變異顯著大於組內變異，從而指示組間存在顯著差異。

要在聽一次嗎？(Y/N)

N

好的，那我們進入實戰環節！

請選擇要做 ANOVA 的資料(輸入對應數字，輸入 -1 結束)：

- 1. score
- 2. course_length

1

已選擇 score

2

已選擇 course_length

-1

ANOVA 分析 - ANOVA Test

總體平方和 (SST): 204458.51488688402

組間平方和 (SSB): 190421.12282211994

組內平方和 (SSW): 14037.392064764077

F 值: 2685.9250026520804

ANOVA (分析變異) 用於比較三個或更多組的平均數是否有顯著差異。它將總變異分解為組間變異和組內變異，並通過 F 統計量來評估組間變異是否顯著大於組內變異。

3.手動+Hyphotesis Testing

請選擇數據來源？ 1:CSV 2:手動輸入 3.使用範例

2

數據欄位名：

data

請問有幾個數據：

5

輸入數據：

1 2 3 4 5

Do you want to enter another set of data? (Y/N)

N

基礎數據分析

以下將把每個數據集的數據進行描述性統計分析，包括平均值、中位數、標準(偏)差、樣本大小、母體方差和母體標準偏差。

數據名稱: data

平均值: 3.0

中位數：3.0

標準偏差：1.4142135623730951

樣本大小：5

母體方差：2.0

母體標準偏差：1.4142135623730951

以下為您講解公式

平均值 = $\Sigma x / n$

中位數 = $(x[n/2] + x[n/2+1]) / 2$

標準偏差 = $\sqrt{(\Sigma(x - x \text{ 平均值})^2 / n)}$

母體方差 = $\Sigma(x - x \text{ 平均值})^2 / n$

母體標準偏差 = $\sqrt{(\Sigma(x - x \text{ 平均值})^2 / n)}$

選擇要進行的分析：

1. HypothesisTest

2. ANOVA

3. Regression

1

歡迎使用「假設檢定」功能，請問有需要為您進行假設檢定之概念講解嗎？(輸入 y/n)
y

假說檢定（英語：hypothesis testing）是推論統計中用於檢定現有數據是否足以支持特定假設的方法。

一旦能估計未知母數，就會希望根據結果對未知的真正母數值做出適當的推論。

欲檢定統計上假設的正確性的為虛無假說，虛無假說通常由研究者決定，反映研究者對未知母數的看法。

相對於虛無假說的其他有關母數之論述是對立假說，它通常反應了執行檢定的研究者對母數可能數值的另一種（對立的）看法

第一個步驟為「確立虛無假設」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)
y

以下說明此步驟之意義：

在檢定之前，我們需要針對情境進行假設。其中包含「虛無假設」以及「對立假設」兩種情況：

「虛無假設」對母體參數提出一個主張，假設此主張為「真實」（除非能證明此主張非真！）。

對立假設是相對於虛無假設所提出的另一個不同(相反)的假設或主張，必須有足夠的證據，才能說明此主張為真。

檢定方式又因虛無假設之不同，分為三種檢定：

雙尾檢定： $H_0 : \theta = \theta_0$; $H_1 : \theta \neq \theta_0$

左尾檢定： $H_0 : \theta \geq \theta_0$; $H_1 : \theta < \theta_0$

右尾檢定： $H_0 : \theta \leq \theta_0$; $H_1 : \theta > \theta_0$

題目是否有提供敘述統計量？(輸入 y/n)

n

輸入資料 (輸入 'q' 結束) : 1 2 3 4 5 q

輸入資料 (輸入 'q' 結束) : 輸入資料 (輸入 'q' 結束) : 輸入資料 (輸入 'q' 結束) : 輸入資料 (輸入 'q' 結束) : 輸入資料 (輸入 'q' 結束) : 請依序輸入「虛無假設值」以及「左尾、右尾、雙尾」

10 雙尾

H0: $\mu=10.0$ H1: $\mu\neq10.0$

您想進行 z 檢定還是 t 檢定 (a.z 檢定 b.t 檢定(僅提供 29 筆以下之數據使用))，輸入 a/b

a

請輸入信心水準

3

第二個步驟為「計算檢定統計量」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)

y

檢定統計量是由樣本所算出來的一個值，用來決定是否接受或拒絕 H0。

不同參數的假設檢定，使用不同的檢定統量。常用的檢定統計量有：Z, t。

在母體平均數的假設檢定裡，不同的情形下使用不同的檢定統計量。

若母體變異數已知，則使用 $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

，其中 \bar{x} 代表檢定量、 μ 代表母體平均數、 σ 代表母體標準差、n 代表母體樣本數量若母體變異數未知，但樣本數超過 30，可將樣本視為一母體，同樣使用 $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

若母體變異數未知，但樣本數小於 30，則同樣使用 $t = (\bar{x} - \mu) / (s / \sqrt{n})$ ，其中 s 代表樣本標準差

可得此情形下，檢定統計量為 -10.0

第三個步驟為「針對結果進行分析」，請問有需要為您進行該步驟之概念講解嗎？(輸入 y/n)

y

研究人員必須決定一個決策法則，以瞭解何時『不拒絕』 H0 ；何時拒絕 H0 。

一般我們說『不拒絕』 H0 ，而不說接受 H0 ，因為我們只是沒有足夠證據拒絕，而不是接受。

決策法則通常是決定一個不拒絕域 (NonrejectionRegion, 或稱接受域) 與拒絕域 (Rejection Region)。

當檢定統計量落入不拒絕域：『不拒絕』 H0 ！

當檢定統計量落入拒絕域：拒絕 H0 ；接受 H1

接受域與拒絕域的接點，稱為臨界點(Critical Point)。

臨界值的決定，是根據顯著水準 α 並利用機率分配計算而得，分成三種形式：

雙尾檢定 (落在兩邊拒絕)、右尾檢定 (落在右邊拒絕)、左尾檢定 (落在左邊拒絕)

在 z 檢定時，我們也會經由計算其 p-value 並與 alpha 值比較，並得到結果

一般來說，當 p-value 小於 alpha，則拒絕 H0

在此情境中，由於 p-value(0.5)大於 alpha(-2.0)，可知此資料無法拒絕 H0 之假設

其他資料

Github: https://github.com/blingblingdong/Java_statistic

Web App: <https://javashinyapp.fly.dev>

Presentation: <https://java.lsyverycute.com>