

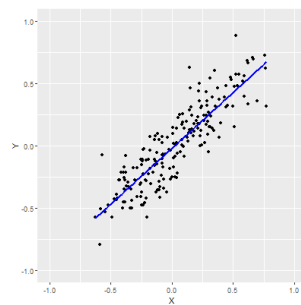
統計分析法 第8週レポート

202212022 田島瑞起

2023/12/12

1 設問 1-1,1-3

week8-data-odd.csv のデータを散布図としてプロットした上に、回帰直線を追加したものが下記の図となる。



2 設問 1-2

```
1  #結果
2  Call:
3  lm(formula = y ~ x)
4
5  Residuals:
6      Min       1Q   Median       3Q      Max
7  -0.37165 -0.10421  0.00426  0.11478  0.51488
8
9  Coefficients:
10             Estimate Std. Error t value Pr(>|t|)
11 (Intercept) -0.01775    0.01116  -1.591   0.113
12 x           0.89836    0.03511  25.588 <2e-16 ***
13 ---
14 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
15
16 Residual standard error: 0.1573 on 198 degrees of freedom
17 Multiple R-squared:  0.7678,    Adjusted R-squared:  0.7666
```

3 設問 1-4

1-2 で示した summary の値を読み取ると、回帰直線の傾き a は 0.89836、切片 b は -0.01775、残差標準偏差 0.1573、自由度については回帰の自由度: 1 残差の自由度: 198、データ数 200、決定係数 0.7678 であることが分かる。

4 設問 1-5

今回のデータセットでは相関係数が 0.8762453 でありこの値は決定係数の 0.7678 と近い値であることが分かる。この二つの係数に直接的な数式的関係があるわけではないが、近い値をとったことには理由が存在し、それは相関係数が二つの変数がどれだけ強く関係しているかを示す値であり、一方決定係数は回帰モデルがデータをどれほど正確に表現しているかを表す値であるからと考えられる。今回の場合はデータが直線に近い形であったため、相関係数は 1 に近い値をとり、回帰モデルによってデータの集合を表現することに問題がなかったため、両者の値が高くなったと考えることが出来る。

5 設問 1-6

相関係数が 0.8762453 であり、また残差標準偏差 0.1573、決定係数が 0.7678 であることから、この回帰直線はデータを良く反映していると判断することが出来るため、x と y には正の相関関係があると判断することが出来る。

6 設問 2-1

帰無仮説を「得点は教科ごとに差がない」、有意水準 0.05 として aov 検定を行う。学生と科目に対応がない場合の aov 分析の結果は下記の図の通りとなった。p 値が優位水準を下回るため、帰無仮説は棄却され、得点は教科ごとに差があると結論付けることが出来る。また分散説明率を計算すると 0.9997588 と算出された。よってモデルがデータの変動をよく説明していることを示している

```

1
2      Df Sum Sq Mean Sq F value    Pr(>F)
3 sub           2    3344   1671.8    5.618 0.00593 **
4 Residuals    57   16962    297.6
5 ---
6 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

7 ソースコード

```
1      #課題1
2
3      #データの下準備
4      setwd('Z:/stats_work')
5      Data <- read.csv("week8-data-odd.csv")
6      x <- Data$X
7      y <- Data$Y
8
9      #散布図の描画
10     library(ggplot2)
11     plot <- ggplot(Data, aes(x = X, y = Y)) +
12       geom_point() +
13       xlim(-1.0, 1.0) +
14       ylim(-1.0, 1.0)
15
16     #散布図に回帰直線を追加した後に画像化
17     model <- lm(Y ~ X, data = Data)
18     png("8-1.png", width = 400, height = 400)
19     plot + geom_smooth(method = "lm", se = FALSE, color = "blue")
20     dev.off()
21
22     #相関係数を求める
23     Correlation <- cor.test(x,y,method="pearson")
24
25     # 単回帰分析の実行
26     model <- lm(y ~ x)
27
28     # 回帰結果の表示
29     summary(model)
30
31     #課題2
32
33     # データの読み込み
34     load("effect-example.dat")
35
36     # aov関数を用いた分散分析
37     model <- aov(score ~ sub, data = scoreDataES)
38
39     # 結果の表示
40     summary(model)
41
42     # 分散分析の結果から必要な情報を取得
43     SS_total <- sum((scoreDataES$score - mean(scoreDataES$score))^2)
44     SS_residual <- sum(residuals(model)^2)
45
46     # 分散説明率の計算
47     variance_explained <- (SS_total - SS_residual) / SS_total * 100
48
49     # 結果の表示
```

```
50 cat("Variance_Explained:", variance_explained, "%\n")
```