

R 入門 (2) 回帰分析

0 準備

ホームページ上の `yabu23.xlsx` の sheet "01" を `01.csv` として working directory (c:\dail または ShortCut"R"(SCR) に保存

R を起動して、作業フォルダを SCR にしてください。

`setwd("~/dail")`

1 データの読み込み

`d <- read.table("01.csv", header = TRUE, sep = ",")`

`d` data の確認

`attach(d)` 変数名から `d$` を省略する

2 回帰分析

`lm(gay~support)` 回帰分析 lm は linear model の略

`lm(gay~0+support)` 切片なし回帰

`result<-lm(gay~support)` 結果を result に保存

`summary(result)` 推定結果の表示

Residuals : 残差の分布

Coefficients : 係数の推定結果

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	係数の	標準誤差	t 値	P 値
support	推定値 b	の推定値 s_b	$t=b/s_b$	

Residual standard error : 残差 e の標準誤差 s

R-Squared : 決定係数 Adjusted R-Squared : 自由度修正済決定係数

F-statistic : すべての係数 = 0 の検定統計量 (F 値) F の p 値

`anova(result)` 分散分析表の表示

	自由度	2 乗和	平均 2 乗和	F 値	P 値
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
support	1	$ESS = \sum (\hat{y} - \bar{y})^2$	ESS/k	$\frac{ESS/(k-1)}{RSS/(n-k)}$	

Residuals $n-k-1$ $RSS = \sum e^2$ $RSS/(n-k-1)$

`yh=predict(result); e=residuals(result)` 予測値と残差を保存

`data.frame(d,yh,e)` データフレーム d に yh と e を追加

`plot(support, gay)` X-Y プロット図

`abline(result)` 回帰直線を追加

`plot(resid(result))` 残差のプロット

演習

1. Sheet "04" 家賃データを利用し、家賃を専有面積、距離、築年数、階数でそれぞれ単回帰をおこなって、それぞれを下記のように整理しなさい (有効数字 4 行)。

家賃 = $2.688 + 0.1603$ 専有面積 決定係数 $R^2 = 0.7566$

(26.72)(47.37) 括弧内は t 値

残差 2 乗和 : 886.3 標準誤差 : 1.108

2. 1 の結果について下記の問い合わせに答えなさい

(1) 回帰係数 b の符号はあなたの想定通りか?

(2) 家賃を最も良く説明する要因はなにか?

(3) ほかにどのような要因を考えるべきか?

```

setwd("/dail")
dat=read.table("04.csv",header = TRUE, sep =",")
attach(dat)
reg1=lm(rent~space)
reg2=lm(rent~distance)
reg3=lm(rent~age)
reg4=lm(rent~floor)
summary(reg1)
anova(reg1)
summary(reg2)
anova(reg2)
summary(reg3)
anova(reg3)
summary(reg4)
anova(reg4)

```

1

家賃=2.688+0.1603 専有面積 決定係数 R2=0.7566

(26.72)(47.37) 括弧内は t 値

残差 2 乗和 : 886.3 標準誤差 : 1.108

家賃=6.032+0.09297 距離 決定係数 R2=0.03905

(29.78)(5.417) 括弧内は t 値

残差 2 乗和 : 3499 標準誤差 : 2.201

家賃=9.028-0.07429 築年数 決定係数 R2=0.1332

(44.15)(-10.53) 括弧内は t 値

残差 2 乗和 : 3156 標準誤差 : **2.091**

家賃=6.145+0.4494 階数 決定係数 R2=0.03948

(33.59)(5.448) 括弧内は t 値

残差 2 乗和 : **3947** 標準誤差 : 2.201

2

(1)専有面積(+)、築年数(-)、階数(+)は想定通りだが、距離(+)は想定外

(2)決定係数が最も大きい専有面積

(3)部屋の日当たり、近隣の施設、部屋の設備など

回帰分析を R で実行できましたか？

單回帰は統計学でも学習した、最も基本的な分析手法です。

まずはご自分で動かせるようにしましょう。

t 値はどれも 2 より大きく、有意であることが判りますが、専有面積以外は、決定係数が小さく、家賃を説明する主要因ではないことが判ります。