インサイト・ファクトリー 社内セミナー

2020年4月

マーケティング・ミックス・モデリング

III. 回帰分析の基礎

小野 滋

INSIGHT FACTORY

2020/04/07

スケジュール

統計学・データ解析	マーケティング・ミックス・モデリングへの適用
I. イントロダクション	
	II. 市場反応モデルとは
III. 回帰分析の基礎	
	IV. 静学的市場反応モデル
V. 時系列分析の基礎	
	VI. 動学的市場反応モデル(1)
VII. 状態空間モデルの基礎	
	VIII. 動学的市場反応モデル(2)

目次

この章の内容

INSIGHT FACTORY

回帰分析についておさらいします

この章の内容は、以降のすべての議論の基礎になります

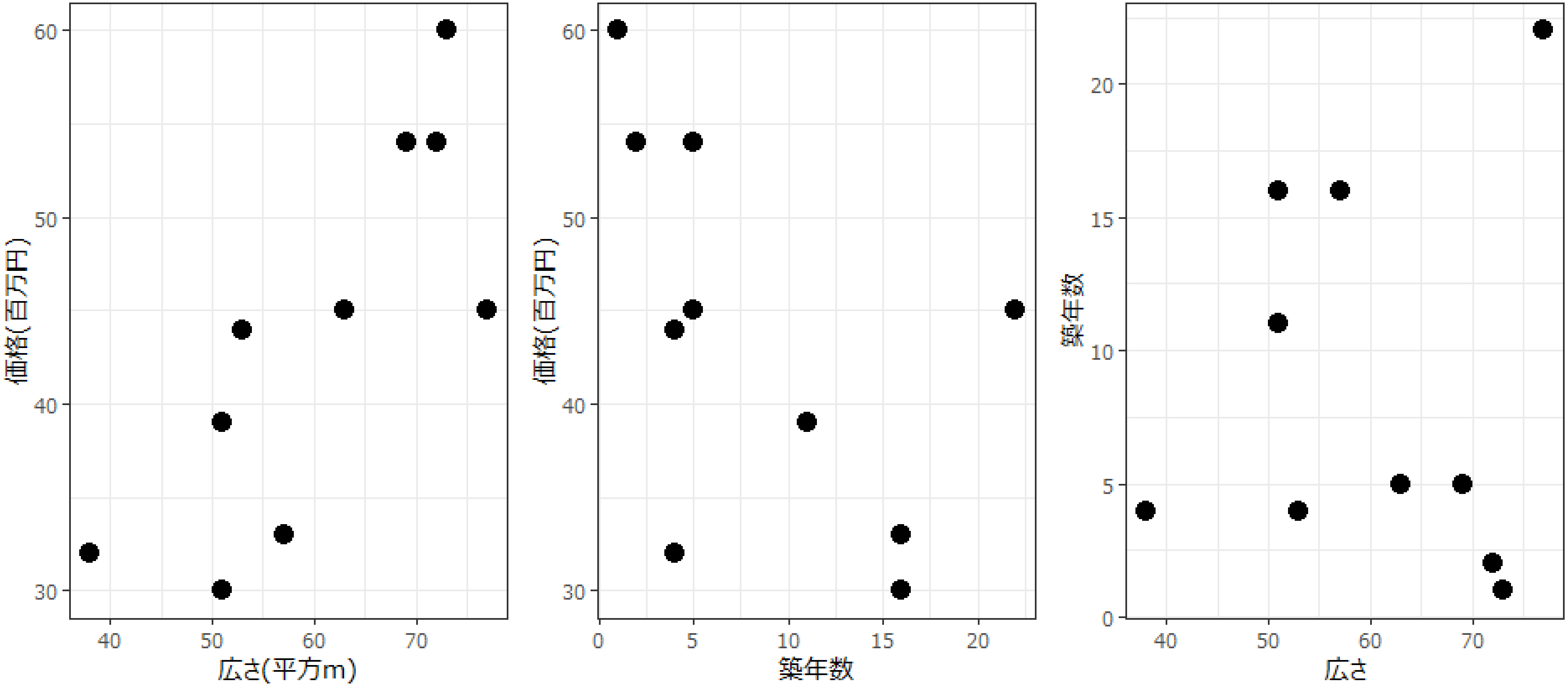
本資料作成に用いたすべてのRコードを、以下で公開しています:
https://rpubs.com/shig_ono/MMM_3

サンプルデータ：中古マンション

Code 1

ある不動産業者は、ある町の中古マンションの価格と、その特徴との関係を知りたいと考えている。

そこで、この町にある中古マンションから10件をランダムに抜き出し、価格・広さ・築年数を調べた。



データ出典: 永田・棟近(2001)

1. 基本的な統計量

1. 基本的な統計量

- **期待値** (平均) $\mu_X = \mathrm{E}(X_i)$
- **分散** $\mathrm{Var}(X_i) = \mathrm{E}[(X_i - \mu_X)^2]$
- **共分散** $\mathrm{Cov}(X_i, Y_i) = \mathrm{E}[(X_i - \mu_X)(Y_i - \mu_Y)]$
- **相関** $\mathrm{Corr}(X_i, Y_i) = \dfrac{\mathrm{Cov}(X_i, Y_i)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(Y_i)}}$

(確率変数)

	X_1	Y_1
	X_2	Y_2
⋮	⋮	⋮

標本統計量

■ **平均**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

■ **分散**

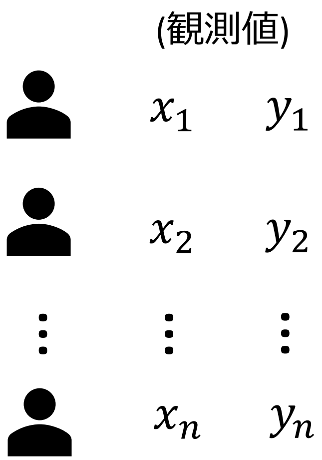
$$s_x^2 = \frac{1}{n-1}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{\text{変動}}$$

■ **共分散**

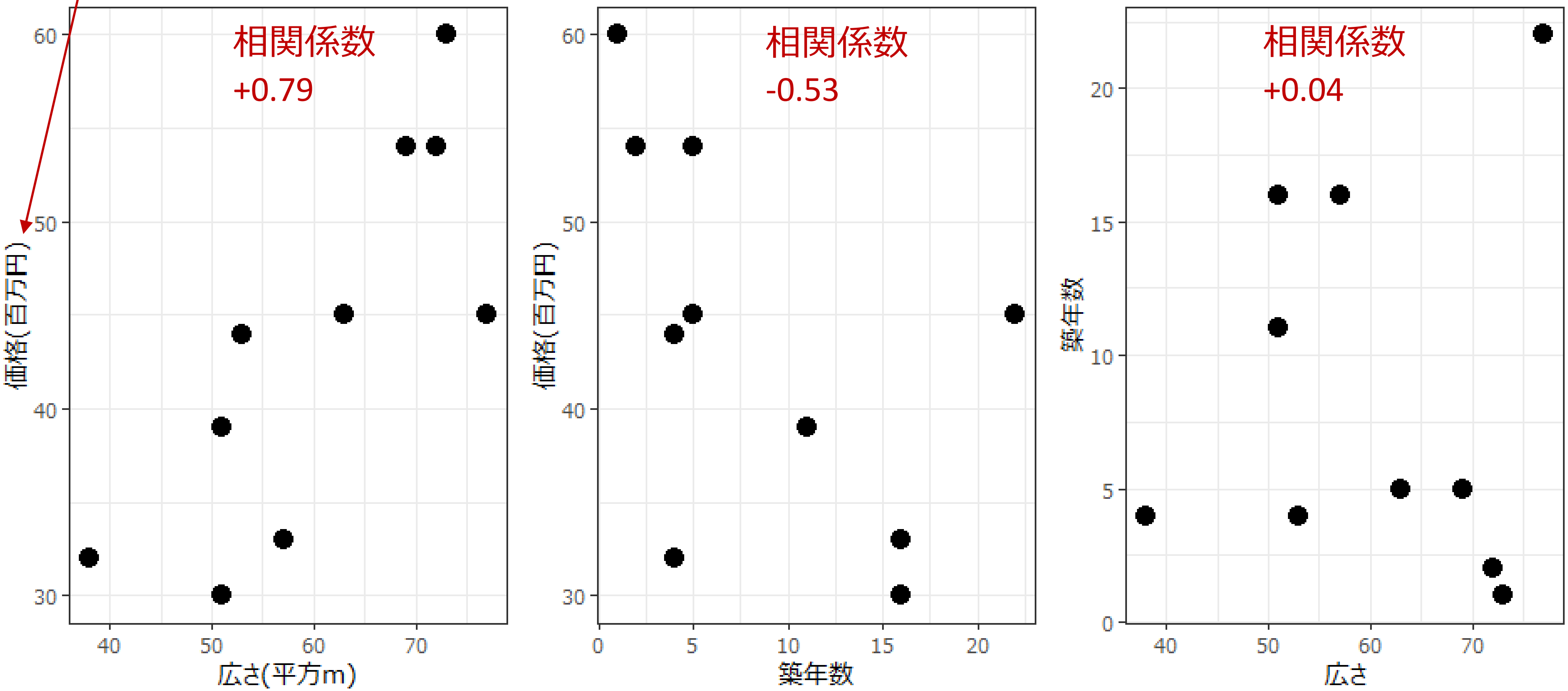
$$\mathrm{cov}_{\mathrm{x,y}} = \frac{1}{n-1}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}_{\text{共変動}}$$

■ **相関**

$$\mathrm{r_{xy}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$



標本平均
43.6 (百万円)



標本平均
60.4(平方m)

標本平均
8.6 (年)

参考：基本的な式変形 (あとで使います)

- 変動の変形

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2$$

$$= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum x_i^2 - n\bar{x}^2$$

$$\sum x_i = n\bar{x}$$

- 共変動の変形

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \bar{y}\sum x_i - \bar{x}\sum y_i + n\bar{x}\bar{y}$$

$$= \sum x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$$

$$= \sum x_i y_i - n\bar{x}\bar{y}$$

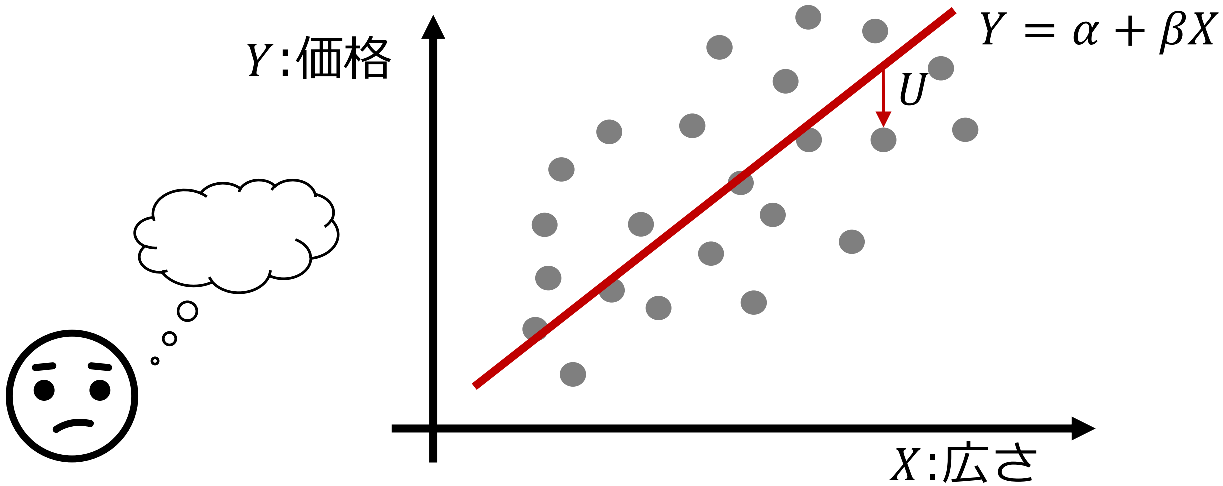
$$\sum x_i = n\bar{x}$$

$$\sum y_i = n\bar{y}$$

2. 単回帰モデル

説明変数がひとつしかない線形回帰モデル

$$\overset{\text{目的変数}}{Y_i} = \underbrace{\underset{\text{切片}}{\alpha} + \underset{\text{回帰係数}}{\beta}}_{\text{パラメータ}} \overset{\text{説明変数}}{X_i} + U_i, \qquad i = 1, \dots, n$$



■ 撹乱項とは？

撹乱項

$$Y_i = \alpha + \beta X_i + U_i, \qquad i = 1, \ldots, n$$

- Yの変動のうち、説明変数 X で説明できないすべての変動を表す項
  - 価格の測定誤差, マンションの築年数の効果, 駅からの距離の効果, … etc.

- それ自体は観測できない
  - パラメータ α, βを推定できれば、その推定を使って推定できる
  - 撹乱項の推定値を残差と呼ぶ

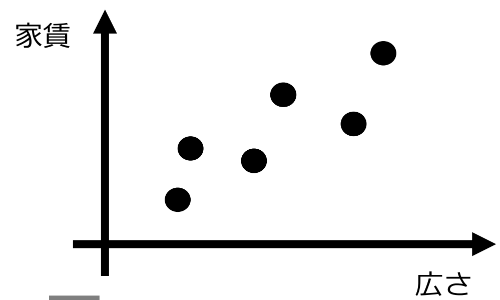
- 誤差項とも呼ばれる

■ 単回帰モデルは、私たちになにを与えてくれるのか？

設計したモデル

$$Y_i = \alpha + \beta X_i + U_i$$

データ



推定したモデル

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + E_i$$

1) **パラメータ推定量** $\hat{\alpha}, \hat{\beta}$

2) 所与のX_iのもとでのY_iの**予測** $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

1) パラメータ推定量

設計したモデル $Y_i = \alpha + \beta X_i + U_i$

推定 推定

推定したモデル $Y_i = \hat{\alpha} + \hat{\beta} X_i + E_i$

回帰係数 β の意味：

- 「説明変数 X が1変化すると、目的変数 Y はどれだけ変化するか」
- β は「因果的効果」か？
  - すなわち、β は X の変化が引き起こす Y の変化の大きさを表しているのか？
  - とても大きな問題。このセミナーでは十分に扱えません

2) 所与のX_iのもとでのY_iの予測

$$\mathrm{E}(Y_i|X_i)$$

↑ **推定**

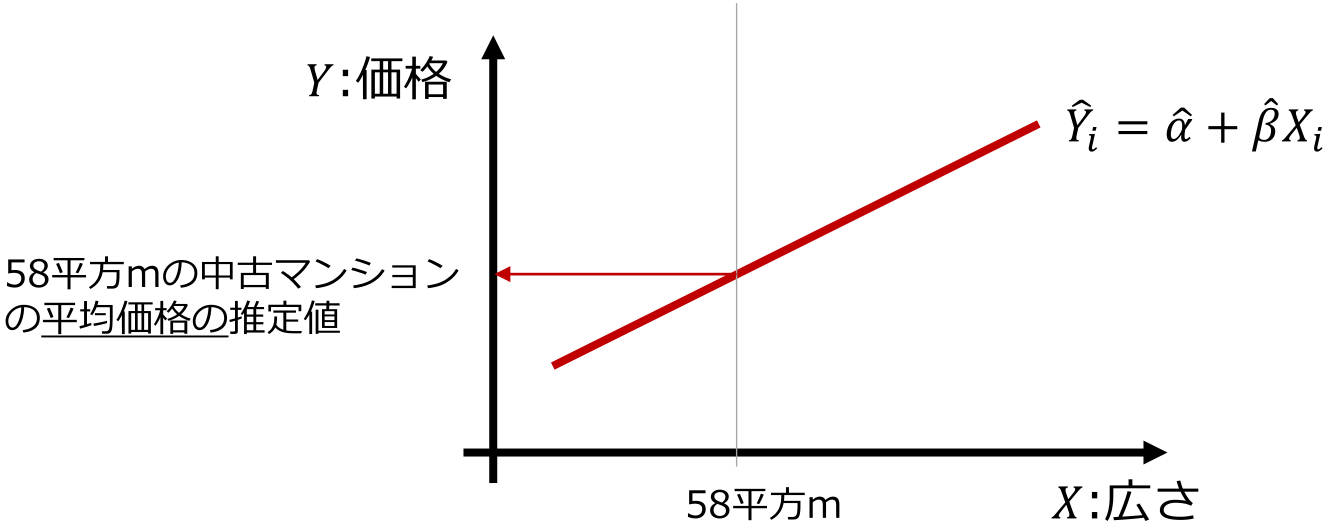
推定したモデル

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

$\hat{Y}_{\mathrm{i}}$の意味:

- 説明変数 X が所与であるときの、目的変数 Y の条件つき期待値の推定値
  - 「説明変数の値がコレコレであるとき、目的変数の値はコレコレになるはずです」という意味ではない
  - 「説明変数の値がコレコレであるとき、目的変数の値は平均するとコレコレになるはずです」という意味

中古マンション


Y:価格
$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
58平方mの中古マンションの平均価格の推定値
58平方m
X:広さ

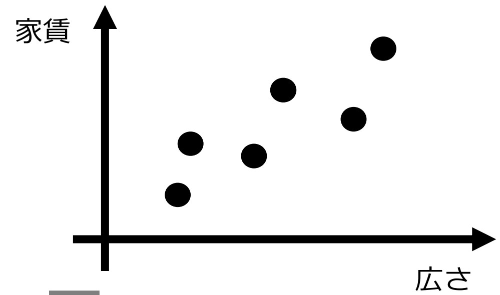
3. 単回帰モデルのパラメータ推定

3. 単回帰モデルのパラメータ推定

設計したモデル

$$Y_i = \alpha + \beta X_i + U_i$$

データ



推定したモデル

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + E_i$$

パラメータ推定量

パラメータ推定の主な方法 (=推定量のタイプ)

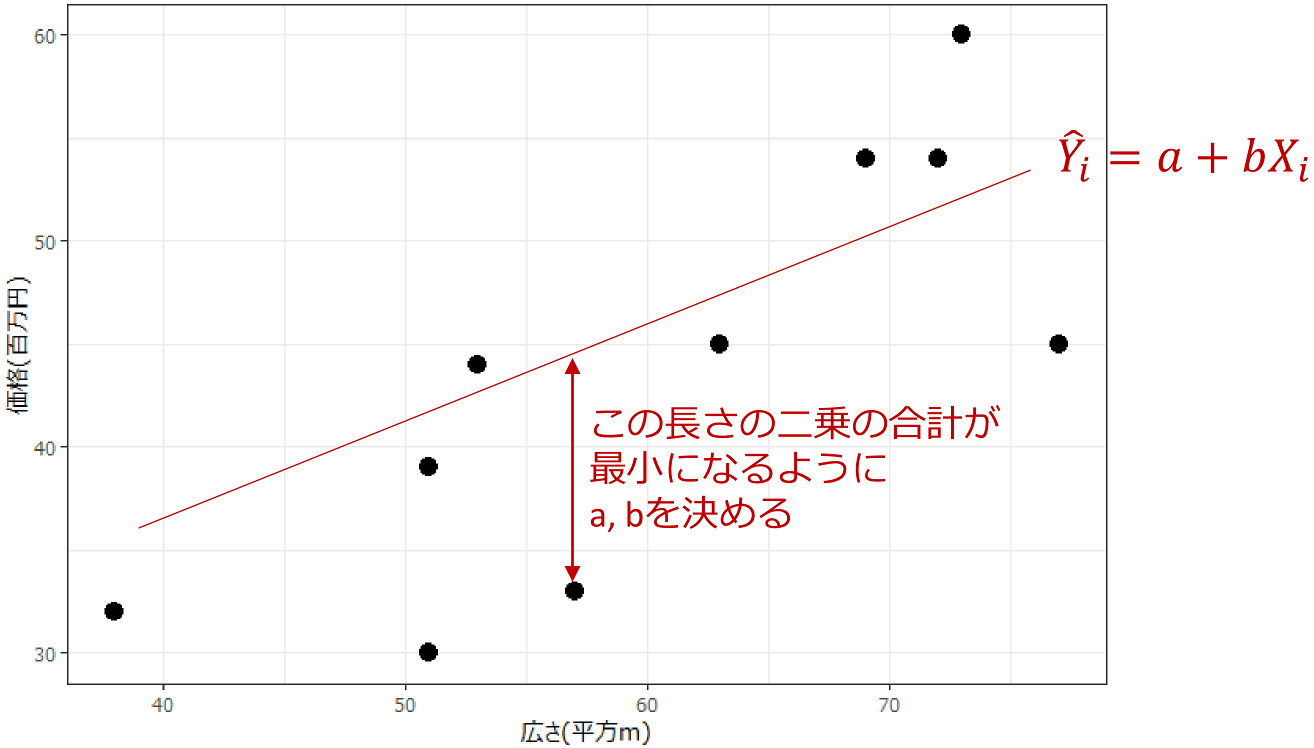
1. **最小二乗法**
2. **最尤法**
3. モーメント法
4. ベイズ法

このセミナーでは扱わない

3-1. 最小二乗法 (oridnary least squares; OLS)

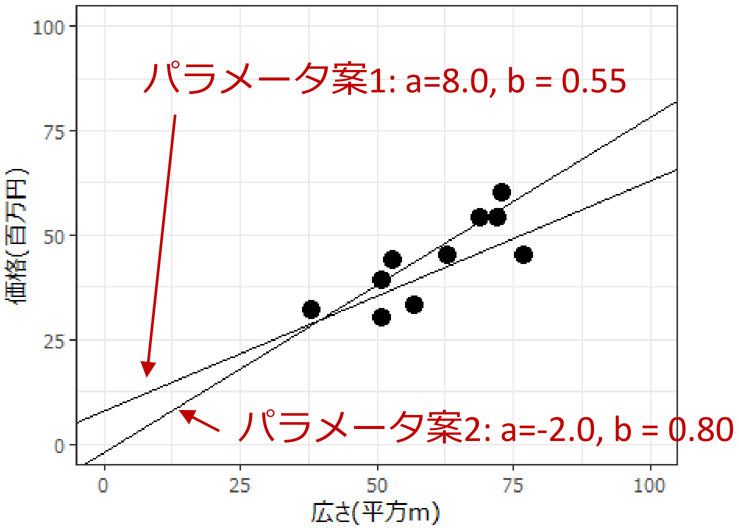
■ 考え方

- Y_i の予測値を $\hat{Y}_i = a + bX_i$ とする
- 残差 (撹乱項の推定値) は $Y_i - \hat{Y}_i$
- 残差二乗和 $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ を最小にする a, b を $\hat{\alpha}, \hat{\beta}$ としよう！



Code 2

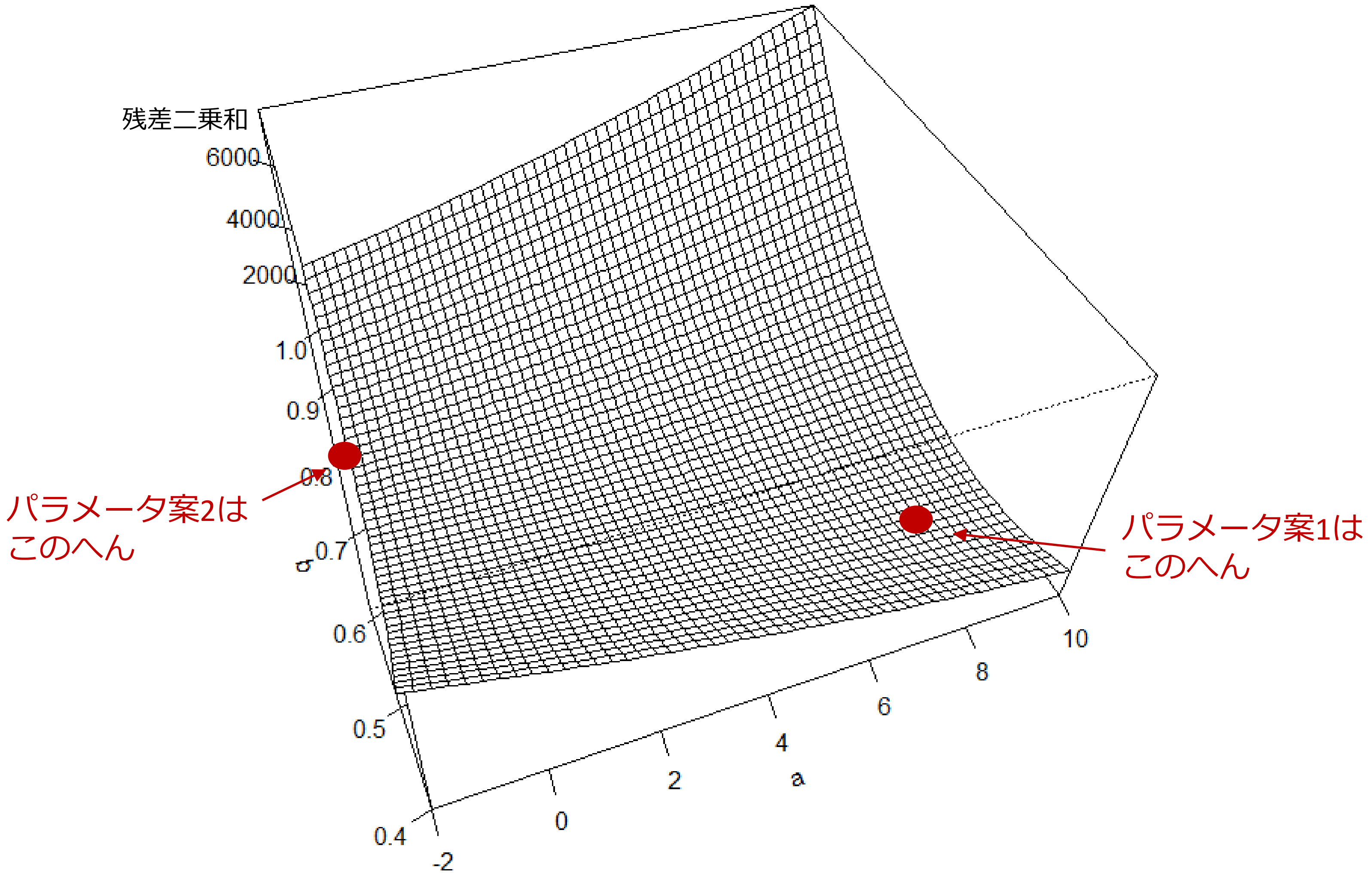
2通りの(a, b)について試してみると...



No	広さ	価格	パラメータ案1 予測値	パラメータ案1 残差の二乗	パラメータ案2 予測値	パラメータ案2 残差の二乗
1	51	30	36.1	36.6	38.8	77.4
2	38	32	28.9	9.6	28.4	13.0
3	57	33	39.4	40.3	43.6	112.4
4	51	39	36.1	8.7	38.8	0.0
5	53	44	37.2	46.9	40.4	13.0
6	77	45	50.4	28.6	59.6	213.2
7	63	45	42.7	5.5	48.4	11.6
8	69	54	46.0	64.8	53.2	0.6
9	72	54	47.6	41.0	55.6	2.6
10	73	60	48.2	140.4	56.4	13.0
		合計		**422.5**		**456.6**

Code 3

さまざまな(a, b)について計算してみると…



■ 推定量の導出 (蓑谷, pp.10-13)

a, bを変数と考え、関数 $Q(a,b) = \sum_{i=1}^{n}\left(Y_i - (a + bX_i)\right)^2$を最小化する。

a, bに関する偏導関数を0とおく。

$$\frac{\partial}{\partial a} Q(a,b) = \frac{\partial}{\partial a}\sum (Y_i - a - bX_i)^2 = \sum [2(Y_i - a - bX_i)(-1)] = 0$$

$$\frac{\partial}{\partial b} Q(a,b) = \frac{\partial}{\partial b}\sum (Y_i - a - bX_i)^2 = \sum [2(Y_i - a - bX_i)(-X_i)] = 0$$

合成関数の偏導関数は

$$\frac{\partial}{\partial a} f\left(g(a,b)\right) = f'(g(a,b))\frac{\partial}{\partial a} g(a,b)$$

両辺を2で割って

$$\sum -(Y_i - a - bX_i) = 0$$

$$\sum -X_i(Y_i - a - bX_i) = 0$$

移項して

$$\left.\begin{aligned} an + b\sum X_i &= \sum Y_i \\ a\sum X_i + b\sum X_i^2 &= \sum X_i Y_i \end{aligned}\right\}$$ **正規方程式**

$\bar{X} = \frac{1}{n}\sum X_i$, $\bar{Y} = \frac{1}{n}\sum Y_i$とする。第1式の両辺を n で割って

$$a = \bar{Y} - b\bar{X}$$

$Y = a + bX$は$(\bar{X}, \bar{Y})$を通る直線

第2式に代入して

$$(\bar{Y} - b\bar{X})\sum X_i + b\sum X_i^2 = \sum X_iY_i$$

$$b\left(\sum X_i^2 - \bar{X}\sum X_i\right) = \sum X_iY_i - \bar{Y}\sum X_i$$

$$b = \frac{\sum X_iY_i - \bar{Y}\sum X_i}{\sum X_i^2 - \bar{X}\sum X_i} = \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

変動・共変動の変形

直線の傾きは
(XとYの共変動) / (Xの変動)

Code 4

中古マンション

Rで実行してみると…

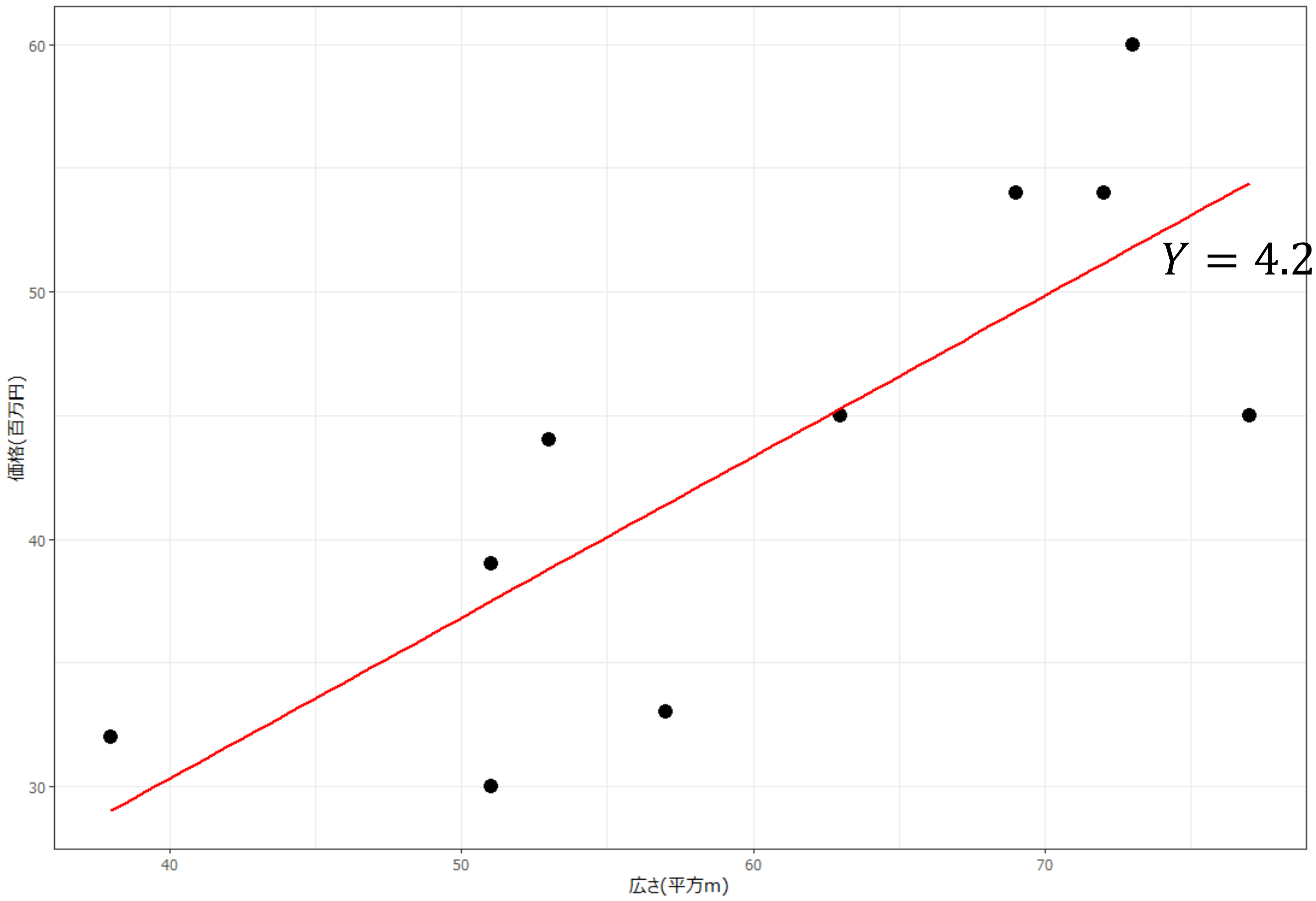
```
> dfMansion <- read_csv("./Mansion.csv")

> x <- dfMansion$Hirosa
> y <- dfMansion$Price

> b <- sum( (x - mean(x))*(y - mean(y))) / sum( (x - mean(x))^2 )
> a <- mean(y) - b * mean(x)

> print(a)
[1] 4.286288

> print(b)
[1] 0.6508893
```

Code 4

lm()関数を使うと...

```
> summary(lm(Price ~ Hirosa, data = dfMansion))

Call:
lm(formula = Price ~ Hirosa, data = dfMansion)

Residuals:
   Min     1Q Median     3Q    Max
-9.405 -5.684  2.184  4.347  8.199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2863    10.9270   0.392  0.70511
Hirosa        0.6509     0.1775   3.666  0.00635 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.63 on 8 degrees of freedom
Multiple R-squared:  0.6269,	Adjusted R-squared:  0.5802
F-statistic: 13.44 on 1 and 8 DF,  p-value: 0.006346
```

Code 5

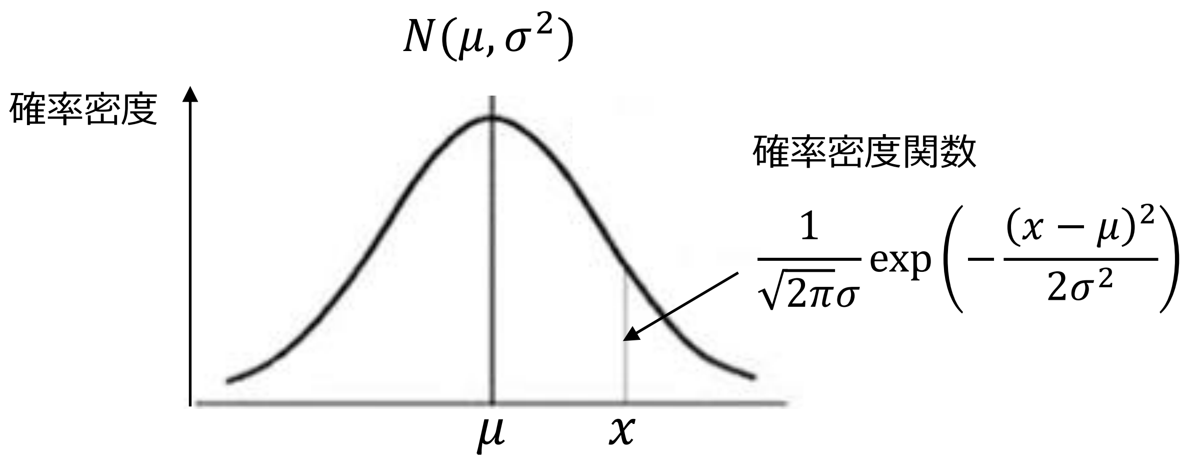


$$Y = 4.29 + 0.65X$$

INSIGHT FACTORY

Code 6

■ 準備1. 正規分布



Quiz:

ある確率変数が平均10, 分散4の正規分布に従うとき、実現値12の確率密度は？

Rで求めてみると…

```
> 1/( sqrt(2*pi) * 2) * exp(-(12-10)^2 / (2*4) )
[1] 0.1209854

> dnorm(12, 10, 2)
[1] 0.1209854
```

Quiz: ある確率変数が平均10, 分散4の正規分布に従うとしよう。互いに独立な実現値を3つ得たとき、それらが(12, 8, 9)となる確率密度 (同時確率密度) は？

互いに独立な出来事が同時に起きる確率は、個々の出来事の確率の積だから、

```
> dnorm(12, 10, 2) * dnorm(8, 10, 2) * dnorm(9, 10, 2)
[1] 0.002576671
```

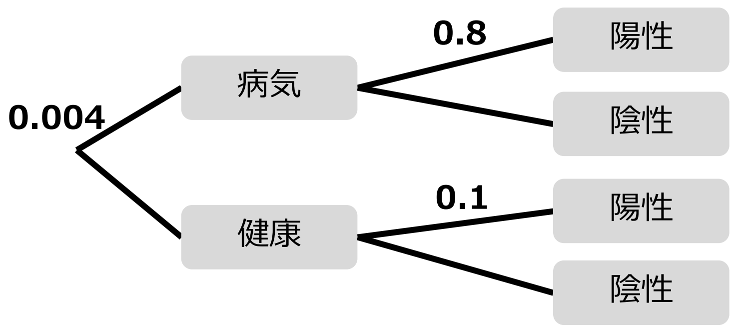
■ 準備2. 尤度とは

- Bであるとき A となる条件つき確率を $P(A|B)$とする
- これを A の関数とみなすこともできる。このときこれを**尤度** $L(B|A)$ と呼ぶ

例) ある国では、 1万人当たり40人が、ある病気にかかっている。

- この病気にかかっている人が検診を受けると、確率0.8で陽性となる。
- この病気にかかっていない人が検診を受けると、確率0.1で陽性となる。

ある人が陽性であるとき、その人が病気にかかっている条件つき確率は？



P(病気&陽性)=0.8x0.004=0.0032

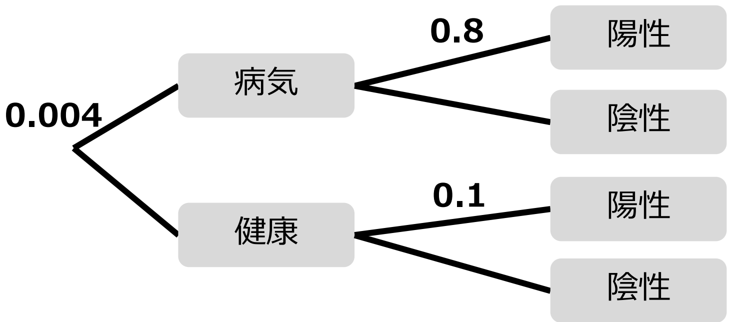
P(健康&陽性)=0.1x0.996=0.0996

$$P(病気|陽性)= \frac{0.0032}{0.0032+0.0996} =0.031$$

これを陽性の下での病気の**条件つき確率**、ないし病気の下での陽性の**尤度**という

なぜわざわざ「尤度」というの？ 「条件つき確率」といえばいいじゃない？

- 「確率」というからには、すべての出来事を通じた合計は1になるべきだ
- $P(A|B)$をBの関数とみなしたとき(=Bが所与だとみなしたとき)
  - すべてのAを通じたその合計は1になる。「確率」と呼ぶにふさわしい
- $P(A|B)$をAの関数とみなしたとき(=Aが所与だとみなしたとき)
  - すべてのBを通じたその合計は1にならない。「確率」と呼ぶわけにはいかない



P(病気&陽性)=0.8x0.004=0.0032

P(病気&陰性)=0.2x0.004=0.0008

P(健康&陽性)=0.1x0.996=0.0996

P(健康&陰性)=0.9x0.996=0.8964

$$\mathrm{P(病気|陽性)}=\frac{0.0032}{0.0032+0.0996}=0.031$$

$$\mathrm{P(健康|陽性)}=\frac{0.0996}{0.0032+0.0996}=0.969$$

合計は1

$$\mathrm{L(陽性|病気)}=\frac{0.0032}{0.0032+0.0996}=0.031$$

$$\mathrm{L(陰性|病気)}=\frac{0.0008}{0.0008+0.8964}=0.0009$$

合計は1にならない

Code 6

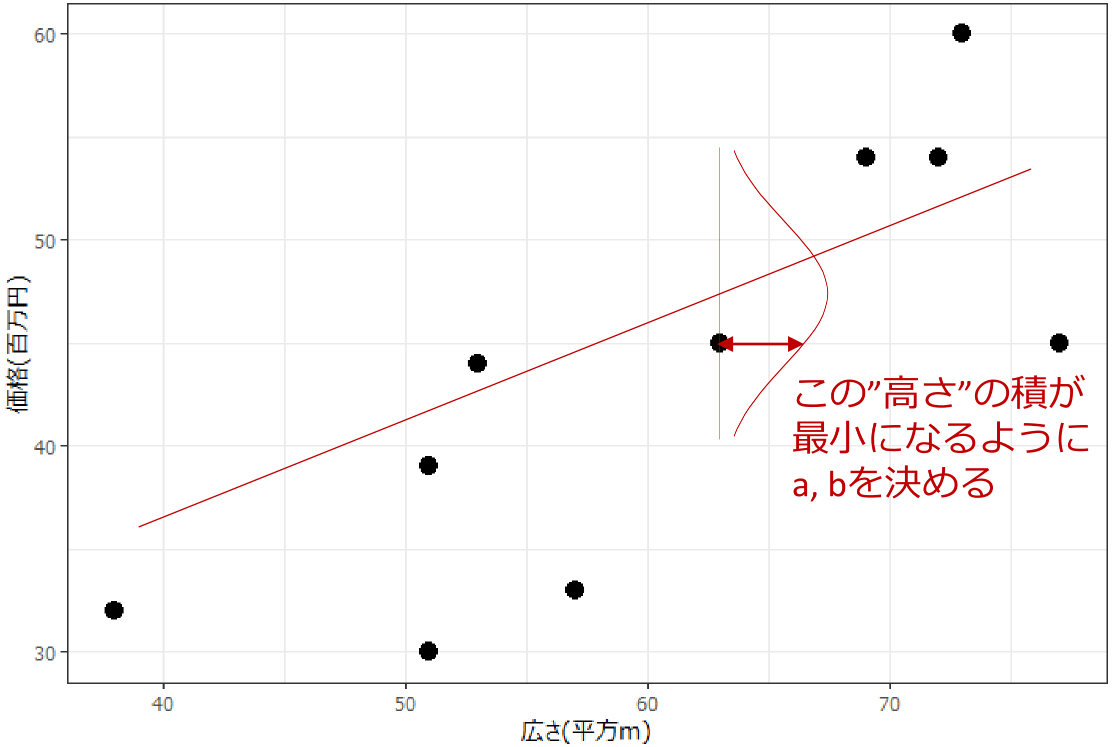
Quiz:

正規分布に従うことはわかっているが、平均と分散は未知の確率変数がある。その平均をμ, 分散をσ^2とする。

- この確率変数の実現値をひとつ得たところ、12であった。$\mu = 10$かつ$\sigma^2 = 4$である尤度は？
- この確率変数の実現値を3つ得たところ、(12, 8, 9)であった。$\mu = 10$かつ$\sigma^2 = 4$である尤度は？

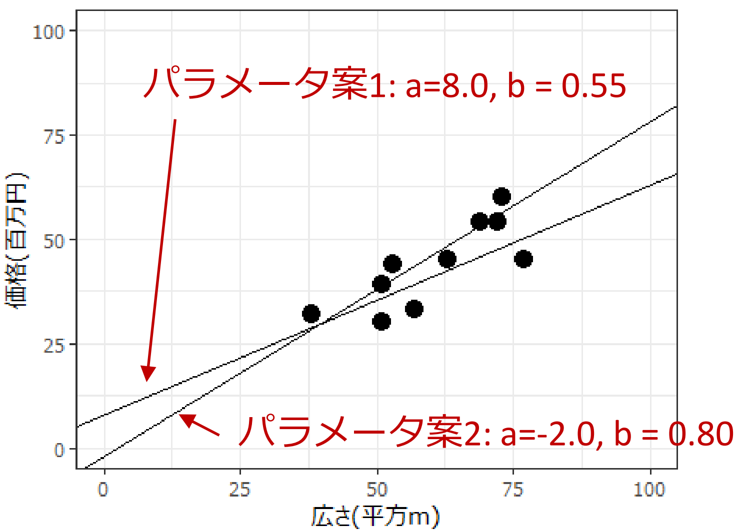
■ 最尤法の考え方

- 撹乱項U_iは、互いに独立に $N(0, s^2)$ に従うと仮定しよう
  このとき、Y_iは互いに独立に$N(a + bX_i, s^2)$ に従う
- $(Y_1, Y_2, \ldots, Y_n)$ の尤度が最大になるa, b, s^2を$\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$としよう
  直観的にいうと: データ $(Y_1, Y_2, \ldots, Y_n)$が生まれる”確率”が一番大きくなるパラメータを求めよう！



Code 7

2通りの(a, b, s^2)について試してみると...



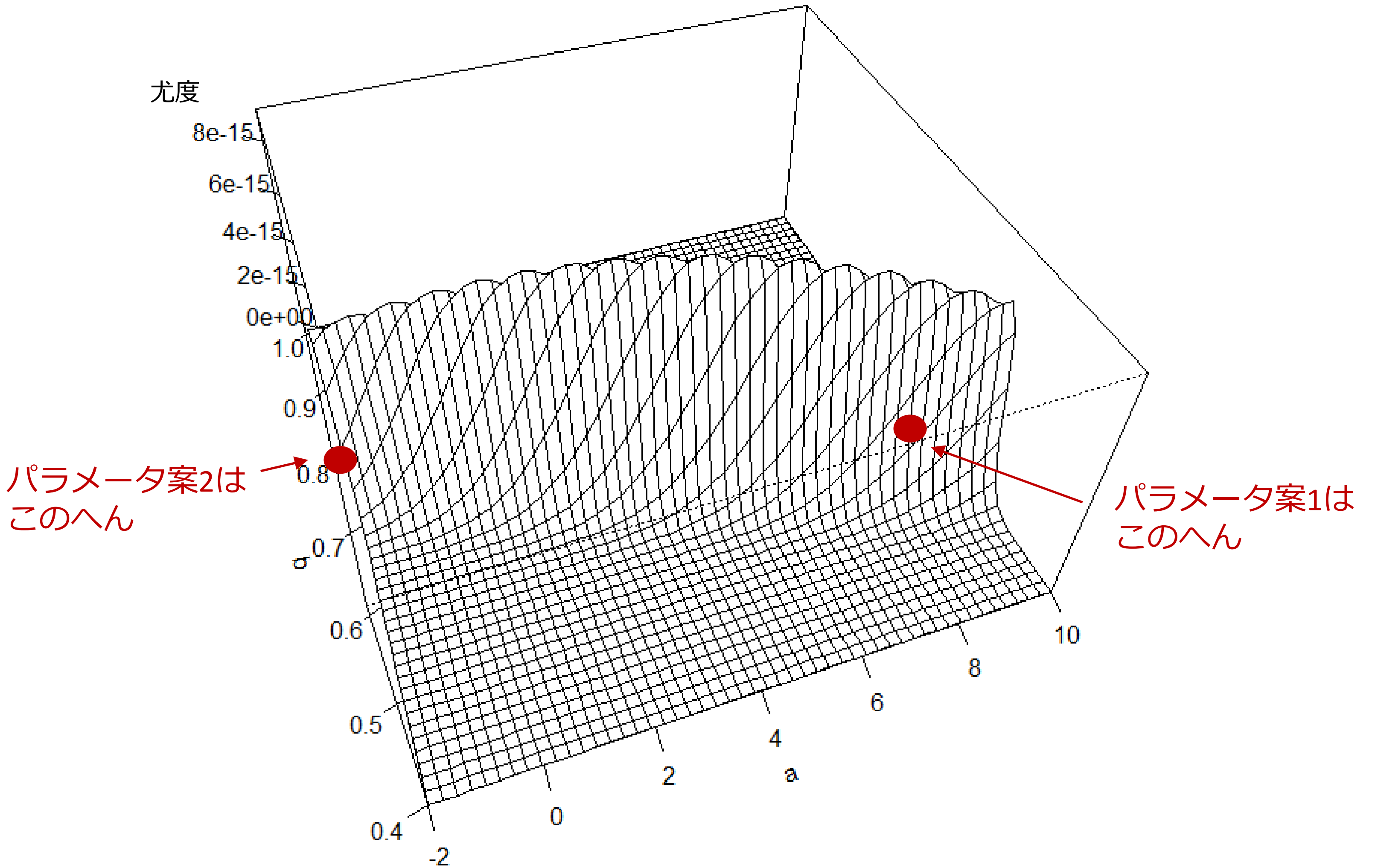
いずれも $s^2 = 25$とする

No	広さ	価格	パラメータ案1		パラメータ案2	
			予測値	データの尤度	予測値	データの尤度
1	51	30	36.1	0.038	38.8	0.017
2	38	32	28.9	0.066	28.4	0.062
3	57	33	39.4	0.036	43.6	0.008
4	51	39	36.1	0.067	38.8	0.080
5	53	44	37.2	0.031	40.4	0.062
6	77	45	50.4	0.045	59.6	0.001
7	63	45	42.7	0.071	48.4	0.063
8	69	54	46.0	0.022	53.2	0.079
9	72	54	47.6	0.035	55.6	0.076
10	73	60	48.2	0.005	56.4	0.062
		積		[illegible]		[illegible]

すいません小さすぎて読めないですね

Code 8

さまざまな(a, b)について計算してみると…



$s^2 = 25$とする

■ 推定量の導出 (蓑谷, pp.21-24)

Y_iの確率密度関数は、正規分布の定義より

$$f(Y_i) = \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(Y_i - a - bX_i)^2}{2s^2}\right)$$

$(Y_1, Y_2, \ldots, Y_n)$の同時確率密度関数は

$$f(Y_1, Y_2, \ldots, Y_n) = (2\pi s^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(Y_i - a - bX_i)^2}{2s^2}\right)$$

$X_1, \ldots, X_n, Y_1, \ldots, Y_n$を所与、$a, b, s^2$を変数と捉えたとき、これを**尤度関数** $L(a, b, s^2)$という

話を簡単にするために、尤度関数の対数について考えよう

$$\log L(a, b, s^2) = -\frac{n}{2}\log(2\pi s^2) - \frac{1}{2s^2}\sum_{i=1}^{n}(Y_i - a - bX_i)^2$$

対数尤度関数 $\log L(a, b, s^2)$ の、a, b に関する偏導関数を0と置く。

$$\frac{\partial \log L(a, b, s^2)}{\partial a} = \frac{1}{s^2}\sum_{i=1}^{n}(Y_i - a - bX_i) = 0$$

$$\frac{\partial \log L(a, b, s^2)}{\partial b} = \frac{1}{s^2}\sum_{i=1}^{n}(Y_i - a - bX_i)X_i = 0$$

これを解いて

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

結局、
最小二乗推定量と同じ

4. 単回帰モデルのパラメータ推定量の分散

4. 単回帰モデルのパラメータ推定量の分散

■ おさらい：パラメータ推定量の分散とは？

- 推定したいパラメータをθ、その推定量を$\hat{\theta}$とする
- どれだけ優れた推定量であるとしても、推定量$\hat{\theta}$は推定対象θと同一ではない。$\hat{\theta}$はθのまわりをばらつく
- そのばらつきの大きさを、分散 $Var(\hat{\theta})$ として表現する
    - $\sqrt{Var(\hat{\theta})}$ の推定値を**標準誤差** (SE) と呼ぶ

■ Quiz

- 推定したいパラメータを平均 μとする
- サイズ n の無作為標本を得た。平均は$\bar{x}$, 標準偏差はsであった
- $\bar{x}$ を μ の推定量として捉えよう。その標準誤差は？

■ 単回帰分析のパラメータ推定量の分散・共分散 (蓑谷, pp.28-30)

撹乱項 U_iの分散をσ^2として、

$$Var(\hat{\alpha}) = \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right)\sigma^2$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2}$$

単回帰係数の推定精度は、
- 撹乱項の分散が大きいとき低く、
- 説明変数の変動が大きいとき高い

$$Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2}$$

Code 4

中古マンション

パラメータ推定量の**標準誤差**

```
> summary(lm(Price ~ Hirosa, data = dfMansion))

Call:
lm(formula = Price ~ Hirosa, data = dfMansion)

Residuals:
   Min     1Q Median     3Q    Max 
-9.405 -5.684  2.184  4.347  8.199 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   4.2863    10.9270   0.392  0.70511  
Hirosa        0.6509     0.1775   3.666  0.00635 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.63 on 8 degrees of freedom
Multiple R-squared:  0.6269,	Adjusted R-squared:  0.5802 
F-statistic: 13.44 on 1 and 8 DF,  p-value: 0.006346
```

5. 単回帰モデルの推定量の性質

INSIGHT FACTORY

■ 推定量に期待される性質とは？

推定したいパラメータをθ, その推定量を$\hat{\theta}$とする

- **不偏性**

$E\left(\hat{\theta}\right) = \theta$

推定量$\hat{\theta}$ の期待値(長い目でみた平均)はθである

- **一致性**

$\lim_{n \to \infty} E\left(\hat{\theta}\right) = \theta, \lim_{n \to \infty} \mathrm{Var}\left(\hat{\theta}\right) = 0$

推定量$\hat{\theta}$は、標本サイズが無限大に大きければθに一致する

- **有効性**

推定量$\hat{\theta}$は、すべての不偏推定量のなかで分散が最小である

■ 単回帰モデルの古典的な仮定 (蓑谷, pp.3-7)

[1] X_{i}は確率変数でないか、 U_iと統計的に独立

「**外生性**」「内生性がない」「交絡がない」などと表現する

[2] $E(U_i) = 0$

期待値は0

[3] $\mathrm{Var}(U_i) = \sigma^2$

分散均一性 homoscedasticity. 撹乱項の分散は一定

[4] $\mathrm{Cov}\left(U_i, U_j\right) = 0, \quad i \neq j$

独立性。異なる観測点の間での相関(自己相関)がない

[5] U_iは正規分布に従う

正規性

■ 最小二乗推定量$\hat{\alpha}, \hat{\beta}$の性質 (蓑谷, pp.27-37)

- Y_iの**線形関数**である (=データY_iの重みつき合計として表現できる)

$$v_i = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}w_i\right), \qquad w_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

として

$$\hat{\alpha} = \sum_{i=1}^{n} v_i Y_i, \qquad \hat{\beta} = \sum_{i=1}^{n} w_i Y_i$$

- 仮定[1][2]のもとで、α, β の**不偏推定量**である
  - 推定量 $\hat{\alpha}, \hat{\beta}$ の期待値はα, β である

- 仮定[1][2] (とXに関するある仮定) の下で、α, βの**一致推定量**である
  - 推定量 $\hat{\alpha}, \hat{\beta}$は、観測点の数が無限に大きければ、$\alpha, \beta$に一致する

- 仮定[1][2][3][4]のもとで、α, β の**最良線形不偏推定量 (BLUE)** である
  - 推定量 $\hat{\alpha}, \hat{\beta}$は、線形関数の形をとる不偏推定量の中で分散が最小である

これをガウス・マルコフ定理という

- 仮定[1][2][3][4][5]のもとで、α, β の**有効推定量** である
  - 推定量 $\hat{\alpha}, \hat{\beta}$は、不偏推定量の中で分散が最小である

- 仮定[1][2][3][4][5]のもとで、α, β の**最尤推定量**である
  - 推定量 $\hat{\alpha}, \hat{\beta}$は、ざっくりいうと「データが得られる"確率"がもっとも高くなるような推定量」である

6. 重回帰モデル

説明変数が複数個ある線形回帰モデルのこと。

説明変数を$X_{2i}, \dots, X_{Ki}$として

$$Y_i = \underbrace{\beta_1}_{\text{切片}} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + U_i \;, \qquad i = 1, \dots, n$$

偏回帰係数

行列で表現すると

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{U}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{K1} \\ 1 & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2,n-1} & \cdots & X_{K,n-1} \\ 1 & X_{2n} & \cdots & X_{Kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{n-1} \\ U_n \end{bmatrix}$$

■ 単回帰係数 β の意味

$$Y_i = \alpha + \beta X_i + U_i, \qquad i = 1, \ldots, n$$

説明変数 X が1変化すると、目的変数 Y はどれだけ変化するか

■ 偏回帰係数 β_2 の意味

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + U_i \ , \qquad i = 1, \ldots, n$$

説明変数$X_3, \ldots, X_K$が変化せず、説明変数 X_2 **だけ**が1変化すると、目的変数 Y はどれだけ変化するか

■ 偏回帰係数の解釈

往々にして困難。問題についての深い理解が必要

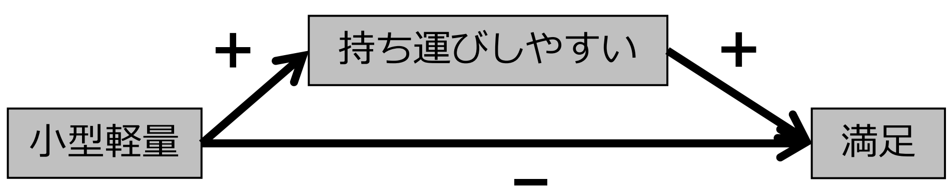
例)

- コンパクトカメラのユーザから、属性評価と満足度評価を得た
- 以下の回帰モデルを構築した:

(満足度) = 1.24 + 0.73 x (持ち運びしやすさ評価) **– 0.32** x (小型軽量さ評価)

- 小型軽量なほうが満足度が高いはずなのに、なぜ偏回帰係数は負なのか？

- 「持ち運びしやすさが一定であれば、小型軽量でないほうが満足度が高い」と解釈できるのでは？



(おそらくは、小型軽量だと操作しにくいから)

事例の出典：小島(2003)

7. 重回帰モデルのパラメータ推定

パラメータ推定量 (最小二乗推定量, 最尤推定量) は下式となる (蓑谷, pp.86-87)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

上式の導出は省略するが、説明変数が1つ(X_i)の場合について確認しておこう：

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix}$$

$\boldsymbol{X}'\boldsymbol{X}$の行列式

$$\boldsymbol{X}'\boldsymbol{Y} = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$= \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i \\ n\sum X_i Y_i - \sum X_i \sum Y_i \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{n\bar{Y}\sum X_i^2 - n\bar{X}\sum X_i Y_i}{n\sum X_i^2 - n^2\bar{X}^2} \\ \dfrac{n\sum X_i Y_i - n^2\bar{X}\bar{Y}}{n\sum X_i^2 - n^2\bar{X}^2} \end{bmatrix} = \begin{bmatrix} \bar{Y} - \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}\bar{X} \\ \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \end{bmatrix}$$

たしかに、
3節と一致している

$\xrightarrow{}$

変動・共変動の変形

Code 10

中古マンション

Rで実行してみると…

```
X <- dfMansion %>%
  mutate(Intercept = 1) %>%
  dplyr::select(Intercept, Hirosa, Nensu) %>%
  as.matrix(.)
Y <- dfMansion %>%
  dplyr::select(Price) %>%
  as.matrix(.)
betahat <- solve(t(X) %*% X) %*% t(X) %*% Y
print(betahat)
```

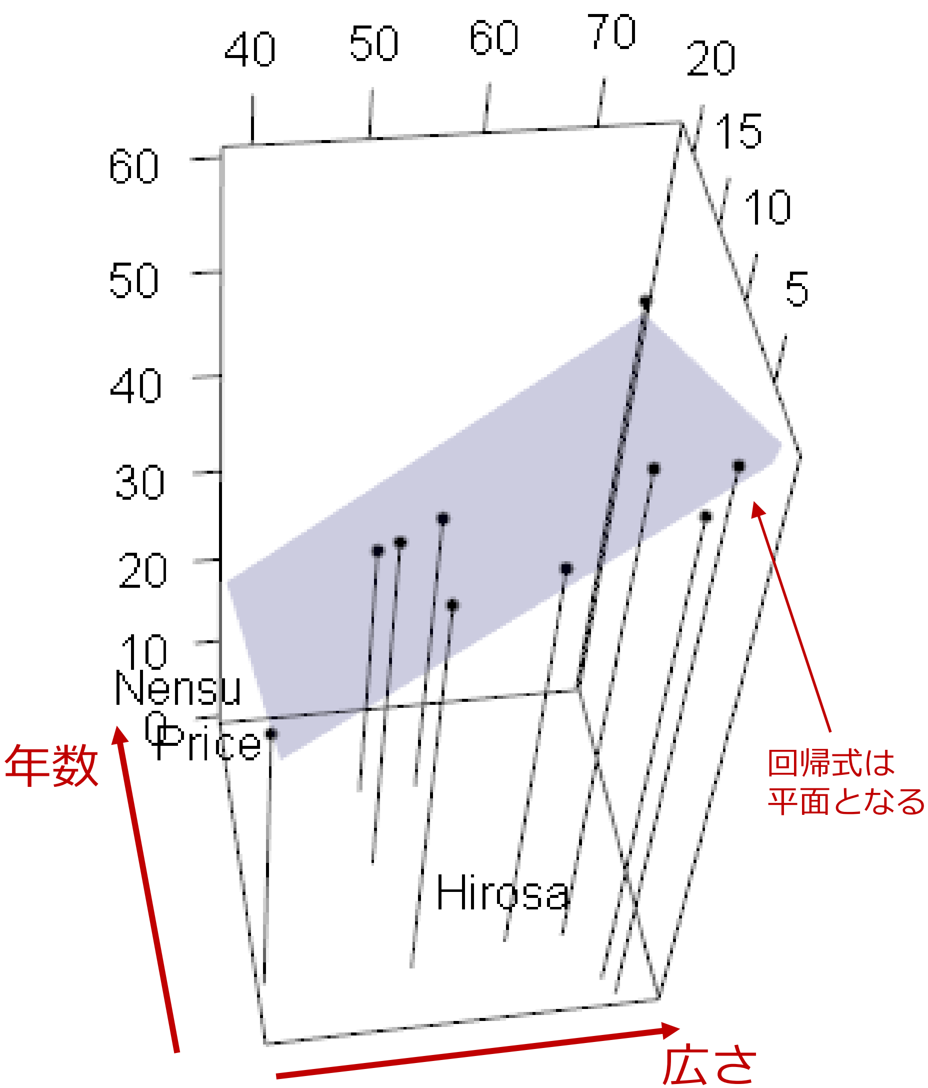
```
               Price
Intercept 10.2012955
Hirosa     0.6680477
Nensu     -0.8082993
```

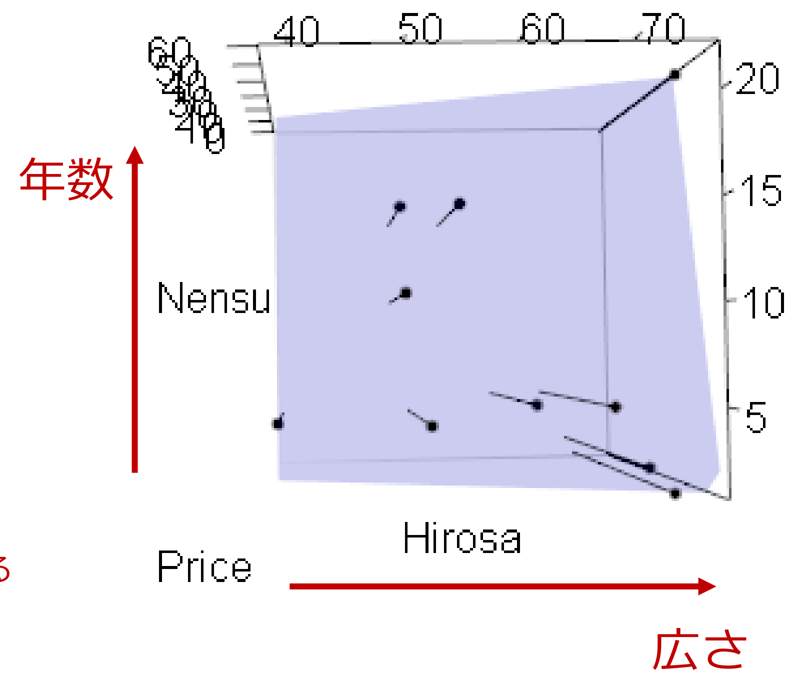
Code 10

lm()関数を使うと...

```
> summary(lm(Price ~ Hirosa + Nensu, data = dfMansion))

Call:
lm(formula = Price ~ Hirosa + Nensu, data = dfMansion)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2468 -2.0952  0.3939  1.7150  3.6196 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.20130    4.43624   2.300 0.055029 .  
Hirosa       0.66805    0.07065   9.456 3.09e-05 ***
Nensu       -0.80830    0.12241  -6.603 0.000303 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.636 on 7 degrees of freedom
Multiple R-squared:  0.9484,	Adjusted R-squared:  0.9336 
F-statistic:  64.3 on 2 and 7 DF,  p-value: 3.126e-05
```

Code 10


40
50
60
70
20
15
10
5
60
50
40
30
20
10
0
Nensu
Price
Hirosa
年数
広さ
回帰式は
平面となる


40
50
60
70
20
15
10
5
年数
Nensu
Hirosa
Price
広さ

8. 重回帰モデルのパラメータ推定量の共分散

8. 重回帰モデルのパラメータ推定量の共分散

パラメータ推定量の共分散は下式となる (蓑谷, pp.91-92)

撹乱項 U_iの分散をσ^2として、

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

偏回帰係数の推定精度は
- 撹乱項の分散が大きいとき低く、
- 説明変数の変動が大きいとき高く、
- 他の説明変数との相関が高いとき低い

- 説明変数が2つ(X_{2i}, X_{3i})の場合は以下となる (蓑谷, p.135)
  撹乱項 U_iの分散をσ^2として、

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum(X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - Corr(X_{2i}, X_{3i})}$$

- 一般に、X_{ji}を目的変数、他のすべての説明変数を説明変数とした回帰モデルの決定係数(後述) をR_j^2として、以下が成り立つ (蓑谷, p.136)

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum\left(X_{ji} - \bar{X}_j\right)^2} \times \frac{1}{1 - R_j^2}$$

第二項 $1/\left(1 - R_j^2\right)$を **VIF** (variance inflation factor) という

9. 重回帰モデルの推定量の性質

INSIGHT FACTORY

■ 重回帰モデルの古典的な仮定　　(行列で書くと)

[1] X_iは確率変数でないか
U_iと統計的に独立

[2] $E(U_i) = 0$　　$E(\boldsymbol{U}) = 0$

[3] $\mathrm{Var}(U_i) = \sigma^2$

[4] $\mathrm{Cov}(U_i, U_j) = 0, \quad i \neq j$

$$\mathrm{E}(\boldsymbol{U}\boldsymbol{U}') = \sigma^2 \boldsymbol{I}$$

[5] U_iは正規分布に従う

[6] $\boldsymbol{X}$の列ベクトルは**一次独立**で、
列数は行数より小さい　　$\mathrm{rank}(\boldsymbol{X}) = K < n$　← 新登場！

[6]が満たされていれば、推定量 $\hat{\boldsymbol{\beta}}$ は単回帰の場合と同様の性質を持つ。つまり、

- 仮定[1][2][6]のもとで、$\boldsymbol{\beta}$ の**不偏推定量**である
- 仮定[1][2][6] (と$\boldsymbol{X}$に関するある仮定) の下で、$\boldsymbol{\beta}$の**一致推定量**である
- 仮定[1][2][3][4][6]のもとで、$\boldsymbol{\beta}$ の**最良線形不偏推定量** (BLUE) である
- 仮定[1][2][3][4][5][6]のもとで、$\boldsymbol{\beta}$ の**有効推定量** である
- 仮定[1][2][3][4][5][6]のもとで、$\boldsymbol{\beta}$ の**最尤推定量**である

■ 一次独立とは？

- ある列が、他の列の重みづけ和になっていないこと

■ なぜ「$\boldsymbol{X}$の列は一次独立で、 列数は行数より小さい」必要があるのか？

- $\boldsymbol{X}'\boldsymbol{X}$が逆行列を持たなくなるから (7節を参照)

直観的にいうと…

- 説明変数がX_2, X_3, X_4の3つで、実は $X_4 = 2X_2 + 3X_3$だとしよう
- このとき、以下の式はいずれも等価である

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$Y = \beta_1 + (\beta_2 + 2\beta_4) X_2 + (\beta_3 + 3\beta_4) X_3 + 0 X_4$$

$$Y = \beta_1 + (\beta_2 + \beta_4) X_2 + (\beta_3 + 1.5\beta_4) X_3 + 0.5\beta_4 X_4$$

- 推定しようとしているパラメータそのものが同定できない

10. 回帰モデルの説明力

10. 回帰モデルの説明力

1. 決定係数
2. 自由度修正済み決定係数
3. AIC

$Y_i = \hat{Y}_i + E_i$ とすると、以下が成り立つ：

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{①観測値の変動}} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}_{\text{②予測値の変動}} + \underbrace{\sum_{i=1}^{n} E_i}_{\text{③残差の二乗和}}$$

そこで

$$R^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 / \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$ ①に占める②の割合

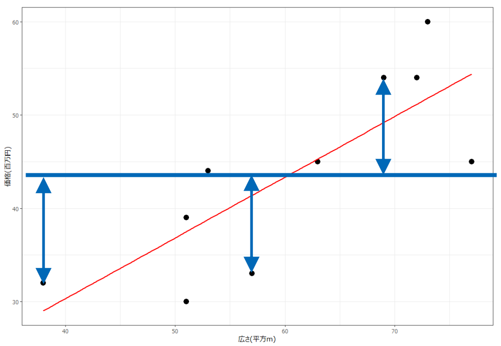
を**決定係数**と呼ぶ。

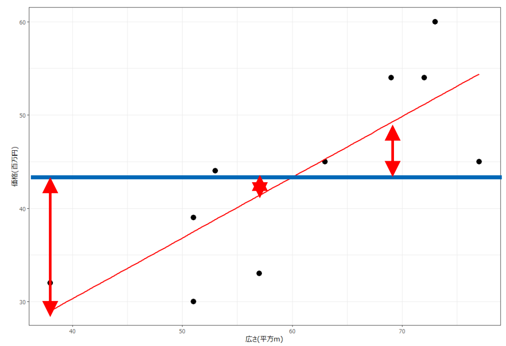
単回帰の場合、決定係数は相関係数 $\mathrm{Corr}(X_i, Y_i)$ の二乗に等しい。

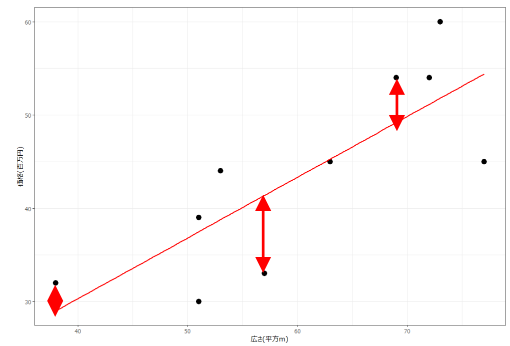
Code 9

①観測値の変動　　②予測値の変動　　③残差の二乗和

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} E_i$$







942.4　　590.7　　351.7

決定係数 = ②/① = 0.63

中古マンション

```
> summary(lm(Price ~ Hirosa + Nensu, data = dfMansion))

Call:
lm(formula = Price ~ Hirosa + Nensu, data = dfMansion)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2468 -2.0952  0.3939  1.7150  3.6196 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.20130    4.43624   2.300 0.055029 .  
Hirosa       0.66805    0.07065   9.456 3.09e-05 ***
Nensu       -0.80830    0.12241  -6.603 0.000303 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.636 on 7 degrees of freedom
Multiple R-squared:  0.9484,      Adjusted R-squared:  0.9336 
F-statistic:  64.3 on 2 and 7 DF,  p-value: 3.126e-05
```

決定係数

10-2. 自由度修正済み決定係数

決定係数は、説明変数の数を増やすと大きくなる

- モデルの説明力をモデル間で比較する際には、説明変数の数も考慮する必要がある

説明変数の数の影響を修正した決定係数

$$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1 - R^2)$$

注意: ここでKは(説明変数の数)＋1

を自由度修正済み決定係数という

Code 10

中古マンション

```
> summary(lm(Price ~ Hirosa + Nensu, data = dfMansion))

Call:
lm(formula = Price ~ Hirosa + Nensu, data = dfMansion)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2468 -2.0952  0.3939  1.7150  3.6196 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.20130    4.43624   2.300 0.055029 .  
Hirosa       0.66805    0.07065   9.456 3.09e-05 ***
Nensu       -0.80830    0.12241  -6.603 0.000303 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.636 on 7 degrees of freedom
Multiple R-squared:  0.9484,      Adjusted R-squared:  0.9336 
F-statistic:  64.3 on 2 and 7 DF,  p-value: 3.126e-05
```

自由度修正済決定係数

最尤法では、対数尤度関数を最大化する $\hat{\boldsymbol{\beta}}$ を求める。最大化された対数尤度は、

$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

として

$$\log L^* = -\frac{n}{2}(\log 2\pi + 1 + \log\hat{\sigma}^2)$$

となる。この値が大きいモデルは、データをよく説明している。

しかし、説明変数の数を増やすと、$\hat{\sigma}^2$は小さくなり、$\log L^*$は大きくなる。
そこで、モデルのパラメータ$(\beta_1, \dots, \beta_K, \sigma^2)$の数を p とし、**AIC (赤池情報量規準)**

$$AIC = -2\log L^* + 2p$$

をモデル選択の基とすることがある。
この値が小さいモデルは、データをよく説明している。

Code 11

中古マンション

```
X <- dfMansion %>%
  mutate(Intercept = 1) %>%
  dplyr::select(Intercept, Hirosa, Nensu) %>%
  as.matrix(.)
Y <- dfMansion %>%
  dplyr::select(Price) %>%
  as.matrix(.)
betahat <- solve(t(X) %*% X) %*% t(X) %*% Y
ehat    <- Y - X %*% betahat
sshat   <- as.vector((1/length(Y)) * t(ehat) %*% ehat)
LL      <- -(length(Y)/2) * (log(sshat) + 1 + log(2 * pi))
AIC     <- -2 * LL + 2 * 4
print(LL)
print(AIC)
```

```
> print(LL)
[1] -22.09959

> print(AIC)
[1] 52.19917
```

Code 11

logLik()関数, AIC()関数を使うと..

```
oModel <- lm(Price ~ Hirosa + Nensu, data = dfMansion)
logLik(oModel)
AIC(oModel)
```

```
> logLik(oModel)
'log Lik.' -22.09959 (df=4)

> AIC(oModel)
[1] 52.19917
```

11. 市場反応モデリングと回帰モデル

11. 市場反応モデリングと回帰モデル

回帰モデルは、市場反応モデルの基盤を提供する。

いっぽう、市場反応の持つ特性は、回帰モデルの前提と多くの点で対立している。

次章からは、この対立を乗り越えるためのさまざまな方法について議論する。

回帰モデル

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + U_i$$

[1] X_iは確率変数でないか
U_iと統計的に独立

[2] $E(U_i) = 0$

[3] $\mathrm{Var}(U_i) = \sigma^2$

[4] $\mathrm{Cov}(U_i, U_j) = 0, \quad i \neq j$

[5] U_iは正規分布に従う

[6] $\boldsymbol{X}$の列ベクトルは一次独立。
列数は行数より小さい

市場反応の特性

A) マーケティング活動と市場反応との関係は、直線的ではないかもしれない

B) 市場反応変数として、売上を用いる場合とシェアを用いる場合がある

C) 市場反応に影響する変数は多様であり、データを入手できない変数も多い

D) 広告・プロモーションの効果は、その内容や媒体によっても異なるかもしれない

E) 異なるマーケティング活動を同時に行うことにより、シナジーが生まれるかもしれない

F) マーケティング活動における消費者の反応には、異質性があるかもしれない

G) 市場反応データは時系列データの形を取ることが多い

H) マーケティング活動が与える効果は、即時的に現れることもあれば、時間的遅延とともに現れることもある

I) マーケティング活動の効率は、時間とともに増大・減衰することがある

J) マーケティング活動は、それまでの市場反応に基づいて計画されることがある

この章の引用文献

永田靖・棟近雅彦(2001) 「多変量解析法入門」, サイエンス社.

小島隆矢 (2003)「Excelで学ぶ共分散構造分析とグラフィカルモデリング」, オーム社.

この章に登場したRの関数

- lm(フォーミュラ, data=データフレーム)
  - 回帰モデルを最小二乗推定し、lmオブジェクトとして返す
  - フォーミュラの書き方

    回帰モデルを指定する。データフレームに含まれている変数の名前をそのまま使うことができる。

    (以下ではデータフレームに変数 Y, X1, X2が含まれているものとする)

    - 例1) 重回帰モデル, 目的変数は Y, 説明変数は X1とX2

      Y ~ X1 + X2
    - 例2) 切片を0に固定する

      Y ~ 0 + X1 + X2
    - 例3) 交互作用項を追加する

      Y ~ X1 + X2 + X1:X2

- summary(lmオブジェクト)
  パラメータ推定値などを返す
- plot(lmオブジェクト)
  残差プロットを表示する
- coef(lmオブジェクト)
  パラメータ推定値をベクトルとして返す
- logLik(lmオブジェクト)
  モデルの最大対数尤度を返す
- AIC(lmオブジェクト)
  モデルのAICを返す
- predict(lmオブジェクト, newdata=データフレーム)
  推定したモデルをデータフレームにあてはめ、予測値を返す
- car::vif(lmオブジェクト)
  VIFを返す
- broom::tidy(lmオブジェクト)
  パラメータ推定値をデータフレームとして返す

注：Rでは、
「hogeパッケージが提供しているfuga関数」を
hoge::fuga(...)
と書くことがあります。
以下この書き方に従います

- dnorm(値, mean=平均, sd=標準偏差)
  正規確率密度を返す
- solve(数値行列)
  逆行列を返す
- outer(ベクトル, ベクトル, 関数)
  2つのベクトルの外積を返す。関数を指定すると積ではなくてその関数が適用される (※説明しにくい…)