統計科学同演習第 4 回レポート

武山尚生 71844768

2020 年 5 月 18 日

1　最小二乗法

問題 1(1) 講義資料 5 枚目の β_0, β_1 を導出しなさい. S を β_0 で偏微分すると。

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \{y_i - \beta_0 + \beta_1 x_i\}^2 \tag{1}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0 \tag{2}$$

$$-2\sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1 x_i\} = 0 \tag{3}$$

$$\frac{1}{n}\sum_{i=1}^{n} \{y_i - \beta_1 x_i\} = \beta_0 \tag{4}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{5}$$

次に β_1 を導出する。

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \{y_i - \beta_0 + \beta_1 x_i\}^2 \tag{6}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0 \tag{7}$$

$$-2\sum_{i=1}^{n} \{x_i(y_i - (\beta_0 + \beta_1 x_i))\} = 0 \tag{8}$$

$$\frac{1}{n}\sum_{i=1}^{n} \left\{x_i y_i - (\bar{y} - \beta_1 \bar{x})x_i - \beta_1 x_i^2\right\} = 0 \tag{9}$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y} = \beta_1 \sum_{i=1}^{n} (x_i^2 - x_i \bar{x}) \tag{10}$$

これより β_1 は、

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{11}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{(x_i - \bar{x})(y_i - \bar{y})\right\}^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{12}$$

故に、$\beta_1 = \frac{S_{xy}}{S_{x^2}}$、$\beta_0 = \bar{y} - \frac{S_{xy}}{S_{x^2}}\bar{x}$ となる。

(2) $S(\beta) = (y - X\beta)^T(y - X\beta)$ を最小にする β を求めなさい。
S を β で偏微分すると、

$$\frac{\partial S(\beta)}{\partial \beta} = 0 \tag{13}$$

$$2X^T(y - X\beta) = 0 \tag{14}$$

$$X^T y = X^T X\beta \tag{15}$$

$$(X^T X)^{-1}X^T y = \beta \tag{16}$$

と β が求まる。
(3) 講義資料 17 枚目の P について、$P^T = P$ を示しなさい。

$$P^T = (X(X^T X)^{-1}X^T)^T \tag{17}$$

$$= ((X^T X)^{-1}X^T)^T X^T \tag{18}$$

$$= X((X^T X)^{-1})^T X^T \tag{19}$$

$$= X((X^T X)^T)^{-1} X^T \tag{20}$$

$$= X(X^T X)^{-1} X^T \tag{21}$$

$$= P \tag{22}$$

(4) 講義資料 17 枚目の P について、$P^2 = P$ を示しなさい。

$$P^2 = X(X^T X)^{-1}X^T X(X^T X)^{-1}X^T \tag{23}$$

$$= X(X^T X)^{-1}X^T \tag{24}$$

$$= P \tag{25}$$

(5) 推定値ベクトル $\hat{y}$ と残差ベクトル $\hat{e}$ は直交する事を示しなさい。

$$\hat{y} = Py, \hat{e} = y - Py \tag{26}$$

$$< \hat{y}, \hat{e} > = (Py)^T(y - Py) \tag{27}$$

$$= y^T P^T y - y^T P^T Py \tag{28}$$

$$= y^T Py - y^T Py \tag{29}$$

$$= 0 \tag{30}$$

2　射影行列

問題 2 $rank(B) = q$ である $m \times q$ 行列 B について，$P = B(B^T B)^{-1} B^T$ とする. 以下の問に答えなさい.

(1) B の列ベクトルの張る空間 M の任意の元 $a(\in \mathbb{R}^m)$ に対し，$Pa = a$ となることを示しなさい.
射影行列の性質より、$P^2 = P$、$P^T = P$。

$$(I-P)^2 = I^2 - 2P + P^2 = I - P \tag{31}$$
$$(I-P)^T = I - P \tag{32}$$

より、$I-P$ も射影行列。また、$P(I-P) = 0$。
背理法より、$Pa_1 = a_2(a_1 \neq a_2)$ とする。

$$P(I-P)a_1 = P(a_1 - a_2) \tag{33}$$
$$= 0 \tag{34}$$

任意の正則な P に対して $Pa = 0$ を満たすベクトルは $a = 0$ の時だけであることから、$a_1 - a_2 = 0$。
仮定と矛盾するため、$Pa = a$ が成り立つ。

(2)M の直交補空間 $M^{\perp}$ の任意の元 $d(\in \mathbb{R}^m)$ に対し，$Pd = 0$ となることを示しなさい.
直交補空間の定義より、任意の M の元 a に対して、$<a, d>= 0$
$Pa = a$ より、

$$d^T Pa = (Pd)^T a \tag{35}$$
$$= 0 \tag{36}$$

$<a, Pd>= 0$ より、$Pd \in M^{\perp}$。射影行列の定義から、$Pd \in M$ は自明である。
$M^{\perp} \bigcap M = \{0\}$ より、$Pd = 0$

(3) 講義資料 19 枚目の性質 3:$\sum_{i=1}^{n} \hat{e}_i = 0$ を示しなさい。

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \left\{ \hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip} \right\}^2 \tag{37}$$
$$= \sum_{i=1}^{n} \left\{ (y_i - \bar{y}) - \beta_1 \sum_{i=1}^{n} (x_{i1} - \bar{x_1}) - \cdots - \beta_p \sum_{i=1}^{n} (x_{ip} - \bar{x_p}) \right\} \tag{38}$$
$$= 0 \tag{39}$$

(4) 講義資料 19 枚目の性質 4(平方和の分解) を示しなさい.

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}\Big\{(\hat{\beta}_0+\hat{\beta}_1 x_{i1}+\cdots+\hat{\beta}_p x_{ip}+\hat{e_i})-(\beta_0+\beta_1\bar{x_1}+\cdots+\beta_p\bar{x_p})\Big\}^2 \tag{40}$$

$$= \sum_{i=1}^{n}\{\hat{\beta}_1(x_{i1}-\bar{x_1})+\cdots+\hat{\beta}_p(x_{ip}-\bar{x_p})+\hat{e_i}\}^2 \tag{41}$$

$$= \sum_{i=1}^{n}\{\hat{\beta}_1(x_{i1}-\bar{x_1})+\cdots+\hat{\beta}_p(x_{ip}-\bar{x_p})\}^2+\sum_{i=1}^{n}\hat{e_i}^2 \tag{42}$$

$$= \sum_{i=1}^{n}(\hat{y_i}-\bar{y})^2+\sum_{i=1}^{n}\hat{e_i}^2 \tag{43}$$

3　最小二乗推定量の性質

問題 3 回帰係数推定量の平均と分散を求めなさい. なお，仮定や表記は講義スライドを参照のこと.

$$E[\hat{\beta}] = (X^TX)^{-1}X^TE[y] \tag{44}$$

$$= \beta \tag{45}$$

$$Var[\beta] = E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] \tag{46}$$

$$= E[((X^TX)^{-1}X^Ty-(X^TX)^{-1}X^TX\beta)((X^TX)^{-1}X^Ty-(X^TX)^{-1}X^TX\beta)^T] \tag{47}$$

$$= (X^TX)^{-1}X^TE[(y-X\beta)(y-X\beta)^T]X(X^TX)^{-1} \tag{48}$$

ここで、y の分散 $E[(y-X\beta)(y-X\beta)^T]$ を σ^2 とすると、

$$Var[\beta] = \sigma^2(X^TX)^{-1} \tag{49}$$

4　母親の身長による娘の身長予測

(1) 母親に身長について、次の統計量を求めて記入しなさい

統計量	母の身長 (cm)	娘の身長 (cm)
(標本) 平均	158.6	161.9
(標本) 標準偏差	5.98	6.60
中央値	158.5	161.5
第 1 四分位	154.4	157.5
第 3 四分位	162.3	166.6
最小値	140.7	140.0
最大値	179.8	185.7

(2b) 最小二乗法による推定値 β^0, β^1 を求め，結果をそれぞれオブジェクト b0, b1 に格納したい．これを実現する R のコマンドと，その値を書きなさい. ただし，lm 関数を用いてはならない
b1 $<-$ var(heightsm)[2] / var(MH)
b0 $<-$ mean(MH) - b1 * mean(DH)

(2c) 次のように入力すると，(i) 母親と娘の身長の散布図，(ii) 推定された回帰直線，(iii) 娘の身長= 母の身長の直線が描かれる. 娘の身長が母親の身長よりも高く予測されるのは，母親の身長がど の範囲の値のときか，計算して答えなさい.

$$y = 70.91 + 0.5417x, \quad y = x \tag{50}$$
$$x = 70.91 + 0.5417x \tag{51}$$
$$x = 154.7 \tag{52}$$

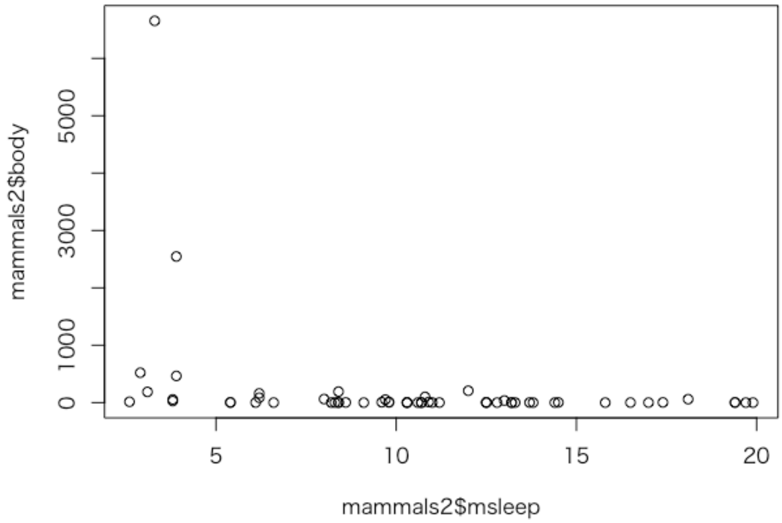
よって、母親の身長が 154.7cm 以下の時

5　哺乳類の睡眠時間

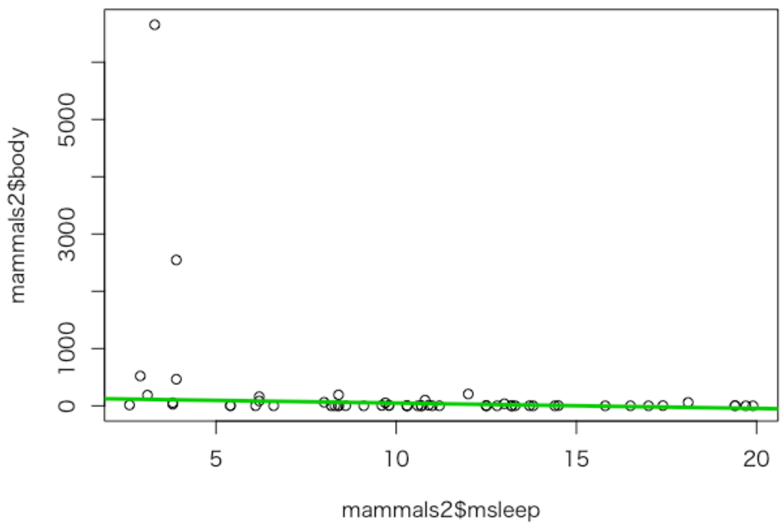
(1) 変数 body, brain, life, gestation の中で，変数 msleep と最も線形の相関の強い変数はどれか. 観測値，散布図行列，相関係数行列のすべてをよく見て答えなさい (ヒント:散布図行列は pairs 関 数，相関係数行列を求める際には cor(na.omit(mammals)[,-1]) と入力).

gestation が最も相関が強い。body や brain はゾウのデータが外れ値となっている。

(2) 次のように入力し，睡眠時間 msleep，体重 body，危険度 danger の 3 つのみ取り出し，さらに欠測値のある個体を除いたデータフレーム mammals2 を作成しなさい. 変数 body を横軸，変数 msleep を縦軸に配した散布図を描きなさい。



(3)(下線部のみ解答しなさい)(2) の散布図の上に，msleep を body で説明するのに適切と思われる回帰直線を計算をせずに目分量で 引きなさい. また，この線を描くときの判断に影響を強く与えた (言い換える と，その点が取り除かれたら「適切と思われる回帰直線が」大きく変化すると予想される) 散布図上 の点に印を付け，その動物名と観測値の特徴を述べなさい.
African elephant, Asia elephant



(4) 睡眠時間 msleep と体重 body は，相関 (線形関係) があるとは言い難い. しかし，適切な単調関数 で body を変換することにより，線形関係が強くなるようにできる. この「適切な単調関数」，すな わち，g(body) が msleep と最も線形関係が強くなるような関数 g を指数関数:exp(− x)，対数関数:log(x)，1 次多項式:1 + 2x，2 次多項式:1 + x + x2 から 1 つ選びなさい. また，変換後の変量 g(body) と msleep のピアソンの積率相関係数を計算しなさい (線形関係があるかどうかは，必ず散布図で確認すること).

関数 g: log(x)
ピアソンの積率相関係数:-0.53

(5) 次の表に，(I)–(III) を書き入れなさい.

		被説明変数	説明変数 1	説明変数 2		
	動物名	msleep	gbody	danger	予測値	残差
推定値が最大	Little brown bat	19.9	-4.61	1	16.5	3.42
残差絶対値が最大	Genet	6.1	0.344	1	13.5	-7.37

感想：LaTeX を使ってレポートを書いてみましたが、想像よりもかなり時間がかかったように感じます。