

▶▶▶▶ 英文での表記

◆ Materials & Methods

Multiple regression was used to determine the variations between serum iron and hemoglobin within the subject (ref) . We used hemoglobin concentration as outcome variable, serum iron concentration and the subjects as predictor variables. Subject was treated as categorical factor using dummy variable with five degrees of freedom. The *P* value from the *t* test for the regression slope of serum iron was used to determine the probability of the analysis. The magnitude of correlation coefficient between serum iron and hemoglobin within subjects was calculated as square root of (sum of squares for serum iron) / (sum of squares for serum iron + residual sum of squares) . The sign of the correlation coefficient was given by that of the regression coefficient for serum iron.

血清鉄とヘモグロビンの患者内での変動は多重回帰によって検討した。ヘモグロビン濃度を目的変数、血清鉄濃度と患者を説明変数として用いた。患者は自由度5の分類変数として扱った。この分析の有意性は、血清鉄の偏回帰係数の*t*検定によって判断した。患者内での血清鉄とヘモグロビンの相関係数の絶対値は(血清鉄の偏差平方和) / (血清鉄の偏差平方和+残差の偏差平方和)の平方根として求めた。また相関係数の符号は血清鉄の偏回帰係数の符号から求めた。

◆ Results

Determining the variations within the subject, the strong positive correlation with correlation coefficient of 0.794 was found between serum iron and hemoglobin (*P* < 0.001) .

患者内での変動を調べてみると、血清鉄とヘモグロビンの間には相関係数 0.794 の強い正の相関関係があることが判った (*P* < 0.001) .

[Reference] Bland JM and Altman DG, Calculating correlation coefficients with repeated observations : Part 1--correlation within subjects. *Brit Med J*, **310**, 446 (1995)



相関関係：繰り返しの数が違う場合

4月の時点で入院していた8名の患者について、4月から11月までの8カ月分の検査結果から、ヘモグロビン(Hb)と血清鉄(Fe)の値を調べてみると表3-7の結果を得た。もう既に転退院してしまった患者さんが多く、4月から11月まで全てのデータが揃っているのは1例のみで、データ数はバラバラだった。どのように解析したら良いだろうか？

表 3-7 8名の患者の血清鉄とヘモグロビン

患者	A		B		C		D		E		F		G		H	
	Fe	Hb	Fe	Hb	Fe	Hb	Fe	Hb	Fe	Hb	Fe	Hb	Fe	Hb	Fe	Hb
4月	116	8.3	140	13.8	102	10.1	87	8.1	130	10.3	55	11.2	90	9.7	153	8.8
5月	102	9.0	115	12.8	112	10.2	111	10.0	123	9.4	77	12.0	84	10.8	141	8.4
6月	123	9.8	135	14.3	121	10.8	118	9.5	137	11.8	78	13.8	99	10.8	147	9.0
7月	137	12.6	158	16.6			100	8.0	119	10.6			80	8.8		
8月	122	11.0	127	13.8			126	10.8	145	13.5			111	10.9		
9月	133	10.7	143	15.8			108	8.4	122	12.1			72	7.3		
10月							85	6.9	126	11.6			100	9.0		
11月							79	6.7					105	9.3		
平均	122	10.2	136	14.5	112	10.4	102	8.6	129	11.3	70	12.3	93	9.6	147	8.7
SD	13	1.5	15	1.4	10	0.4	17	1.5	9	1.3	13	1.3	13	1.3	6	0.3
n	6		6		3		8		7		3		8		3	

(単位 Fe : μg/dl, Hb : g/dl)

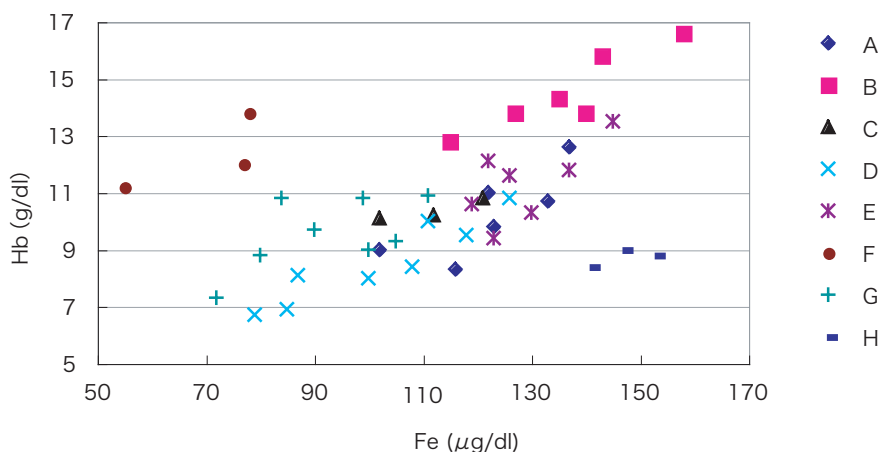


図 3-9 8名の患者の血清鉄とヘモグロビン

▶▶▶▶ してはいけない解析

例題 3-4 (p.39) では、血清鉄とヘモグロビン一般的な関係を調べたいときには、平均値の相関関係を調べていた。この場合も同じように解析してみよう。表 3-7 から、患者ごとの平均値を抜き出して表 3-8 が得られた。

表 3-8 患者ごとの測定値の平均値

	測定値数	Fe	Hb
A	6	122	10.2
B	6	136	14.5
C	3	112	10.4
D	8	102	8.6
E	7	129	11.3
F	3	70	12.3
G	8	93	9.6
H	3	147	8.7
合計	44		

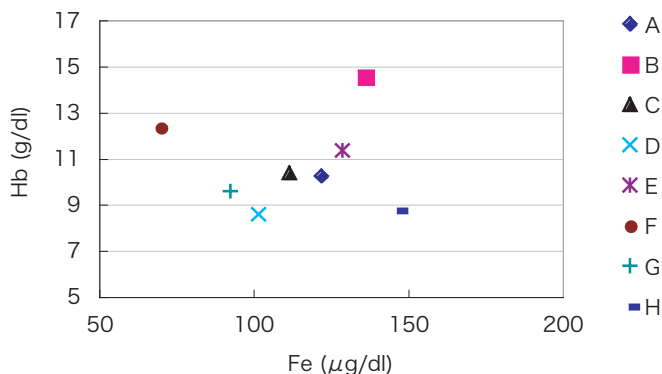


図3-10 患者ごとの測定値の平均値

この表 3-8 のデータをピアソンの相関係数によって解析し、解析結果 3-10 の結果を得た。

■ 解析結果 3-10 相関分析

	データ数	相関係数	t 値	P 値 (両側確率)	t(0.975)	95%下限	95%上限
Fe, Hb	8	0.01476	0.0362	0.9723	2.4469	-0.6972	0.71203

この結果から、血清鉄とヘモグロビンの間には相関関係が認められないと結論してよいだろうか？ 実は、このように解析をしてはいけない。

1. 問題点と正しい解析方法の選択

表 3-8 をみて、患者は 8 人なのだからデータ数は $n = 8$ で良いと思うかもしれないが、患者 F は 3 点の測定値しかないのに対して、患者 G は 8 点の測定値がある。このように測定値の数がバラバラであることが、上の解析ではまったく考慮されていない。それが問題なのである。

そこで、その点も考慮した解析方法を用いなければならない。この例題の場合も例題 3-4 と同様に下記の 2 つの解析が考えられる。

- ①' 一般的に、総タンパクとアルブミンの濃度にはどのような関係があるかを調べたい場合
 - ②' ある個人の総タンパクとアルブミンにはどのような関係があるかを調べたい場合
- それでは両者について解析してみよう。

▶▶▶▶ 解析 1：一般的な関係

このように、各患者の測定値数に違いがある場合には、その測定値数を考慮して、重み付けを行って解析しなければならない。重み付けのためには、表 3-8 を表 3-9 のように書き変えればよい。

表 3-9 重み付けを行った相関係数の解析のためのデータ

Fe	Hb		Fe	Hb		Fe	Hb	
122	10.2	A n = 6	102	8.6	D n = 8	70	12.3	F n = 3
122	10.2		102	8.6		70	12.3	
122	10.2		102	8.6		70	12.3	
122	10.2		102	8.6		93	9.6	G n = 8
122	10.2		102	8.6		93	9.6	
122	10.2		102	8.6		93	9.6	
136	14.5	102	8.6	93	9.6			
136	14.5	102	8.6	93	9.6			
136	14.5	102	8.6	93	9.6			
136	14.5	B n = 6	129	11.3	E n = 7	147	8.7	H n = 3
136	14.5		129	11.3		147	8.7	
136	14.5		129	11.3		147	8.7	
136	14.5		129	11.3		129	11.3	
136	14.5		129	11.3		129	11.3	
136	14.5		129	11.3		129	11.3	
112	10.4	C n = 3	129	11.3	129	11.3		
112	10.4		129	11.3	129	11.3		
112	10.4		129	11.3	129	11.3		

表 3-9 には、各患者の血清鉄とヘモグロビンの平均値が、測定値の数 n だけ記載されている (ex. 患者 A と B は 6, 患者 C は 3, …)。この表 3-9 の合計 44 個のデータについてピアソンの相関係数を調べると解析結果 3-11 のようになる。

■ 解析結果 3-11 表 3-9 のデータについてピアソンの相関係数を求めた結果

	データ数	相関係数	t 値	P 値 (両側確率)	t(0.975)	95%下限	95%上限
Fe, Hb	44	0.31666	2.1636	0.0362	2.0181	0.0218	0.5608

解析結果 3-11 では、相関係数 $r = 0.317$, $P = 0.036$ であるので、この結果から、血清鉄とヘモグロビンが正に相関すると言えるのだろうか？

実はまだダメなのだ。この結果では相関係数と t 値は良いのだが、元々の患者が 8 人しかいないのに、データ数が 44 になっている。つまり 44 人の患者を解析したことになってしまっている。そのため P 値が $n = 44$ のときの値になってしまっているのだ。

そこで、コンピューターが計算した P 値は無視して、 $n = 8$ であることを考慮して、 t 値に対応する P 値を求めなければならない。これはエクセルで計算できる。データ数が 8 の場合は統計量 t は自由度 $8 - 2 = 6$ の t 分布に従うので、これを利用する。

エクセルの TDIST 関数では、 t 値の絶対値、自由度、両側か片側かの 3 つの数字を指定する

と、 P 値が計算できる。まず、 t 値の絶対値は2.1636、そして自由度は $8 - 2 = 6$ 。そして、3番目の数字として、両側検定を示す2（片側検定の場合は1）を使う。エクセルのセルに " $=TDIST(2.163555, 6, 2)$ " を入力すると、 P 値（両側）として0.0737を求めることができる。

以上の修正を行うと、解析結果3-11を解析結果3-12のように書き直すことができる。

■ 解析結果3-12 表3-13のデータを正しく重み付けしてピアソンの相関係数を解析した結果

	データ数	相関係数	t 値	P 値 (両側確率)	t(0.975)	95%下限	95%上限
Fe, Hb	44	0.3167	2.1636	0.0362	2.0181	0.0218	0.5608
	↓			↓			
	8			0.0737			

表3-18の結果より、血清鉄とヘモグロビンの間には相関関係は認められないという解析結果が得られた。

▶▶▶▶ 英文での表記（解析1）

◆ Materials & Methods

The relation between serum iron and hemoglobin was determined by weighted correlation coefficient using the numbers of observations on each subject as weight

血清鉄とヘモグロビンの関係は、患者ごとの測定値数を重みとし、重み付けを行った相関係数により確認した。

◆ Results

The correlation coefficient between serum iron and hemoglobin was $r = 0.317$, $P = 0.074$. There was no statistically significant evidence that subjects with a high total protein concentration had a high (or low) albumin level.

総タンパクとアルブミンの相関係数は、 $r = 0.317$, $P = 0.074$ であり、血清鉄濃度が高い患者はヘモグロビンが高い（あるいは低い）という統計的に有意な証拠は得られはなかった。

▶▶▶▶ 解析2：個人の血清鉄とヘモグロビンの関係を調べたい場合

この例題も例題3-4と同様に、「同一の患者では血清鉄とヘモグロビンに相関関係があるかどうか」を、

$$Hb = a \times Fe + b_i \times (\text{患者})_i + c$$

a : 血清鉄の係数

b_i : i 番目の患者の係数

c : 定数項

という一次式で近似して解析することにする。患者ごとの測定値の数は違うが、手順はまったく同じである。以下に手順を示す。

手順1：データを整理する

まず、表3-7を表3-10のように書き換えてみる。この表でも例題3-4の表3-6 (p.43)と同様、患者AではA欄が1でB～Hまでが0。他の患者についてもすべて同様にどの患者のデータかを示すA～Hの欄に1か0が入れられている。

表 3-10 表 3-7 のデータを線形モデルにあてはめるために書き直したもの

目的変数 Hb	説明変数									除外 H
	Fe	A	B	C	D	E	F	G		
8.3	116	1	0	0	0	0	0	0	0	患者 A の データ
9.0	102	1	0	0	0	0	0	0	0	
9.8	123	1	0	0	0	0	0	0	0	
12.6	137	1	0	0	0	0	0	0	0	
11.0	122	1	0	0	0	0	0	0	0	
10.7	133	1	0	0	0	0	0	0	0	
13.8	140	0	1	0	0	0	0	0	0	
12.8	115	0	1	0	0	0	0	0	0	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
9.0	100	0	0	0	0	0	0	1	0	
9.3	105	0	0	0	0	0	0	1	0	
8.8	153	0	0	0	0	0	0	0	1	患者 H の データ
8.4	141	0	0	0	0	0	0	0	1	
9.0	147	0	0	0	0	0	0	0	1	

手順 2：重回帰分析

いよいよ解析に取り掛かる。目的変数を Hb とし、血清鉄濃度 (Fe) および患者を説明変数として重回帰分析を行うが、このときダミー変数である患者の自由度を 1 つ減らして 7 にする。具体的には患者の変数 A ~ H の内どれか 1 つの列を除外する。どの患者を除外してもかまわないが、ここでは、H を除外して、Hb を目的変数、Fe、A ~ G を説明変数として重回帰分析を行い、解析結果 3-13 が得られた。(解析手順は「4Steps エクセル統計」(第二版) p.179 を参照)

■ 解析結果 3-13 表 3-10 のデータを重回帰分析した結果

・分散分析表						
要因	偏差平方和	自由度	不偏分散	F 値	P 値	F(0.95)
回帰	196.38	8	24.547	35.32705	1.53×10^{-14}	2.216673
残差	24.32	35	0.6949			
計	220.7	43				

・回帰係数の有意性の検定と信頼区間								
	回帰係数	標準誤差	標準回帰係数	t 値	P 値	t(0.975)	95%下限	95%上限
定数項	-2.7943	1.6281	-2.7943	-1.7163	0.0949	2.0301	-6.0995	0.5109
Fe	0.0784	0.0106	0.836	7.4117	1.13×10^{-08}	2.0301	0.0569	0.0999
A	3.4474	0.6453	0.5282	5.342	5.69×10^{-06}	2.0301	2.1373	4.7575
B	6.6198	0.6001	1.0143	11.03	6.13×10^{-13}	2.0301	5.4015	7.8382
C	4.4041	0.7765	0.4957	5.6716	2.09×10^{-06}	2.0301	2.8277	5.9806
D	3.3651	0.7401	0.5795	4.5471	6.25×10^{-05}	2.0301	1.8627	4.8676
E	4.018	0.6064	0.6562	6.6258	1.17×10^{-07}	2.0301	2.7869	5.2491
F	9.6383	1.0616	1.0847	9.0791	9.97×10^{-11}	2.0301	7.4831	11.793
G	5.1057	0.8059	0.8793	6.3355	2.8×10^{-07}	2.0301	3.4697	6.7418

まず、解析結果 3-13 の「回帰係数の有意性の検定と信頼区間」の表より、Fe の P 値を見てみる。この P 値から、この線形モデルに当てはめた場合の Fe の偏回帰係数が統計的に意味があるかどうか判定される。この例題では、 $P = 1.13 \times 10^{-08}$ なので、統計的に有意であることがわかる。

手順3：相関係数を求める

そこで、この線形モデルでの相関係数を求めることにする。

解析結果 3-13 の「分散分析表」では回帰と残差が記載されているが、回帰の欄には Fe とダミー変数である患者を合わせた解析結果が表示されている。そこで、ここから Fe の要因を分離する。

まず、Fe に関わる不偏分散 (Mean Square) を下の式によって求める。

$$\text{Fe の不偏分散 (Mean Square)} = (\text{Fe の } t \text{ 値})^2 \times (\text{誤差の不偏分散})$$

解析結果 3-13 の「回帰係数の有意性の検定と信頼区間」の表より、Fe の t 値は 7.4117。「分散分析表」から、誤差の不偏分散は 0.6949。これを代入して

$$\text{Fe の不偏分散 (Mean Square)} = (7.4119)^2 \times 0.6949 = 38.171$$

これを元に Fe の偏差平方和を求めるが、(偏差平方和) = (自由度) × (不偏分散) であり、Fe の自由度は 1 であるので、

$$(\text{Fe の偏差平方和}) = (\text{Fe の不偏分散}) = 38.171$$

として、Fe に関する偏差平方和が求められる。ここから、Fe とヘモグロビンの相関係数を下記のように求める。

$$\begin{aligned} |\text{相関係数}| &= \sqrt{\frac{\text{Fe の偏差平方和}}{\text{Fe の偏差平方和} + \text{残差の偏差平方和}}} \\ &= \sqrt{\frac{38.171}{38.171 + 24.32}} \\ &= 0.781551159 \end{aligned}$$

この相関係数の符号は、解析結果 3-13 の「回帰係数の有意性の検定と信頼区間」の表の Fe の回帰係数 (0.0784) の符号と同じでなので+である。

以上の結果から、ある患者に着目すると血清鉄とヘモグロビンには相関係数 0.7816 の相関関係があることが統計的に有意 ($P = 1.13 \times 10^{-08}$) に確認された。

おまけ

例題 3-4 と同様に、解析結果 3-13 の分散分析表から血清鉄と患者の要因を分けて解析結果 3-14 の分散分析表を作ることができる。P 値はエクセルの FDIST 関数で計算できる。

→エクセルの関数や数式の入力方法は、付録 A「エクセルでの数式の入力方法」(p.135) を参照

■ 解析結果 3-14 分散分析表

・分散分析表					
要因	偏差平方和	自由度	不偏分散	F 値	P 値
Fe	38.171	1	38.171	54.933	1.13×10^{-08}
Subject	158.21	7	22.601	32.526	1.67×10^{-13}
残差	24.32	35	0.6949		
計	220.7	43			

▶▶▶▶ 英文での表記

⇒ 例題 3-4 の英文での表記 (p.46) を参照