実験計画学
2004 年 12 月 21 日

第１１回　分散分析その４　多元配置・多重検定

Ａ．多元配置の分散分析と実験計画

１．要因（処理）の数を３つ以上にした実験の分散分析

①　分散分析の原理は同じである

要因（処理）の数が３つ以上になっても

総変動＝個々の処理による変動＋交互作用による変動＋誤差変動

例：４つの要因について，ハムスターの成長を調べた．

Ａ：餌因子　植物の種子　ヒマワリ，クルミ，米，落花生

Ｂ：温度因子　高温，中温，低温

Ｃ：光因子　明　暗

Ｄ：密度因子　密　疎

②　３つ以上の交互作用

この場合, A, B, C, D各因子の効果だけでなく, (　　　　　), (　　　　　), (　　　　　), (　　　　　), (　　　　　), (　　　　　) という２つの因子の絡む交互作用（２次の交互作用), (　　　　　), (　　　　　), (　　　　　), (　　　　　) という３つの因子の絡む交互作用（３次の交互作用), (　　　　　) という４つの因子の絡む交互作用（４次の交互作用）の解析も分散分析で可能になる.

③　データの変動

総変動＝

２．実験を効率よく行う

①　まともに多因子実験すると，交互作用ばかり測定することになる

要因を増やせば増やすほど，実験が困難になる

実験する数が激増する.

２水準の実験について２反復行うとしよう．ハムスターを２匹ずつ各処理に割り当てることにすると,

２処理だと（　　）×（　　）×（　　）＝（　　）匹のハムスターが必要になる.

８処理だと（　　）$^{(\quad)}$×（　　）＝（　　）匹のハムスターが必要になる.

３水準の実験を２反復行うとしよう．４処理（因子）だと（　　）匹のハムスターが必要である.

表　2^8 計画の要因とその自由度

要因	自由度
主効果	8
２因子交互作用	28
３因子交互作用	56
４因子交互作用	70
５因子交互作用	56
６因子交互作用	28
７因子交互作用	8
８因子交互作用	1
全体	255

このように処理（要因）の数が増えると２水準の組み合わせでも大量の実験をする必要が出てくる．処理内の水準数を３や４にしたら，処理がそれほど多くなくても実験の数は著しく増える．

どうにかして実験を減らせないか？

８因子の実験（反復１）について考えてみると　Ａという因子について２つの水準しか設定していないから，その各水準についてだけ着目すると，反復数は（　　）になる．

このように８因子２水準の実験をまともにやっても，個々の因子についての精度は大して向上しない（反復数は５ないし６より多くなると，精度の向上はあまり大きなものでなくなってくる）．

これだけ多くの実験をして，いったいどこの精度がよくなったのか？

自由度を比べてみよう（右上の表）．

全体の自由度は（　　）ある．そのうち主効果に振り分けられる自由度は（　　），２因子の交互作用に振り分けられる自由度は（　　）に対し，４因子以上の交互作用に振り分けられる自由度は合計（　　）にもなる．ゆえに交互作用を考えなければ，実験の数を減らせる．

高次の交互作用は無視しても通常かまわない．

理由　①　高次の交互作用はあまり大きくない．

ひとつにはふつうの実験では主効果に効果のありそうな因子を選んで実験する

さらに交互作用の定義は総変動から主効果による変動と誤差変動を引いた残りであること．

②　３つ以上の要因が関与する交互作用の解釈が難しい．

３次以上の交互作用を無視すると考えると，主効果と２因子交互作用の自由度の和は（　　）なので，2^8（＝（　　））の１／４である（　　）区だけの実験でも，その計画をうまく組めば，必要な情報は得られると考えられる．

②　ラテン方格法による一部実施の見本

ラテン方格法による一部実施の具体例として３処理２水準の実験を考える．

水稲の施肥実験

窒素（無施肥N_0・施肥N_1），リン酸（無施肥P_0・施肥P_1），カリ（無施肥K_0・施肥K_1）の３処理を組み合わせた８つの実験を圃場の関係から４つの実験に減らしたい．

まともにすべての組み合わせについて実験すると

窒素＼リン酸	P_0	P_1
N_0	K_0，K_1	K_0，K_1
N_1	K_0，K_1	K_0，K_1

ラテン方格法に従い，右のように組むとどの処理も２回ずつ現れる（K_0＝Ａ，K_1＝Ｂとみなす）.

窒素＼リン酸	P_0	P_1
N_0	A	B
N_1	B	A

窒素＼リン酸	P_0	P_1
N_0	K_0	K_1
N_1	K_1	K_0

このようにラテン方格法にすると８つの組み合わせがあった実験を４つの処理の組み合わせに減らすことができる．しかし，この場合，交互作用は検出できない.

窒素，リン酸，カリをそれぞれ５水準にすると，反復がなくても 5^3＝１２５の実験がいる．右のようにラテン方格法を用いれば，２５ですむ.

	1	2	3	4	5
1	A	E	C	D	B
2	E	C	A	B	D
3	B	D	E	C	A
4	D	A	B	E	C
5	C	B	D	A	E

③　直交表による一部実施

ラテン方格法は３因子の実験でしか適用できない．４因子以上では直交表を用いて，一部実施することができる．授業の時間的に直交表を説明する時間は十分にないので，次のごく簡単な例だけを紹介する.

例：ハムスターに４種の栄養剤（Ａ，Ｂ，Ｃ，Ｄ）を与えるか与えないかによる成長の違いを実験したい.

与えるかあるいは与えないかの２水準であるから，そのまま実験を組み合わせると 2^4=16 の処理の組み合わせになる．しかし，ハムスター用のかごが８つしかなく，ハムスターも各かご５匹ずつとして，４０匹しかいなかったとする.

次のように８つのかごに実験を配置すれば，４種の栄養剤のそれぞれを均等に割り付けて，しかも実験の組み合わせを 8 つに減らすことができる（添え字の１は投与，２は無投与）．このような実験配置に利用するのが直交表である.

No.＼列	(1)	(2)	(4)	(7)
1	1	1	1	1
2	1	1	2	2
3	1	2	1	2
4	1	2	2	1
5	2	1	1	2
6	2	1	2	1
7	2	2	1	1
8	2	2	2	2
因子	K	L	M	N

No.1 $\boxed{A_1B_1C_1D_1}$

No.2 $\boxed{A_1B_1C_2D_2}$

No.3 $\boxed{A_1B_2C_1D_2}$

No.4 $\boxed{A_1B_2C_2D_1}$

No.5 $\boxed{A_2B_1C_1D_2}$

No.6 $\boxed{A_2B_1C_2D_1}$

No.7 $\boxed{A_2B_2C_1D_1}$

No.8 $\boxed{A_2B_2C_2D_2}$

直交表によって実験数を減らせるのは変動の一部を分離しないからである．

交絡：変動を分離できないでからまってしまっていること

繰り返しのない２元配置では交互作用が評価できないが，これは誤差変動と交互作用による変動が交絡して，分離できないからである．評価の必要のない交互作用は主効果などと交絡させると実験回数を少なくできる．

直交表には以下の原則で割り付けるとよい．

（１）　主効果は３次以上（なるべく高次）の交互作用と交絡させる．３次以上の交互作用は通常，無視できる．高次の交互作用ほど小さい．

（２）　２次の交互作用のうち，評価したい２次の交互作用を評価できるように割り付ける．

④　直交表へのわりつけの仕組み

$L_8(2^7)$ 直交表

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

直交表の各列は以下の主効果と交互作用を表している

(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	B	A×B	C	A×C	B×C	A×B×C

もしＡ×Ｂ×Ｃの交互作用を無視するなら（７）に因子Ｄを入れることができる．

(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	B	A×B	C	A×C	B×C	D

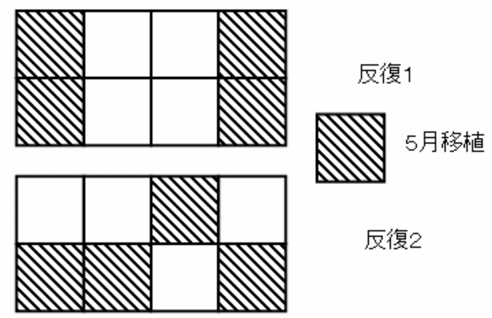
もしすべての交互作用を無視するなら７つの因子を入れることができる．

(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	B	D	C	E	F	G

B．分割区法

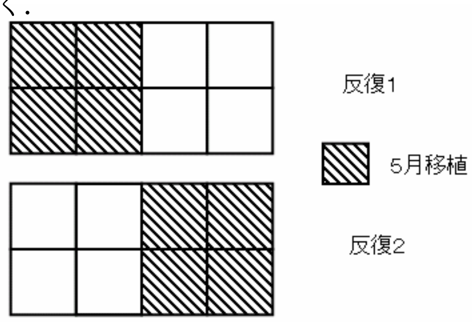
いくつかの要因を組み合わせて実験を行う場合に，因子の種類によっては，その水準を試験区ごとに完全にランダムに配置することが難しいことがある．例えば，水稲の苗について，移植時期（５月移植，６月移植)，苗の性質（二酸化炭素処理苗，対照苗)，栽植密度（密植，疎植)，を組み合わせて，水稲の苗の性質と収量の違いを主に調べ，苗の性質と交互作用のありそうな作期と栽植密度の２つを標示因子として取り上げた実験を考える．３反復の実験を行うことにし，各ブロックに８つの処理の組み合わせを配置することにした．

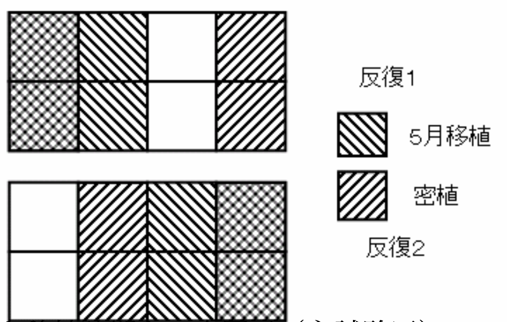
下の図は乱塊法によって，各ブロックに無作為に８つの処理を配置した例である．



こうすると移植時期が異なる２つの水田の間には水が進入しないように区切りを必ず入れなければならないので，区切りがたくさん必要になって，煩雑である．さらに栽植密度についても，密度の変わるたびに植え綱を交換しなければならないからこれも煩雑である．そこで，作期あるいは密度についてはひとまとまりにして配置することにする．これを分割区法という．

下の例では，左では移植時期についてひとまとめにしている．残りの２つの要因はそれぞれ４つのマスにランダムに配置される．右の例ではまず移植時期についてひとまとめにし，さらに栽植密度についてもひとまとめにしている．残りの要因である苗の性質だけを２つのマスにランダムに配置していく．





左の例では移植時期をひとまとめにしている．このとき移植時期を１次因子（主試験区）といい，苗の性質と栽植密度を２次因子（副試験区）という．右の例ではまず移植時期でまとめ，次に栽植密度でまとめて，試験区を配置している．この場合，移植時期を１次因子（主試験区)，栽植密度を２次因子（副試験区)，苗の性質を３次因子（副々試験区）という．

分割区法では１次因子と２次因子の組み合わせで分散分析の方法が変わる（ここでは説明しない)．１次因子に比べ，より高次の因子ほど検出力が高く，さらにその因子と他の因子との交互作用の検出力も高い．すなわち左の場合，移植時期の違いについての検出力はあまり高くないが，３次因子である苗の性質の違いや苗の性質と移植時期との交互作用の検出力は高い．実験の目的に応じて，分割区法ではどの因子を１次因子に持っていき，どの因子を高次の因子に配置するかを考えることが実験の精度を高める上で重要である．

C．多重検定

１．分散分析の結果でわかるのは，処理間に有意差があることだけである

下の例は，４種類の薬と対照区（薬を与えない処理）でヤギを育てたときの成長量を調べたものである．

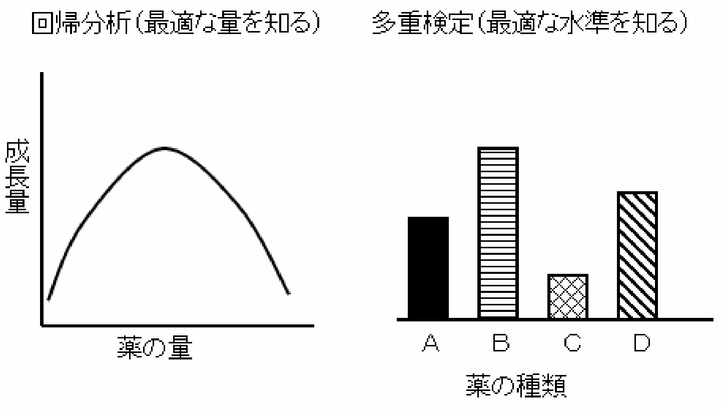
処理	対照区	A	B	C	D
	100	105	96	100	115
	102	108	97	97	112
	104	110	100	95	100
	105	106	102	104	105
	103	104	99	103	106
平均	102.8	106.6	98.8	99.8	107.6
標準偏差	1.92	2.41	2.39	3.83	5.94

分散分析表

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	309.84	4	77.46	5.940184049	0.002556	2.866081
グループ内	260.8	20	13.04			
合計	570.64	24				

分散分析の結果，薬を与える処理は 1%の有意水準で有意であった．したがって，薬には効果があることがわかる．しかし，上の場合，A，B，C，Dそれぞれの薬が対照区より効果があるのか，あるいはそれぞれの薬どうしで差があるのかは分散分析では検定できない．そのため対照区よりA，Dの薬が効果が有意に高いかは分散分析ではわからない．

３水準以上を同時に行った実験について，それぞれの水準について個々に差があるかどうかを検定する方法が多重検定である．なお水準が量的因子である場合には第１３回で学ぶ回帰分析をする．量的因子であれば，実験で採用した水準そのものが最適な水準とは限らない．採用した水準と水準の間に最適な水準がある方がふつうである．そのような場合，水準と反応を回帰分析し，最適な水準をみつける．一方，質的因子では最適な水準は実験で採用した水準から選ぶことができる．



2．多重検定

古くから使われた LSD 法は t 検定を繰り返して行う方法なので, 実際にはすべての水準の組み合わせについて LSD 法で有意差を検定すると, その中のあるペアについて誤った判定をする確率は高くなる．このことからすべての水準の組み合わせについて同時に検定しながら，宣言した有意水準の範囲で第１種の誤りを抑えるような検定法が考えられている．これを多重検定という．しかし，多重検定には多数の手法があり，状況に応じて使い分ける必要がある．その上，専門家の間でもどれを使ったらよいのかの意見の一致をかならずしも見ていない（Clewer and Scarisbrick, 2001）．ここではよく聞かれる多重検定を紹介するにとどめる．

① LSD 法　古くからよく使われた方法である．水準が 3 つのときは分散分析の結果，処理に有意な効果があると認められたときに LSD 法を使うことができる．

② テューキー法（テューキーのＨＳＤ法ともいう）
基本的な多重検定法で，水準間のすべての組み合わせについて検定を行う．

③ ダネット法
対照区と残りの水準という比較をするときに用いる方法．例えば，Ａ～Ｃの３種類の新薬の効果を調べるときに共通の対照区を１つ設けた実験ならば，ダネット法で検定できる．

①から③はどれも各水準間で分散が等しいことを仮定している．

多重検定した結果，各水準にそれぞれ有意差がある，あるいは有意差が認められなかった場合は以下のように記述するのが一般的である．

概要

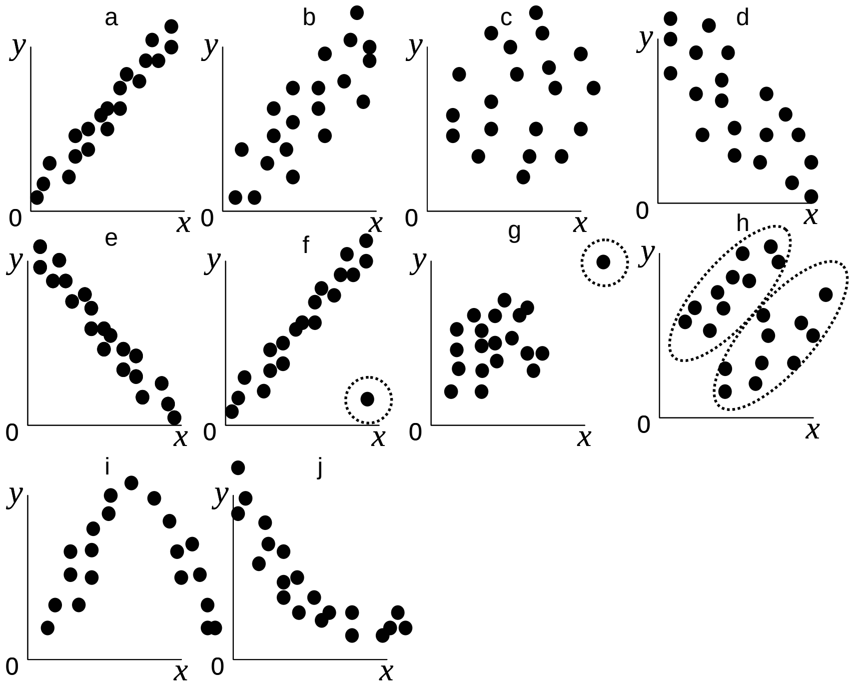
ｸﾞﾙｰﾌﾟ	標本数	合計	平均	
対照区	5	514	102.8	b
A	5	533	106.6	c
B	5	494	98.8	a
C	5	499	99.8	a
D	5	538	107.6	c

同じ符号を記した値の間には，○○法 で有意水準 5% において有意差がない

D．宿題

１．次回からは相関分析および回帰分析について学ぶ．

相関分析は２つの変量間の関係の強さを相関係数という値を求めて調べる方法である．回帰分析とは２つの変量の関係がどのくらいあるかを定量的に見積もり，さらに２つの変量の間の関係をある式に表現する方法である．下には散布図（２つの変量のうち，一方をｘ軸に，もう一方をｙ軸にして，両者の関係を図示した図）のパターンがいくつかある．ａはｘが増えれば，ｙも増加する傾向がある関係を示す．ａの例としては，世界各地点での一年の最高気温と最低気温にはこのような関係があるだろう．ｅはｘが増えれば，ｙは減る傾向のある関係を示す．ｅの例としては，一年間の晴天の日数と降水量の関係が挙げられる．ｃはｘとｙの間に関係がないことを示す．ｃの例としては，北極でのオーロラの数と東京駅で売れる弁当の数があるかもしれない（たぶん関係はないとは思うが･･･）．



①　以上のａ，ｃ，ｅの３つのパターンに当てはまる２つの変量を考えよ．

②　**ａもしくはｅのパターン**にあてはまりそうなデータの組み合わせについて，相関，回帰それぞれに当てはまる 20 組以上のデータを集め，散布図を書け（もとのデータも必ずノートに載せること）．

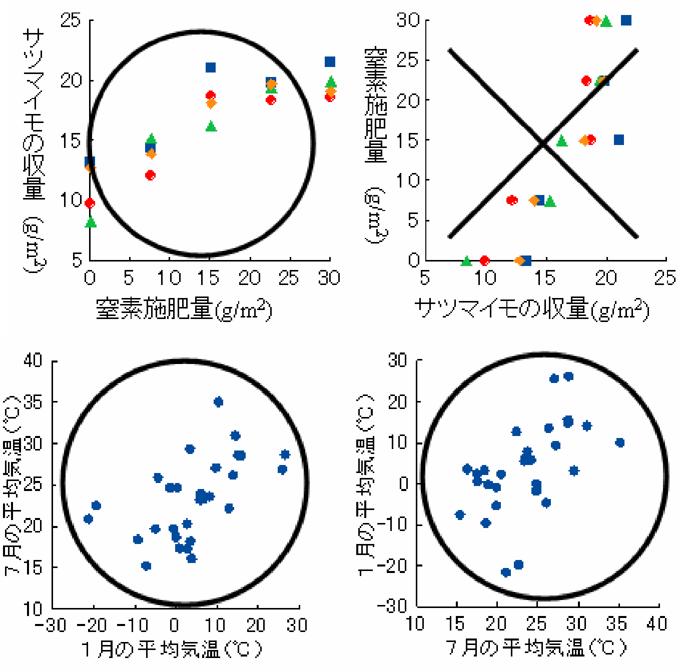
③　②で書いた散布図を見て，２つの変量の関係がどの程度強いのかを考えよ．すなわち下の図でもｂよりはａの方が両変数の関係が強い．自分の集めたデータを右のパターンと比較し，どれに近いかを考えよ．

各自が異なるデータを調査すること．そのために事前にどういうデータを調査するかを 204 室前のホワイトボードに貼った紙に報告すること（早い者勝ちとする）．今回の宿題の提出は 1 月 10 日（月）午後 1 時（厳守）とする．

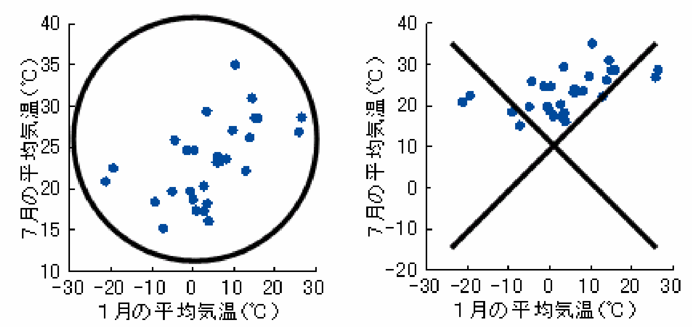
相関と回帰の２つとも探してくること．なお距離と運賃のようにばらつきがないデータはこのような解析にはふさわしくない．国別，都道府県別データの場合，面積や人口などでは極端に多いデータが１つか２つ入るようなものは望ましくない．なお調査テーマが相関，回帰にふさわしくないときはテーマを却下するので注意すること．

2．散布図の書き方

① ｘ軸（横軸）には原因となる変量を，ｙ軸（縦軸）には結果となる変量をふつうとる．



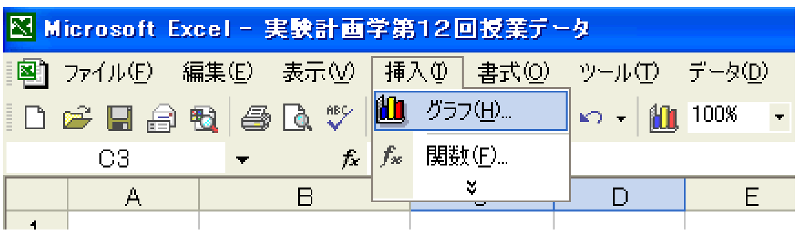
② できるだけ点が図全体にばらつくように軸の上下限を決める・

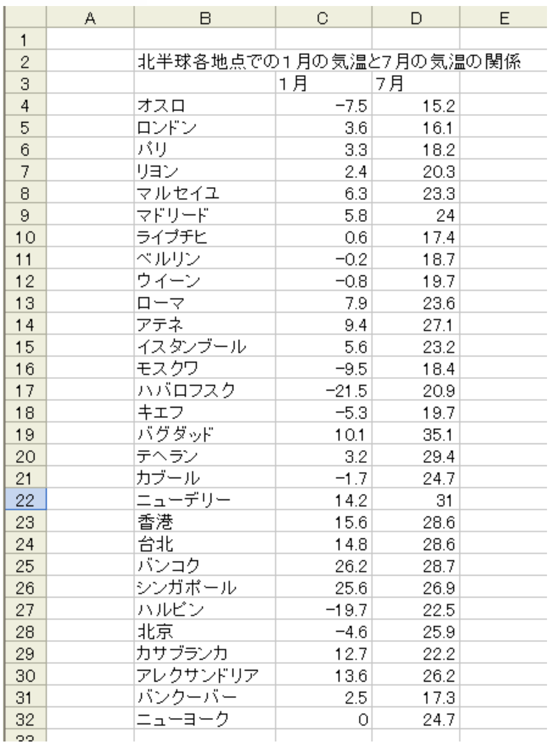


3．エクセルでの散布図の書き方

① 右のデータを散布図にしてみよう．

② 挿入→グラフを選ぶ．あるいはグラフの絵が描いてあるアイコンを選ぶ

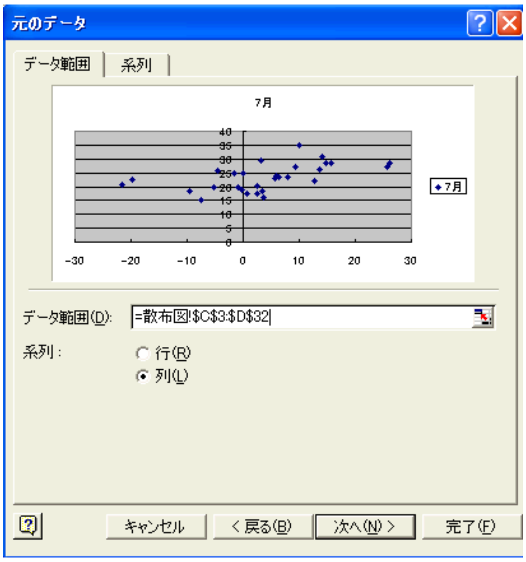




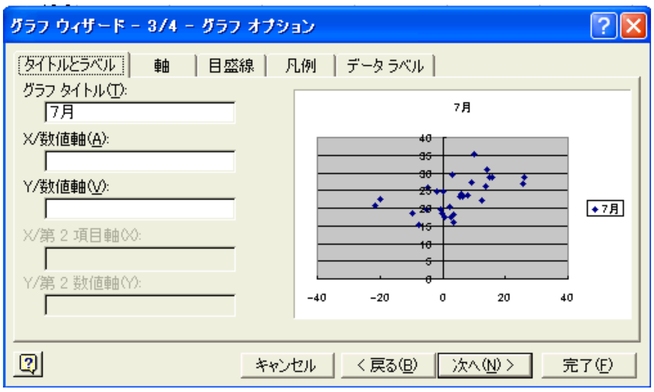
	A	B	C	D	E
1					
2		北半球各地点での１月の気温と7月の気温の関係			
3			1月	7月	
4		オスロ	-7.5	15.2	
5		ロンドン	3.6	16.1	
6		パリ	3.3	18.2	
7		リヨン	2.4	20.3	
8		マルセイユ	6.3	23.3	
9		マドリード	5.8	24	
10		ライプチヒ	0.6	17.4	
11		ベルリン	-0.2	18.7	
12		ウィーン	-0.8	19.7	
13		ローマ	7.9	23.6	
14		アテネ	9.4	27.1	
15		イスタンブール	5.6	23.2	
16		モスクワ	-9.5	18.4	
17		ハバロフスク	-21.5	20.9	
18		キエフ	-5.3	19.7	
19		バグダッド	10.1	35.1	
20		テヘラン	3.2	29.4	
21		カブール	-1.7	24.7	
22		ニューデリー	14.2	31	
23		香港	15.6	28.6	
24		台北	14.8	28.6	
25		バンコク	26.2	28.7	
26		シンガポール	25.6	26.9	
27		ハルビン	-19.7	22.5	
28		北京	-4.6	25.9	
29		カサブランカ	12.7	22.2	
30		アレクサンドリア	13.6	26.2	
31		バンクーバー	2.5	17.3	
32		ニューヨーク	0	24.7	

③　グラフの種類は散布図を選ぶ．右側の形式は一番上の形式，すなわち点を線で結ばない形式を選ぶ．

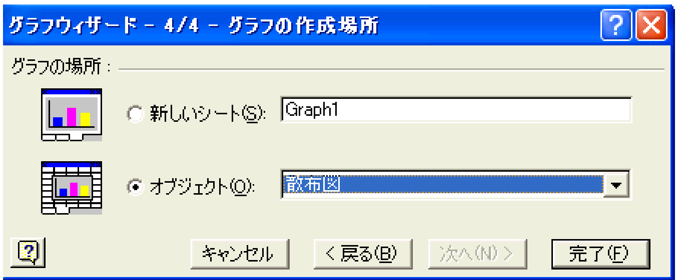
④　データの範囲を入力する．データの方向を系列で指定する．



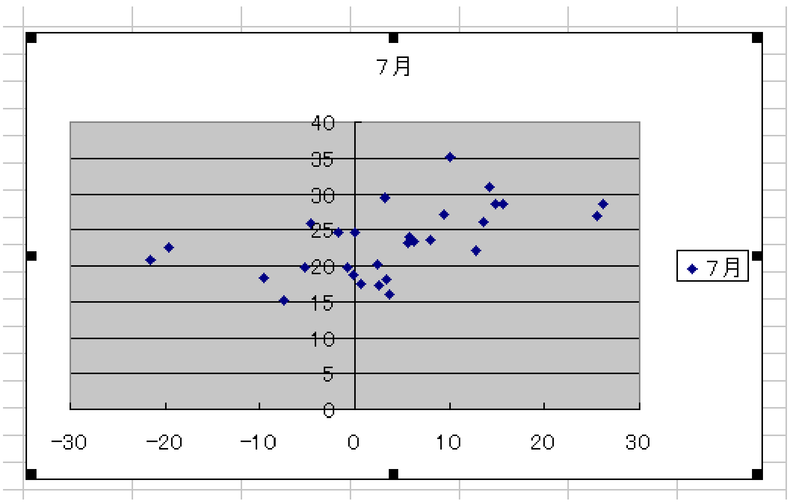
⑤　必要に応じてタイトル，軸のラベルなどを指定する．



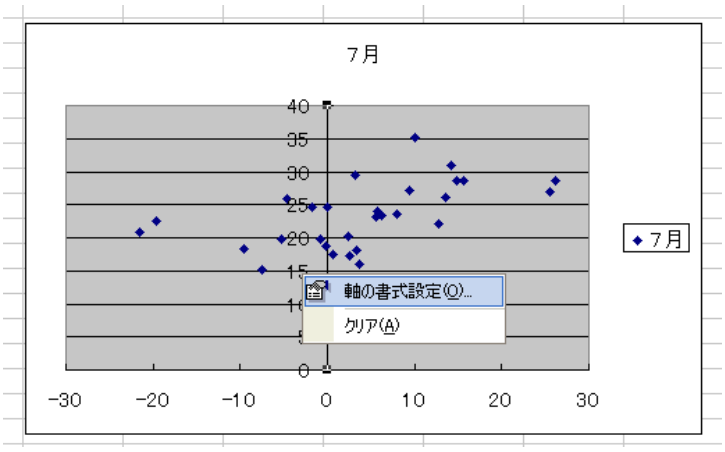
⑥　グラフを作成する場所を指定する．

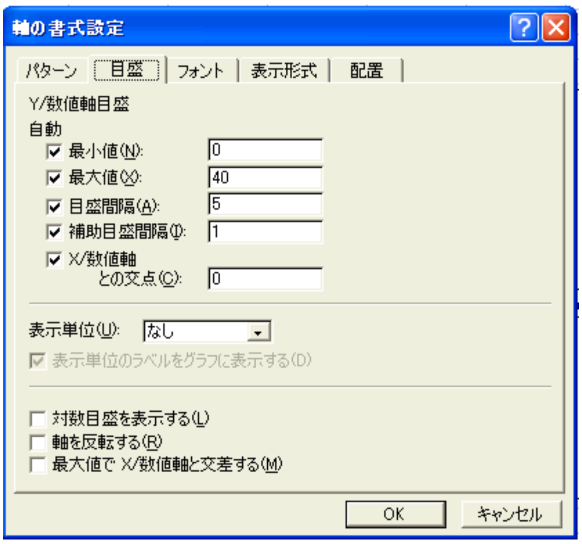


⑦　右のようなグラフが完成する．さらに必要に応じて，グラフを修正する．

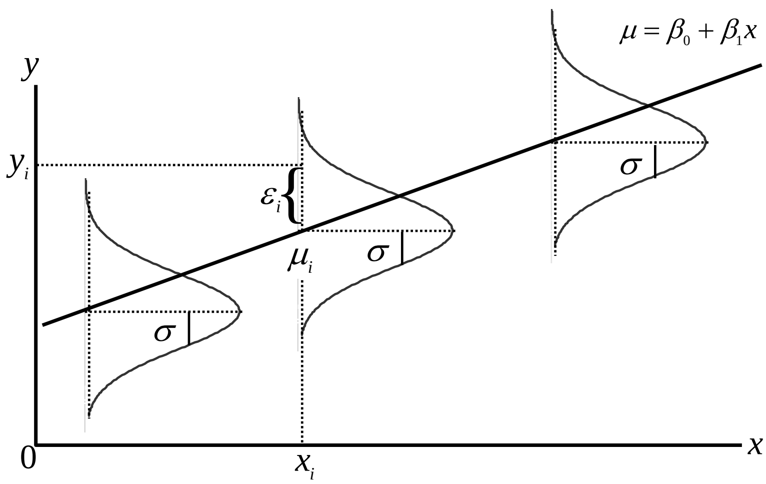


⑧　例えば，グラフのデータの範囲に合わせて，軸の範囲を設定するには下のようにする．

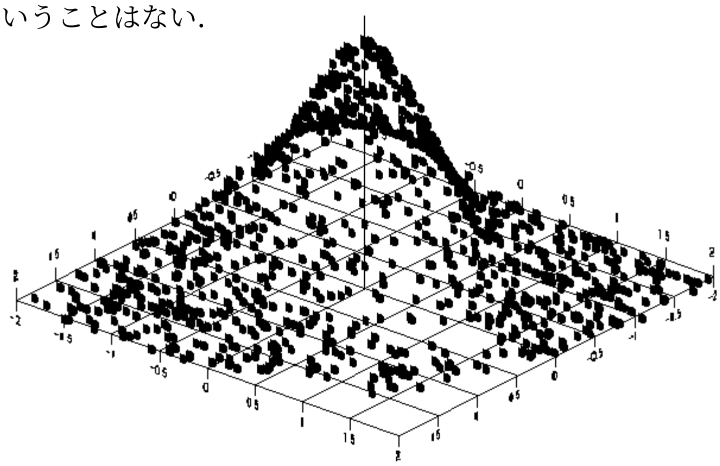




回帰　２つの変量の関係を考えるときに，片方を指定してやるともう一方が（ある一定の誤差ばらつきを含んで）決まる場合は回帰分析を行う．



相関　２つの変量がともにあるばらつきを持っている（正規分布する）ときは相関分析する．
　２つの変量のどちらが原因とか結果ということはない．



　兄と弟の身長の場合はどちらが原因でどちらが結果とは考えられないから相関分析する．ところが父と子の身長であれば父の身長が遺伝的に子の身長をある程度支配すると考えられるから，父の身長を原因，子の身長を結果とみて，回帰分析できる．一日の最高気温と最低気温は相関分析するべきであり，一日の太陽エネルギーと平均気温は回帰分析するべきである．